NATIONAL QUALITY FORUM

National Voluntary Consensus Standards for Patient Outcomes
Measure Summary

Measure number: OT2-002-09

Measure name: Risk Adjusted Colorectal Surgery Outcome Measure

Description: This is a hospital based, risk adjusted, case mix adjusted morbidity and mortality aggregate outcome measure of adults 18+ years undergoing colorectal surgery.

Numerator statement: The outcome of interest is 30-day, hospital-specific risk-adjusted (all cause) mortality, a return to the operating room, or any of the following morbidities as defined by American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP): Cardiac Arrest requiring CPR, Myocardial Infarction, DVT requiring therapy, Sepsis, Septic Shock, Deep Incisional ssi, Organ/Space SSI, Wound Disruption, Unplanned Reintubation without prior ventilator dependence, Pneumonia without pre-operative pneumonia, Pulmonary Embolism, progressive Renal Insufficiency or Acute Renal Failure without pre-operative renal failure or dialysis, or UTI. All outcomes are definitively resolved within 30 days of any ACS NSQIP listed (CPT) surgical procedure. All variables (fields) are explicitly defined in the tradition of the ACS NSQIP and definitions are also submitted in these materials.

The current set of mortality and major complications for this measure was chosen based on prior work revealing that these complications are related to other important criteria such as large contributions to excess length of stay, large complication burdens, or correlations with mortality. Of note, the measure does specifically include return to the operating room within 30 days as a dependent outcome. In addition, the desire to limit the outcomes to significant events (i.e. - some degree of severity according to certain criteria) is the reason that superficial wound infection is excluded from the measure.

Denominator statement: Patients undergoing any ACS NSQIP listed (primary CPT) colorectal surgical procedure. (44140, 44141, 44143, 44144, 44145, 44146, 44147, 44150, 44151, 44155, 44156, 44157, 44158, 44160, 44204, 44205, 44206, 44207, 44208, 44210, 44211, 44212, 45110, 45111, 45112, 45113, 45114, 45116, 45119, 45120, 45121, 45123, 45126, 45130, 45135, 45160, 45395, 45397, 45402, 45550) Notes: following codes are not included in this denominator list: 44152 (not found), 44153 (not found), 44239 (not found), 45540 (proctopexy without resection), 45499 (unlisted laparoscopy, rectum).

Level of Analysis: Facility/Agency , Population: national, Population: regional/network, Population: states

Type of Measure: Outcome

Data Source: Documentation of original self-assessment, Management data, Electronic clinical data, electronic Health/Medical Record, lab data, paper medical record/flowsheet

Measure developer: ACS

Type of Endorsement: (full or time-limited): Full Endorsement (Recommend-22, Do not Recommend-0, April 20-21, 2010 Meeting)

NATIONAL QUALITY FORUM

National Voluntary Consensus Standards for Patient Outcomes

Measure Summary

Summary table of TAP ratings of sub criteria and comments:

<table>
<tr><td colspan="3">IMPORTANCE TO MEASURE AND REPORT</td></tr>
<tr><td>1a Impact</td><td>Overall ranking: Completely

Completely: 6
Partially: 3
Minimally: 0
Did not meet: 0</td><td rowspan="3">This measure is meaningful as mortality and severe morbidity are important. The information provided and intent of the measure clearly meets the subcriteria for importance.</td></tr>
<tr><td>1b gap</td><td>Overall ranking: Completely

Completely: 6
Partially: 3
Minimally: 0
Did not meet: 0</td></tr>
<tr><td>1c relation to outcomes</td><td>Overall ranking: Partially

Completely: 2
Partially: 7
Minimally: 0
Did not meet: 0</td></tr>
<tr><td colspan="2">SCIENTIFIC ACCEPTABILTY</td><td>Overall ranking: Partially

Completely: 0
Partially: 9
Minimally: 0
Did not meet: 0</td></tr>
<tr><td>2a specs</td><td>Overall ranking: Partially

Completely: 1
Partially: 3
Minimally: 0
Did not meet: 0</td><td rowspan="3">Information on reliability and validity of the measure is provided. The ability to compare across hospitals is a real strength. One concern raised was on how smaller hospitals will be able to publicly report on this measure given the need for a sufficient number of cases. The measure developer clarified that approximately 65 cases were needed each year, which means that the measure may only apply to 40-45% of all hospitals but would cover 85% of all colorectal surgery.

As included in the forms, reliability was found to be moderate. The measure developer clarified that while the findings were found to be moderate, the information is more than typically provided and meet acceptable standards proposed in the literature. A member also questioned whether the measure has been validated outside of NSQIP but the developer supports that the measure can be implemented in other programs and by other organizations. It is</td></tr>
<tr><td>2b reliability</td><td>Overall ranking: Completely

Completely: 2
Partially: 1
Minimally: 1
Did not meet: 0</td></tr>
<tr><td>2c validity</td><td>Overall ranking:</td></tr>
</table>

	Completely: 1 Partially: 1 Minimally: 1 Did not meet: 1	estimated that it would require about a 20th of a full-time employee to abstract the data needed for the measure. ***Additional Cancer TAP Comments following call:*** Strong methodology. Lack of validation outside NSQIP data platform. Reliance on x numbers of patients per hospital still yields poor reliability. Learner curve for data abstractors outside of NSQIP platform. These issues must be addressed before endorsement can be considered. ***April 13 follow-up call:*** TAP members asked for additional information on the procedure for selecting cases for the sample size. The measure developer clarified that the sampling mechanism would be specified by the organization implementing the measure. Currently, in NSQIP a temporal systematic sampling is used but for the specifications in the proposed measure they chose to not limit it to one approach but recommend that it be a systematic sample. The concern with the low reliability results was also discussed. The ability to be able to act upon the results is difficult due to these results; although, it was appreciated that the only way to increase reliability would be to increase the number of cases that must be included in the sample. It is not an issue that can be easily solved and should be noted as a weakness to be considered during this evaluation. The developer noted that the information provided on this measure often exceeds what is found with other measures that have been endorsed by NQF. In addition, one would expect that **inter-class correlations would increase with national implementation.** In addition, when the measure is publicly reported, the advantage to drill down for quality improvement would be lost. The developer noted that any institution who implemented this measure would still be able to analyze the data at that more detailed level. NSQIP experience has shown that even measures with low reliability have still driven quality improvement.
2d exclusions	Overall ranking: Partially Completely: 0 Partially: 3 Minimally: 1 Did not meet: 0	
2e risk adjustment	Overall ranking: Completely Completely: 4 Partially: 0 Minimally: 0 Did not meet: 0	
2f meaningful differences	Overall ranking: Partially Completely: 1 Partially: 2 Minimally: 1 Did not meet: 0	
2g comparability	Overall ranking: Minimally Completely: 0 Partially: 1 Minimally: 3 Did not meet: 0	
2h disparities	Overall ranking: Partially Completely: 0 Partially: 3 Minimally: 0 Did not meet: 1	
USEABILITY	Overall ranking: Partially Completely: 0 Partially: 7	

	Minimally: 1 Did not meet: 1	
3a distinctive	Overall ranking: Completely: 0 Partially: 2 Minimally: 2 Did not meet: 0	Members of the TAP were unsure of whether the public would understand the composite nature of the measure. In addition, the performance data provided appeared to show that the improvement curve is relatively flat. The developer clarified that those who have implemented the measure have demonstrated improvement over time. Consumer understanding of the odds ratio is sometimes difficult and may impact its usability for patients. ***Additional Cancer TAP Comments following call:*** Does not specifically address why 30-day mortality or failure to rescue are not suitable for measurement (Already NQF endorsed). Innovations appear to be risk adjustment. However, this measure is more difficult to collect, measure, and report. I contend that NSQIP mortality rates are not uniformly improving. ***April 13 follow-up call:*** TAP members requested clarification on the total amount of FTE required to abstract the measure and what training and education would be available. The developer confirmed that a conservative estimate of .3 FTE was included but it is assumed that it will take less than .1 FTE when the measure is implemented. One assumption would be that an organization would replace an existing measure that is being collected with this measure; thus, reducing the associated time and costs. Implementation would include education on auditing but that level of detail would need to be provided by the organization that has chosen to implement this measure. The American College of Surgeons (ACS) is committed to sharing its experience and expertise based on NSQIP to others implementing this measures that may not use NSQIP. Currently, there are no plans to pilot the measure outside of NSQIP.
3b harmonization	Overall ranking: Completely: 1 Partially: 1 Minimally: 1 Did not meet: 0	
3c Added value	Overall ranking: Completely: 0 Partially: 2 Minimally: 2 Did not meet: 0	
FEASIBILITY	Overall ranking: Minimally Completely: 0 Partially: 3 Minimally: 6 Did not meet: 0	
4a Data a byproduct of care	Overall ranking: Did not meet Completely: 0 Partially: 0	One of the key concerns for the TAP was feasibility and the associated costs to implement the measure. All agreed that the developer has demonstrated how the measure works but it is not yet clear whether it can also be implemented outside of a hospital that currently participates in NSQIP. The data required must be

	Minimally: 1 Did not meet: 3	generated by abstraction and is not readily available through electronic data sources. The measure developer clarified that it is a parsimonious algorithm that hospitals would apply and other organizations would be able to implement this measure without participation in NSQIP. ***Additional Cancer TAP Comments following call:*** Risk of mis-measurement largely unknown as collected currently in NSQIP. Validation study in non-NSQIP hospitals on feasibility, usability, reliability, and validity recommended prior to final endorsement. ***April 13 follow-up call:*** It is no question that the measure in the NSQIP environment is useful and meaningful. It is important to ensure that the training of the abstractors be consistent across the organizations implementing this measure. The developer confirmed that based on the NSQIP experience abstraction of the data for this measure used to be completed by a registered nurse but they have now found that anyone who receives training is able to do this.
4b Electronic	Overall ranking: Minimally Completely: 0 Partially: 2 Minimally: 2 Did not meet: 0	
4c Exclusions	Overall ranking: Partially Completely: 0 Partially: 2 Minimally: 1 Did not meet: 1	
4d Inaccuracies/errors	Overall ranking: Minimally Completely: 1 Partially: 1 Minimally: 2 Did not meet: 0	
4e Implementation	Overall ranking: Minimally Completely: 0 Partially: 0 Minimally: 3 Did not meet: 1	

Summary of SC ratings of sub criteria and comments:

IMPORTANCE TO MEASURE AND REPORT	
N/A	SC Vote on Importance Yes - 22 No - 0
SCIENTIFIC ACCEPTABILITY	

The risk-adjustment model uses a parsimonious set of clinical risk factors collected in the database. The measure has been specified for broader implementation by hospitals who do not participate in NSQIP.	SC vote on scientific acceptability Completely - 11 Partially – 11 Minimally – 0 Not at all – 0
USABILITY	
N/A	SC vote on usability Completely - 11 Partially – 11 Minimally – 0 Not at all – 0
FEASIBILITY	
Includes capability for non-NSQIP hospitals to participate.	SC vote on feasibility Completely - 15 Partially – 7 Minimally – 0 Not at all – 0

Summary of Biostatistical review:

Type of Risk Model : *Marginal (i.e. not hierarchical) logistic regression. Hospital results were summarized as observed-to-expected ratios (O/E ratios).*
RISK FACTORS Are the risk factors clearly identified in the submission information? *YES. (But definitions were not provided.)*

Model covariates are: ASA Class, pre-operative Functional Status, Indication, Log Odds CPT (CPT Risk), Emergent, and Wound Class

Does the model include risk factors associated with differences/inequalities with care such as race, socioeconomic status or gender? *NO*

Are the conceptual and quantitative criteria for inclusion or exclusion or combining of risk factors explained and appropriate?

Variables were selected by a stepwise (forward, I believe) variable selection algorithm with an entry criterion of p<.05. From an initial list of 26 candidate predictors, 20 were selected. The final model was chosen by keeping only the first 6 variables to enter the model. The full list of 26 candidate variables is not given. It is not clear why the developers selected 6 variables for the final model.

Is quantitative assessment of the relative contribution of the model components described in detail?

NO.

Does the measure have exclusions that influence outcomes that should be included as risk factors?

The exclusions seem to be well justified.

Comments on risk factors:

The list of candidate variable was not provided.
Was preoperative creatinine considered as a risk factor?
Why did the developers opt for a highly parsimonious model?
Are risk factors NOT in the model randomly distributed across hospitals?
A limitation of automated variable selection is that the choice of variables can be sensitive to chance. Would the same predictors be identified if the variable selection algorithm was repeated in the validation sample? If not, would this have a large impact on hospital performance estimates?

VALIDATION OF THE RISK MODEL

Is there information provided on the cross-validation of the model comparing a development sample and a validation sample provided? *YES*

Is there information on independent, external validation of the model in another data set? *NO*

Are the results supportive of a valid model? *YES. (But see comments.)*

RISK MODEL PERFORMANCE (2e)

DISCRIMINATION: *C = 0.727 (C = 0.721 in validation set)*

Does the statistic support good discrimination? *YES*

CALIBRATION: Is a calibration curve included? *NO*
Is a risk decile plot included? *NO*
Hosmer-Lemeshow statistic: p = 0.177
Does the data support good model calibration? *YES*

The developers state: "Because of the very large sample sizes studied here, a statistically significant Hosmer=Lemeshow statistic is not considered informative with respect to calibration."

Although the HL statistic is uninformative, model calibration could still be assessed graphically. This could be done by comparing observed vs. predicted event rates within deciles of predicted risk. The analysis population is relatively broad. Conceptually, with a large enough sample size, each type of surgery could have its own custom risk model which may perform better than an overall global model. Given the broad population, it would be reasonable to assess the fit of the model in specific subgroups, for example, by CRT group or indication for surgery.

Comments on Risk Model Performance:

The developers opted for a model with 6 variables and a discrimination of C=0.727 over a model with 20 variables and discrimination of 0.738. Why was the more parsimonious model selected?

Reliability testing (2b):
Is the reliability of the key data elements, such as risk factors and the outcome demonstrated?

Data fields are well defined, but rigorous reliability testing has not been reported. In Section 4b, the developers mention that NSQIP data collectors undergo extensive training and are audited.

Is there information about the reliability of the measure score, such as signal to noise ratio?

YES. This was analyzed in detail. The developers estimated the intraclass correlation coefficient (ICC) and used this result to project the number of patients per hospital needed to achieve a specified signal to noise ratio (reliability). The developers report that 63 eligible operations per hospital would be sufficient to achieve a reliability coefficient of 0.4. The developers estimate that 43% of US hospitals and 69% of NSQIP participants have volumes large enough to meet this minimum sample size threshold.

Note. The following calculations (based on the binomial distribution) may provide some additional context for assessing precision:

- *If exactly 14 of 63 patients (22.2%) at a hospital experienced the endpoint of interest, the 95% binomial confidence interval estimate would extend from 12.7% to 34.5%.*
- *A sample size of 63 patients yields approximately 34% power to detect a 5 percentage point increase (27.2% vs. 22.2%) and 85% power to detect a 10 percentage point increase (37.2% vs. 22.2%) in a hospital's event rate compared to the national average.*

Has a sensitivity analysis been performed for problem or missing data?

YES. This is not described in the material provided, but a reference is provided. NSQIP investigators compared a variety of strategies for handling missing data and concluded that their approach is reasonable. Although inferences about specific risk factors did change depending on the missing data method, the differences did not substantially impact inferences about relative hospital performance.

Does the data demonstrate that the risk model is reliable? *YES*

The developers performed sensitivity analysis to determine whether inferences about hospital performance were sensitive to the type of risk adjustment model (conventional vs. hierarchical). Results of each were similar.

Comments on reliability testing:

Validity testing (2c):

Is validity testing of the measure to demonstrate results can be used to make conclusions about quality provided?

YES. Validity testing focused on assessing the performance of the risk-adjustment model.

Are the results supportive of a valid measure? *YES. (More details on model calibration would be helpful.)*

Comments on validity testing:

It would be useful to have more information about the protocol for insuring consistent and reliable 30-day endpoint ascertainment. Observed differences in mortality & morbidity could conceivably reflect differences in protocols for following patients post-discharge during the 30-day window. Are patients who are lost to follow-up excluded from the calculations? Or, are they included and assumed not to have an event?

Did the investigators examine the relative frequency of individual endpoints in the composite endpoint? Did they verify that no single item dominates?

Scoring Method Justification (2f):
Is the choice of method for computing risk-adjusted scores and identifying statistically significant differences justified? *YES*

Comments on scoring methods:

The O/E point estimates may be somewhat noisy. Reporting confidence intervals (as the developers propose to do) will be helpful.

NATIONAL QUALITY FORUM

National Voluntary Consensus Standards for Patient Outcomes
Measure Summary

Summary comments:
Reviewer: *Sean O'Brien, PhD* *Assistant Professor, Department of Biostatistics and Bioinformatics* *Duke University Medical Center, Duke Clinical Research Institute,* *Durham, NC*

Attachments:

NQF_Asthma_Pediatric_Risk_Adjustment_10 30 09.xls

NQF_CAD_all_codes_10.30.09.xls

NQF_CAD_Risk_Adjustment_10.30.09.xls

NQF_CHF_all_codes_10.30.09.xls

NQF_CHF_Risk_Adjustment_10.30.09.xls

NQF_GERD_all_codes_10.29.09.xls

NQF_GERD_Risk_Adjustment_10.29.09.xls

NQF_HTN_all_codes_10.30.09.xls

NQF_HTN_Risk_Adjustment_10.30.09.xls

NQF_Pediatric_ ASTHMA_all_codes_10.29.09.xls

NQF_PNE_Risk_Adjustment_10.30.09.xls

NQF_Pneumonia_all_codes_10.30.09.xls

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** #: OT2-002-09 **NQF Project:** Patient Outcomes Measures: Phases I and II
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Risk Adjusted Colorectal Surgery Outcome Measure
De.2 **Brief description of measure***:* This is a hospital based, risk adjusted, case mix adjusted morbidity and mortality aggregate outcome measure of adults 18+ years undergoing colorectal surgery.
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**
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*)**:** proprietary complex measure with fees 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 Accountability	**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:** Yes, fully developed and tested 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:** Colon and rectal resections constitute approximately 10% of all general surgery procedures performed in the United States accounting for over half a million procedures performed annually.[1] Estimates from the Centers for Medicare & Medicaid Serives estimate that colorectal surgery will exceed $77 million in 2009 and over 3,700 hospitals in the U.S. perform colorectal surgery. The most common indications for colon and rectal resections include cancer, diverticular disease, trauma, bowel infarction, and inflammatory bowel disease including Ulcerative Colitis or Crohn's disease. Large bowel cancer is the fourth most common malignancy in the United States, with more than 150,000 new cases in 2009. With 49,920 estimated deaths, it is second only to lung cancer as the leading cause of cancer-related deaths.[2]. Upwards of 80% of colon cancer patients are appropriate candidates for curative resection at presentation. Diverticular disease accounts for approximately 130,000 hospitalizations annually in the U.S. and poses significant cost to the health care system.[3] Though less than 10% of patients require emergent sigmoid resection for complications of acute diverticulitis,[4] a much higher proportion of patients undergo colon resection under elective settings. With the aging US population, there is an expectation that diverticular disease requiring surgically management as well as patients requiring colorectal surgery for oncologic procedures are on the rise.[5]	**1a** C☐ P☐ M☐ N☐

The postoperative complication rate for colon and rectal procedures approaches 30% ranking them among the most morbid of all surgical procedures. Failures of adherence to best practices in colorectal surgery are associated with increased complications.[6] Seventeen percent of patients who undergo colon resections have postoperative ileus and therefore have prolonged hospitalizations (13.8 days as opposed to 8.9 days without ileus, $P < 0.001$).[7] Ten to thirty percent of patients undergoing elective colorectal resections develop surgical site infections.[8, 9] There are an estimated 600,000 surgical site infections per year for major surgery in the United States, at an estimated cost of $1.8 billion.[10] The estimated additional costs per surgical site infection in published studies differ widely, from <$400 for superficial infections to $30,000 for serious intra-abdominal infections.[11] A study published in 2004 found that approximately half of surgical site infections were detected in the outpatient setting following discharge and accumulated a mean of $6200/patient of home health expenses related to wound care.[9] Another review of 1,127 patients undergoing elective colon resections found that those with postoperative deep and organ space infections had a longer hospitalizations as well as markedly higher costs (mean length of stay 21 days versus 6 days and and $42,516 versus $10,999, both $P < .001$).[12]

1a.4 Citations for Evidence of High Impact:

1. Owings, M.F. and L.J. Kozak, Ambulatory and inpatient procedures in the United States, 1996. Vital Health Stat 13, 1998(139): p. 1-119.
2. Jemal, A., et al., Cancer statistics, 2009. CA Cancer J Clin, 2009. 59(4): p. 225-49.
3. Munson, K.D., et al., Diverticulitis. A comprehensive follow-up. Dis Colon Rectum, 1996. 39(3): p. 318-22.
4. Stollman, N.H. and J.B. Raskin, Diagnosis and management of diverticular disease of the colon in adults. Ad Hoc Practice Parameters Committee of the American College of Gastroenterology. Am J Gastroenterol, 1999. 94(11): p. 3110-21.
5. Etzioni, D.A., et al., Impact of the aging population on the demand for colorectal procedures. Dis Colon Rectum, 2009. 52(4): p. 583-90; discussion 590-1.
6. Arriaga, A.F., et al., The Better Colectomy Project: Association of Evidence-Based Best-Practice Adherence Rates to Outcomes in Colorectal Surgery. Ann Surg, 2009.
7. Iyer, S., W.B. Saunders, and S. Stemkowski, Economic burden of postoperative ileus associated with colectomy in the United States. J Manag Care Pharm, 2009. 15(6): p. 485-94.
8. Prospero, E., et al., Surveillance for surgical site infection after hospital discharge: a surgical procedure-specific perspective. Infect Control Hosp Epidemiol, 2006. 27(12): p. 1313-7.
9. Smith, R.L., et al., Wound infection after elective colorectal resection. Ann Surg, 2004. 239(5): p. 599-605; discussion 605-7.
10. Bratzler, D.W., et al., Use of antimicrobial prophylaxis for major surgery: baseline results from the National Surgical Infection Prevention Project. Arch Surg, 2005. 140(2): p. 174-82.
11. Urban, J.A., Cost analysis of surgical site infections. Surg Infect (Larchmt), 2006. 7 Suppl 1: p. S19-22.
12. Eagye, K.J. and D.P. Nicolau, Deep and organ/space infections in patients undergoing elective colorectal surgery: incidence and impact on hospital length of stay and costs. Am J Surg, 2009. 198(3): p. 359-67.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: 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 success in achieving results with information similar to this outcome measure.

This risk-adjusted and benchmarked measure provides enormous motivation for hospitals to see their outcomes improve. A recent analysis (Hall BL, 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. 2009;250:363-376) 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. Other research has shown NSQIP improvements as well- many are referenced in the above citation.

1b
C ☐
P ☐
M ☐
N ☐

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:

Morbidity and mortality rates vary by institution for colorectal surgery. The 2006 to 2007 NSQIP dataset yielded 28,863 colorectal procedures at 182 hospitals. The overall morbidity rate was 24.3%, the serious morbidity rate was 11.4%, and the mortality rate was 3.9%.[13] An analysis of ACS NSQIP data calculated the risk-adjusted observed to expected (O/E) ratios for mortality and serious morbidity for patients undergoing colorectal resections using the methodology for the measure proposed herein. The results show that O/E ratios for mortality and serious morbidity range from 0 to 2.27 for all participating hospitals demonstrating a wide gap between those performing better and worse than expected after risk and case mix adjustment. The interquartile range for O/E ratios is 0.84-1.17, and the 10th percentile and 90th percentile O/E ratios were 0.64 and 1.39, respectively.

1b.3 Citations for data on performance gap:

13. Cohen, M.E., et al., 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): p. 1009-16.

1b.4 Summary of Data on disparities by population group:

There is wide variation in the care and outcomes of patients undergoing colorectal resections in the U.S. In 2003 McGlynn et al. showed stark deficits in the adherence to standard processes and quality of health care being delivered to the American public with only 55% of colorectal cancer patients receiving the recommended care.[14]. Obese patients have been shown to have higher rates of death due to cancer of the colon and rectum.[15]. In patients undergoing rectal resections, obesity is associated with a higher anastomotic leakage rates (16% versus 6% for non-obese patients, P <0. 05).[16]. Race also plays an important role in outcomes after colorectal resections. Black patients, compared with white patients, have lower 5-year overall survival rates after surgery for colon cancer (41.3% v 45.4%, respectively; P < .001)[17] and are less likely to receive adjuvant therapy after rectal cancer resection (48.6% versus 60.9%, p < 0.0001).[18] Compared to non-Hispanic whites, blacks, American Indians, Chinese, Filipinos, Koreans, Hawaiians, Mexicans, South/Central Americans, and Puerto Ricans are 10-60% more likely to be diagnosed with late stage colorectal cancer.[19] Hardiman et al. demonstrated through a retrospective review of prospectively collected data on 10,433 patients diagnosed with primary colon tumors that individuals who were at least 80 years old were less likely to have colectomy for advanced or metastatic disease, have fewer lymph nodes removed, receive chemotherapy for every stage than those who were younger than 80 years old.[20] Disparities in socioeconomic factors have also been identified for colorectal surgery. A study of 7,160 patients from Denmark, found that postoperative mortality after elective colorectal cancer surgery was significantly lower in patients with high income, higher education, and home ownership compared to home rental.[21]

1b.5 Citations for data on Disparities:

14. McGlynn, E.A., et al., The quality of health care delivered to adults in the United States. N Engl J Med, 2003. 348(26): p. 2635-45.
15. Calle, E.E., et al., Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med, 2003. 348(17): p. 1625-38.
16. Benoist, S., et al., Impact of obesity on surgical outcomes after colorectal resection. Am J Surg, 2000. 179(4): p. 275-81.
17. Breslin, T.M., et al., Hospital factors and racial disparities in mortality after surgery for breast and colon cancer. J Clin Oncol, 2009. 27(24): p. 3945-50.
18. Morris, A.M., et al., Racial disparities in late survival after rectal cancer surgery. J Am Coll Surg, 2006. 203(6): p. 787-94.
19. Chien, C., et al., Differences in colorectal carcinoma stage and survival by race and ethnicity. Cancer, 2005. 104(3): p. 629-39.
20. Hardiman, K.M., et al., Disparities in the treatment of colon cancer in octogenarians. Am J Surg, 2009. 197(5): p. 624-8.
21. Frederiksen, B.L., et al., The impact of socioeconomic factors on 30-day mortality following elective colorectal cancer surgery: a nationwide study. Eur J Cancer, 2009. 45(7): p. 1248-56.

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*)**:** This is a risk-adjusted outcome measure. Approximately 500,000 inpatient colorectal procedures are performed annually in the United States. Much of the excess length of stay (LOS), charges, and death can be attributed to postoperative complications.[22] The LOS for patients who have any postoperative complication is 3–11 days longer than that for patients who do not experience complications.[22, 23] In an analysis of data from the Veterans Administration (VA) NSQIP, the occurrence of a complication 30 days in duration reduced median patient survival by 69% independent of preoperative patient risk.[24]

Higher hospital volume has not been correlated with improved outcomes for colorectal surgery as is the case for pancreatic and esophageal surgeries.[25, 26] Nevertheless, higher surgeon volume has been correlated with improved outcomes with lower volume surgeons benefiting from performance of colorectal procedures at high volume hospitals.[27, 28] These findings suggest that hospital level improvements in quality of care can influence postoperative outcomes in colorectal surgery regardless of hospital volume.

Ultimately, a variety of process measures exist that correlate to improved outcomes in colorectal surgery but there are no guidelines tied directly to outcome measurement. The NSQIP program has shown that a composite outcome that includes serious morbidity and mortality for patients undergoing colorectal surgery is measureable. Current reports provided back to individual hospitals based on colorectal surgerical models include 30-day mortality, 30-day morbidity, length of stay, and surgical site infections.

22. 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.
23. Coello, R., et al., Adverse impact of surgical site infections in English hospitals. J Hosp Infect, 2005. 60(2): p. 93-103.
24. Khuri, S.F., et al., Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. Ann Surg, 2005. 242(3): p. 326-41; discussion 341-3.
25. Birkmeyer, J.D., et al., Volume and process of care in high-risk cancer surgery. Cancer, 2006. 106(11): p. 2476-81.
26. Finlayson, E.V. and J.D. Birkmeyer, Effects of hospital volume on life expectancy after selected cancer operations in older adults: a decision analysis. J Am Coll Surg, 2003. 196(3): p. 410-7.
27. Karanicolas, P.J., et al., The more the better?: the impact of surgeon and hospital volume on in-hospital mortality following colorectal resection. Ann Surg, 2009. 249(6): p. 954-9.
28. Harmon, J.W., et al., Hospital volume can serve as a surrogate for surgeon volume for achieving excellent outcomes in colorectal resection. Ann Surg, 1999. 230(3): p. 404-11; discussion 411-3.

1c.2-3. Type of Evidence: observational study, cohort study

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Observational evidence is based on prospectively collected rigorously controlled data (ACS NSQIP). As noted, a variety of processes influence outcomes but no guidelines address outcome measurement as a means toward quality improvement. McGlynn et al. showed stark deficits in the adherence to standard processes and quality of health care being delivered to the American public with only 55% of colorectal cancer patients receiving the recommended care.[14] The morbidity and mortality associated with colorectal operations are common and pose considerable risk to patients.[29] Several statistical models have been proposed to predict surgical risk for these operations.[30-34] Based on NSQIP data from 2006-2007, the overall morbidity rate for colorectal surgery was 24.3%, the serious morbidity rate was 11.4%, and the mortality rate was 3.9% and prompted the creation of a colorectal risk calculator.[13]

Several Surgical Care Improvement Project (SCIP) measures have been developed and have been tied to outcomes specific to colorectal surgery.[35] SCIP measures are related to decreasing infections, specifically postoperative wound infections, urinary tract infections, and pneumonia. SCIP-Inf-1, -2, and -3 addresses antibiotic use within one hour prior to surgical incision for colon surgery, appropriate antibiotic selection,

1c
C☐
P☐
M☐
N☐

discontinuation of antibiotics within 24 hours of surgery end time. Appropriate antibiotic use in colorectal surgery has been studied in several randomized controlled trails and meta-analysis.[36] SCIP-Inf-10 addresses normothermia. Compliance with these measures along with increased attention at maintaining normoglycemia have decreased surgical site infection rates.[37] SCIP-VTE-1 and -2 address appropriate venous thromboembolism prophylaxis use and timing. Abdominal surgery for colorectal cancer has been classified as a high-risk procedure for VTE and thromboprophylaxis is strongly recommended.[38]

Despite identification of several process measures for colorectal surgery, adherence to these measures alone is rapidly achievable while complications and poor outcomes remain a significant issue.[39, 40] Following outcomes as an end point in quality improvement may fuel identification of novel process measures as well as provide a more clinically meaningful endpoint for patients to follow.

There are few randomized controlled trials comparing overall outcomes in colorectal surgery beyond small studies examining the influence of antibiotic prophylaxis on individual outcomes such as wound infection. Hence, there are few process measures developed specific to colectomy alone. This is evidenced further by two articles that provide a comprehensive review of colorectal surgical quality.[41, 42] In both articles, the lack of level 1/trail based evidence forced expert opinion and a systematic review of the literature were used to identify candidate quality indicators. More specifically, Gagliardi et al. used a three-step modified Delphi approach, whereas McGory et al. used the RAND/University of California-Los Angeles appropriateness method.[43, 44] Nevertheless, these reviews of the best available literature focus on various process measures and do not provide feedback about a composite outcome measure based on colorectal surgery.

Gagliardi and colleagues identified 45 candidate indicators, of which 37 (82%) were considered valid by the panel. McGory and colleagues identified 142 candidate indicators, of which 92 (65%) were considered valid. In the former study, panelists were asked to rank candidate indicators; the investigators reported only the top 15 prioritized quality indicators as their final recommendation for improving the quality of colorectal cancer surgery. Although such reporting does present a parsimonious set of quality indicators, the detailed list of 92 quality indicators presented in the latter study is comprehensive and encompasses the entire perioperative time period, including preparation of the patient for surgery, intraoperative issues, and postoperative processes of care. In that regard, the focus of the quality indicators is an interesting difference between the two studies.

Gagliardi and colleagues focused on four outcome measures (e.g., 30-day mortality) and four measures evaluated at the province level (e.g., 5-year survival). In contrast, McGory and colleagues focused on process and structural measures and did not include any outcome measures. The latter investigators presented six quality domains spanning the 92 indicators: surgeon privileging (e.g., credentialing for laparoscopic colectomy), preoperative evaluation (e.g., staging), patient-provider discussions (e.g., informed consent), medications (e.g., antibiotic prophylaxis), intraoperative care (e.g., prevention of ureteral injury), and postoperative management (e.g., control of blood glucose). In addition to focusing on process rather than outcomes, all of their measures were recorded at the provider level, not the hospital, county, or state level. It seems that the two studies developed indicators with different agendas in mind, and as such, the potential application of these studies for quality improvement will be broad.

Though comprehensive, these reviews of available literature focus on processes of care in order to influence outcome.

As already noted above, Hall et al (2009), showed that using outcomes feedback analagous to this proposed measure, 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. Other research has shown NSQIP improvements as well- many are referenced in the above citation.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
No Level 1A/Randomized Controlled Trials for overall colorectal surgery outcomes. This measure is developed based on prospectively collected rigorously controlled data. Curerent literature is primarily level 2 and 3 evidence being validated by multiple expert panels with trials exploring several process related compnents.

1c.6 Method for rating evidence: RAND/UCLA appropriateness methodology.[43, 44]

1c.7 Summary of Controversy/Contradictory Evidence: Colon and rectal surgeries are a common procedure group that are performed at a high number of hospitals across the U.S., and are associated with significant morbidity and mortality. Current process measures are not specific to colorectal surgery. Current process measures do not correlate with clinical risk adjusted outcomes as discussed elsewhere in this measure proposal. Specific process measures for colectomy do not have high enough evidence base for accountability. With this measure we propose to evaluate colorectal surgery quality of care using a novel, clinical, risk-adjusted outcomes measure. As described, several process measures exist and additional groups and panels have described outcomes of interest in colorectal surgery. While further studies are needed to identify additional indicators and those that are most important to monitor on a regular basis, no true controversy or contradictory evidence currently exists when observing a composite outcome that spans a wide array of clinically significant outcomes.

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

14. McGlynn, E.A., et al., The quality of health care delivered to adults in the United States. N Engl J Med, 2003. 348(26): p. 2635-45.
29. Schilling, P.L., J.B. Dimick, and J.D. Birkmeyer, Prioritizing quality improvement in general surgery. J Am Coll Surg, 2008. 207(5): p. 698-704.
30. Bromage, S.J. and W.J. Cunliffe, Validation of the CR-POSSUM risk-adjusted scoring system for major colorectal cancer surgery in a single center. Dis Colon Rectum, 2007. 50(2): p. 192-6.
31. Tekkis, P.P., et al., Development of a dedicated risk-adjustment scoring system for colorectal surgery (colorectal POSSUM). Br J Surg, 2004. 91(9): p. 1174-82.
32. Duval, H., et al., [The Association Francaise de Chirurgie (AFC) colorectal index: a reliable preoperative prognostic index in colorectal surgery]. Ann Chir, 2006. 131(1): p. 34-8.
33. Alves, A., et al., The AFC score: validation of a 4-item predicting score of postoperative mortality after colorectal resection for cancer or diverticulitis: results of a prospective multicenter study in 1049 patients. Ann Surg, 2007. 246(1): p. 91-6.
34. Fazio, V.W., et al., Assessment of operative risk in colorectal cancer surgery: the Cleveland Clinic Foundation colorectal cancer model. Dis Colon Rectum, 2004. 47(12): p. 2015-24.
35. Surgical Care Improvement Project Core Measure Set Available at http://www.jointcommission.org/PerformanceMeasurement/PerformanceMeasurement/SCIP+Core+Measure+Set.htm 2008 [cited January 26, 2009.
36. Woodfield, J.C., N. Beshay, and A.M. van Rij, A Meta-Analysis of Randomized, Controlled Trials Assessing the Prophylactic Use of Ceftriaxone. A Study of Wound, Chest, and Urinary Infections. World J Surg, 2009.
37. Kurz, A., D.I. Sessler, and R. Lenhardt, Perioperative normothermia to reduce the incidence of surgical-wound infection and shorten hospitalization. Study of Wound Infection and Temperature Group. N Engl J Med, 1996. 334(19): p. 1209-15.
38. Geerts, W.H., et al., Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest, 2008. 133(6 Suppl): p. 381S-453S.
39. Forbes, S.S., et al., Implementation of evidence-based practices for surgical site infection prophylaxis: results of a pre- and postintervention study. J Am Coll Surg, 2008. 207(3): p. 336-41.
40. Hedrick, T.L., et al., Efficacy of protocol implementation on incidence of wound infection in colorectal operations. J Am Coll Surg, 2007. 205(3): p. 432-8.
41. McGory, M.L., P.G. Shekelle, and C.Y. Ko, Development of quality indicators for patients undergoing colorectal cancer surgery. J Natl Cancer Inst, 2006. 98(22): p. 1623-33.
42. Gagliardi, A.R., et al., Development of quality indicators for colorectal cancer surgery, using a 3-step modified Delphi approach. Can J Surg, 2005. 48(6): p. 441-52.
43. Fink, A., et al., Consensus methods: characteristics and guidelines for use. Am J Public Health, 1984. 74(9): p. 979-83.
44. Brook, R.H., The RAND/UCLA Appropriateness Method, in Methodology perspectives, K.A. McCormick, S.R. Moore, and R.A. Siegel, Editors. 1994, Public Health Services, U.S. Department of Health and Human Services: Rockville, MD. p. 59-70.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
While quality indicators for patients undergoing colorectal surgical procedures exist (as described above),

there are no national clinical practice guidelines specific for colorectal surgical procedures (http://www.guideline.gov/; accessed 9/10/09). **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	
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 30-day, hospital-specific risk-adjusted (all cause) mortality, a return to the operating room, or any of the following morbidities as defined by American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP): Cardiac Arrest requiring CPR, Myocardial Infarction, DVT requiring therapy, Sepsis, Septic Shock, Deep Incisional ssi, Organ/Space SSI, Wound Disruption, Unplanned Reintubation without prior ventilator dependence, Pneumonia without pre-operative pneumonia, Pulmonary Embolism, progressive Renal Insufficiency or Acute Renal Failure without pre-operative renal failure or dialysis, or UTI. All outcomes are definitively resolved within 30 days of any ACS NSQIP listed (CPT) surgical procedure. All variables (fields) are explicitly defined in the tradition of the ACS NSQIP and definitions are also submitted in these materials. The current set of mortality and major complications for this measure was chosen based on prior work revealing that these complications are related to other important criteria such as large contributions to excess length of stay, large complication burdens, or correlations with mortality. Of note, the measure does specifically include return to the operating room within 30 days as a dependent outcome. In addition, the desire to limit the outcomes to significant events (ie- some degree of severity according to certain criteria) is the reason that superficial wound infection is excluded from the measure. **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 index operation are included. **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes,*	**2a-specs** **C☐** **P☐** **M☐** **N☐**

logic, and definitions):
Mortality- "All cause" Death within 30 day follow-up period: Any death occurring through midnight on the 30th day after the date of the procedure, regardless of cause, in or out of the hospital.

All other outcome fields also defined explicitly in the tradition of ACS NSQIP:

Return to the Operating Room within Thirty Days after the Assessed Procedure: Return to the operating room includes all major surgical procedures that required the patient to be taken to the surgical operating room for intervention of any kind. "Major surgical procedures" are defined as those cases in any and all surgical subspecialties that meet Program criteria for inclusion.

Cardiac Arrest Requiring CPR: The absence of cardiac rhythm or presence of chaotic cardiac rhythm that results in loss of consciousness requiring the initiation of any component of basic and/or advanced cardiac life support. Patients with automatic implantable cardioverter defibrillator (AICD) that fire but the patient has no loss of consciousness should be excluded.

Myocardial Infarction: An acute myocardial infarction occurring within 30 days following surgery as manifested by one of the following three criteria:
a. Documentation of ECG changes indicative of acute MI(one or more of the following):
- ST elevation > 1 mm in two or more contiguous leads
- New left bundle branch
- New q-wave in two of more contiguous leads

b. New elevation in troponin greater than 3 times upper level of the reference range in the setting of suspected myocardial ischemia
c. Physician diagnosis of myocardial infarction.

Deep Vein Thrombosis (DVT)/Requiring Therapy: The identification of a new blood clot or thrombus within the venous system, which may be coupled with inflammation. This diagnosis is confirmed by a duplex, venogram or CT scan. The patient must be treated with anticoagulation therapy and/or placement of a vena cava filter or clipping of the vena cava.

Sepsis: Sepsis is the systemic response to infection. Report this variable if the patient has TWO OR MORE of the following five clinical signs and symptoms of Systemic Inflammatory Response Syndrome (SIRS):
a. Temp >38 degrees C (100.4 degrees F) or < 36 degrees C (96.8 degrees F)
b. HR >90 bpm
c. RR >20 breaths/min or PaCO2 <32 mmHg(<4.3 kPa)
d. WBC >12,000 cell/mm3, <4000 cells/mm3, or >10% immature (band) forms
e. Anion gap acidosis: this is defined by either:
- [Na + K] – [Cl + HCO3 (or serum CO2)]. If this number is greater than 16, then an anion gap acidosis is present.
- Na – [Cl + HCO3 (or serum CO2)]. If this number is greater than 12, then an anion gap acidosis is present.

AND one of the following TWO:
a. positive blood culture
b. clinical documentation of purulence or positive culture from any site thought to be causative

Severe Sepsis/Septic Shock: Sepsis is considered severe when it is associated with organ and/or circulatory dysfunction. Report this variable if the patient has sepsis AND documented organ and/or circulatory dysfunction. Examples of organ dysfunction include: oliguria, acute alteration in mental status, acute respiratory distress. Examples of circulatory dysfunction include: hypotension, requirement of inotropic or vasopressor agents. Severe Sepsis/Septic Shock is assigned when it appears to be related to Sepsis and not a Cardiogenic or Hypovolemic etiology.

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.

Wound Disruption: Separation of the layers of a surgical wound, which may be partial or complete, with disruption of the fascia.

Unplanned Intubation for Respiratory/Cardiac Failure (without preoperative ventilator dependent): Patient required placement of an endotracheal tube and mechanical or assisted ventilation because of the onset of respiratory or cardiac failure manifested by severe respiratory distress, hypoxia, hypercarbia, or respiratory acidosis. In patients who were intubated for their surgery, unplanned intubation occurs after they have been extubated after surgery. In patients who were not intubated during surgery, intubation at any time after their surgery is considered unplanned.

Pneumonia (without preoperative pneumonia): if the patient has pneumonia meeting the definition below AND pneumonia was not present preoperatively. Patients with pneumonia must meet criteria from both Radiology and Signs/Symptoms/Laboratory sections listed as follows:

Radiology: One definitive chest radiological exam (x-ray or CT) with at least one of the following: New or progressive and persistent infiltrate, Consolidation or opacity, Cavitation. In patients with underlying pulmonary or cardiac disease (e.g. respiratory distress syndrome, bronchopulmonary dysplasia, pulmonary edema, or chronic obstructive pulmonary disease), two or more serial chest radiological exams (x-ray or CT) are required.

Signs/Symptoms/Laboratory

FOR ANY PATIENT, at least one of the following three:
a. Fever (>38 degrees C or >100.4 degrees F) with no other recognized cause
b. Leukopenia (<4000 WBC/mm3) or leukocytosis(=12,000 WBC/mm3)
c. For adults = 70 years old, altered mental status with no other recognized cause

AND
At least one of the following four:

a. 5% Bronchoalveolar lavage (BAL) -obtained cells contain intracellular bacteria on direct microscopic exam (e.g., Gram stain)
b. Positive growth in blood culture not related to another source of infection
c. Positive growth in culture of pleural fluid
d. Positive quantitative culture from minimally contaminated lower respiratory tract (LRT) specimen (e.g. BAL or protected specimen brushing)

OR
At least two of the following four:
a. New onset of purulent sputum, or change in character of sputum, or increased respiratory secretions, or increased suctioning requirements
b. New onset or worsening cough, or dyspnea, or tachypnea
c. Rales or bronchial breath sounds
d. Worsening gas exchange (e.g. O2 desaturations (e.g., PaO2/FiO2 = 240), increased oxygen requirements, or increased ventilator demand)

Pulmonary Embolism: Lodging of a blood clot in a pulmonary artery with subsequent obstruction of blood supply to the lung parenchyma. The blood clots usually originate from the deep leg veins or the pelvic venous system. Pulmonary embolism is recorded if the patient has a V-Q scan interpreted as high probability of pulmonary embolism or a positive CT spiral exam, pulmonary arteriogram or CT angiogram. Treatment usually consists of: Initiation of anticoagulation therapy, Placement of mechanical interruption (for example Greenfield Filter), for patients in whom anticoagulation is contraindicated or already instituted.

Progressive Renal Insufficiency (without preoperative renal failure or dialysis): The reduced capacity of the kidney to perform its function as evidenced by a rise in creatinine of >2 mg/dl from preoperative value, but with no requirement for dialysis.

Acute Renal Failure Requiring Dialysis (without preoperative renal failure or dialysis): In a patient who did not require dialysis preoperatively, worsening of renal dysfunction postoperatively requiring hemodialysis, peritoneal dialysis, hemofiltration, hemodiafiltration, or ultrafiltration.

Urinary Tract Infection: Postoperative symptomatic urinary tract infection must meet ONE of the following TWO criteria:
Criterion One:
One of the following five:
a. fever (>38 degrees C),
b. urgency,
c. frequency,
d. dysuria,
e. suprapubic tenderness
AND a urine culture of > 100,000 colonies/ml urine with no more than two species of organisms.

OR

Criterion Two:
Two of the following five:
a. fever (>38 degrees C),
b. urgency,
c. frequency,
d. dysuria,
e. suprapubic tenderness
AND ANY ONE or MORE of the following seven:
f. Dipstick test positive for leukocyte esterase and/or nitrate,
g. Pyuria (>10 WBCs/mm3 or > 3 WBC/hpf of unspun urine),
h. Organisms seen on Gram stain of unspun urine,
i. Two urine cultures with repeated isolation of the same uropathogen with >100 colonies/ml urine in non-voided specimen,
j. Urine culture with < 100,000 colonies/ml urine of single uropathogen in patient being treated with appropriate antimicrobial therapy,
k. Physician's diagnosis,
l. Physician institutes appropriate antimicrobial therapy.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
Patients undergoing any ACS NSQIP listed (primary CPT) colorectal surgical procedure. (44140, 44141, 44143, 44144, 44145, 44146, 44147, 44150, 44151, 44155, 44156, 44157, 44158, 44160, 44204, 44205, 44206, 44207, 44208, 44210, 44211, 44212, 45110, 45111, 45112, 45113, 45114, 45116, 45119, 45120, 45121, 45123, 45126, 45130, 45135, 45160, 45395, 45397, 45402, 45550)
Notes: following codes are not included in this denominator list: 44152 (not found), 44153 (not found), 44239 (not found), 45540 (proctopexy without resection), 45499 (unlisted laparoscopy, rectum).

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. See also exclusions below.

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** As noted above, cases are collected so as to match ACS NSQIP inclusion and exclusion criteria, thereby permitting valid application of ACS NSQIP model-based risk adjustment. Therefore, trauma and transplant surgeries are excluded as are surgeries not on the ACS NSQIP CPT list as eligible for selection (see details in next item). Patients who are ASA 6 (brain-death organ donor) are not eligible surgical cases. Of note, the measure excludes patients identified as having had prior surgical procedures within 30 days of a potential index procedure, since this measure is based on 30 day outcomes. A patient who is identified as having had a prior surgical procedure within 30 days of the index case being considered is excluded from accrual. A patient who has a second surgical procedure performed within 30 days after an index procedure has the second procedure recorded as a "Return to the operating room within 30 days" (one of the outcomes defined), but the second procedure cannot be accrued into the program as a new index procedure.

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 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, though any operation performed after the patient has been discharged from the transplant stay is eligible for selection. Donor procedures on living donors are not excluded unless meeting other exclusion criteria. 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.
As noted above, the measure excludes patients identified as having had prior surgical procedures within 30 days of a potential index procedure, since this measure is based on 30 day outcomes. A patient who is identified as having had a prior surgical procedure within 30 days of the index case being considered is excluded from accrual. A patient who has a second surgical procedure performed within 30 days after an index procedure has the second procedure recorded as a "Return to the operating room within 30 days" (one of the outcomes defined), but the second procedure cannot be accrued into the program as a new index procedure.

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 this risk-adjusted measure.
Note: if an implementation required stratification by race or ethnicity post-hoc, then race/ethnicity variables could be added to the implementation with no other changes necessary under 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 ; 21,694 acceptable records from 211 hospitals (mean/hospital=103) 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 “Risk" groups. These categorizations have been defined and implemented for risk adjustment in previously published research.* Missing variables within the ACS NSQIP framework are traditionally handled by imputation, generally invoked mainly for laboratory variables since case inclusion typically requires complete data (For a discussion of imputation issues within the program approach see J Am Coll Surg 2010;210:125-139).

An outcome was defined as 30-day mortality or any serious morbidity including: cardiac arrest requiring CPR, myocardial infarction, DVT requiring therapy, sepsis, septic shock, organ space SSI, deep incisional SSI, wound disruption, unplanned reintubation without prior ventilator dependence, pneumonia without pre-operative pneumonia, pulmonary embolism, progressive renal insufficiency or acute renal failure without pre-operative renal failure or dialysis, urinary tract infection, or return to the operating room, according to ACS NSQIP definitions. Of the 21,694 patients, 4,862 (22.4%) experienced death or a serious morbidity event.

To control for procedure-specific effects, CPT code was originally considered a categorical variable but, to maintain methodological consistency with other proposed measures, CPT code was converted to a continuous scalar risk variable: "CPT Risk". This was accomplished by making the categorical CPT code variable a single predictor for mortality/morbidity and invoking the Firth penalized likelihood method in the logistic modeling software (SAS PROC LOGISTIC). The patient-based predicted log odds from this model for each CPT code was then used as a continuous predictor in subsequent logistic models which also included all other specified risk predictors. The result is that the scalar "CPT Risk" variable included in the subsequent regressions provides a very high level of control for "procedure" or "procedure mix" within the measure. This alleviates the majority of concern over the measure being dominated by unique, procedure-specific effects. This control is further enhanced by the limited CPT code set for the measure focusing on colon and rectal surgery.

Step-wise logistic regression ($P<0.05$ for inclusion), which selected from a total of 26 NSQIP predictors, identified 20 predictors for inclusion in the model. In order of inclusion these variables were: ASA Class, pre-operative Functional Status, Indication, Log Odds CPT (CPT Risk), Emergent, Wound Class, Dyspnea, Weight Loss, Steroid Use, Smoking, Disseminated Cancer, History of COPD, Ascites, Hypertension, Ventilator Dependent, Age Group, Radio Therapy, Alcohol Use, Bleeding Disorder, and Previous Vascular Event/Disease. The c-statistic was 0.738 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 six selected variables (ASA Class, pre-operative Functional Status, Indication, Log Odds CPT (CPT Risk), Emergent, and Wound Class), the c-statistic was 0.727 and the Hosmer-Lemeshow was 0.177). The use of these six predictors for modeling was further evaluated. Using a 95% confidence interval for the ratio of observed to expected events (O/E), this six-variable logistic model identified 16 statistical outliers (10 low outliers and 6 high outliers). When the same six 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), 17 outliers were detected (11 low outliers and 6 high outliers). Thus, using a 95% confidence interval, logistic and hierarchical models identified 3% of hospitals as high outliers. When the logistic model parameters were applied to an independent validation data set (the 2007 Data file composed of 18,098 patients) after coding CPT Risk with log odds derived from the original 1-variable model on 2008 data, the c-statistic was essentially unchanged (c-statistic=0.721).

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 0.0106. 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 41, 63, 94, 141, and 219, respectively.

For the table detailing risk factors, odds ratios, and parameters for the logistic model, please see attachment (Parsimonious Model for Colorectal.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 Colorectal.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. See also the risk adjustment methodology section.

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 greater than, and does not include 1.0, the hospital is identified as having performance significantly worse than expected. If the interval is less than, and does not include 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: estimated sample size for reliability 0.4 ~=63 cases), 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, Electronic clinical data, electronic Health/Medical Record, lab data, paper medical record/flowsheet

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.***):**
Model is based on historical ACS NSQIP Data file. Data sources are as above- collection is consistent with historical ACS NSQIP approaches to data collection. Model is based on ACS NSQIP but measure would not require participation in ACS NSQIP.

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, Population: national, Population: regional/network, Population: states

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*
Hospital, Ambulatory Care: Hospital Outpatient

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. Data sample for hospitals would be one year sampling according to systematic algorithm and sample and reliability information supplied: e.g. reliability of 0.4 would require roughly 63 cases/annum.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)***:**
See Risk-adjustment Methodology in Specifications for hierarchical risk adjustment model, incorporating procedure risk score. 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 mortality/serious morbidity in patients =65. 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	INTEPRETATION
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 0.0106. Using a minimum acceptable reliability for mortality/serious morbidity in patients =65 of 0.4, the proportions of hospitals likely to have a "minimally acceptable" reliability estimate are as follows. 42.9% of all U.S. hospitals and 68.7% of ACS NSQIP hospitals meet the 0.4 reliability requirement. These ~40% of US hospitals perform roughly 85% of all colectomies in the country. This level of reliability is comparable to or exceeds published figures for other approved measures- the ACS provides this reliability data on all submitted measures despite the fact that many measure developers do not submit comparable data. Furthermore, it is also expected that as the population and diversity of institutions participating in this measure increases, the reliability will increase as well- making our initial estimate a conservative one. [This is related to the ACS NSQIP having some bias toward larger academic institutions.] Furthermore, we do provide in our results below information on increasing the reliability by increasing the sample size, which would be considered for any implementation. However, there will always be a trade-off between drafting a measure with higher reliability but having it apply to fewer institutions (since requiring

2b
C☐
P☐
M☐
N☐

increasing sample size will exclude more and more institutions).

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	41	55.8	79.6
0.4	63	42.9	68.7
0.5	94	31.6	53.5
0.6	141	20.1	27.5
0.7	219	9.6	3.8

* Based on volume data from the 2005 National Inpatient Survey and inflated to account for outpatient procedures.
+ 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. See 2b1 above.

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. For this measure, only the denominator codes submitted above (Measure specifications- denominator statement) are eligible and all others are excluded. 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). In addition, these special procedures have extensive documentation of risk assessment evidence and approaches that are outside of the current scope of ACS NSQIP. Procedures within 30 days prior to index procedure are excluded to reduce risk adjustment challenges associated with early reoperation and to eliminate the dilemma of assigning outcomes to different procedures. Procedures subsequent to an index procedure, within 30 days, are treated only as "reoperation" outcome events and are excluded from consideration as new index events (see Measure specifications- denominator exclusions).

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)*: same

2d
C☐ P☐ M☐ N☐ NA☐

2d.4 Analytic Method *(type analysis & rationale)*:
same

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
same

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 (2008). The Colorectal model used 21,694 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. See all other details on risk adjustment described elsewhere above, including generation of CPT risk score, above (Measure specifications- risk adjustment methodology) and following (2e3).

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 20 predictors for inclusion in the model. In order of inclusion these variables were: ASA Class, pre-operative Functional Status, Indication, (Log Odds CPT) "CPT Risk", Emergent, Wound Class, Dyspnea, Weight Loss, Steroid Use, Smoking, Disseminated Cancer, History of COPD, Ascites, Hypertension, Ventilator Dependent, Age Group, Radio Therapy, Alcohol Use, Bleeding Disorder, and Previous Vascular Event/Disease. The c-statistic was 0.738 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 six selected variables (ASA Class, pre-operative Functional Status, Indication, (Log Odds CPT) "CPT Risk", Emergent, and Wound Class), the c-statistic was 0.727 and the Hosmer-Lemeshow was 0.177). The use of these six predictors for modeling was further evaluated. Using a 95% confidence interval for the ratio of observed to expected events (O/E), this six-variable logistic model identified 16 statistical outliers (10 low outliers and 6 high outliers). When the same six 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), 17 outliers were detected (11 low outliers and 6 high outliers). Thus, using a 95% confidence interval, logistic and hierarchical models identified 3% of hospitals as high outliers. See additional data on reliability and sample size estimation provided above (Scientific Acceptability- reliability testing).

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

2e
C☐
P☐
M☐
N☐
NA☐

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 Strategy Data Sample Section. See also "Importance-Opportunity for Improvement" for data indicating interquartile range of 0.84-1.17 for O/E- a difference in performance of roughly 40% from 25th to 75th percentiles.

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

2f
C☐
P☐
M☐
N☐

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. 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)*: See above	**2g** C☐ P☐ M☐ N☐ NA☐
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. As mentioned above, if straticiation by race or ethnicity is required in implementation then those variables can easily be added to the data specification without changing any other specification of the model. Race and ethnicity are not utilized in risk adjustment, per NQF guidelines. 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*	**3a** C☐ P☐ M☐ N☐

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 being used as a public reporting initiative. Currently results are reported confidentially only to participating institutions, currently numbering ~250, which use the information to drive quality improvement.

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 (https://acsnsqip.org/login/default.aspx)

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.5 **Methods** *(e.g., focus group, survey, QI project)***:**
Analyses of differences in performance O/E ratios and longitudinal changes in O/E ratios.

3a.6 **Results** *(qualitative and/or quantitative results and conclusions)***:**
See above section on 'Testing of interpretability".

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

3b.1 **NQF # and Title of similar or related measures:**

(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?**

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:**

3c
C☐ P☐ M☐

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: No competing measure identified	**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 implemented for performance measurement. (evaluation criteria)	Eval 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, other dedicated abstraction personnel	**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 would be needed to capture all 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**	**4e** **C☐** **P☐** **M☐** **N☐**

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 measures*)**:**
Using a conservative estimate, approximately .125 to .333 of a FTE might be needed to collect the data for the measure. The requirement might realistically be as low as 0.05 FTE based on ACS NSQIP experience. 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 65+ cases.

4e.4 **Business case documentation:** Business case has not been developed for this measure; however, literature results show that the costs for each post surgical complication can reach up to $57,000.

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 Surgeon | 633 N. Saint Clair St, 22nd Floor | Chicago | Illinois | 60611-3211

Co.2 **Point of Contact**

Karen | Richards, Administrative Director, Division of Research and Optimal Patient Care | krichards@facs.org | 312-202-5011

Measure Developer If different from Measure Steward
Co.3 Organization
American College of Surgeon | 633 N. Saint Clair St, 22nd Floor | Chicago | Illinois | 60611-3211

Co.4 Point of Contact
Karen | Richards, Administrative Director, Division of Research and Optimal Patient Care | krichards@facs.org | 312-202-5011

Co.5 Submitter If different from Measure Steward POC
Karen | Richards, Administrative Director, Division of Research and Optimal Patient Care | krichards@facs.org | 312-202-5011- |American College of Surgeon

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.
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:

Ad.11 -13 Additional Information web page URL or attachment: Attachment Updated Conditions Addendum-633888776131131575.doc

Date of Submission (*MM/DD/YY*)**:** 03/23/2010

Parsimonious Hierarchical Model for Mortality or Serious Morbidity Following Colorectal Surgery Data from 1/1/2008 – 12/31/2008 (2008 Data File)					
Predictor Set	Risk Factor	Odds Ratio	Lower 95% Boundary	Upper 95% Boundary	Parameter Estimate
	Intercept				-1.8542
1	ASA Class: 3-Severe Disturbance vs. 1/2-No/Mild Disturbance 4/5- Life Threatening/Moribund vs. 1/2-No/Mild Disturbance	 1.815 3.929	 1.676 3.453	 1.966 4.470	 0.5962 1.3683
2	Preoperative Functional Status: Partially Dependent vs. Independent Totally Dependent vs. Independent	 1.960 3.076	 1.733 2.551	 2.217 3.708	 0.6730 1.1236
3	Indication: Enteritis/Colitis vs. Diverticulitis Hemorrhage vs. Diverticulitis Neoplasm vs. Diverticulitis Obstruction/Perforation vs. Diverticulitis Other vs. Diverticulitis Rectal Prolapse vs. Diverticulitis Vascular Insufficiency vs. Diverticulitis Volvulus vs. Diverticulitis	 1.833 1.548 1.315 2.022 1.428 0.867 2.030 1.199	 1.566 1.184 1.167 1.715 1.250 0.650 1.606 0.912	 2.146 2.025 1.481 2.384 1.632 1.158 2.565 1.577	 0.6061 0.4370 0.2738 0.7042 0.3563 -0.1424 0.7078 0.1814
4	Log Odds CPT Group (continuous)	1.603	1.473	1.745	0.4721
5	Emergent: Yes vs. No	 1.477	 1.327	 1.644	 0.3902
6	Wound Class: 3/4 Contaminated/Dirty/Infected vs. 1/2 Clean/Clean Contaminated	 1.392	 1.273	 1.523	 0.3308
Observations = 21,694; Hospitals = 211; Patients with Mortality or Serious Morbidity event = 4,862, Rate = 22.4%, (logistic) c-statistic = 0.727, HL = 0.177.					