NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 0301 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Surgery patients with appropriate hair removal
De.2 **Brief description of measure:** Percentage of surgery patients with surgical hair site removal with clippers or depilatory or no surgical site hair removal.
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** N/A
De.4 **National Priority Partners Priority Area:** Safety De.5 **IOM Quality Domain:** Safety De.6 **Consumer Care Need:** Staying healthy

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary 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	**B** Y☐

every 3 years. Yes, information provided in contact section	N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Accountability, Payment incentive, Accreditation	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, Frequently performed procedure, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** There are over 40 million surgeries performed in the United States each year. Surgical site infection (SSIs) are the second most common cause of healthcare associated infections.SSIs account for 14-16% of all hospital-acquired infections and are among the most common complications of care, occurring in 2 to 5% of patients after clean extra-abdominal operations and up to 20 % of intra-abdominal procedures. Among surgical patients, SSIs account for 40% of all such hospital-acquired infections. By reducing SSIs, hospitals on average could recognize a savings of $3.152 and a reductions in extended length of stay by seven days on each patient developing an infection. 1a.4 **Citations for Evidence of High Impact:** Selected References: Zhan C, Miller MR. Excess length of stay, charges and mortality attributable to medical injuries during hospitalization. JAMA 2003; 290: 1868-1874. Delgado-Rodriguez M, Sillero-Arenas M, Medina-Cuadros M, Martinez-Gallego G. Nosocomial infections in surgical patients: comparison of two measures of intrinsic patient risk. Infect Control Hosp Epidemiol 1997; 18: 19-23. Polk HC, Christmas AB. Prophylactic antibiotics in surgery and surgical wound infections. Am Surg 200; 66: 105-111.	1a C☐ P☐ M☐ N☐

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: A reduction in the number of patients having hair removal performed by shaving (razors) may reduce the incidence of surgical site infection.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
In a national sample of 19,497 Medicare patients undergoing surgery in US hospitals in the first quarter of 2005, the rate of appropriate hair removal was 91.5%. In the second quarter of 2010 (most recent data available), the rate was 99.6%.

1b.3 Citations for data on performance gap:
Number of hospitals reporting: 3629
Denominator:401,573
Numerator:399,933

1b.4 Summary of Data on disparities by population group:
A disparities report is attached to this submission.

1b.5 Citations for data on Disparities:
The attached disparities report uses 2009 data from the clinical data warehouse.

1b
C☐
P☐
M☐
N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** The desired outcome would be fewer surgical site infections. Since this is only one process in the care of surgery patients, it would be difficult to attribute a reduction in SSI to this one measure.

1c.2-3. Type of Evidence: Evidence-based guideline, Randomized controlled trial, Expert opinion, Systematic synthesis of research, Meta-analysis

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
There are over 40 million surgeries performed in the United States each year. Surgical site infection (SSIs) are the second most common cause of healthcare associated infections.Studies show that shaving causes multiple skin abrasions that later may become infected. In a randomized study of 1,980 adult patients undergoing cardiopulmonary bypass surgeries, Ko, et al (1992), reported a significantly higher rate of infection among patients who were shaved with a razor than those who had hair removal by electric clippers before skin incision. In another randomized trial of 200 patients undergoing elective inguinal herniorraphy, Balthazar, et al (1982), concluded that hair removal with electric clippers immediately prior to the procedures “did not increase the risk of postoperative wound infection” (p. 799). In a systematic literature review by Kjonniksen, et al (2002), there was no strong evidence to contraindicate preoperative hair removal; however, there was strong evidence against hair removal with a razor. This review recommended depilatory or electric clippers immediately prior to surgery when hair removal was required. Alexander, et al (1983), reported that clippers, used on the morning of surgery, resulted in reduced surgical site infections and healthcare expenditures.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom*)**:**
RCTs performed and evidence supporting recommendation.

1c.6 Method for rating evidence: Classes and levels
Level A: Data derived from multiple randomized clinical trials
Level B: Data derived from a single randomized trial or from nonrandomized trials
Level C: Consensus expert opinion
Classification of Recommendations
Class I: Conditions for which there is evidence and/or general agreement that a given procedure is useful and effective

1c
C☐
P☐
M☐
N☐

Class II: Conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure
IIa: Weight of evidence favors usefulness/efficacy.
IIb: Usefulness/efficacy is less well established by evidence.
Class III: Conditions for which there is evidence and/or general agreement that the procedure is not useful/effective

1c.7 Summary of Controversy/Contradictory Evidence: There have been no studies that contradict the guidelines for surgical site hair removal.

1c.8 Citations for Evidence (*other than guidelines***):** Kjonniksen I, Andersen BM, Sondenaa VG, et al. Preoperative hair removal-a systematic literature review. AORN J. 2002 May;75 (5):928-938,940. PMID:12063942.

Ko W, Lazenby WD, Zelano JA, et al. Effects of shaving methods and intraoperative irrigation on suppurative mediastinitis after bypass operations. Ann Thorac Surg. 1992 Feb;53 (2):301-305. PMID: 1731672.

Alexander, JW. Influence of hair removal methods on wound infections. Archives of Surgery Vol 118, March 1983. PMID: 6824435.

Balthazar ER, Colt JD, Nichols RL. Preoperative hair removal: a random prospective study of shaving versus clipping. South Med J. 1982 Jul:75(7): 799-801. PMID: 7089645.
Uckay I, Harbarth S, Peter R, Lew D, Hoffmeyer P, Pittet D. Preventing Surgical Site Infections. Expert Reviews. Anti Infect Ther 2010; 8(6): 657-670.
Tanner J, Woodings D, Moncaster K. Preoperative hair removal to reduce surgical site infection. Cochrane Database of Systematic Reviews 2006, Issue 3.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
From CDC HICPAC:
Category 1A
Do not remove hair from operative site unless necessary to facilitate surgery
If hair is removed, do immediately before surgery, preferably with electric clippers

1c.10 Clinical Practice Guideline Citation: Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR, the Hospital Infection Control Practices Advisory Committee. Guideline for prevention of surgical site infection 1999. Infect Control Hosp Epidemiol 1999;20:247-80.

AORN. Recommended practices for skin preparation of patients. AORN Journal 2002;
1c.11 National Guideline Clearinghouse or other URL:
http://www.guideline.gov/search/search.aspx?term=prevention+of+surgical+site+infection

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
CDC HICPAC gave a Category 1A to this recommendation.

1c.13 Method for rating strength of recommendation (*If different from USPSTF system, also describe rating and how it relates to USPSTF)***:**
RANKINGS
Category I recommendations, including IA and IB, are those recommendations that are viewed as effective by HICPAC and experts in the fields of surgery, infectious diseases, and infection control. Both Category IA and IB recommendations are applicable for, and should be adopted by,all healthcare facilities; IA and IB recommendations differ only in the strength of the supporting scientific evidence.
Category II recommendations are supported by less scientific data than Category I recommendations; such recommendations may be appropriate for addressing specific nosocomial problems or specific patient populations.
No recommendation is offered for some practices, either because there is a lack of consensus regarding their efficacy or because the available scientific evidence is insufficient to support their adoption. For such unresolved

issues, practitioners should use judgement to determine a policy regarding these practices within their organization. 1c.14 **Rationale for using this guideline over others:** "The Guideline for Prevention of Surgical Site Infection, 1999, provides recommendations concerning reduction of surgical site infection risk. Each recommendation is categorized on the basis of existing scientific data, theoretical rationale, and applicability." Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR, the Hospital Infection Control Practices Advisory Committee. Guideline for prevention of surgical site infection 1999. Infect Control Hosp Epidemiol 1999;20:247-80.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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***):** Surgery patients with surgical hair site removal with clippers or depilatory or no surgical site hair removal 2a.2 **Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** Admission to discharge. 2a.3 **Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Data Elements: Preoperative Hair Removal Included Populations: An ICD-9-CM Principal Procedure Code of selected surgeries (as defined in Appendix A, Table 5.10 for ICD-9-CM codes).	
2a.4 **Denominator Statement (***Brief, text description of the denominator - target population being measured***):** All selected surgery patients Include patients with an ICD-9-CM Principal Procedure Codes of selected surgeries. 2a.5 **Target population gender:** Female, Male 2a.6 **Target population age range:** 18 years of age and older 2a.7 **Denominator Time Window (***The time period in which cases are eligible for inclusion in the denominator***):** Admission to discharge 2a.8 **Denominator Details (***All information required to collect/calculate the denominator - the target*	**2a-specs** **C☐** **P☐** **M☐** **N☐**

population being measured - including all codes, logic, and definitions):
Data Elements:
Admission Date
Anesthesia Start Date
Birthdate
Clinical Trial
Discharge Date
ICD-9-CM Principal Procedure Code
Laparoscope
Include patients with an ICD-9-CM Principal Procedure code or ICD-9-CM Other Procedure Codes of selected surgeries.

2a.9 **Denominator Exclusions** (*Brief text description of exclusions from the target population*): Excluded Populations:
Patients less than 18 years of age
Patients who have a length of Stay greater than 120 days
Patients whose ICD-9-CM principal procedure was performed entirely by laparoscope.
Patients enrolled in clinical trials
Patients whose ICD-9-CM principal procedure occurred prior to the date of admission
Patients who performed their own hair removal

2a.10 **Denominator Exclusion Details** (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions*):
The data elements include:
Clinical Trial and Laparoscope.
Affirmative answers to these data elements excludes the patient from the measure.

2a.11 **Stratification Details/Variables** (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions*):
NA

2a.12-13 **Risk Adjustment Type:** No risk adjustment necessary

2a.14 **Risk Adjustment Methodology/Variables** (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method*):
N/A

2a.15-17 **Detailed risk model available Web page URL or attachment:**

2a.18-19 **Type of Score:** Rate/proportion
2a.20 **Interpretation of Score:** Better quality = Higher score
2a.21 **Calculation Algorithm** (*Describe the calculation of the measure as a flowchart or series of steps*):
SCIP-Infection (Inf)-6: Surgery Patients with Appropriate Hair Removal
Variable Key: Patient Age, Surgery Days
1.Start processing. Run cases that are included in the Surgical Care Improvement Project (SCIP) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.
2.Calculate Patient Age. The Patient Age, in years, is equal to the Admission Date minus the Birthdate. Use the month and day portion of admission date and birthdate to yield the most accurate age.
3.Check Patient Age
a.If Patient Age is less than 18 years, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b.If Patient Age is greater than or equal to 18 years, continue processing and proceed to Laparoscope.
4.Check Laparoscope
a.If Laparoscope is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b.If Laparoscope equals 1 or 3, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c.If Laparoscope equals 2, continue processing and proceed to Clinical Trial.
5.Check Clinical Trial

a.If Clinical Trial is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b.If Clinical Trial equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c.If Clinical Trial equals No, continue processing and proceed to Anesthesia Start Date.
6.Check Anesthesia Start Date
a.If the Anesthesia Start Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b.If the Anesthesia Start Date equals Unable To Determine, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c.If Anesthesia Start Date equals a Non Unable To Determine Value, continue processing and proceed to the Surgery Days calculation.
7.Calculate Surgery Days. Surgery Days, in days, is equal to the Anesthesia Start Date minus the Admission Date.
8.Check Surgery Days
a.If the Surgery Days is less than zero, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b.If the Surgery Days is greater than or equal to zero, continue processing and proceed to Preoperative Hair Removal.
9.Check Preoperative Hair Removal – Note: No allowable value can occur more than once. Allowable values of '1' or '7' cannot be combined with each other or with any of the other allowable values.
a.If Preoperative Hair Removal is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b.If Any Preoperative Hair Removal equals 6, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c.If Any Preoperative Hair Removal equals 1, 2, 3, 4, 5, 7, or 8 and None equals 6, continue processing and recheck Preoperative Hair Removal.
10.Recheck Preoperative Hair Removal
a.If Any Preoperative Hair Removal equals 2, 5, or 7, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
b.If Any Preoperative Hair Removal equals 1, 3, 4, or 8 and None equals 2, 5, or 7, the case will proceed to a Measure Category Assignment of E and will be in the Numerator Population.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)*:
Benchmarks are established using the ABC methodology, based on the actual performance of the top facilities. ABC benchmarks identify superior performance and encourage poorer performers to improve. It is data-driven, peer-group performance feedback.
Achievable Benchmarks of Care TM: developed at the University of Alabama at Birmingham for AHRQ. This methodology identifies benchmark care levels already achieved by "best-in-class" care givers. Development of benchmarks that are realistic and achievable may help to motivate providers that are having difficulty improving care. The benchmarks represent a measureable level of excellence that always exceeds average performance. It ensures that all superior providers contribute to the benchmark but also ensures that providers with high performance but very low numbers of cases do not unduly influence benchmark levels. Additional information can be found at http://main.uab.edu/show.asp?durki=14527

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)*:
The SCIP Topic Population (common to all SCIP measures) is defined as patients admitted to the hospital for inpatient acute care with an ICD-9-CM Principal Procedure Code for SCIP as defined in Appendix A, Table 5.10 and a Length of Stay (Discharge Date - Admission Date) <= 120 days.
Appendix A (PDF) can be found at this link:
http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228754600169
There are eight distinct strata or sub-populations within the SCIP Topic Population, each identified by a specific group of procedure codes. The patients in each stratum are counted in the Initial Patient Population of multiple measures.

The following sample size tables for each option automatically build in the number of cases needed to obtain the required sample sizes.

Quarterly Sampling
For hospitals selecting sample cases for SCIP, a modified sampling procedure is required. Hospitals selecting sample cases for this set must ensure that each individual stratum's population and quarterly sample size meets the following conditions:
• Select within each of the seven individual measure stratum (e.g., colorectal surgery, hip arthroplasty, etc.) and the 8th SCIP stratum (Table 5.25 in Appendix A).

Quarterly Sample Size
Based on Initial Patient Population Size for the SCIP Measure Set

Hospital's Measure
Average Quarterly
Stratum Initial Patient Population Size
"N" Minimum Required
Stratum Sample Size
"n"
>/= 481 49
171-480 10% of Initial Patient Population size
17-170 17
< 17 No sampling; 100% Initial Patient Population required

Monthly Sampling
For hospitals selecting sample cases for SCIP, a modified sampling procedure is required. Hospitals selecting sample cases for this set must ensure that each individual strata population and monthly sample size meets the following conditions:
• Select within each of the seven individual measure stratum (e.g., colorectal surgery, hip arthroplasty, etc.) and the 8th SCIP stratum (Table 5.25 in Appendix A).

Monthly Sample Size
Based on Initial Patient Population Size for the SCIP Measure Set

Hospital's Measure
Average Monthly
Stratum Initial Patient Population Size
"N" Minimum Required
Stratum Sample Size
"n"
>/= 151 16
61-150 10% of Initial Patient Population size
6-60 6
<6 No sampling; 100% Initial Patient Population required

All of the SCIP measures´ specific exclusion criteria are used to filter out cases that do not belong in the measure denominator.

2a.24 **Data Source** (*Check the source(s) for which the measure is specified and tested)*
Paper medical record/flow-sheet, Electronic administrative data/claims, Electronic Health/Medical Record

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.*)**:**
Most facilities use vendors to collect the data electronically. CMS provides a free, downloadable tool called CART. A paper tool modeled after the data collected electronically is provided as an attachment. CART downloads can be found on QualityNet.org at http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1138900279093

2a.26-28 **Data source/data collection instrument reference web page URL or attachment:** Attachment

SCIPCARTpapertool_10.01.10.doc

2a.29-31 Data dictionary/code table web page URL or attachment: URL
http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228754600169

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency, Population: national, Program: QIO, Can be measured at all levels

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

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: For Q2 2010, the national rate was 99.6%. The number of facilities reporting: 3,629. The number of cases in the denominator: 401,573. The number of cases in the numerator: 399, 933.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
Measure has been in use since 2001 and has been continually collected nationally for the Hospital Inpatient Quality Reporting Program since Jan 2007. A predetermined number of charts are requested and submitted to an independent abstraction/validation contractor quarterly. Mismatches are calculated and reported to facilities and are used to determine eligibility for incentives. Facilities must achieve an 80% agreement with CDAC abstractors in addition to agreeing to report measure rates on Hospital Compare.

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Feedback from the hospital abstractors and the independent validation team is collected and incorporated. Reports on mismatches between national abstractors and the independent abstraction/validation contractor are reviewed quarterly. Revisions to data elements are made accordingly. Mismatches for 1Q2010: 143 mismatches out of 2269 cases validated equals a percentage of 6.3%.

2b C☐ P☐ M☐ N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: Review of relevant guidelines and studies is performed quarterly with a Technical Expert Panel. Antibiotic selection guidelines are reviewed during quarterly TEP teleconfereces. Specifications (including codes and data elements) are modified every six months according to feedback provided by clinicians and hospital staff collecting data for the measure. National performance of the measure is monitored by the measure steward with quarterly benchmarks of hospital submitted data developed for distribution to QIOs. Trend reports are also prepared and reviewed. The measure is collecting the information it was designed to collect.

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
Face validity is systematically assessed by the Technical Expert Panels and the measure is judged to assess the provision of appropriate care for the target population.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
The measure is collecting the information it was designed to collect, according to expert panel review.

2c C☐ P☐ M☐ N☐

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
The exclusions used in this measure are the exclusions used for all SCIP measures and are reviewed by the Technical Expert Panel as needed.

2d C☐ P☐ M☐ N☐

2d.2 Citations for Evidence: NA **2d.3 Data/sample** *(description of data/sample and size)*: NA **2d.4 Analytic Method** *(type analysis & rationale)*: NA **2d.5 Testing Results** *(e.g., frequency, variability, sensitivity analyses)*: NA	**NA** ☐
2e. Risk Adjustment for Outcomes/ Resource Use Measures **2e.1 Data/sample** *(description of data/sample and size)*: NA **2e.2 Analytic Method** *(type of risk adjustment, analysis, & rationale)*: NA **2e.3 Testing Results** *(risk model performance metrics)*: NA **2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:** This is a process measure.	**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)*: Measure rate trends are reviewed every quarter, using a rolling 5 quarters of national hospital submitted data. **2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*: Analysts review quarterly benchmarks and trends to identify differences in performance scores and investigate the possible causes. If measure specifications (algorithms, data elements) are causing the difference in performance, they are reviewed for possible updates by the subject matter experts. This measure has had consistently high rates of performance the last several quarters. **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)*: A trends report is provided with this submission.	**2f** **C** ☐ **P** ☐ **M** ☐ **N** ☐
2g. Comparability of Multiple Data Sources/Methods **2g.1 Data/sample** *(description of data/sample and size)*: Currently, this measure is collected from the medical record. The medical record can be paper or an EHR. No analysis between chart-abstracted and eMeasure collection has been performed because the eMeasure specifications have not been implemented at this time. **2g.2 Analytic Method** *(type of analysis & rationale)*: NA **2g.3 Testing Results** *(e.g., correlation statistics, comparison of rankings)*: NA	**2g** **C** ☐ **P** ☐ **M** ☐ **N** ☐ **NA** ☐
2h. Disparities in Care **2h.1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: An updated disparities report has been submitted to NQF for review. Data on the range of performance values by decile for the hospital process measures was provided also.	**2h** **C** ☐ **P** ☐ **M** ☐ **N** ☐

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: All of the inpatient quality reporting measures collect this information: Birthdate, Hispanic Ethnicity, Payment Source, Race and Sex. Additional analysis was performed to determine disparities in US region and urban vs rural.	**NA** ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, met? **Rationale:**	***2*** **C☐** **P☐** **M☐** **N☐**
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Eval Rating**
3a. Meaningful, Understandable, and Useful Information **3a.1 Current Use:** In use **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** The measure is currently in use for the Hospital Inpatient Quality Reporting Program under CMS. To receive the APU from Medicare, hospitals agree to report their data and have their measure rates reported on Hospital Compare. http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier1&cid=1121785350606 **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***):** This measure is also used in the accreditation process for the Joint Commission. It is part of the SCIP measure set, which facilities can choose to report for accreditation purposes. **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)***:** The measures rates are reported on the website Hospital Compare. **3a.5 Methods** *(e.g., focus group, survey, QI project)***:** Data about interpretability of reported measure rates are collected by the CMS contractor responsible for maintaining HOspital Compare. **3a.6 Results** *(qualitative and/or quantitative results and conclusions)***:**	**3a** **C☐** **P☐** **M☐** **N☐**
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** **C☐** **P☐**

3b.2 Are the measure specifications harmonized? If not, why?	**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:** **5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:**	**3c C☐ P☐ M☐ N☐ NA ☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**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.** There are plans to provide e-specifications for all measures in the near future.	**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.** Interpretation of data elements will always be a factor, since the instructions for obtaining the data are written by the measure developers. No unintended consequences have been identified with the hair removal measure.	**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**	**4e C☐ P☐**

measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues: The measure specifications have been stable for the last 2 manuals. 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** The cost associated with measure use is that of data collection only. Many facilities employ quality improvement staff to perform data abstraction and entry. The same employees may develop reports and provide information to clinicians and hospital administration. 4e.3 **Evidence for costs:** No studies have been performed on the cost of implementation. 4e.4 **Business case documentation:**	**M☐ N☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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** Centers for Medicare & Medicaid Services, 7500 Security Blvd, Mail Stop S3-02-01, Baltimore, Maryland, 21244 Co.2 **Point of Contact** Kristie, Baus, RN, MS, kristie.baus@cms.hhs.gov, 410-786-8161-	
Measure Developer If different from Measure Steward Co.3 **Organization** Centers for Medicare & Medicaid Services, 7500 Security Blvd, Mail Stop S3-02-01, Baltimore, Maryland, 21244 Co.4 **Point of Contact** Kristie, Baus, RN, MS, kristie.baus@cms.hhs.gov, 410-786-8161-	
Co.5 **Submitter If different from Measure Steward POC** Wanda, Johnson, RN, wjohnson@ofmq.com, 405-302-3278-, Oklahoma Foundation for Medical Quality	
Co.6 **Additional organizations that sponsored/participated in measure development** This measure is aligned with the Joint Commission. NOTE: This measure is being considered for retirement by CMS due to high rates of performance. If not collected in the Hospital Inpatient Quality Reporting Program, the Joint Commission will become the measure steward.	
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.** The Surgical Care Improvement Project´s Infection TEP was involved in this measure´s development and remains	

involved in its maintenance.
Ad.2 If adapted, provide name of original measure: NOTE: This measure is being considered for retirement by CMS due to high rates of performance. If not collected in the Hospital Inpatient Quality Reporting Program, the Joint Commission will become the measure steward. **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:** 2008 **Ad.7 Month and Year of most recent revision:** 10, 2010 **Ad.8 What is your frequency for review/update of this measure?** every 6 months **Ad.9 When is the next scheduled review/update for this measure?** 04, 2011
Ad.10 Copyright statement/disclaimers: Trend Report (BM= Benchmark, rate = national score) Q209 BM: 100 Rate: 99.1 Q309 BM: 100 Rate 99.3 Q409 BM: 100 Rate 99.4 Q110 BM: 100 Rate 99.5 Q210 BM: 100 Rate 99.6
Ad.11 -13 Additional Information web page URL or attachment: Attachment IP Measures Disp_2009.xls
Date of Submission (*MM/DD/YY*)**:** 03/28/2011

Disparities analysis for 26 performance measures using 2009 Clinical Data Warehouse

By Race/Ethnicity (3% of cases were excluded due to missing data on race/ethnicity)

Measures and Race/ethnicity group	Num	Den	Percent	Unadjusted OR (95%CI)	p-value
AMI1: Aspirin at arrival					
Caucasian	247,145	251,158	98.4	ref.	ref.
African-American	36,868	37,747	97.7	0.68 (0.63-0.73)	<0.001
Hispanic	26,561	27,316	97.2	0.57 (0.53-0.62)	<0.001
Asian/Pacific Islander	7,346	7,472	98.3	0.95 (0.79-1.13)	0.548
Native American	1,074	1,087	98.8	1.34 (0.78-2.32)	0.293
AMI2: Aspirin at discharge					
Caucasian	305,754	310,489	98.5	ref.	ref.
African-American	39,545	40,591	97.4	0.59 (0.55-0.63)	<0.001
Hispanic	27,791	28,805	96.5	0.42 (0.40-0.45)	<0.001
Asian/Pacific Islander	7,694	7,854	98.0	0.74 (0.64-0.87)	<0.001
Native American	1,908	1,935	98.6	1.09 (0.75-1.60)	0.643
AMI3: ACEI or ARB for LVSD					
Caucasian	54,767	57,482	95.3	ref.	ref.
African-American	8,642	9,024	95.8	1.12 (1.01-1.25)	0.040
Hispanic	5,591	5,896	94.8	0.91 (0.80-1.03)	0.123
Asian/Pacific Islander	1,302	1,372	94.9	0.92 (0.72-1.18)	0.514
Native American	371	393	94.4	0.84 (0.54-1.29)	0.416
AMI4: Smoking cessation counseling					
Caucasian	103,977	104,611	99.4	ref.	ref.
African-American	16,611	16,741	99.2	0.78 (0.64-0.94)	0.010
Hispanic	7,671	7,757	98.9	0.54 (0.43-0.68)	<0.001
Asian/Pacific Islander	1,720	1,747	98.5	0.39 (0.26-0.57)	<0.001
Native American	753	767	98.2	0.33 (0.19-0.56)	<0.001
AMI5: Beta-blocker at discharge					
Caucasian	298,954	304,013	98.3	ref.	ref.
African-American	39,112	40,008	97.8	0.74 (0.69-0.79)	<0.001
Hispanic	27,331	28,382	96.3	0.44 (0.41-0.47)	<0.001

Asian/Pacific Islander	7,602	7,738	98.2	0.95 (0.80-1.12)	0.526
Native American	1,841	1,882	97.8	0.76 (0.56-1.04)	0.083
AMI7a: Fibrinolytic within 30 minutes					
Caucasian	651	1,169	55.7	ref.	ref.
African-American	73	157	46.5	0.69 (0.50-0.97)	0.030
Hispanic	190	417	45.6	0.67 (0.53-0.83)	<0.001
Asian/Pacific Islander	36	61	59.0	1.15 (0.68-1.93)	0.610
Native American	1	3	33.3	0.40 (0.04-4.40)	0.452
AMI8a: PCI within 90 minutes					
Caucasian	38,044	43,171	88.1	ref.	ref.
African-American	3,448	4,234	81.4	0.59 (0.54-0.64)	<0.001
Hispanic	3,297	3,936	83.8	0.70 (0.64-0.76)	<0.001
Asian/Pacific Islander	1,079	1,237	87.2	0.92 (0.78-1.09)	0.337
Native American	160	189	84.7	0.74 (0.50-1.11)	0.143
HF1: Discharge instructions					
Caucasian	357,746	414,742	86.3	ref.	ref.
African-American	124,070	143,689	86.3	1.01 (0.99-1.03)	0.400
Hispanic	44,786	51,690	86.6	1.03 (1.01-1.06)	0.016
Asian/Pacific Islander	9,895	11,375	87.0	1.07 (1.01-1.13)	0.025
Native American	2,351	3,083	76.3	0.51 (0.47-0.56)	<0.001
HF2: Evaluation of LV function					
Caucasian	521,142	535,940	97.2	ref.	ref.
African-American	159,661	163,219	97.8	1.27 (1.23-1.32)	<0.001
Hispanic	55,388	57,714	96.0	0.68 (0.65-0.71)	<0.001
Asian/Pacific Islander	12,720	13,004	97.8	1.27 (1.13-1.43)	<0.001
Native American	3,201	3,416	93.7	0.42 (0.37-0.49)	<0.001
HF3: ACEI or ARB for LVSD					
Caucasian	145,067	155,808	93.1	ref.	ref.
African-American	66,217	69,597	95.1	1.45 (1.39-1.51)	<0.001
Hispanic	18,769	20,068	93.5	1.07 (1.01-1.14)	0.026
Asian/Pacific Islander	3,777	3,962	95.3	1.51 (1.30-1.75)	<0.001
Native American	1,173	1,278	91.8	0.83 (0.68-1.01)	0.064
HF4: Smoking cessation counseling					
Caucasian	76,177	77,858	97.8	ref.	ref.

African-American	44,071	44,760	98.5	1.41 (1.29-1.54)	<0.001
Hispanic	7,273	7,423	98.0	1.07 (0.90-1.27)	0.432
Asian/Pacific Islander	1,375	1,413	97.3	0.80 (0.58-1.11)	0.176
Native American	692	732	94.5	0.38 (0.28-0.53)	<0.001
PN2: Pnemococal vaccination given or screened for					
Caucasian	378,259	408,034	92.7	ref.	ref.
African-American	34,705	39,186	88.6	0.61 (0.59-0.63)	<0.001
Hispanic	24,135	28,528	84.6	0.43 (0.42-0.45)	<0.001
Asian/Pacific Islander	8,804	9,900	88.9	0.63 (0.59-0.67)	<0.001
Native American	2,310	2,640	87.5	0.55 (0.49-0.62)	<0.001
PN3a: Initial blood culture within 24 hours - ICU only					
Caucasian	78,108	82,387	94.8	ref.	ref.
African-American	12,551	13,078	96.0	1.30 (1.19-1.43)	<0.001
Hispanic	7,338	7,863	93.3	0.77 (0.70-0.84)	<0.001
Asian/Pacific Islander	2,199	2,271	96.8	1.67 (1.32-2.12)	<0.001
Native American	776	846	91.7	0.61 (0.47-0.78)	<0.001
PN3b: Initial blood culture before first antibiotic dose - ED only					
Caucasian	361,802	380,083	95.2	ref.	ref.
African-American	56,541	60,416	93.6	0.74 (0.71-0.76)	<0.001
Hispanic	34,169	37,132	92.0	0.58 (0.56-0.61)	<0.001
Asian/Pacific Islander	9,388	9,889	94.9	0.95 (0.86-1.04)	0.240
Native American	3,058	3,402	89.9	0.45 (0.40-0.50)	<0.001
PN4: Smoking cessation counseling					
Caucasian	153,759	158,876	96.8	ref.	ref.
African-American	30,859	31,710	97.3	1.21 (1.12-1.30)	<0.001
Hispanic	9,885	10,230	96.6	0.95 (0.85-1.07)	0.400
Asian/Pacific Islander	1,689	1,759	96.0	0.80 (0.63-1.02)	0.074
Native American	1,722	1,940	88.8	0.26 (0.23-0.30)	<0.001
PN5c: First antibiotic dose within 6 hours					
Caucasian	402,180	421,893	95.3	ref.	ref.
African-American	60,989	66,036	92.4	0.59 (0.57-0.61)	<0.001
Hispanic	35,145	39,094	89.9	0.44 (0.42-0.45)	<0.001
Asian/Pacific Islander	9,399	9,865	95.3	0.99 (0.90-1.09)	0.812
Native American	3,430	3,752	91.4	0.52 (0.47-0.59)	<0.001

PN6: Antibioti selection consistent with guidelines					
Caucasian	254,116	279,291	91.0	ref.	ref.
African-American	35,023	38,201	91.7	1.09 (1.05-1.13)	<0.001
Hispanic	25,350	28,361	89.4	0.83 (0.80-0.87)	<0.001
Asian/Pacific Islander	6,093	6,689	91.1	1.01 (0.93-1.10)	0.770
Native American	2,570	2,922	88.0	0.72 (0.65-0.81)	<0.001
PN7: Influenza vaccination given or screened for					
Caucasian	266,920	293,208	91.0	ref.	ref.
African-American	31,910	37,007	86.2	0.62 (0.60-0.64)	<0.001
Hispanic	18,854	22,505	83.8	0.51 (0.49-0.53)	<0.001
Asian/Pacific Islander	5,702	6,539	87.2	0.67 (0.62-0.72)	<0.001
Native American	1,927	2,405	80.1	0.40 (0.36-0.44)	<0.001
SCIP1: Antibiotic within 1 hour before incision or 2 hours for vancomycin or quinolone					
Caucasian	827,536	860,067	96.2	ref.	ref.
African-American	95,484	99,527	95.9	0.93 (0.90-0.96)	<0.001
Hispanic	60,439	64,806	93.3	0.54 (0.53-0.56)	<0.001
Asian/Pacific Islander	14,743	15,282	96.5	1.08 (0.99-1.17)	0.101
Native American	4,037	4,325	93.3	0.55 (0.49-0.62)	<0.001
SCIP2: Prophylactic antibiotic consistent with guidelines					
Caucasian	848,411	868,974	97.6	ref.	ref.
African-American	97,576	100,464	97.1	0.82 (0.79-0.85)	<0.001
Hispanic	62,778	64,991	96.6	0.69 (0.66-0.72)	<0.001
Asian/Pacific Islander	15,171	15,547	97.6	0.98 (0.88-1.08)	0.672
Native American	4,230	4,360	97.0	0.79 (0.66-0.94)	0.008
SCIP3: Prophylactic ABX discontinued within 24 h. of surgery end time or 48 h. for cardiac surgery					
Caucasian	766,551	819,715	93.5	ref.	ref.
African-American	87,315	94,468	92.4	0.85 (0.83-0.87)	<0.001
Hispanic	54,461	61,420	88.7	0.54 (0.53-0.56)	<0.001
Asian/Pacific Islander	13,218	14,358	92.1	0.80 (0.76-0.85)	<0.001
Native American	3,812	4,103	92.9	0.91 (0.81-1.02)	0.116
SCIP4: Controlled 6 AM postoperative serum glucose - cardiac surgery					
Caucasian	134,822	144,908	93.0	ref.	ref.
African-American	10,742	11,722	91.6	0.82 (0.77-0.88)	<0.001
Hispanic	11,031	12,520	88.1	0.55 (0.52-0.59)	<0.001

Asian/Pacific Islander	3,437	3,773	91.1	0.77 (0.68-0.86)	<0.001
Native American	706	766	92.2	0.88 (0.68-1.15)	0.344
SCIP6: appropriate hair removal					
Caucasian	1,222,603	1,232,305	99.2	ref.	ref.
African-American	149,984	151,395	99.1	0.84 (0.80-0.89)	<0.001
Hispanic	95,326	97,273	98.0	0.39 (0.37-0.41)	<0.001
Asian/Pacific Islander	23,368	23,575	99.1	0.90 (0.78-1.03)	0.119
Native American	6,390	6,543	97.7	0.33 (0.28-0.39)	<0.001
SCIPCARD2: Perioperative period beta blocker					
Caucasian	327,860	359,462	91.2	ref.	ref.
African-American	34,505	38,004	90.8	0.95 (0.92-0.99)	0.007
Hispanic	17,805	20,128	88.5	0.74 (0.71-0.77)	<0.001
Asian/Pacific Islander	5,128	5,770	88.9	0.77 (0.71-0.84)	<0.001
Native American	1,312	1,493	87.9	0.70 (0.60-0.82)	<0.001
SCIPVTE1: Recommended VTE prophylaxis ordered during admission					
Caucasian	343,547	367,129	93.6	ref.	ref.
African-American	49,075	52,658	93.2	0.94 (0.91-0.98)	<0.001
Hispanic	27,199	30,224	90.0	0.62 (0.59-0.64)	<0.001
Asian/Pacific Islander	7,406	8,195	90.4	0.64 (0.60-0.69)	<0.001
Native American	1,999	2,208	90.5	0.66 (0.57-0.76)	<0.001
SCIPVTE2: Received VTE prophylaxis within 24 hours prior to or after surgery					
Caucasian	334,443	365,471	91.5	ref.	ref.
African-American	47,804	52,220	91.5	1.00 (0.97-1.04)	0.798
Hispanic	26,376	29,811	88.5	0.71 (0.69-0.74)	<0.001
Asian/Pacific Islander	7,241	8,126	89.1	0.76 (0.71-0.81)	<0.001
Native American	1,942	2,183	89.0	0.75 (0.65-0.86)	<0.001

Disparities analysis for 26 performance measures using 2009 Clinical Data Warehouse

By Gender (less than 0.1% of cases were excluded due to missing data on gender)

Measures and gender	Num	Den	Percent	Unadjusted OR (95%CI)	p-value
AMI1: Aspirin at arrival					
Female	132,222	135,450	97.6	ref.	ref.
Male	197,136	199,829	98.7	1.79 (1.70-1.88)	<0.001
AMI2: Aspirin at discharge					
Female	150,930	154,577	97.6	ref.	ref.
Male	247,653	251,152	98.6	1.71 (1.63-1.79)	<0.001
AMI3: ACEI or ARB for LVSD					
Female	26,127	27,376	95.4	ref.	ref.
Male	47,156	49,502	95.3	0.96 (0.90-1.03)	0.269
AMI4: Smoking cessation counseling					
Female	42,885	43,241	99.2	ref.	ref.
Male	93,180	93,741	99.4	1.38 (1.21-1.58)	<0.001
AMI5: Beta-blocker at discharge					
Female	149,171	152,804	97.6	ref.	ref.
Male	240,965	244,715	98.5	1.56 (1.49-1.64)	<0.001
AMI7a: Fibrinolytic within 30 minutes					
Female	254	523	48.6	ref.	ref.
Male	730	1,347	54.2	1.25 (1.02-1.53)	0.029
AMI8a: PCI within 90 minutes					
Female	12,629	15,029	84.0	ref.	ref.
Male	35,545	40,118	88.6	1.48 (1.40-1.56)	<0.001
HF1: Discharge instructions					
Female	264,674	308,679	85.7	ref.	ref.
Male	286,692	330,544	86.7	1.09 (1.07-1.10)	<0.001
HF2: Evaluation of LV function					
Female	391,232	403,675	96.9	ref.	ref.
Male	378,142	387,472	97.6	1.29 (1.25-1.32)	<0.001
HF3: ACEI or ARB for LVSD					
Female	92,111	98,257	93.7	ref.	ref.
Male	148,513	158,409	93.8	1.00 (0.97-1.03)	0.936
HF4: Smoking cessation counseling					

Female	51,445	52,630	97.7	ref.	ref.
Male	80,801	82,294	98.2	1.25 (1.15-1.35)	<0.001
PN2: Pnemococal vaccination given or screened for					
Female	247,221	269,382	91.8	ref.	ref.
Male	212,145	231,563	91.6	0.98 (0.96-1.00)	0.042
PN3a: Initial blood culture within 24 hours - ICU only					
Female	50,079	52,932	94.6	ref.	ref.
Male	53,544	56,305	95.1	1.10 (1.05-1.17)	<0.001
PN3b: Initial blood culture before first antibiotic dose - ED only					
Female	246,104	260,181	94.6	ref.	ref.
Male	230,916	243,503	94.8	1.05 (1.02-1.08)	<0.001
PN4: Smoking cessation counseling					
Female	103,237	106,615	96.8	ref.	ref.
Male	99,296	102,754	96.6	0.94 (0.90-0.99)	0.011
PN5c: First antibiotic dose within 6 hours					
Female	272,016	288,698	94.2	ref.	ref.
Male	252,643	266,222	94.9	1.14 (1.11-1.17)	<0.001
PN6: Antibioti selection consistent with guidelines					
Female	175,954	193,373	91.0	ref.	ref.
Male	156,410	172,235	90.8	0.98 (0.96-1.00)	0.059
PN7: Influenza vaccination given or screened for					
Female	180,348	200,180	90.1	ref.	ref.
Male	153,242	170,972	89.6	0.95 (0.93-0.97)	<0.001
SCIP1: Antibiotic within 1 hour before incision or 2 hours for vancomycin or quinolone					
Female	660,133	687,675	96.0	ref.	ref.
Male	383,816	399,901	96.0	1.00 (0.98-1.02)	0.660
SCIP2: Prophylactic antibiotic consistent with guidelines					
Female	672,428	691,674	97.2	ref.	ref.
Male	398,658	406,588	98.0	1.44 (1.40-1.48)	<0.001
SCIP3: Prophylactic ABX discontinued within 24 h. of surgery end time or 48 h. for cardiac surgery					
Female	613,378	657,129	93.3	ref.	ref.
Male	351,165	378,744	92.7	0.91 (0.89-0.92)	<0.001
SCIP4: Controlled 6 AM postoperative serum glucose - cardiac surgery					
Female	52,328	56,457	92.7	ref.	ref.
Male	114,589	124,004	92.4	0.96 (0.92-1.00)	0.038

SCIP6: appropriate hair removal					
Female	944,375	951,265	99.3	ref.	ref.
Male	613,124	620,263	98.8	0.63 (0.61-0.65)	<0.001
SCIPCARD2: Perioperative period beta blocker					
Female	210,810	232,468	90.7	ref.	ref.
Male	189,354	207,438	91.3	1.08 (1.05-1.10)	<0.001
SCIPVTE1: Recommended VTE prophylaxis ordered during admission					
Female	266,908	284,212	93.9	ref.	ref.
Male	177,139	192,153	92.2	0.76 (0.75-0.78)	<0.001
SCIPVTE2: Received VTE prophylaxis within 24 hours prior to or after surgery					
Female	260,379	282,821	92.1	ref.	ref.
Male	171,935	190,847	90.1	0.78 (0.77-0.80)	<0.001

Disparities analysis for 26 performance measures using 2009 Clinical Data Warehouse

By Age-Group

Measures and age group	Num	Den	Percent	Unadjusted OR (95%CI)	p-value
AMI1: Aspirin at arrival					
under 65 years	141,150	142,677	98.9	ref.	ref.
65 to 74 years	69,462	70,636	98.3	0.64 (0.59-0.69)	<0.001
75 to 84 years	68,661	70,270	97.7	0.46 (0.43-0.50)	<0.001
85 or older	50,094	51,705	96.9	0.34 (0.31-0.36)	<0.001
AMI2: Aspirin at discharge					
under 65 years	188,910	191,432	98.7	ref.	ref.
65 to 74 years	86,865	88,378	98.3	0.77 (0.72-0.82)	<0.001
75 to 84 years	76,528	78,185	97.9	0.62 (0.58-0.66)	<0.001
85 or older	46,290	47,744	97.0	0.42 (0.40-0.45)	<0.001
AMI3: ACEI or ARB for LVSD					
under 65 years	30,729	31,955	96.2	ref.	ref.
65 to 74 years	16,782	17,608	95.3	0.81 (0.74-0.89)	<0.001
75 to 84 years	16,144	17,053	94.7	0.71 (0.65-0.77)	<0.001
85 or older	9,631	10,265	93.8	0.61 (0.55-0.67)	<0.001
AMI4: Smoking cessation counseling					
under 65 years	101,819	102,305	99.5	ref.	ref.
65 to 74 years	23,569	23,794	99.1	0.50 (0.43-0.59)	<0.001
75 to 84 years	8,919	9,074	98.3	0.27 (0.23-0.33)	<0.001
85 or older	1,762	1,813	97.2	0.16 (0.12-0.22)	<0.001
AMI5: Beta-blocker at discharge					
under 65 years	181,451	184,294	98.5	ref.	ref.
65 to 74 years	85,291	86,894	98.2	0.83 (0.78-0.89)	<0.001
75 to 84 years	76,749	78,361	97.9	0.75 (0.70-0.79)	<0.001
85 or older	46,654	47,979	97.2	0.55 (0.52-0.59)	<0.001
AMI7a: Fibrinolytic within 30 minutes					
under 65 years	648	1,212	53.5	ref.	ref.
65 to 74 years	194	358	54.2	1.03 (0.81-1.30)	0.810
75 to 84 years	93	202	46.0	0.74 (0.55-1.00)	0.051
85 or older	49	98	50.0	0.87 (0.58-1.31)	0.508
AMI8a: PCI within 90 minutes					
under 65 years	31,621	35,686	88.6	ref.	ref.
65 to 74 years	9,116	10,546	86.4	0.82 (0.77-0.87)	<0.001
75 to 84 years	5,398	6,466	83.5	0.65 (0.60-0.70)	<0.001
85 or older	2,040	2,451	83.2	0.64 (0.57-0.71)	<0.001
HF1: Discharge instructions					
under 65 years	178,658	207,594	86.1	ref.	ref.
65 to 74 years	123,528	143,712	86.0	0.99 (0.97-1.01)	0.373
75 to 84 years	151,451	175,244	86.4	1.03 (1.01-1.05)	0.001
85 or older	97,755	112,707	86.7	1.06 (1.04-1.08)	<0.001
HF2: Evaluation of LV function					

under 65 years	216,443	221,533	97.7	ref.	ref.
65 to 74 years	162,507	166,888	97.4	0.87 (0.84-0.91)	<0.001
75 to 84 years	220,926	227,028	97.3	0.85 (0.82-0.88)	<0.001
85 or older	169,548	175,750	96.5	0.64 (0.62-0.67)	<0.001
HF3: ACEI or ARB for LVSD					
under 65 years	95,238	99,651	95.6	ref.	ref.
65 to 74 years	52,803	56,622	93.3	0.64 (0.61-0.67)	<0.001
75 to 84 years	58,917	63,666	92.5	0.57 (0.55-0.60)	<0.001
85 or older	33,681	36,742	91.7	0.51 (0.49-0.53)	<0.001
HF4: Smoking cessation counseling					
under 65 years	78,879	80,061	98.5	ref.	ref.
65 to 74 years	31,278	32,007	97.7	0.64 (0.59-0.71)	<0.001
75 to 84 years	17,689	18,260	96.9	0.46 (0.42-0.51)	<0.001
85 or older	4,402	4,599	95.7	0.33 (0.29-0.39)	<0.001
PN2: Pnemococal vaccination given or screened for					
under 65 years	--	--	--	--	--
65 to 74 years	154,049	168,347	91.5	ref.	ref.
75 to 84 years	180,579	195,787	92.2	1.10 (1.08-1.13)	<0.001
85 or older	124,772	136,849	91.2	0.96 (0.93-0.98)	0.001
PN3a: Initial blood culture within 24 hours - ICU only					
under 65 years	43,154	45,370	95.1	ref.	ref.
65 to 74 years	23,165	24,488	94.6	0.90 (0.84-0.96)	0.003
75 to 84 years	23,777	25,070	94.8	0.94 (0.88-1.01)	0.111
85 or older	13,530	14,312	94.5	0.89 (0.82-0.97)	0.006
PN3b: Initial blood culture before first antibiotic dose - ED only					
under 65 years	180,506	192,602	93.7	ref.	ref.
65 to 74 years	92,223	97,052	95.0	1.28 (1.24-1.32)	<0.001
75 to 84 years	116,268	121,901	95.4	1.38 (1.34-1.43)	<0.001
85 or older	88,051	92,159	95.5	1.44 (1.39-1.49)	<0.001
PN4: Smoking cessation counseling					
under 65 years	138,481	142,258	97.3	ref.	ref.
65 to 74 years	39,066	40,713	96.0	0.65 (0.61-0.69)	<0.001
75 to 84 years	20,330	21,389	95.0	0.52 (0.49-0.56)	<0.001
85 or older	4,673	5,027	93.0	0.36 (0.32-0.40)	<0.001
PN5c: First antibiotic dose within 6 hours					
under 65 years	196,974	210,170	93.7	ref.	ref.
65 to 74 years	103,529	109,243	94.8	1.21 (1.18-1.25)	<0.001
75 to 84 years	128,404	134,912	95.2	1.32 (1.28-1.36)	<0.001
85 or older	95,798	100,641	95.2	1.33 (1.28-1.37)	<0.001
PN6: Antibioti selection consistent with guidelines					
under 65 years	145,078	158,844	91.3	ref.	ref.
65 to 74 years	60,719	67,599	89.8	0.84 (0.81-0.86)	<0.001
75 to 84 years	74,042	81,558	90.8	0.93 (0.91-0.96)	<0.001
85 or older	52,553	57,638	91.2	0.98 (0.95-1.01)	0.255
PN7: Influenza vaccination given or screened for					
under 65 years	92,150	105,920	87.0	ref.	ref.
65 to 74 years	80,824	89,267	90.5	1.43 (1.39-1.47)	<0.001

75 to 84 years	94,637	103,395	91.5	1.61 (1.57-1.66)	<0.001
85 or older	65,988	72,586	90.9	1.49 (1.45-1.54)	<0.001
SCIP1: Antibiotic within 1 hour before incision or 2 hours for vancomycin or quinolone					
under 65 years	543,747	565,392	96.2	ref.	ref.
65 to 74 years	264,596	275,189	96.2	0.99 (0.97-1.02)	0.637
75 to 84 years	185,731	194,018	95.7	0.89 (0.87-0.92)	<0.001
85 or older	49,930	53,035	94.1	0.64 (0.62-0.67)	<0.001
SCIP2: Prophylactic antibiotic consistent with guidelines					
under 65 years	554,132	569,841	97.2	ref.	ref.
65 to 74 years	272,719	278,267	98.0	1.39 (1.35-1.44)	<0.001
75 to 84 years	192,365	196,738	97.8	1.25 (1.21-1.29)	<0.001
85 or older	51,927	53,474	97.1	0.95 (0.90-1.00)	0.066
SCIP3: Prophylactic ABX discontinued within 24 h. of surgery end time or 48 h. for cardiac surgery					
under 65 years	509,115	543,621	93.7	ref.	ref.
65 to 74 years	243,668	262,144	93.0	0.89 (0.88-0.91)	<0.001
75 to 84 years	168,265	182,048	92.4	0.83 (0.81-0.84)	<0.001
85 or older	43,548	48,116	90.5	0.65 (0.63-0.67)	<0.001
SCIP4: Controlled 6 AM postoperative serum glucose - cardiac surgery					
under 65 years	72,979	79,327	92.0	ref.	ref.
65 to 74 years	52,359	56,792	92.2	1.03 (0.99-1.07)	0.185
75 to 84 years	36,879	39,404	93.6	1.27 (1.21-1.33)	<0.001
85 or older	4,704	4,942	95.2	1.72 (1.51-1.96)	<0.001
SCIP6: appropriate hair removal					
under 65 years	810,303	818,220	99.0	ref.	ref.
65 to 74 years	380,445	383,750	99.1	1.12 (1.08-1.17)	<0.001
75 to 84 years	279,516	281,752	99.2	1.22 (1.17-1.28)	<0.001
85 or older	87,319	87,891	99.3	1.49 (1.37-1.62)	<0.001
SCIPCARD2: Perioperative period beta blocker					
under 65 years	143,202	157,742	90.8	ref.	ref.
65 to 74 years	125,183	136,865	91.5	1.09 (1.06-1.12)	<0.001
75 to 84 years	101,842	111,827	91.1	1.04 (1.01-1.06)	0.010
85 or older	29,959	33,499	89.4	0.86 (0.83-0.89)	<0.001
SCIPVTE1: Recommended VTE prophylaxis ordered during admission					
under 65 years	204,866	222,992	91.9	ref.	ref.
65 to 74 years	111,168	117,886	94.3	1.46 (1.42-1.51)	<0.001
75 to 84 years	92,459	97,769	94.6	1.54 (1.49-1.59)	<0.001
85 or older	35,581	37,747	94.3	1.45 (1.39-1.52)	<0.001
SCIPVTE2: Received VTE prophylaxis within 24 hours prior to or after surgery					
under 65 years	199,284	221,436	90.0	ref.	ref.
65 to 74 years	108,467	117,367	92.4	1.35 (1.32-1.39)	<0.001
75 to 84 years	90,083	97,336	92.5	1.38 (1.34-1.42)	<0.001
85 or older	34,507	37,557	91.9	1.26 (1.21-1.31)	<0.001

Disparities analysis for 26 performance measures using 2009 Clinical Data Warehouse

By Census Region

Measures and census region	Num	Den	Percent	Unadjusted OR (95%CI)	p-value
AMI1: Aspirin at arrival					
South	126,608	129,145	98.0	ref.	ref.
Midwest	75,072	76,242	98.5	1.29 (1.20-1.38)	<0.001
Northeast	62,335	63,302	98.5	1.29 (1.20-1.39)	<0.001
West	61,600	62,432	98.7	1.48 (1.37-1.61)	<0.001
US Territories	3,752	4,167	90.0	0.18 (0.16-0.20)	<0.001
AMI2: Aspirin at discharge					
South	154,361	157,475	98.0	ref.	ref.
Midwest	96,702	98,082	98.6	1.41 (1.33-1.51)	<0.001
Northeast	72,945	73,951	98.6	1.46 (1.36-1.57)	<0.001
West	71,443	72,548	98.5	1.30 (1.22-1.40)	<0.001
US Territories	3,142	3,683	85.3	0.12 (0.11-0.13)	<0.001
AMI3: ACEI or ARB for LVSD					
South	30,162	31,629	95.4	ref.	ref.
Midwest	17,573	18,369	95.7	1.07 (0.98-1.17)	0.114
Northeast	13,443	14,124	95.2	0.96 (0.87-1.05)	0.392
West	11,325	11,875	95.4	1.00 (0.91-1.11)	0.977
US Territories	783	884	88.6	0.38 (0.30-0.47)	<0.001
AMI4: Smoking cessation counseling					
South	59,052	59,326	99.5	ref.	ref.
Midwest	34,282	34,529	99.3	0.64 (0.54-0.77)	<0.001
Northeast	21,314	21,497	99.1	0.54 (0.45-0.65)	<0.001
West	20,782	20,940	99.2	0.61 (0.50-0.74)	<0.001
US Territories	639	694	92.1	0.05 (0.04-0.07)	<0.001
AMI5: Beta-blocker at discharge					
South	150,602	153,698	98.0	ref.	ref.
Midwest	94,600	96,058	98.5	1.33 (1.25-1.42)	<0.001
Northeast	72,919	73,919	98.6	1.50 (1.40-1.61)	<0.001
West	68,776	70,048	98.2	1.11 (1.04-1.19)	0.002
US Territories	3,248	3,805	85.4	0.12 (0.11-0.13)	<0.001
AMI7a: Fibrinolytic within 30 minutes					
South	386	691	55.9	ref.	ref.
Midwest	71	157	45.2	0.65 (0.46-0.92)	0.016
Northeast	114	221	51.6	0.84 (0.62-1.14)	0.266
West	325	577	56.3	1.02 (0.82-1.27)	0.868
US Territories	88	224	39.3	0.51 (0.38-0.70)	<0.001
AMI8a: PCI within 90 minutes					
South	18,249	21,033	86.8	ref.	ref.
Midwest	12,047	13,530	89.0	1.24 (1.16-1.33)	<0.001
Northeast	7,776	8,945	86.9	1.01 (0.94-1.09)	0.695
West	10,077	11,545	87.3	1.05 (0.98-1.12)	0.182

US Territories	26	96	27.1	0.06 (0.04-0.09)	<0.001
HF1: Discharge instructions					
South	230,620	268,753	85.8	ref.	ref.
Midwest	123,214	142,800	86.3	1.04 (1.02-1.06)	<0.001
Northeast	104,441	118,681	88.0	1.21 (1.19-1.24)	<0.001
West	87,789	101,987	86.1	1.02 (1.00-1.04)	0.037
US Territories	5,328	7,036	75.7	0.52 (0.49-0.55)	<0.001
HF2: Evaluation of LV function					
South	313,881	323,530	97.0	ref.	ref.
Midwest	177,519	182,711	97.2	1.05 (1.02-1.09)	0.004
Northeast	154,546	157,057	98.4	1.89 (1.81-1.98)	<0.001
West	117,503	120,882	97.2	1.07 (1.03-1.11)	0.001
US Territories	5,975	7,019	85.1	0.18 (0.16-0.19)	<0.001
HF3: ACEI or ARB for LVSD					
South	102,341	109,272	93.7	ref.	ref.
Midwest	54,335	57,985	93.7	1.01 (0.97-1.05)	0.700
Northeast	44,314	47,239	93.8	1.03 (0.98-1.07)	0.259
West	37,449	39,660	94.4	1.15 (1.09-1.21)	<0.001
US Territories	2,200	2,525	87.1	0.46 (0.41-0.52)	<0.001
HF4: Smoking cessation counseling					
South	60,779	61,825	98.3	ref.	ref.
Midwest	30,645	31,366	97.7	0.73 (0.66-0.81)	<0.001
Northeast	20,880	21,315	98.0	0.83 (0.74-0.92)	<0.001
West	19,359	19,792	97.8	0.77 (0.69-0.86)	<0.001
US Territories	585	629	93.0	0.23 (0.17-0.31)	<0.001
PN2: Pnemococal vaccination given or screened for					
South	179,960	194,612	92.5	ref.	ref.
Midwest	114,202	124,453	91.8	0.91 (0.88-0.93)	<0.001
Northeast	88,746	95,893	92.5	1.01 (0.98-1.04)	0.466
West	75,360	83,017	90.8	0.80 (0.78-0.82)	<0.001
US Territories	1,132	3,008	37.6	0.05 (0.05-0.05)	<0.001
PN3a: Initial blood culture within 24 hours - ICU only					
South	41,731	43,940	95.0	ref.	ref.
Midwest	24,196	25,563	94.7	0.94 (0.87-1.00)	0.065
Northeast	16,787	17,632	95.2	1.05 (0.97-1.14)	0.225
West	20,703	21,725	95.3	1.07 (0.99-1.16)	0.072
US Territories	209	380	55.0	0.06 (0.05-0.08)	<0.001
PN3b: Initial blood culture before first antibiotic dose - ED only					
South	187,438	197,520	94.9	ref.	ref.
Midwest	110,172	115,477	95.4	1.12 (1.08-1.16)	<0.001
Northeast	93,600	98,873	94.7	0.95 (0.92-0.99)	0.008
West	83,935	89,171	94.1	0.86 (0.83-0.89)	<0.001
US Territories	1,903	2,673	71.2	0.13 (0.12-0.14)	<0.001
PN4: Smoking cessation counseling					
South	91,072	93,604	97.3	ref.	ref.
Midwest	48,987	51,087	95.9	0.65 (0.61-0.69)	<0.001
Northeast	32,410	33,325	97.3	0.98 (0.91-1.06)	0.695

West	29,466	30,694	96.0	0.67 (0.62-0.72)	<0.001
US Territories	615	677	90.8	0.28 (0.21-0.36)	<0.001
PN5c: First antibiotic dose within 6 hours					
South	208,883	220,861	94.6	ref.	ref.
Midwest	128,036	134,173	95.4	1.20 (1.16-1.23)	<0.001
Northeast	96,895	102,680	94.4	0.96 (0.93-0.99)	0.014
West	88,422	93,297	94.8	1.04 (1.01-1.08)	0.024
US Territories	2,469	3,955	62.4	0.10 (0.09-0.10)	<0.001
PN6: Antibioti selection consistent with guidelines					
South	134,164	147,904	90.7	ref.	ref.
Midwest	78,294	86,405	90.6	0.99 (0.96-1.02)	0.434
Northeast	59,152	63,980	92.5	1.25 (1.21-1.30)	<0.001
West	58,295	63,887	91.2	1.07 (1.03-1.10)	<0.001
US Territories	2,487	3,463	71.8	0.26 (0.24-0.28)	<0.001
PN7: Influenza vaccination given or screened for					
South	136,798	151,103	90.5	ref.	ref.
Midwest	82,023	90,887	90.2	0.97 (0.94-0.99)	0.021
Northeast	60,341	66,389	90.9	1.04 (1.01-1.08)	0.008
West	53,674	60,817	88.3	0.79 (0.76-0.81)	<0.001
US Territories	763	1,972	38.7	0.07 (0.06-0.07)	<0.001
SCIP1: Antibiotic within 1 hour before incision or 2 hours for vancomycin or quinolone					
South	394,545	409,842	96.3	ref.	ref.
Midwest	266,459	276,954	96.2	0.98 (0.96-1.01)	0.223
Northeast	193,461	200,392	96.5	1.08 (1.05-1.11)	<0.001
West	183,368	192,227	95.4	0.80 (0.78-0.82)	<0.001
US Territories	6,171	8,219	75.1	0.12 (0.11-0.12)	<0.001
SCIP2: Prophylactic antibiotic consistent with guidelines					
South	403,132	414,194	97.3	ref.	ref.
Midwest	273,589	279,578	97.9	1.25 (1.21-1.29)	<0.001
Northeast	197,917	202,575	97.7	1.17 (1.13-1.21)	<0.001
West	189,102	194,077	97.4	1.04 (1.01-1.08)	0.015
US Territories	7,403	7,896	93.8	0.41 (0.38-0.45)	<0.001
SCIP3: Prophylactic ABX discontinued within 24 h. of surgery end time or 48 h. for cardiac surgery					
South	361,060	388,513	92.9	ref.	ref.
Midwest	248,442	264,681	93.9	1.16 (1.14-1.19)	<0.001
Northeast	180,683	191,769	94.2	1.24 (1.21-1.27)	<0.001
West	169,118	183,133	92.3	0.92 (0.90-0.94)	<0.001
US Territories	5,293	7,833	67.6	0.16 (0.15-0.17)	<0.001
SCIP4: Controlled 6 AM postoperative serum glucose - cardiac surgery					
South	66,018	71,829	91.9	ref.	ref.
Midwest	40,808	44,136	92.5	1.08 (1.03-1.13)	<0.001
Northeast	29,288	30,993	94.5	1.51 (1.43-1.60)	<0.001
West	29,005	31,251	92.8	1.14 (1.08-1.20)	<0.001
US Territories	1,802	2,256	79.9	0.35 (0.31-0.39)	<0.001
SCIP6: appropriate hair removal					
South	587,629	592,145	99.2	ref.	ref.
Midwest	385,646	388,859	99.2	0.92 (0.88-0.97)	<0.001

Northeast	297,284	299,532	99.2	1.02 (0.97-1.07)	0.532
West	279,180	282,116	99.0	0.73 (0.70-0.77)	<0.001
US Territories	7,844	8,961	87.5	0.05 (0.05-0.06)	<0.001
SCIPCARD2: Perioperative period beta blocker					
South	147,784	162,051	91.2	ref.	ref.
Midwest	106,546	117,054	91.0	0.98 (0.95-1.01)	0.113
Northeast	85,381	92,184	92.6	1.21 (1.18-1.25)	<0.001
West	59,482	67,099	88.6	0.75 (0.73-0.78)	<0.001
US Territories	993	1,545	64.3	0.17 (0.16-0.19)	<0.001
SCIPVTE1: Recommended VTE prophylaxis ordered during admission					
South	169,988	182,774	93.0	ref.	ref.
Midwest	99,327	106,377	93.4	1.06 (1.03-1.09)	<0.001
Northeast	96,401	100,803	95.6	1.65 (1.59-1.71)	<0.001
West	76,837	84,597	90.8	0.74 (0.72-0.77)	<0.001
US Territories	1,521	1,843	82.5	0.36 (0.31-0.40)	<0.001
SCIPVTE2: Received VTE prophylaxis within 24 hours prior to or after surgery					
South	164,922	181,622	90.8	ref.	ref.
Midwest	96,639	105,893	91.3	1.06 (1.03-1.09)	<0.001
Northeast	94,639	100,532	94.1	1.63 (1.58-1.68)	<0.001
West	74,698	83,964	89.0	0.82 (0.79-0.84)	<0.001
US Territories	1,443	1,685	85.6	0.60 (0.53-0.69)	<0.001

Disparities analysis for 26 performance measures using 2009 Clinical Data Warehouse

By Hospital Rural/Urban Location (less than 0.1 of cases were excluded due to missing data on hospital rural/urban location)

Measures and hospital rural/urban location	Num	Den	Percent	Unadjusted OR (95%CI)	p-value
AMI1: Aspirin at arrival					
Urban	291,143	295,802	98.4	ref.	ref.
Rural	38,206	39,467	96.8	0.48 (0.46-0.52)	<0.001
AMI2: Aspirin at discharge					
Urban	358,943	364,751	98.4	ref.	ref.
Rural	39,639	40,973	96.7	0.48 (0.45-0.51)	<0.001
AMI3: ACEI or ARB for LVSD					
Urban	65,715	68,816	95.5	ref.	ref.
Rural	7,570	8,064	93.9	0.72 (0.66-0.80)	<0.001
AMI4: Smoking cessation counseling					
Urban	122,296	123,021	99.4	ref.	ref.
Rural	13,772	13,964	98.6	0.43 (0.36-0.50)	<0.001
AMI5: Beta-blocker at discharge					
Urban	350,908	356,917	98.3	ref.	ref.
Rural	39,223	40,596	96.6	0.49 (0.46-0.52)	<0.001
AMI7a: Fibrinolytic within 30 minutes					
Urban	743	1,378	53.9	ref.	ref.
Rural	241	491	49.1	0.82 (0.67-1.01)	0.066
AMI8a: PCI within 90 minutes					
Urban	44,330	50,581	87.6	ref.	ref.
Rural	3,845	4,568	84.2	0.75 (0.69-0.82)	<0.001
HF1: Discharge instructions					
Urban	462,198	530,366	87.1	ref.	ref.
Rural	89,161	108,850	81.9	0.67 (0.66-0.68)	<0.001
HF2: Evaluation of LV function					
Urban	640,201	651,626	98.2	ref.	ref.
Rural	129,180	139,524	92.6	0.22 (0.22-0.23)	<0.001
HF3: ACEI or ARB for LVSD					
Urban	204,835	216,883	94.4	ref.	ref.
Rural	35,794	39,788	90.0	0.53 (0.51-0.55)	<0.001

HF4: Smoking cessation counseling					
Urban	109,946	111,420	98.7	ref.	ref.
Rural	22,294	23,495	94.9	0.25 (0.23-0.27)	<0.001
PN2: Pnemococal vaccination given or screened for					
Urban	343,445	372,029	92.3	ref.	ref.
Rural	115,907	128,899	89.9	0.74 (0.73-0.76)	<0.001
PN3a: Initial blood culture within 24 hours - ICU only					
Urban	82,609	86,195	95.8	ref.	ref.
Rural	21,017	23,045	91.2	0.45 (0.43-0.48)	<0.001
PN3b: Initial blood culture before first antibiotic dose - ED only					
Urban	370,713	390,752	94.9	ref.	ref.
Rural	106,285	112,910	94.1	0.87 (0.84-0.89)	<0.001
PN4: Smoking cessation counseling					
Urban	153,343	157,007	97.7	ref.	ref.
Rural	49,195	52,364	93.9	0.37 (0.35-0.39)	<0.001
PN5c: First antibiotic dose within 6 hours					
Urban	391,112	414,535	94.3	ref.	ref.
Rural	133,539	140,375	95.1	1.17 (1.14-1.20)	<0.001
PN6: Antibioti selection consistent with guidelines					
Urban	244,813	267,228	91.6	ref.	ref.
Rural	87,548	98,376	89.0	0.74 (0.72-0.76)	<0.001
PN7: Influenza vaccination given or screened for					
Urban	250,927	277,437	90.4	ref.	ref.
Rural	82,639	93,694	88.2	0.79 (0.77-0.81)	<0.001
SCIP1: Antibiotic within 1 hour before incision or 2 hours for vancomycin or quinolone					
Urban	873,006	907,766	96.2	ref.	ref.
Rural	170,887	179,749	95.1	0.77 (0.75-0.79)	<0.001
SCIP2: Prophylactic antibiotic consistent with guidelines					
Urban	895,997	917,696	97.6	ref.	ref.
Rural	175,035	180,505	97.0	0.77 (0.75-0.80)	<0.001
SCIP3: Prophylactic ABX discontinued within 24 h. of surgery end time or 48 h. for cardiac surgery					
Urban	805,137	863,438	93.2	ref.	ref.
Rural	159,351	172,373	92.4	0.89 (0.87-0.90)	<0.001
SCIP4: Controlled 6 AM postoperative serum glucose - cardiac surgery					
Urban	155,675	168,209	92.5	ref.	ref.
Rural	11,246	12,256	91.8	0.90 (0.84-0.96)	0.001

SCIP6: appropriate hair removal					
Urban	1,304,767	1,316,311	99.1	ref.	ref.
Rural	252,581	255,064	99.0	0.90 (0.86-0.94)	<0.001
SCIPCARD2: Perioperative period beta blocker					
Urban	341,816	374,870	91.2	ref.	ref.
Rural	58,327	65,020	89.7	0.84 (0.82-0.87)	<0.001
SCIPVTE1: Recommended VTE prophylaxis ordered during admission					
Urban	368,551	393,488	93.7	ref.	ref.
Rural	75,501	82,880	91.1	0.69 (0.67-0.71)	<0.001
SCIPVTE2: Received VTE prophylaxis within 24 hours prior to or after surgery					
Urban	358,864	391,436	91.7	ref.	ref.
Rural	73,455	82,235	89.3	0.76 (0.74-0.78)	<0.001

SURGICAL IMPROVEMENT PROJECT (SCIP) CART PAPER TOOL

Provider Name: _____

CMS Certification Number (CCN): _____

National Provider Identifier (NPI): _____

Health Care Organization Identifier (HCOID): (Joint Commission Required) _____

First Name: _____

Last Name: _____

Sex: ☐ Female ☐ Male ☐ Unknown

Birthdate: _____
Dates are MM-DD-YYYY. UTD is not an allowable entry.

Race: (Select one option)
- ☐ White
- ☐ Black or African American
- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Native Hawaiian or Pacific Islander
- ☐ UTD

Hispanic Ethnicity:
- ☐ No
- ☐ Yes

Hospital Patient ID: _____
Up to 40 letters, numbers, and/or characters.

Admission Date: _____
Dates are MM-DD-YYYY. UTD is not an allowable entry.

Discharge Date: _____
Dates are MM-DD-YYYY. UTD is not an allowable entry.

Abstractor ID: _____

Abstraction Date: _____
Dates are MM-DD-YYYY. UTD is not an allowable entry.

Vendor Tracking ID:
(Joint Commission Required) _____

1. **Would you like the questions to be enabled or disabled appropriately per the measure algorithms, or do you want all questions enabled? (SKIPPATTERN**) (Data Entry Question Only)
2. **What was the ICD-9-CM code selected as the principal diagnosis for this record? (PRINDX)** (Format three digits period two digits):

   _____

3. **Were there ICD-9-CM Other Diagnosis Codes?(OTHRDX#A)**
   (Format three digits period two digits):

   _____ _____ _____

   _____ _____ _____

   _____ _____ _____

   _____ _____ _____

4. **Was there an ICD-9-CM code selected as the principal procedure for this record?**

ICD-9-CM Principal Procedure Code (PRINPXA) (Format three digits period two digits):	**Date Performed (PRINPXDATE)** Dates are (MM-DD-YYYY or UTD)
_____	_____

5. **Were there ICD-9-CM other Procedure Codes?**

ICD-9-CM Other Procedure Code(s) (OTHERPX#A) (Format three digits period two digits):	**Date Performed (OTHERPX#DT)** (Dates are MM-DD-YYYY or UTD)

6. **What is the patient's source of payment for this Episode of Care? (PMTSRCE)**

☐ Source of payment is Medicare
☐ Source of payment is Non-Medicare

7. **What is the patient's Medicare/HIC number? (PTHIC)** (Required for data transmission of all cases that have a standard HIC#, All alpha characters must be upper case)

_____

8. **What is the postal code of the patient's residence? (POSTALCODE)**
(Five or nine digits, HOMELESS or NON-US)

_____

9. **Does this case represent part of a sample? (SAMPLE)**

☐ Yes
☐ No

10. What was the patient's discharge disposition? (DISCHGSTAT)

- ☐ 01 Discharged to home care or self care (routine discharge)
- ☐ 02 Discharged/transferred to a short term general hospital for inpatient care
- ☐ 03 Discharged/transferred to skilled nursing facility (SNF) with Medicare certification in anticipation of skilled care
- ☐ 04 Discharged/transferred to a facility that provides custodial or supportive care
- ☐ 05 Discharged/transferred to a designated cancer center or children's hospital
- ☐ 06 Discharged/transferred to home under care of organized home health service organization in anticipation of covered skilled care
- ☐ 07 Left against medical advice or discontinued care
- ☐ 20 Expired
- ☐ 21 Discharged/transferred to court/law enforcement
- ☐ 43 Discharged/transferred to a federal health care facility
- ☐ 50 Hospice - home
- ☐ 51 Hospice - medical facility (certified) providing hospice level of care
- ☐ 61 Discharged/transferred to hospital-based Medicare approved swing bed
- ☐ 62 Discharged/transferred to an inpatient rehabilitation facility (IRF) including rehabilitation distinct part units of a hospital
- ☐ 63 Discharged/transferred to a Medicare certified long term care hospital (LTCH)
- ☐ 64 Discharged/transferred to a nursing facility certified under Medicaid but not certified under Medicare
- ☐ 65 Discharged/transferred to a psychiatric distinct part unit of a hospital
- ☐ 66 Discharged/transferred to a Critical Access Hospital (CAH)
- ☐ 70 Discharged/transferred to another type of health care institution not defined elsewhere in this code list (See Code 05)

11. Was the procedure performed entirely by laparoscope or other fiber optic scope? (LAPAROSCOPE)

- ☐ Yes
- ☐ No
- ☐ UTD

12. During this hospital stay, was the patient enrolled in a clinical trial in which patients with the same condition as the measure set were being studied (CLNCLTRIAL)

- ☐ Yes
- ☐ No

13. Is there documentation that the patient was on continuous warfarin prior to admission? (PREADWARFARIN)

- ☐ Yes
- ☐ No

14. On what date did the anesthesia for the procedure start? (ANESTSTARTDT)
Dates are in MM-DD-YYYY format unless specified

_____

☐ UTD

15. Did the patient have an infection during this hospitalization prior to the principal procedure? (INFECPTA)

☐ Yes
☐ No

16. Is there documentation that the patient expired during the timeframe from surgical incision through discharge from the post anesthesia care/recovery area? (PERIOPDEATH)

☐ Yes
☐ No

17. Were there any other procedures requiring general or spinal/epidural anesthesia that occurred within three days (four days for CABG or Other Cardiac Surgery) prior to or after the principal procedure during this hospital stay? (OTHERSURG)

☐ Yes
☐ No

18. Did the patient receive antibiotics within 24 hours of arrival or the day prior to arrival and/or during this hospital stay? (ANTIBIRCVD)

☐ Antibiotic received only within 24 hours of arrival or the day prior to arrival and not during hospital stay.
☐ Antibiotic received within 24 hours of arrival or the day prior to arrival and during hospital stay (arrival through 24 hours for PN and arrival through 48 hours postop [72 hours post op for CABG or Other Cardiac Surgery] for SCIP-Inf).
☐ Antibiotic received only during hospital stay (arrival through 24 hours for PN and arrival through 48 hours postop [72 hours post op for CABG or Other Cardiac Surgery] for SCIP-Inf).
☐ Antibiotic not received (within 24 hours of arrival or arrival through 24 hours for PN and arrival through 48 hours postop [72 hours post op for CABG or Other Cardiac Surgery] for SCIP-Inf), or unable to determine from medical record documentation.

19. What were the antibiotics administered any time after hospital arrival and within the specified timeframe? (ABXDETAILS)

Antibiotic Name (NAMEABX) (trade or generic) see Appendix C, Table 2.1.	**Antibiotic Administration Date (DTABX)** Dates are MM-DD-YYYY or UTD	**Antibiotic Administration Time (TMABX)** Times are military format HH:MM or UTD	**Antibiotic Administration Route (ROUTEABX)** Format: 1=PO/NG/PEG tube (Oral) 2=IV (Intravenous) 3=IM (Intramuscular) 10=UTD

20. **Were the only antibiotic combinations administered prior to hospital arrival or more than 24 hours prior to incision either oral Neomycin Sulfate + Erythromycin Base or oral Neomycin Sulfate + Metronidazole? (ORALANTIBIOTIC)**

☐ Yes
☐ No

21. **At what time was the anesthesia initiated for the principal procedure? (ANESTSTARTTM)HH:MM military format**

_____

☐ UTD

22. **At what time was the initial incision made for the principal procedure? (SURGINCISTM) HH:MM military format**

_____

☐ UTD

23. **On what date was the incision for the principal procedure made? (SURGINCISDT)** Dates are in MM-DD-YYYY format unless specified

_____

☐ UTD

24. **On what date did the anesthesia for the for the principal procedure end? (ANESTHENDDATE)** Dates are in MM-DD-YYYY format unless specified

_____

☐ UTD

25. **At what time did the anesthesia for the principal procedure end? (ANESTHENDTIME) HH:MM military format**

_____

☐ UTD

26. **What reason was documented postoperatively by the physician/APN/PA for extending the duration of the antibiotic administration past 24 hours (48 hours for CABG or Other Cardiac Surgery) after *Anesthesia End Time*?(RSNEXTABX)** ***(Select all that apply)***

☐ There is physician/advanced practice nurse/physician assistant (physician/APN/PA) documentation within 2 days (3 days for CABG or Other Cardiac Surgery) following the principal procedure with the day of surgery being day zero that erythromycin was administered postoperatively for the purpose of increasing gastric motility.

- ☐ There is physician/APN/PA documentation within 2 days (3 days for CABG or Other Cardiac Surgery) following the principal procedure with the day of surgery being day zero that an antibiotic was administered postoperatively for the treatment of hepatic encephalopathy.
- ☐ There is physician/APN/PA documentation within 2 days (3 days for CABG or Other Cardiac Surgery) following the principal procedure with the day of surgery being day zero that an antibiotic was administered postoperatively as prophylaxis of Pneumocystis pneumonia (PCP) to a patient with a diagnosis of AIDS.
- ☐ There is physician/APN/PA documentation within 2 days (3 days for CABG or Other Cardiac Surgery) following the principal procedure with the day of surgery being day zero that the patient had an infection.
- ☐ There is physician/APN/PA documentation within 2 days following the principal procedure with the day of surgery being day zero that the patient has a current malignancy of the lower extremity involving the same extremity as the principal procedure that was an original arthroplasty or a joint revision surgery.
- ☐ There is documentation within 2 days following the principal procedure with the day of surgery being day zero that the principal procedure was a joint revision surgery.
- ☐ No documented reason/Unable to Determine.

27. **What method of surgical site hair removal was performed prior to the principal procedure? (PREOPHRREM)** ***(Select all that apply)***

- ☐ No documented hair removal or no hair removal performed
- ☐ Razor
- ☐ Clippers/Scissors
- ☐ Depilatory
- ☐ Other
- ☐ Patient performed their own hair removal
- ☐ Unable to determine method
- ☐ Hair removal with a razor from the scrotal area OR from the scalp after a current traumatic head injury

28. Was there documentation that the procedure was performed using general or neuraxial anesthesia? (ANESTTYPE)

- ☐ There is documentation that the procedure was performed using general anesthesia.
- ☐ There is documentation that the procedure was performed using neuraxial anesthesia.
- ☐ There is documentation that the procedure was performed using **both** neuraxial and general anesthesia.
- ☐ There is no documentation that the procedure was performed using either general or neuraxial anesthesia or unable to determine from the medical record documentation.

29. Was there documentation that intentional hypothermia was utilized during the perioperative period? (INTENTHYPO)

☐ Yes
☐ No

30. Was there documentation of active warming used intraoperatively OR at least one body temperature equal to or greater than 96.8 degrees F/36 degrees C within the 30 minutes immediately prior to or the 15 minutes immediately after *Anesthesia End Time* in the medical record?(TEMPERATURE) *(Select all that apply)*

☐ 1 Active warming was performed intraoperatively.
☐ 2 There is documentation of at least one body temperature greater than or equal to 96.8 degrees F/36 degrees C within the 30 minutes immediately prior to or the 15 minutes immediately after Anesthesia End Time.
☐ 3 There is no documentation of Allowable Values 1 AND 2.
☐ 4 Unable to determine from the medical record documentation.

31. Is there documentation that the patient had a urinary catheter paced in the perioperative timeframe and that it was still in place at the time of discharge from the recovery/post-anesthesia care area? (URINECATH)

☐ There is documentation that an indwelling urethral catheter was placed perioperatively and was still in place at the time of discharge from the recovery/post-anesthesia care area.
☐ There is no documentation that an indwelling urethral catheter was placed perioperatively and was still in place at the time of discharge from the recovery/post-anesthesia care area.
☐ There is documentation that the patient had an indwelling urethral or suprapubic catheter or was being intermittently catheterized prior to the perioperative timeframe.
☐ There is documentation that the patient had a suprapubic catheter placed perioperatively and was still in place at the time of discharge from the recovery/post-anesthesia care area or the patient was being intermittently catheterized during the perioperative period.
☐ Unable to determine whether the patient had a catheter in place from medical record documentation.

32. Is there documentation that the urinary catheter was removed on POD 0 through POD 2 with the *Anesthesia End Date* being POD 0? (CATHREMOVE)

☐ There is documentation that the urinary catheter was removed on POD 0 through POD 2.
☐ There is no documentation that the urinary catheter was removed on POD 0 through POD 2.
☐ Unable to determine (UTD) from medical record documentation whether the urinary catheter was removed on POD 0 through POD 2.

33. Was there documentation of reason(s) for not removing the urinary catheter postoperatively? (REASONCNTCATH)

- ☐ There is documentation that the patient was in the intensive care unit (ICU) AND receiving diuretics.
- ☐ There is physician/advanced practice nurse/physician assistant (physician/APN/PA) documentation of reasons for not removing the urinary catheter postoperatively.
- ☐ There is no physician/APN/PA documentation of reasons for not removing the urinary catheter postoperatively or unable to determine from medical record documentation.

34. Is there documentation that the patient was on a daily beta-blocker therapy prior to arrival? (BBLKRCURRENT)

- ☐ Yes
- ☐ No

35. Was the patient taking the beta-blocker prior to arrival pregnant? (BBLKRPREG)

- ☐ Yes
- ☐ No
- ☐ UTD

36. Is there documentation that a beta-blocker was received during the perioperative period? (BBLKRPERIOP)

- ☐ Yes
- ☐ No

37. Was there documentation of reasons for not administering a beta-blocker during the perioperative period? (CTRBBLKPERIOP)

- ☐ Yes
- ☐ No

38. Is there documentation by a physician/advanced practice nurse/physician assistant (physician/APN/PA) or pharmacist in the medical record of a reason for not administering pharmacological and/or mechanical VTE prophylaxis? (CONTRAVTEPRO)

- ☐ There is physician/APN/PA or pharmacist documentation of a reason for not administering mechanical VTE prophylaxis.
- ☐ There is physician/APN/PA or pharmacist documentation of a reason for not administering pharmacological VTE prophylaxis.
- ☐ There is physician/APN/PA or pharmacist documentation of a reason for not administering both mechanical and pharmacological VTE prophylaxis.
- ☐ There is no physician/APN/PA or pharmacist documentation of a reason for not administering either mechanical or pharmacological VTE prophylaxis or unable to determine from medical record documentation.

39. **What type of VTE prophylaxis was documented in the medical record? (Collect any VTE prophylaxis that was ordered at anytime from hospital arrival to 24 hours after Anesthesia End time). (VTEPROA)**

VTE Prophylaxis Ordered (VTEPROPH) *(Select all that apply)*	**Was VTE Prophylaxis Timely? (VTETIMELY)**
☐ Low dose unfractionated heparin (LDUH)	☐ Yes ☐ No
☐ Low molecular weight heparin (LMWH)	☐ Yes ☐ No
☐ Intermittent pneumatic compression devices (IPC)	☐ Yes ☐ No
☐ Graduated compression stocking (GCS)	☐ Yes ☐ No
☐ Factor Xa Inhibitor	☐ Yes ☐ No
☐ Warfarin	☐ Yes ☐ No
☐ Venous foot pumps (VFP)	☐ Yes ☐ No
☐ Oral Factor Xa Inhibitor	☐ Yes ☐ No
☐ None of the above or not documented or unable to determine from medical record documentation	☐ Yes ☐ No

40. Did the patient have any allergies, sensitivities or intolerance to beta-lactam/penicillin antibiotic or cephalosporin medications? (ANTIALLERGY)

☐ Yes
☐ No

41. **What reason was documented for using vancomycin? (VANCO)**

(Select all that apply)

- ☐ Documentation of beta-lactam (penicillin or cephalosporin) allergy.
- ☐ Physician/APN/PA or pharmacist documentation of MRSA colonization or infection.
- ☐ Documentation of patient being high-risk due to acute inpatient hospitalization within the last year.
- ☐ Documentation of patient being high-risk due to nursing home or extended care facility setting within the last year, prior to admission.
- ☐ Physician/APN/PA or pharmacist documentation of increased MRSA rate, either facility-wide or operation-specific.
- ☐ Physician/APN/PA or pharmacist documentation of chronic wound care or dialysis.
- ☐ Documentation of continuous inpatient stay more than 24 hours prior to the principal procedure.
- ☐ Other Physician/APN/PA or pharmacist documented reason.
- ☐ No documented reason/Unable to Determine.
- ☐ Physician/APN/PA or pharmacist documentation of patient undergoing valve surgery.
- ☐ Documentation of patient being transferred from another inpatient hospitalization after a 3-day stay.

42. What was the patient's blood glucose level on postoperative day one (POD 1) closest to 6:00 A.M.? (GLUPOD1)

_____ (1-3000 mg per dL)

☐ UTD

43. What was the patient's blood glucose level on postoperative day two (POD 2) closest to 6:00 A.M.? (GLUPOD2)

_____ (1-3000 mg per dL)

☐ UTD

44. What is the first physician identifier? (PHYSICIAN_1)

_____

45. What is the second physician identifier? (PHYSICIAN_2)

_____

This material was prepared by the IFMC (Hospital Inpatient Quality Reporting Program Contractor) under contract with the Centers for Medicare & Medicaid Service (CMS), an agency of the US Department of Health and Human Services. It is based on *The Specifications Manual for National Hospital Inpatient Quality Measures*, which is a collaborative effort of CMS, The Joint Commission, SDPS, and the Hospital Inpatient Quality Reporting Program Contractor. 9SoW-IA-HIQRP-09/10-106

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 0339 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Pediatric Heart Surgery Mortality (PDI 6)
De.2 **Brief description of measure:** Percentage of cases undergoing surgery for congenital heart disease with an in-hospital death.
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** Pediatric Heart Surgery Volume (PDI 7) (NQF #0340)
De.4 **National Priority Partners Priority Area:** Population health, Safety De.5 **IOM Quality Domain:** Effectiveness De.6 **Consumer Care Need:** Getting better

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary 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	**B** Y☐

every 3 years. Yes, information provided in contact section	N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	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:** Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** According to Odegard et al [1] despite advances in perioperative care, including monitoring and drugs, unexpected cardiac arrest remains a significant hazard during anesthesia [2- 5]. Anesthesia-related morbidity and mortality is more frequent in children than in adults, and is more frequent in infants and younger children than in older children [2,4,5,7 – 11]. Using a multivariate model that included age, complexity category, and four comorbidities, Hannan et al. found 8.26% risk-adjusted mortality at hospitals with fewer than 100 cases per year, versus 5.95% at higher volume hospitals (an effect limited to surgeons who performed at least 75 cases per year). [12] For additional material on this topic, see: URL:http://www.qualityindicators.ahrq.gov/downloads/pdi/pdi_measures_v31.pdf 1a.4 **Citations for Evidence of High Impact:** Updated citations will be presented in the May Steering Committee meeting [1] Odegard KC, DiNardo JA, Kussman BD, Shukla A, Harrington J, Casta A, McGowan FX Jr, Hickey PR, Bacha EA, Thiagarajan RR, Laussen PC. The frequency of anesthesia-related cardiac arrests in patients with congenital heart disease undergoing cardiac surgery. Anesth Analg. 2007 Aug;105(2):335-43. PMID: 17646487 [2] Cohen MM, Cameron CB, Duncan PG. Pediatric anesthesia morbidity and mortality in the perioperative	1a C☐ P☐ M☐ N☐

period. Anesth Analg 1990;70:160-7Abstract/FREE Full Text2.?
[3] Keenan RL, Boyan CP. Cardiac arrest due to anesthesia. A study of incidence and causes. JAMA 1985;253:2373-7Abstract/FREE Full Text3.
[4] Morray JP, Geiduschek JM, Ramamoorthy C, Haberkern CM, Hackel A, Caplan RA, Domino KB, Posner K, Cheney FW. Anesthesia-related cardiac arrest in children: initial findings of the Pediatric Perioperative Cardiac Arrest (POCA) Registry. Anesthesiology 2000;93:6-14Medline4.
[5] Olsson GL, Hallen B. Cardiac arrest during anaesthesia. A computer-aided study in 250,543 anaesthetics. Acta Anaesthesiol Scand 1988;32:653-64Medline5.
[6] Posner KL, Geiduschek J, Haberkern CM, Ramamoorthy C, Hackel A, Morray JP. Unexpected cardiac arrest among children during surgery: a North American registry to elucidate the incidence and causes of anesthesia related cardiac arrest. Qual Saf Health Care 2002;11:252-7Medline6.
[7] Morray JP. Anesthesia-related cardiac arrest in children. An update. Anesthesiol Clin North America 2002;20:1-287.
[8] Rackow H, Salanitre E, Green LT. Frequency of cardiac arrest associated with anesthesia in infants and children. Pediatrics 1961;28:697-704Medline8.?
[9] Murat I, Constant I, Maud´huy H. Perioperative anaesthetic morbidity in children: a database of 24,165 anaesthetics over a 30-month period. Paediatr Anaesth 2004;14:158-66CrossRefMedline9.
[10] Tay CL, Tan GM, Ng SB. Critical incidents in paediatric anaesthesia: an audit of 10 000 anaesthetics in Singapore. Paediatr Anaesth 2001;11:711-18Medline10.
[11] Braz LG, Modolo NS, do Nascimento P Jr, Bruschi BA, Castiglia YM, Ganem EM, de Carvalho LR, Braz JR. Perioperative cardiac arrest: a study of 53,718 anaesthetics over 9 yr from a Brazilian teaching hospital. Br J Anaesth 2006;96:569-75Abstract/FREE Full Text
[12] Hannan EL, Racz M, Kavey RE, Quaegebeur JM, Williams R. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. Pediatrics 1998;101(6):963-9

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Higher volume is associeted with reduced mortality and morbidity.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
Adjusted per 1,000 rates by patient and hospital characteristics, 2007

Mean	Standard error	Location	P-value: Relative to Northeast
63.931	7.946	Northeast	1.000
30.730	2.637	Midwest	0.000
44.326	1.760	South	0.016
33.496	3.316	West	0.000

1b.3 Citations for data on performance gap:
See the following report for a complete treatment of the methodology: “Methods: Applying AHRQ Quality Indicators to Healthcare Cost and Utilization Project (HCUP) Data for the National Healthcare Quality Report” [URL: http://hcupnet.ahrq.gov/QI%20Methods.pdf?JS=Y]

1b.4 Summary of Data on disparities by population group:
1) Estimate 2) Standard error 3) P-value: Relative to marked group-c 4) P-value: 2007 relative to 2006

Median income of patient´s ZIP code:
First quartile (lowest income) 44.830 2.315 0.394 0.112
Second quartile 39.643 2.577 0.671 0.053
Third quartile 32.492 2.639 0.034 0.679
Fourth quartile (highest income)c 41.414 3.276 0.043

Expected payment source:
Private insurancec 29.862 2.198 0.297
Medicare * * * DNC
Medicaid 45.617 1.707 0.000 0.129

1b
C☐
P☐
M☐
N☐

Other insurance 52.447 8.437 0.010 0.494
Uninsured / self-pay / no charge 44.691 10.293 0.159 0.182

1b.5 Citations for data on Disparities:
AHRQ 2007 Nationwide Inpatient Sample (NIS)

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*)**:** The measure focus is an outcome (mortality) that is relevant to a neonatal population with a diagnosis of congenital heart defect or procedure for congenital heart repair.

1c.2-3. Type of Evidence: Expert opinion, Systematic synthesis of research

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Using a multivariate model that included age, complexity category, and four comorbidities, Hannan et al. found 8.26% risk-adjusted mortality at hospitals with fewer than 100 cases per year, versus 5.95% at higher volume hospitals (an effect limited to surgeons who performed at least 75 cases per year). [1] Two other studies using hospital discharge data from California and Massachusetts found similar effects of hospital volume. [2] [3]
Another source of evidence is that cardiopulmonary bypass or aortic crossclamp time has been repeatedly associated with postoperative mortality, adjusting for a variety of patient characteristics.[4-7] This relationship has been demonstrated not just for the Fontan procedure, but also for the Norwood procedure for hypoplastic left heart syndrome. [8] Experienced surgeons and surgical teams should be able to reduce cardiopulmonary bypass or aortic cross-clamp time, thereby improving postoperative mortality.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
B there is moderate certainty that the net benefit is moderate to substantial (review by project team)

1c.6 Method for rating evidence: U.S. Preventive Services Task Force (USPSTF) assigns one of five letter grades to each of its recommendations (A, B, C, D, or I).

1c.7 Summary of Controversy/Contradictory Evidence: Quality-of-care evaluation must take into account variations in "case mix." One study reviewed the application of two case-mix complexity-adjustment tools in the Society of Thoracic Surgeons (STS) Congenital Heart Surgery Database: the Aristotle Basic Complexity (ABC) score and the Risk Adjustment in Congenital Heart Surgery (RACHS-1) method. With both RACHS-1 and ABC, as complexity increases, discharge mortality also ncreases. The ABC approach allows classification of more operations, whereas the RACHS-1 discriminates better at the higher end of complexity. Complexity stratification is a useful method for analyzing the impact of case mix on pediatric cardiac surgical outcomes. Both the RACHS-1 and ABC methods facilitate complexity stratification in the STS database.

1c.8 Citations for Evidence (*other than guidelines***):** Updated citations will be presented in the May Steering Committee meeting

[1] Hannan EL, Racz M, Kavey RE, Quaegebeur JM, Williams R. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. Pediatrics 1998;101(6):963-9.
[2] Jenkins KJ, Newburger JW, Lock JE, Davis RB, Coffman GA, Iezzoni LI. In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. Pediatrics 1995;95(3):323-30.
[3] Sollano JA, Gelijns AC, Moskowitz AJ, Heitjan DF, Cullinane S, Saha T, et al. Volume-outcome relationships in cardiovascular operations: New York State, 1990-1995. J Thorac Cardiovasc Surg 1999;117(3):419-28.
[4] Cetta F, Feldt RH, O´Leary PW, Mair DD, Warnes CA, Driscoll DJ, et al. Improved early morbidity and mortality after Fontan operation: the Mayo Clinic experience, 1987 to 1992. J Am Coll Cardiol 1996;28(2):480-6.
[5] Gentles TL, Gauvreau K, Mayer JE, Jr., Fishberger SB, Burnett J, Colan SD, et al. Functional outcome after the Fontan operation: factors influencing late morbidity. J Thorac Cardiovasc Surg 1997;114(3):392-403; discussion 404-5.

1c
C☐
P☐
M☐
N☐

[6] Kaulitz R, Ziemer G, Luhmer I, Kallfelz HC. Modified Fontan operation in functionally univentricular hearts: preoperative risk factors and intermediate results. J Thorac Cardiovasc Surg 1996;112(3):658-64.
[7] Fontan F, Kirklin JW, Fernandez G, Costa F, Naftel DC, Tritto F, et al. Outcome after a "perfect" Fontan operation. Circulation 1990;81(5):1520-36.
[8] Kern JH, Hayes CJ, Michler RE, Gersony WM, Quaegebeur JM. Survival and risk factor analysis for the Norwood procedure for hypoplastic left heart syndrome. Am J Cardiol 1997;80(2):170-4.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
Surgery for congenital heart disease, especially in infants, requires a setting that readily meets the complex and special needs of this group of patients. These requirements include a cardiac surgeon experienced in the operative and perioperative management of such patients. There should be a pediatric cardiologist, an anesthesia team, perfusionists, intensive care nurses, and appropriate intensive care facilities for the treatment of infants and children. At a hospital where congenital heart operations are performed, a total of 100 congenital heart operations (both open and closed, not including neonatal ductus ligations) should be done. The occasional management of an infant or child with congenital heart disease by an otherwise busy and well-functioning adult cardiac surgical team is strongly discouraged.

1c.10 Clinical Practice Guideline Citation: http://www.facs.org/fellows_info/guidelines/cardiac.html
1c.11 National Guideline Clearinghouse or other URL: Not Applicable.

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
Not Applicable.

1c.13 Method for rating strength of recommendation (*If different from USPSTF system, also describe rating and how it relates to USPSTF)*:
Not Applicable.

1c.14 Rationale for using this guideline over others:
No competing measures found.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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***):**
Number of deaths (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator with a code of pediatric heart surgery with ICD-9-CM diagnosis of congenital heart disease in any field.

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator***):**
Time window can be determined by user, but is generally a calendar year.

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions***):**

2a-specs
C☐
P☐
M☐
N☐

Number of deaths (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator with a code of pediatric heart surgery with ICD-9-CM diagnosis of congenital heart disease in any field.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
Discharges under age 18 with ICD-9-CM procedure codes for congenital heart disease (1P) in any field or non-specific heart surgery (2P) in any field with ICD-9-CM diagnosis of congenital heart disease (2D) in any field.

2a.5 Target population gender: Female, Male
2a.6 Target population age range: Age less than 18 years

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
Time window can be determined by user, but is generally a calendar year.

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Discharges under age 18 with ICD-9-CM procedure codes for congenital heart disease (1P) or non-specific heart surgery (2P) with ICD-9-CM diagnosis of congenital heart disease (2D) in any field.

Congenital heart disease procedures (1P):
3500
CLOSED VALVOTOMY NOS
3501
CLOSED AORTIC VALVOTOMY
3502
CLOSED MITRAL VALVOTOMY
3503
CLOSED PULMON VALVOTOMY
3504
CLOSED TRICUSP VALVOTOMY
3510
OPEN VALVULOPLASTY NOS
3511
OPN AORTIC VALVULOPLASTY
3512
OPN MITRAL VALVULOPLASTY
3513
OPN PULMON VALVULOPLASTY
3514
OPN TRICUS VALVULOPLASTY
3520
REPLACE HEART VALVE NOS
3521
REPLACE AORT VALV-TISSUE
3522
REPLACE AORTIC VALVE NEC
3523
REPLACE MITR VALV-TISSUE
3524
REPLACE MITRAL VALVE NEC
3525
REPLACE PULM VALV-TISSUE
3526
REPLACE PULMON VALVE NEC
3527
REPLACE TRIC VALV-TISSUE
3528
REPLACE TRICUSP VALV NEC

3531
PAPILLARY MUSCLE OPS
3532
CHORDAE TENDINEAE OPS
3533
ANNULOPLASTY
3534
INFUNDIBULECTOMY
3535
TRABECUL CARNEAE CORD OP
3539
TISS ADJ TO VALV OPS NEC
3541
ENLARGE EXISTING SEP DEF
3542
CREATE SEPTAL DEFECT
3550
PROSTH REP HRT SEPTA NOS
3551
PROS REP ATRIAL DEF-OPN
3552
PROS REPAIR ATRIA DEF-CL
3553
PROST REPAIR VENTRIC DEF
3554
PROS REP ENDOCAR CUSHION
3560
GRFT REPAIR HRT SEPT NOS
3561
GRAFT REPAIR ATRIAL DEF
3562
GRAFT REPAIR VENTRIC DEF
3563
GRFT REP ENDOCAR CUSHION
3570
HEART SEPTA REPAIR NOS
3571
ATRIA SEPTA DEF REP NEC
3572
VENTR SEPTA DEF REP NEC
3573
ENDOCAR CUSHION REP NEC
3581
TOT REPAIR TETRAL FALLOT
3582
TOTAL REPAIR OF TAPVC
3583
TOT REP TRUNCUS ARTERIOS
3584
TOT COR TRANSPOS GRT VES
3591
INTERAT VEN RETRN TRANSP
3592
CONDUIT RT VENT-PUL ART
3593
CONDUIT LEFT VENTR-AORTA
3594
CONDUIT ARTIUM-PULM ART

3595
HEART REPAIR REVISION
3598
OTHER HEART SEPTA OPS
3599
OTHER OP ON HRT VALVES
3699
OTHER OPERATIONS ON VESSEL OF HEART
3733
EXCISION OR DESTRUCTION OF OTHER LESION OR TISSUE OF HEART
3736
EXCISION OR DESTRUCTION OF LEFT ATRIAL APPENDAGE (LAA) OCT08-
375
HEART TRANSPLANTATION (invalid as of OCT03)
3751
HEART TRANSPLANTATION OCT03-
3752
IMPLANT TOT REP HRT SYS OCT03-
390
SYSTEMIC-PULM ART SHUNT
3921
CAVAL-PULMON ART ANASTOM

Non-specific cardiac procedures (2P):
3834
RESECTION OF ABDOMINAL AORTA WITH ANASTOMOSIS
3835
THOR VESSEL RESECT/ANAST
3844
RESECTION OF ABDOMINAL AORTA WITH REPLACEMENT
3845
RESECT THORAC VES W REPL
3864
OTHER EXCISION OF ABDOMINAL AORTA
3865
OTHER EXCISION OF THORACIC VESSEL
3884
OTHER SURGICAL OCCLUSION OF ABDOMINAL AORTA
3885
OCCLUDE THORACIC VES NEC
3949
OTHER REVISION OF VASCULAR PROCEDURE
3956
REPAIR OF BLOOD VESSEL WITH TISSUE PATCH GRAFT
3957
REPAIR OF BLOOD VESSEL WITH SYNTHETIC PATCH GRAFT
3958
REPAIR OF BLOOD VESSEL WITH UNSPECIFIED TYPE OF PATCH GRAFT
3959
REPAIR OF VESSEL NEC

Congenital heart disease diagnoses (2D):
7450
COMMON TRUNCUS
74510
COMPL TRANSPOS GREAT VES
74511
DOUBLE OUTLET RT VENTRIC

74512
CORRECT TRANSPOS GRT VES
74519
TRANSPOS GREAT VESS NEC
7452
TETRALOGY OF FALLOT
7453
COMMON VENTRICLE
7454
VENTRICULAR SEPT DEFECT
7455
SECUNDUM ATRIAL SEPT DEF
74560
ENDOCARD CUSHION DEF NOS
74561
OSTIUM PRIMUM DEFECT
74569
ENDOCARD CUSHION DEF NEC
7457
COR BILOCULARE
7458
SEPTAL CLOSURE ANOM NEC
7459
SEPTAL CLOSURE ANOM NOS
74600
PULMONARY VALVE ANOM NOS
74601
CONG PULMON VALV ATRESIA
74602
CONG PULMON VALVE STENOS
74609
PULMONARY VALVE ANOM NEC
7461
CONG TRICUSP ATRES/STEN
7462
EBSTEIN'S ANOMALY
7463
CONG AORTA VALV STENOSIS
7464
CONG AORTA VALV INSUFFIC
7465
CONGEN MITRAL STENOSIS
7466
CONG MITRAL INSUFFICIENC
7467
HYPOPLAS LEFT HEART SYND
74681
CONG SUBAORTIC STENOSIS
74682
COR TRIATRIATUM
74683
INFUNDIB PULMON STENOSIS
74684
OBSTRUCT HEART ANOM NEC
74685
CORONARY ARTERY ANOMALY
74687
MALPOSITION OF HEART

74689
CONG HEART ANOMALY NEC
7469
CONG HEART ANOMALY NOS
7470
PATENT DUCTUS ARTERIOSUS
74710
COARCTATION OF AORTA
74711
INTERRUPT OF AORTIC ARCH
74720
CONG ANOM OF AORTA NOS
74721
ANOMALIES OF AORTIC ARCH
74722
AORTIC ATRESIA/STENOSIS
74729
CONG ANOM OF AORTA NEC
7473
PULMONARY ARTERY ANOM
74740
GREAT VEIN ANOMALY NOS
74741
TOT ANOM PULM VEN CONNEC
74742
PART ANOM PULM VEN CONN
74749
GREAT VEIN ANOMALY NEC

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** Exclude cases:
• MDC 14 (pregnancy, childbirth and pueperium)
• with transcatheter interventions (either 3AP, 3BP, 3CP, 3DP, 3EP with 3D, or 3FP) as single cardiac procedures, performed without bypass (5P) but with catheterization (6P)
• with septal defects (4P) as single cardiac procedures without bypass (5P)
• with diagnosis of ASD or VSD (5D) with PDA as the only cardiac procedure
• heart transplant (7P)
• premature infants (4D) with PDA closure (3D and 3EP) as only cardiac procedure;
• age less than or equal to 30 days with PDA closure as only cardiac procedure
• missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1 =missing)
• transferring to another short-term hospital (DISP=2)
• neonates with birth weight less than 500 grams (Birth Weight Category 1)

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
Exclude cases:
• MDC 14 (pregnancy, childbirth and pueperium)
• with transcatheter interventions (either 3AP, 3BP, 3CP, 3DP, 3EP with 3D, or 3FP) as single cardiac procedures, performed without bypass (5P) but with catheterization (6P)
• with septal defects (4P) as single cardiac procedures without bypass (5P)
• with diagnosis of ASD or VSD (5D) with PDA as the only cardiac procedure
• heart transplant (7P)
• premature infants (4D) with PDA closure (3D and 3EP) as only cardiac procedure;
• age less than or equal to 30 days with PDA closure as only cardiac procedure
• missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1 =missing)
• transferring to another short-term hospital (DISP=2)
• neonates with birth weight less than 500 grams (Birth Weight Category 1)

A neonate is defined as any discharge with age in days at admission between zero and 28 days (inclusive). If age in days is missing, then a neonate is defined as an admission type of newborn (SID ATYPE=4) OR an ICD-9-CM code for either in-hospital live birth or neonate observation and evaluation.

Newborn in Hospital Live Birth Codes
V3000
SINGLE LB IN-HOSP W/O CS OCT05-
V3001
SINGLE LB IN-HOSP W CS OCT05-
V3100
TWIN-MATE LB-HOSP W/O CS OCT05-
V3101
TWIN-MATE LB-IN HOS W CS OCT05-
V3200
TWIN-MATE SB-HOSP W/O CS OCT05-
V3201
TWIN-MATE SB-HOSP W CS OCT05-
V3300
TWIN-NOS-IN HOSP W/O CS OCT05-
V3301
TWIN-NOS-IN HOSP W CS OCT05-
V3400
OTH MULT LB-HOSP W/O CS OCT05-
V3401
OTH MULT LB-IN HOSP W CS OCT05-
V3500
OTH MULT SB-HOSP W/O CS OCT05-
V3501
OTH MULT SB-IN HOSP W CS OCT05-
V3600
MULT LB/SB-IN HOS W/O CS OCT05-
V3601
MULT LB/SB-IN HOSP W CS OCT05-
V3700
MULT BRTH NOS-HOS W/O CS OCT05-
V3701
MULT BIRTH NOS-HOSP W CS OCT05-
V3900
LIVEBORN NOS-HOSP W/O CS OCT05-
V3901
LIVEBORN NOS-HOSP W CS OCT05-

Neonate Observation and Evaluation codes:
V290
NB OBSRV SUSPCT INFECT
V291
NB OBSRV SUSPCT NEURLGCL
V292
OBSRV NB SUSPC RESP COND
V293
NB OBS GENETC/METABL CND
V298
NB OBSRV OTH SUSPCT COND
V299
NB OBSRV UNSP SUSPCT CND

Less than 500 grams - Birth Weight Category 1

76401
LIGHT-FOR-DATES <500G
76411
LT-FOR-DATE W/MAL <500G
76421
FETAL MALNUTRITION <500G
76491
FET GROWTH RETARD <500G
76501
EXTREME IMMATUR <500G
76511
PRETERM NEC <500G
V2131
LOW BIRTHWT STATUS <500G

Closed heart valvotomy (3AP):
3500
CLOSED HEART VALVOTOMY, UNSPECIFIED VALUE
3501
CLOSED HEART VALVOTOMY, AORTIC VALUE
3502
CLOSED HEART VALVOTOMY, MITRAL VALUE
3503
CLOSED HEART VALVOTOMY, PULMONARY VALUE
3504
CLOSED HEART VALVOTOMY, TRICUSPID VALUE
Atrial septal enlargement (3BP)
3541
ENLARGEMENT OF EXISTING ATRIAL SEPTAL DEFECT
3542
CREATION OF SEPTAL DEFECT IN HEART
Atrial septal defect repair (3CP)
3551
REPAIR OF ATIAL SEPTAL DEFECT WITH PROSTHESIS, OPEN TECHNIQUE
3571
OTHER AND UNSPECIFIED REPAIR OF ATRIAL SEPTAL DEFECT

Ventricular septal defect repair (3DP):
3553
REPAIR OF VENTRICULAR SEPTAL DEFECT WITH PROSTHESIS
3572
OTHER AND UNSPECIFIED REPAIR OF VENTRICULAR SEPTAL DEFECT

Occlusion of thoracic vessel (3EP):
3885
OCCLUDE THORACIC VES NEC

PDA closure diagnosis code (3D):
7470
PATENT DUCTUS ARTERIOSUS

Other surgical occlusion (3FP):
3884
OTHER SURGICAL OCCLUSION OF AORTA, ABDOMINAL
3885
OTHER SURGICAL OCCLUSION OF THORACIC VESSEL
3959
OTHER REPAIR OF VESSEL

Atrial septal defect repair and enlargement (4P):
3541
ENLARGE EXISTING SEP DEF
3552
PROS REPAIR ATRIA DEF-CL

Extracorporeal circulation (5P):
3961
EXTRACORPOREAL CIRCULAT

Atrial Septal Defect or Ventricular Septal Defect diagnosis (5D):
7454
VENTRICULAR SEPT DEFECT
7455
SECUNDUM ATRIAL SEPT DEF

Catheterization (6P):
3721
RT HEART CARDIAC CATH
3722
LEFT HEART CARDIAC CATH
3723
RT/LEFT HEART CARD CATH
8842
CONTRAST AORTOGRAM
8843
CONTR PULMON ARTERIOGRAM
8844
ARTERIOGRAPHY OF OTHER INTRATHORACIC VESSELS
8850
ANGIOCARDIOGRAPHY, NOT OTHERWISE SPECIFIED
8851
ANGIOCARDIOGRAPHY OF VENAE CAVAE
8852
ANGIOCARDIOGRAPHY OF RIGHT HEART STRUCTURES
8853
ANGIOCARDIOGRAPHY OF LEFT HEART STRUCTURES
8854
COMBINED RIGHT AND LEFT HEART ANGIOCARDIOGRAPHY
8855
CORONARY ARTERIOGRAPHY USING A SINGLE CATHETER
8856
CORONARY ARTERIOGRAPHY USING TWO CATHETERS
8857
OTHER AND UNSPECIFIED CORONARY ARTERIOGRAPHY
8858
NEGATIVE-CONTRAST CARDIAC ROENTGENOGRAPHY

Heart Transplant (7P):
375
HEART TRANSPLANTATION (invalid as of OCT03)
3751
HEART TRANSPLANTATION OCT03-
3752
IMPLANT TOT REP HRT SYS OCT03-

Premature infants (4D):

76500
EXTREME IMMATUR WTNOS
76501
EXTREME IMMATUR <500G
76502
EXTREME IMMATUR 500-749G
76503
EXTREME IMMATUR 750-999G
76504
EXTREME IMMAT 1000-1249G
76505
EXTREME IMMAT 1250-1499G
76506
EXTREME IMMAT 1500-1749G
76507
EXTREME IMMAT 1750-1999G
76508
EXTREME IMMAT 2000-2499G
76509
EXTREME IMMAT 2500+G
76510
PRETERM INFANT NEC WTNOS
76511
PRETERM NEC <500G
76512
PRETERM NEC 500-749G
76513
PRETERM NEC 750-999G
76514
PRETERM NEC 1000-1249G
76515
PRETERM NEC 1250-1499G
76516
PRETERM NEC 1500-1749G
76517
PRETERM NEC 1750-1999G
76518
PRETERM NEC 2000-2499G
76519
PRETERM NEC 2500+G

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
The user has the option to stratify by Gender, birthweight, age in days, age in years, race / ethnicity, primary payer, and custom stratifiers.

2a.12-13 Risk Adjustment Type: Risk adjustment method widely or commercially available

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
PQI: The predicted value for each case is computed using a logistic regression model and covariates for gender and age in years (in 5-year age groups). The reference population used in the model is the universe of discharges for states that participate in the HCUP State Inpatient Databases (SID) for the year 2007 (updated annually), a database consisting of 43 states and approximately 30 million adult discharges. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., county, state, and region). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate
The model includes additional covariates for RACHS-1 risk categories.

Required data elements: CMS Diagnosis Related Group (DRG); CMS Major Diagnostic Category (MDC); age in days up to 364, then age years at admission; International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) principal and secondary diagnosis codes.

2a.15-17 **Detailed risk model available Web page URL or attachment:** URL None
http://qualityindicators.ahrq.gov/downloads/pd/PDI_Risk_Adjustment_Tables_(Version_4_2).pdf

2a.18-19 **Type of Score:** Rate/proportion
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***):**
Each indicator is expressed as a rate, is defined as outcome of interest / population at risk or numerator / denominator. The AHRQ Quality Indicators (AHRQ QI) software performs five steps to produce the rates. 1) Discharge-level data is used to mark inpatient records containing the outcome of interest and 2) the population at risk. For provider indicators, the population at risk is also derived from hospital discharge records; for area indicators, the population at risk is derived from U.S. Census data. 3) Calculate observed rates. Using output from steps 1 and 2, rates are calculated for user-specified combinations of stratifiers. 4) Calculate expected rates. Regression coefficients from a reference population database are applied to the discharge records and aggregated to the provider or area level. 5) Calculate risk-adjusted rate. Use the indirect standardization to account for case-mix. 6) Calculate smoothed rate. A Univariate shrinkage factor is applied to the risk-adjusted rates. The shrinkage estimate reflects a reliability adjustment unique to each indicator. Full information on calculation algorithms and specifications can be found at http://qualityindicators.ahrq.gov/PDI_download.htm

2a.22 **Describe the method for discriminating performance** *(e.g., significance testing)***:**
Significance testing is not prescribed by the software. Users may calculate a confidence interval for the risk-adjusted rates and a posterior probability interval for the smoothed rates at a 95% or 99% level. Users may define the relevant benchmark and the methods of discriminating performance according to their application.

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)*:
Not applicable

2a.24 **Data Source (***Check the source(s) for which the measure is specified and tested)*
Electronic administrative data/claims

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.***):**
The data source is hospital discharge data such as the HCUP State Inpatient Databases (SID) or equivalent using UB-04 coding standards. The data collection instrument is public-use AHRQ QI software available in SAS or Windows versions

2a.26-28 **Data source/data collection instrument reference web page URL or attachment:** URL None
http://www.qualityindicators.ahrq.gov/software.htm

2a.29-31 **Data dictionary/code table web page URL or attachment:** URL None
http://www.qualityindicators.ahrq.gov/downloads/winqi/AHRQ_QI_Windows_Software_Documentation_V41a.pdf

2a.32-35 **Level of Measurement/Analysis** (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency

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

2a.38-41 **Clinical Services** (*Healthcare services being measured, check all that apply*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing **2b**

2b.1 Data/sample *(description of data/sample and size)*: The Pediatric Health Information System (PHIS) dataset was used for these analyses. This dataset represents detailed hospital-based inpatient information from all discharges (n = 385,157) from 34 independent, academic, free-standing, pediatric hospitals in the United States (PHIS). They are heterogeneous with respect to geographic location, bedsize, and average daily census. Data are submitted to PHIS and tested for reliability and validity before inclusion. [1]

References
[1] Slonim AD, Marcin JP, Turenne W, Hall M, Joseph JG. Pediatric patient safety events during hospitalization: approaches to accounting for institution-level effects. Health Serv Res. 2007 Dec;42(6 Pt 1):2275-93; discussion 2294-323. PMID: 17995566.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
The rates of PSIs were computed for all discharges. The patient and institutional characteristics associated with these PSIs were calculated. The analyses sequentially applied three increasingly conservative methods to control for the institution-level effects robust standard error estimation, a fixed effects model, and a random effects model. The degree of difference from a "base state," which excluded institution-level variables, and between the models was calculated. The effects of these analyses on the interpretation of the PSIs are presented. [1]
References
[1] Slonim AD, Marcin JP, Turenne W, Hall M, Joseph JG. Pediatric patient safety events during hospitalization: approaches to accounting for institution-level effects. Health Serv Res. 2007 Dec;42(6 Pt 1):2275-93; discussion 2294-323. PMID: 17995566.

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
PRINCIPAL FINDINGS: PSIs are relatively infrequent events in hospitalized children ranging from 0 per 10,000 (postoperative hip fracture) to 87 per 10,000 (postoperative respiratory failure). Significant variables associated PSIs included age (neonates), race (Caucasians), payor status (public insurance), severity of illness (extreme), and hospital size (>300 beds), which all had higher rates of PSIs than their reference groups in the bivariable logistic regression results. The three different approaches of adjusting for institution-level effects demonstrated that there were similarities in both the clinical and statistical significance across each of the models. [1]

References
[1] Slonim AD, Marcin JP, Turenne W, Hall M, Joseph JG. Pediatric patient safety events during hospitalization: approaches to accounting for institution-level effects. Health Serv Res. 2007 Dec;42(6 Pt 1):2275-93; discussion 2294-323. PMID: 17995566.

C☐ P☐ M☐ N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: We performed a cross-sectional analysis of California hospital discharges from 2005–2007 for patients aged <18 years. [1]

Agency for Healthcare Research and Quality pediatric-specific quality indicators were used to identify adverse events in 431524 discharges from 38 freestanding, academic, not-for-profit, tertiary care pediatric hospitals in the United States participating in the Pediatric Health Information System database in 2006. [2]

References
[1] Bardach NS, Chien AT, Dudley RA. Small numbers limit the use of the inpatient pediatric quality indicators for hospital comparison. Acad Pediatr. 2010 Jul-Aug;10(4):266-73. PMID: 20599180; doi:10.1016/j.acap.2010.04.025.
[2] Kronman MP, Hall M, Slonim AD, Shah SS. Charges and lengths of stay attributable to adverse patient-care events using pediatric-specific quality indicators: a multicenter study of freestanding children´s hospitals. Pediatrics. 2008 Jun;121(6):e1653-9. PMID: 18519468; DOI: http://dx.doi.org/10.1542/peds.2007-2831.

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
After excluding discharges with PDIs indicated as present on admission, we determined for each PDI the volume of eligible pediatric patients for each measure at each hospital, the statewide mean rate, and the

2c C☐ P☐ M☐ N☐

percentage of hospitals with adequate volume to identify an adverse event rate twice the statewide mean. [2]

In this study, we matched each case subject with 3 control subjects within the same all-patient refined diagnosis-related group (APR-DRG [3M Corporation, St Paul, MN]) severity level, age group (as defined by the American Academy of Pediatrics as <30 days, 30–364 days, 1–4 years, 5–12 years, 13–17 years, and 18 years), and hospital. If >3 control subjects were available on the basis of these restrictions, we used propensity scores to minimize the bias in selecting matched control subjects. Statistical significance for the difference in use between the case and control subjects was determined by using Wilcoxon´s signed rank test, a nonparametric alternative to the 1-sample t test. [2]

References
[1] Bardach NS, Chien AT, Dudley RA. Small numbers limit the use of the inpatient pediatric quality indicators for hospital comparison. Acad Pediatr. 2010 Jul-Aug;10(4):266-73. PMID: 20599180; doi:10.1016/j.acap.2010.04.025.
[2] Kronman MP, Hall M, Slonim AD, Shah SS. Charges and lengths of stay attributable to adverse patient-care events using pediatric-specific quality indicators: a multicenter study of freestanding children´s hospitals. Pediatrics. 2008 Jun;121(6):e1653-9. PMID: 18519468; DOI: http://dx.doi.org/10.1542/peds.2007-2831.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
Event rates for pediatric heart surgery mortality were 38 per 1000, requiring patient volumes of 201 to detect an event rate twice the statewide average; 25% of California hospitals had this pediatric volume. Using these AHRQ-developed, nationally endorsed measures of the quality of inpatient pediatric care, one would not be able to identify many hospitals with performance 2 times worse than the statewide average due to extremely low event rates and inadequate pediatric hospital volume. [1]

Age was the only demographic factor with any statistically significant differences between matched and unmatched case subjects for accidental puncture and laceration. The demographic variables race, gender, payer, disposition, and census region had no differences in any of the PDIs. The occurrence of In-hospital mortality after pediatric heart surgery was not associated with a statistically significant increase in LOS but was associated with an increase in overall charges (P < .006 after the Bonferroni correction). [2]

References
[1] Bardach NS, Chien AT, Dudley RA. Small numbers limit the use of the inpatient pediatric quality indicators for hospital comparison. Acad Pediatr. 2010 Jul-Aug;10(4):266-73. PMID: 20599180; doi:10.1016/j.acap.2010.04.025.
[2] Kronman MP, Hall M, Slonim AD, Shah SS. Charges and lengths of stay attributable to adverse patient-care events using pediatric-specific quality indicators: a multicenter study of freestanding children´s hospitals. Pediatrics. 2008 Jun;121(6):e1653-9. PMID: 18519468; DOI: http://dx.doi.org/10.1542/peds.2007-2831.

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
Exclusions remove cases where the outcome of interest is less likely to be preventable or more likely to be preventable or with no or very low risk

2d.2 Citations for Evidence:
Updated citations will be presented in the May Steering Committee meeting

Measures of Pediatric Health Care Quality Based on Hospital Administrative Data, The Pediatric Quality Indicators. Ver 3.1 March 2007
http://qualityindicators.ahrq.gov/downloads/pdi/pdi_measures_v31.pdf

2d.3 Data/sample *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 3,500 hospitals and 6 million pediatric discharges

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

2d
C☐
P☐
M☐
N☐
NA ☐

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
Measures of Pediatric Health Care Quality Based on Hospital Administrative Data, The Pediatric Quality Indicators. Ver 3.1 March 2007
http://qualityindicators.ahrq.gov/downloads/pdi/pdi_measures_v31.pdf

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 3,500 hospitals and 6 million pediatric discharges

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
Risk-adjustment models use a standard set of categories based on readily available classification systems for demographics, severity of illness and comorbidities. Within each category, covariates are initially selected based on a minimum of 30 cases in the outcome of interest. Then a stepwise regression process on a development sample is used to select a parsimonious set of covariates where p<.05. Model is then tested on a validation sample

2e.3 Testing Results *(risk model performance metrics)*:
C-statistic 0.8750

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

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)*: AHRQ 2007 State Inpatient Databases (SID) with 3,500 hospitals and 6 million pediatric discharges

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
Posterior probability distribution parameterized using the Gamma distribution

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)*:

5th	25th	Median	75th	95th
0.025200	0.037077	0.047287	0.059225	0.079624

2f C☐ P☐ M☐ N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample *(description of data/sample and size)*: Not applicable

2g.2 Analytic Method *(type of analysis & rationale)*:
Not applicable

2g.3 Testing Results *(e.g., correlation statistics, comparison of rankings)*:
Not applicable

2g C☐ P☐ M☐ N☐ NA☐

2h. Disparities in Care

2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)*: Median income of patient´s ZIP code:
1) Estimate 2) Standard error 3) P-value: Relative to marked group-c 4) P-value:
2007 relative to 2006
First quartile (lowest income) 44.830 2.315 0.394 0.112
Second quartile 39.643 2.577 0.671 0.053
Third quartile 32.492 2.639 0.034 0.679
Fourth quartile (highest income)c 41.414 3.276 0.043

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities,

2h C☐ P☐ M☐ N☐ NA☐

provide follow-up plans: Users may stratify based on gender and race/ethnicity	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, met? **Rationale:**	**2** **C** ☐ **P** ☐ **M** ☐ **N** ☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Eval Rating**
3a. Meaningful, Understandable, and Useful Information 3a.1 **Current Use:** In use 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** Florida (state) Florida Health Finder http://www.floridahealthfinder.gov/ Kentucky (Norton Healthcare, a hospital system) Norton Healthcare Quality Report http://www.nortonhealthcare.com/body.cfm?id=157 Texas (state) Reports on Hospital Performance http://www.dshs.state.tx.us/thcic/ Vermont (state) Dept of Banking, Insurance, Securities & Health Care Administration Comparison Report http://www.bishca.state.vt.us/health-care/hospitals-health-care-practitioners/2009-vermont-hospital-report-card The measure is also reported on HCUPnet: http://hcupnet.ahrq.gov/HCUPnet.jsp?Id=EB57801381F71C41&Form=MAINSEL&JS=Y&Action=%3E%3ENext%3E%3E&_MAINSEL=AHRQ%20Quality%20Indicators This measure will be used in the MONAHRQ system that is provided for public reporting and quality improvement throughout the United States: http://monahrq.ahrq.gov/ 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***):** University Healthcare Consortium (UHC) - An alliance of 103 academic medical centers and 219 of their affiliated hosptials. UCH reports the AHRQ QIs to their member hospitals. (See www.uhc.edu. Note that meaure results are reported to hospitals; not reported on the UHC site). National Association of Children´s Hospitals and Related Institutions (NACHRI) reports all provider level PDIs to its approximately 85 member children´s hospitals. (See http://www.childrenshospitals.net. Note that	**3a** **C** ☐ **P** ☐ **M** ☐ **N** ☐

meaure results are reported to hospitals; not reported on the NACHRI site).

Norton Healthcare - a multi-hospital system in Kentucky (see http://www.nortonhealthcare.com/about/Our_Performance/index.aspx)

Ministry Health Care - a multi-hospital system in Wisconsin (see http://ministryhealth.org/display/router.aspx. Note: measure results reported to hospitals; not reported on site).

Child Health Corporation of America (CHCA) reports all PDIs to its 42 member hospitals, which are large freestanding pediatric hospitals. (See http://www.chca.com/. Note that meaure results are reported to hospitals; not reported on the CHCA site).

This measure will be used in the MONAHRQ system that is provide for public reporting and quality improvement throughout the United States: http://monahrq.ahrq.gov/

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)*: AHRQ 2007 State Inpatient Databases (SID) with 3,500 hospitals and 6 million pediatric discharges

3a.5 **Methods** *(e.g., focus group, survey, QI project)*:
A research team from the School of Public Affairs, Baruch College, under contracts with the Department of Public Health, Weill Medical College and Battelle, Inc., has developed a pair of Hospital Quality Model Reports at the request of the Agency for Healthcare Research & Quality (AHRQ). These reports are designed specifically to report comparative information on hospital performance based on the AHRQ Quality Indicators (QIs). The work was done in close collaboration with AHRQ staff and the AHRQ Quality Indicators team.
The Model Reports (discussed immediately above) are based on:
• Extensive search and analysis of the literature on hospital quality measurement and reporting, as well as public reporting on health care quality more broadly;
• Interviews with quality measurement and reporting experts, purchasers, staff of purchasing coalitions, and executives of integrated health care delivery systems who are responsible for quality in their facilities;
• Two focus groups with chief medical officers of hospitals and/or systems and two focus groups with quality managers from a broad mix of hospitals;
• Four focus groups with members of the public who had recently experienced a hospital admission; and
• Four rounds of cognitive interviews (a total of 62 interviews) to test draft versions of the two Model Reports with members of the public with recent hospital experience, basic computer literacy but widely varying levels of education.

3a.6 **Results** *(qualitative and/or quantitative results and conclusions)*:
Given the above review of the literature and original research that was conducted, a Model report was the result that could help sponsors use the best evidence on public reports so they are most likely to have the desired effects on quality.

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:** **5.1 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 measures found.	**3c** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**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*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	**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.** Coding professionals follow detail guidelines, are subject to training and credentialing requirements, peer review and audit.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** None **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** Administrative data are collected as part of the routine operations. Some staff time is required to download	**4e** C☐ P☐ M☐ N☐

and execute the software from the AHRQ webs site, which is available at no cost. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm 4e.3 **Evidence for costs:** All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm 4e.4 **Business case documentation:** All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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**
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Co.2 **Point of Contact**
John, Bott, MSSW, MBA, John.Bott@AHRQ.hhs.gov, 301-427-1317-

Measure Developer If different from Measure Steward
Co.3 **Organization**
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Co.4 **Point of Contact**
John, Bott, MSSW, MBA, John.Bott@AHRQ.hhs.gov, 301-427-1317-

Co.5 **Submitter If different from Measure Steward POC**
John, Bott, MSSW, MBA, John.Bott@AHRQ.hhs.gov, 301-427-1317-, Agency for Healthcare Research and Quality

Co.6 **Additional organizations that sponsored/participated in measure development**
UC Davis,
Stanford University,
Battelle Memorial Institute

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.**
None

Ad.2 **If adapted, provide name of original measure:** None
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:** 2006 **Ad.7 Month and Year of most recent revision:** 10, 2009 **Ad.8 What is your frequency for review/update of this measure?** Annual **Ad.9 When is the next scheduled review/update for this measure?** 05, 2011
Ad.10 Copyright statement/disclaimers: The AHRQ QI software is publicly available; no copyright disclaimers
Ad.11 -13 Additional Information web page URL or attachment:
Date of Submission (*MM/DD/YY*)**:** 04/05/2011

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 0340 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Pediatric Heart Surgery Volume (PDI 7)
De.2 **Brief description of measure:** Number of discharges with procedure for pediatric heart surgery
1.1-2 **Type of Measure:** Structure/management De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** Pediatric Heart Surgery Mortality (PDI 6) (NQF #0339))
De.4 **National Priority Partners Priority Area:** Population health, Safety De.5 **IOM Quality Domain**: Effectiveness, Safety De.6 **Consumer Care Need:** Getting better

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	NQF Staff
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary 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, Internal quality improvement	**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:** Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Pending update. Using a multivariate model that included age, complexity category, and four comorbidities, Hannan et al. found 8.26% risk-adjusted mortality at hospitals with fewer than 100 cases per year, versus 5.95% at higher volume hospitals (an effect limited to surgeons who performed at least 75 cases per year). [1] For a more complete review of this topic, see: URL:http://www.qualityindicators.ahrq.gov/downloads/pdi/pdi_measures_v31 1a.4 **Citations for Evidence of High Impact:** Updated citations will be presented in the May Steering Committee meeting [1] Hannan EL, Racz M, Kavey RE, Quaegebeur JM, Williams R. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. Pediatrics 1998;101(6):963-9	**1a** C☐ P☐ M☐ N☐
1b. Opportunity for Improvement 1b.1 **Benefits (improvements in quality) envisioned by use of this measure:** Higher volume is associated with reduced mortality and morbidity. 1b.2 **Summary of data demonstrating performance gap (variation or overall poor performance) across providers:** The number of pediatric cardiac procedures is measured accurately with discharge data; in fact, discharge	**1b** C☐ P☐ M☐ N☐

data are probably the best available source for hospital volume information. Previous studies suggest that pediatric cardiac surgery is already highly concentrated at a relatively small number of facilities (e.g., 16 hospitals in New York, 37 in California and Massachusetts together). Although some of these facilities have very high volumes, a significant number (e.g., 16 hospitals in California and Massachusetts) perform fewer than 10 cases per year. The highly skewed volume distribution may have an adverse effect on the precision of this measure.

1b.3 Citations for data on performance gap:
AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

1b.4 Summary of Data on disparities by population group:
Across a broad set of 23 quality indicators, findings indicate that racial/ethnic disparities vary by income levels and types of insurance. Key highlights include the finding that racial/ethnic differences within income or insurance/payer groups are more pronounced for some racial/ethnic groups than others. Hispanic children followed by Asian children had worse quality than whites as measured by the majority of quality indicators. Exceptions included rates of admissions for diabetes, admissions for gastroenteritis, accidental puncture during procedures, and decubitus ulcers . Many indicators showed less than ideal quality for all subgroups of children, even whites with private insurance. [1]

References
[1] Berdahl T, Owens PL, Dougherty D, McCormick MC, Pylypchuk Y, Simpson LA. Annual report on health care for children and youth in the United States: racial/ethnic and socioeconomic disparities in children´s health care quality. Acad Pediatr. 2010 Mar-Apr;10(2):95-118. PMID: 20206909.

1b.5 Citations for data on Disparities:
The analyses are based on data from a nationally representative random sample of children in the United States in 2004 and 2005 from the Medical Expenditure Panel Survey (MEPS) and pediatric hospitalizations from a nationwide sample of hospitals in 2005 from the State Inpatient Databases disparities analysis file from the Healthcare Cost and Utilization Project (HCUP). [1]

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*)**:** The measure focus is an structure (volume) that is associated with an outcome (mortality) relevant to a neonatal population with a diagnosis of congenital heart defect or procedure for congenital heart repair.

1c.2-3. Type of Evidence: Expert opinion, Systematic synthesis of research

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Using a multivariate model that included age, complexity category, and four comorbidities, Hannan et al. found 8.26% risk-adjusted mortality at hospitals with fewer than 100 cases per year, versus 5.95% at higher volume hospitals (an effect limited to surgeons who performed at least 75 cases per year). [1] Two other studies using hospital discharge data from California and Massachusetts found similar effects of hospital volume. [2] [3]
Another source of evidence is that cardiopulmonary bypass or aortic crossclamp time has been repeatedly associated with postoperative mortality, adjusting for a variety of patient characteristics.[4-7] This relationship has been demonstrated not just for the Fontan procedure, but also for the Norwood procedure for hypoplastic left heart syndrome. [8] Experienced surgeons and surgical teams should be able to reduce cardiopulmonary bypass or aortic cross-clamp time, thereby improving postoperative mortality.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
B there is moderate certainty that the net benefit is moderate to substantial (review by project team)

1c.6 Method for rating evidence: U.S. Preventive Services Task Force (USPSTF) assigns one of five letter grades to each of its recommendations (A, B, C, D, or I).

1c.7 Summary of Controversy/Contradictory Evidence: A study reviewed the application of two case-mix

1c
C☐
P☐
M☐
N☐

complexity-adjustment tools in the Society of Thoracic Surgeons (STS) Congenital Heart Surgery Database: the Aristotle Basic Complexity (ABC) score and the Risk Adjustment in Congenital Heart Surgery (RACHS-1) method. With both RACHS-1 and ABC, as complexity increases, discharge mortality also ncreases. The ABC approach allows classification of more operations, whereas the RACHS-1 discriminates better at the higher end of complexity. Complexity stratification is a useful method for analyzing the impact of case mix on pediatric cardiac surgical outcomes. Both the RACHS-1 and ABC methods facilitate complexity stratification in the STS database.

1c.8 Citations for Evidence (*other than guidelines*): Updated citations will be presented in the May Steering Committee meeting

[1] Hannan EL, Racz M, Kavey RE, Quaegebeur JM, Williams R. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. Pediatrics 1998;101(6):963-9.
[2] Jenkins KJ, Newburger JW, Lock JE, Davis RB, Coffman GA, Iezzoni LI. In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. Pediatrics 1995;95(3):323-30.
[3] Sollano JA, Gelijns AC, Moskowitz AJ, Heitjan DF, Cullinane S, Saha T, et al. Volume-outcome relationships in cardiovascular operations: New York State, 1990-1995. J Thorac Cardiovasc Surg 1999;117(3):419-28.
[4] Cetta F, Feldt RH, O´Leary PW, Mair DD, Warnes CA, Driscoll DJ, et al. Improved early morbidity and mortality after Fontan operation: the Mayo Clinic experience, 1987 to 1992. J Am Coll Cardiol 1996;28(2):480-6.
[5] Gentles TL, Gauvreau K, Mayer JE, Jr., Fishberger SB, Burnett J, Colan SD, et al. Functional outcome after the Fontan operation: factors influencing late morbidity. J Thorac Cardiovasc Surg 1997;114(3):392-403; discussion 404-5.
[6] Kaulitz R, Ziemer G, Luhmer I, Kallfelz HC. Modified Fontan operation in functionally univentricular hearts: preoperative risk factors and intermediate results. J Thorac Cardiovasc Surg 1996;112(3):658-64.
[7] Fontan F, Kirklin JW, Fernandez G, Costa F, Naftel DC, Tritto F, et al. Outcome after a "perfect" Fontan operation. Circulation 1990;81(5):1520-36.
[8] Kern JH, Hayes CJ, Michler RE, Gersony WM, Quaegebeur JM. Survival and risk factor analysis for the Norwood procedure for hypoplastic left heart syndrome. Am J Cardiol 1997;80(2):170-4.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number*):
Surgery for congenital heart disease, especially in infants, requires a setting that readily meets the complex and special needs of this group of patients. These requirements include a cardiac surgeon experienced in the operative and perioperative management of such patients. There should be a pediatric cardiologist, an anesthesia team, perfusionists, intensive care nurses, and appropriate intensive care facilities for the treatment of infants and children. At a hospital where congenital heart operations are performed, a total of 100 congenital heart operations (both open and closed, not including neonatal ductus ligations) should be done. The occasional management of an infant or child with congenital heart disease by an otherwise busy and well-functioning adult cardiac surgical team is strongly discouraged.

1c.10 Clinical Practice Guideline Citation: http://www.facs.org/fellows_info/guidelines/cardiac.html
1c.11 National Guideline Clearinghouse or other URL: Not Applicable.

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom*):
Not Applicable.

1c.13 Method for rating strength of recommendation (*If different from USPSTF system, also describe rating and how it relates to USPSTF)*:
Not Applicable.

1c.14 Rationale for using this guideline over others:
No competing measures found.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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***):** Discharges under age 18 with ICD-9-CM procedure codes for either congenital heart disease (1P) in any field or non-specific heart surgery (2P) with ICD-9-CM diagnosis of congenital heart disease (2D) in any field. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** Time window can be determined by user, but is generally a calendar year. **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Discharges under age 18 with ICD-9-CM procedure codes for either congenital heart disease (1P) or non-specific heart surgery (2P) with ICD-9-CM diagnosis of congenital heart disease (2D) in any field. Congenital heart disease procedures (1P): 3500 CLOSED VALVOTOMY NOS 3501 CLOSED AORTIC VALVOTOMY 3502 CLOSED MITRAL VALVOTOMY 3503 CLOSED PULMON VALVOTOMY 3504 CLOSED TRICUSP VALVOTOMY 3510 OPEN VALVULOPLASTY NOS 3511 OPN AORTIC VALVULOPLASTY 3512 OPN MITRAL VALVULOPLASTY 3513 OPN PULMON VALVULOPLASTY 3514 OPN TRICUS VALVULOPLASTY 3520 REPLACE HEART VALVE NOS 3521 REPLACE AORT VALV-TISSUE 3522 REPLACE AORTIC VALVE NEC 3523 REPLACE MITR VALV-TISSUE 3524 REPLACE MITRAL VALVE NEC	**2a-specs** C☐ P☐ M☐ N☐

3525
REPLACE PULM VALV-TISSUE
3526
REPLACE PULMON VALVE NEC
3527
REPLACE TRIC VALV-TISSUE
3528
REPLACE TRICUSP VALV NEC
3531
PAPILLARY MUSCLE OPS
3532
CHORDAE TENDINEAE OPS
3533
ANNULOPLASTY
3534
INFUNDIBULECTOMY
3535
TRABECUL CARNEAE CORD OP
3539
TISS ADJ TO VALV OPS NEC
3541
ENLARGE EXISTING SEP DEF
3542
CREATE SEPTAL DEFECT
3550
PROSTH REP HRT SEPTA NOS
3551
PROS REP ATRIAL DEF-OPN
3552
PROS REPAIR ATRIA DEF-CL
3553
PROST REPAIR VENTRIC DEF
3554
PROS REP ENDOCAR CUSHION
3560
GRFT REPAIR HRT SEPT NOS
3561
GRAFT REPAIR ATRIAL DEF
3562
GRAFT REPAIR VENTRIC DEF
3563
GRFT REP ENDOCAR CUSHION
3570
HEART SEPTA REPAIR NOS
3571
ATRIA SEPTA DEF REP NEC
3572
VENTR SEPTA DEF REP NEC
3573
ENDOCAR CUSHION REP NEC
3581
TOT REPAIR TETRAL FALLOT
3582
TOTAL REPAIR OF TAPVC
3583
TOT REP TRUNCUS ARTERIOS
3584
TOT COR TRANSPOS GRT VES

3591
INTERAT VEN RETRN TRANSP
3592
CONDUIT RT VENT-PUL ART
3593
CONDUIT LEFT VENTR-AORTA
3594
CONDUIT ARTIUM-PULM ART
3595
HEART REPAIR REVISION
3598
OTHER HEART SEPTA OPS
3599
OTHER OP ON HRT VALVES
3699
OTHER OPERATIONS ON VESSEL OF HEART
3733
EXCISION OR DESTRUCTION OF OTHER LESION OR TISSUE OF HEART
3736
EXCISION OR DESTRUCTION OF LEFT ATRIAL APPENDAGE (LAA) OCT08-
375
HEART TRANSPLANTATION (invalid as of OCT03)
3751
HEART TRANSPLANTATION OCT03-
3752
IMPLANT TOT REP HRT SYS OCT03-
390
SYSTEMIC-PULM ART SHUNT
3921
CAVAL-PULMON ART ANASTOM

Non-specific cardiac procedures (2P):
3834
RESECTION OF ABDOMINAL AORTA WITH ANASTOMOSIS
3835
THOR VESSEL RESECT/ANAST
3844
RESECTION OF ABDOMINAL AORTA WITH REPLACEMENT
3845
RESECT THORAC VES W REPL
3864
OTHER EXCISION OF ABDOMINAL AORTA
3865
OTHER EXCISION OF THORACIC VESSEL
3884
OTHER SURGICAL OCCLUSION OF ABDOMINAL AORTA
3885
OCCLUDE THORACIC VES NEC
3949
OTHER REVISION OF VASCULAR PROCEDURE
3956
REPAIR OF BLOOD VESSEL WITH TISSUE PATCH GRAFT
3957
REPAIR OF BLOOD VESSEL WITH SYNTHETIC PATCH GRAFT
3958
REPAIR OF BLOOD VESSEL WITH UNSPECIFIED TYPE OF PATCH GRAFT
3959
REPAIR OF VESSEL NEC

Congenital heart disease diagnoses (2D):
7450
COMMON TRUNCUS
74510
COMPL TRANSPOS GREAT VES
74511
DOUBLE OUTLET RT VENTRIC
74512
CORRECT TRANSPOS GRT VES
74519
TRANSPOS GREAT VESS NEC
7452
TETRALOGY OF FALLOT
7453
COMMON VENTRICLE
7454
VENTRICULAR SEPT DEFECT
7455
SECUNDUM ATRIAL SEPT DEF
74560
ENDOCARD CUSHION DEF NOS
74561
OSTIUM PRIMUM DEFECT
74569
ENDOCARD CUSHION DEF NEC
7457
COR BILOCULARE
7458
SEPTAL CLOSURE ANOM NEC
7459
SEPTAL CLOSURE ANOM NOS
74600
PULMONARY VALVE ANOM NOS
74601
CONG PULMON VALV ATRESIA
74602
CONG PULMON VALVE STENOS
74609
PULMONARY VALVE ANOM NEC
7461
CONG TRICUSP ATRES/STEN
7462
EBSTEIN'S ANOMALY
7463
CONG AORTA VALV STENOSIS
7464
CONG AORTA VALV INSUFFIC
7465
CONGEN MITRAL STENOSIS
7466
CONG MITRAL INSUFFICIENC
7467
HYPOPLAS LEFT HEART SYND
74681
CONG SUBAORTIC STENOSIS
74682
COR TRIATRIATUM

74683
INFUNDIB PULMON STENOSIS
74684
OBSTRUCT HEART ANOM NEC
74685
CORONARY ARTERY ANOMALY
74687
MALPOSITION OF HEART
74689
CONG HEART ANOMALY NEC
7469
CONG HEART ANOMALY NOS
7470
PATENT DUCTUS ARTERIOSUS
74710
COARCTATION OF AORTA
74711
INTERRUPT OF AORTIC ARCH
74720
CONG ANOM OF AORTA NOS
74721
ANOMALIES OF AORTIC ARCH
74722
AORTIC ATRESIA/STENOSIS
74729
CONG ANOM OF AORTA NEC
7473
PULMONARY ARTERY ANOM
74740
GREAT VEIN ANOMALY NOS
74741
TOT ANOM PULM VEN CONNEC
74742
PART ANOM PULM VEN CONN
74749
GREAT VEIN ANOMALY NEC

Exclude cases:
• MDC 14 (pregnancy, childbirth and pueperium)
• with transcatheter interventions (either 3AP, 3BP, 3CP, 3DP, 3EP with 3D, or 3FP) as single cardiac procedures, performed without bypass (5P) but with catheterization (6P);
• with septal defects (4P) as single cardiac procedures without bypass (5P)

Transcatheter interventions procedure codes:

Closed heart valvotomy (3AP):
3500
CLOSED HEART VALVOTOMY, UNSPECIFIED VALUE
3501
CLOSED HEART VALVOTOMY, AORTIC VALUE
3502
CLOSED HEART VALVOTOMY, MITRAL VALUE
3503
CLOSED HEART VALVOTOMY, PULMONARY VALUE
3504
CLOSED HEART VALVOTOMY, TRICUSPID VALUE

Atrial septal enlargement (3BP):

3541
ENLARGEMENT OF EXISTING ATRIAL SEPTAL DEFECT
3542
CREATION OF SEPTAL DEFECT IN HEART

Atrial septal defect repair (3CP):
3551
REPAIR OF ATIAL SEPTAL DEFECT WITH PROSTHESIS, OPEN TECHNIQUE
3571
OTHER AND UNSPECIFIED REPAIR OF ATRIAL SEPTAL DEFECT

Ventricular septal defect repair (3DP):
3553
REPAIR OF VENTRICULAR SEPTAL DEFECT WITH PROSTHESIS
3572
OTHER AND UNSPECIFIED REPAIR OF VENTRICULAR SEPTAL DEFECT

Occlusion of thoracic vessel (3EP):
3885
OCCLUDE THORACIC VES NEC

PDA closure diagnosis code (3D):
7470
PATENT DUCTUS ARTERIOSUS

Other surgical occlusion (3FP):
3884
OTHER SURGICAL OCCLUSION OF AORTA, ABDOMINAL
3885
OTHER SURGICAL OCCLUSION OF THORACIC VESSEL
3959
OTHER REPAIR OF VESSEL

Extracorporeal circulation (5P):
3961
EXTRACORPOREAL CIRCULAT

Catheterization (6P):
3721
RT HEART CARDIAC CATH
3722
LEFT HEART CARDIAC CATH
3723
RT/LEFT HEART CARD CATH
8842
CONTRAST AORTOGRAM
8843
CONTR PULMON ARTERIOGRAM
8844
ARTERIOGRAPHY OF OTHER INTRATHORACIC VESSELS
8850
ANGIOCARDIOGRAPHY, NOT OTHERWISE SPECIFIED
8851
ANGIOCARDIOGRAPHY OF VENAE CAVAE
8852
ANGIOCARDIOGRAPHY OF RIGHT HEART STRUCTURES
8853
ANGIOCARDIOGRAPHY OF LEFT HEART STRUCTURES

8854
COMBINED RIGHT AND LEFT HEART ANGIOCARDIOGRAPHY
8855
CORONARY ARTERIOGRAPHY USING A SINGLE CATHETER
8856
CORONARY ARTERIOGRAPHY USING TWO CATHETERS
8857
OTHER AND UNSPECIFIED CORONARY ARTERIOGRAPHY
8858
NEGATIVE-CONTRAST CARDIAC ROENTGENOGRAPHY

Atrial septal defect repair and enlargement (4P):
3541
ENLARGE EXISTING SEP DEF
3552
PROS REPAIR ATRIA DEF-CL

2a.4 **Denominator Statement (***Brief, text description of the denominator - target population being measured***):**
This measure does not have a denominator due to the fact it is a volume measure.

2a.5 **Target population gender:** Female, Male
2a.6 **Target population age range:** Age less than 18 years

2a.7 **Denominator Time Window (***The time period in which cases are eligible for inclusion in the denominator***):**
Not applicable

2a.8 **Denominator Details (***All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Not applicable

2a.9 **Denominator Exclusions (***Brief text description of exclusions from the target population***):** Not applicable. This measure does not have a denominator due to the fact it is a volume measure.

2a.10 **Denominator Exclusion Details (***All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
Not applicable. This measure does not have a denominator due to the fact it is a volume measure.

2a.11 **Stratification Details/Variables (***All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
Not applicable

2a.12-13 **Risk Adjustment Type:** No risk adjustment necessary

2a.14 **Risk Adjustment Methodology/Variables (***List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
Not applicable

2a.15-17 **Detailed risk model available Web page URL or attachment:**

2a.18-19 **Type of Score:** Count
2a.20 **Interpretation of Score:** Better quality = Higher score
2a.21 **Calculation Algorithm (***Describe the calculation of the measure as a flowchart or series of steps***):**
The volume is the number of discharges with a procedure for pediatric heart surgery.

2a.22 **Describe the method for discriminating performance** *(e.g., significance testing)***:**
Not applicable

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)*:
Not applicable

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Electronic administrative data/claims

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.*)**:**
The data source is hospital discharge data such as the HCUP State Inpatient Databases (SID) or equivalent using UB-04 coding standards. The data collection instrument is public-use AHRQ QI software available in SAS or Windows versions.

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL None
http://www.qualityindicators.ahrq.gov/software.htm

2a.29-31 Data dictionary/code table web page URL or attachment: URL None
http://www.qualityindicators.ahrq.gov/downloads/winqi/AHRQ_QI_Windows_Software_Documentation_V41a.pdf

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency

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

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
Literature review, clinical panels and empirical analysis

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Pediatric heart surgery procedure codes are based on physician documentation; no evidence has been suggested that these codes are not reliably reported.

2b C☐ P☐ M☐ N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
Literature review, clinical panels and empirical analysis

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
Volume is not a direct measure of the quality or outcomes of care. Although higher volumes have been repeatedly associated with better outcomes after pediatric cardiac surgery, these findings may be limited by inadequate risk adjustment.
Only one study used prospectively collected clinical data to estimate the association between hospital volume and mortality following pediatric cardiac surgery.(55) Hannan et al. ordered all cardiac surgical procedures by their actual mortality rates in the 1992-95 Cardiac Surgery Reporting System database. Expert clinicians then grouped the procedures into four clinically sensible subgroups, designed to achieve maximal separation of crude mortality rates (from 1.4% for Category I to 20.1% for Category IV). A multivariate model that included age, complexity category, and four comorbidities (preoperative cyanosis or hypoxia, barotrauma, pulmonary

2c C☐ P☐ M☐ N☐

hypertension, major extracardiac anomalies) achieved excellent calibration and discrimination (c=0.818). Using this model to estimate risk-adjusted mortality, Hannan et al. found a statistically significant hospital effect (8.26% risk-adjusted mortality at hospitals with fewer than 100 cases per year, versus 5.95% at higher volume hospitals), which was limited to surgeons who performed at least 75 cases per year. Lower volume surgeons experienced relatively high mortality, regardless of total hospital volume. Risk-adjusted mortality differed between low and high-volume hospitals for all 4 complexity categories, although the smallest difference occurred for the highest risk procedures.
Two other studies using hospital discharge data found similar effects of hospital volume. Using aggregated data from California (1988) and Massachusetts (1989), Jenkins et al.(54) estimated risk-adjusted mortality rates of 8.35% and 5.95% at low-volume (100 or fewer cases) and high-volume (more than 100 cases), respectively. However, they also demonstrated especially high risk-adjusted mortality (18.5%) at very low-volume hospitals with fewer than 10 annual cases, and especially low mortality (3.0%) at very high-volume hospitals with more than 300 annual cases. Jenkins et al. could not evaluate the impact of surgeon volume, but they did report stronger volume effects for higher-risk procedures (e.g., OR=12.1 and 3.2 for Category III-IV procedures at hospitals with <10 and 10-100 annual cases, versus OR=2.4 for Category I-II procedures at hospitals with 10-100 annual cases). Finally, Sollano et al. (Sollano, Gelijns et al. 1999) applied the same 4-category risk adjustment procedure developed by Jenkins to hospital discharge data from New York State in 1990-95. They reported a modest but statistically significant effect (OR=0.944 for each additional 100 annual cases), which was limited to neonates (OR=0.636) and post-neonatal infants (OR=0.720) in stratified analyses. Although volume-outcome associations have been demonstrated for pediatric cardiac surgery, volume seems likely to both insensitive and nonspecific as a measure of quality. In addition, pediatric cardiac care is already regionalized, so most procedures are performed in medium-to-high volume hospitals. It has been estimated that shifting patients in California from low-volume to high-volume hospitals would avert only 7 deaths per year.(65)

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
Exclusions remove cases where the outcome of interest is less likely to be preventable or more likely to be preventable or with no or very low risk.

2d.2 Citations for Evidence:
Updated citations will be presented in the May Steering Committee meeting

Jenkins KJ, Newburger JW, Lock JE, Davis RB, Coffman GA, Iezzoni LI. In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. Pediatrics 1995;95(3):323-30.

2d.3 Data/sample *(description of data/sample and size)*: Not applicable

2d.4 Analytic Method *(type analysis & rationale)*:
Not applicable

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
Not applicable

2d
C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: Not applicable

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
Not applicable

2e.3 Testing Results *(risk model performance metrics)*:
Not applicable

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

2e
C☐ P☐ M☐ N☐ NA☐

2f. Identification of Meaningful Differences in Performance

2f
C☐

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges **2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*: Descriptive analysis **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)*: The number of pediatric cardiac procedures is measured accurately with discharge data. In fact, discharge data are probably the best available source for hospital volume information. Previous studies suggest that pediatric cardiac surgery is already highly concentrated at a relatively small number of facilities (e.g., 16 hospitals in New York, 37 in California and Massachusetts together). Although some of these facilities have very high volumes, a significant number (e.g., 16 hospitals in California and Massachusetts) perform fewer than 10 cases per year. The highly skewed volume distribution may have an adverse effect on the precision of this measure.	P☐ M☐ N☐
2g. Comparability of Multiple Data Sources/Methods **2g.1 Data/sample** *(description of data/sample and size)*: Not applicable **2g.2 Analytic Method** *(type of analysis & rationale)*: Not applicable **2g.3 Testing Results** *(e.g., correlation statistics, comparison of rankings)*: Not applicable	**2g** C☐ P☐ M☐ N☐ NA ☐
2h. Disparities in Care **2h.1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: Not applicable **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** Not applicable	**2h** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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 Rati ng
3a. Meaningful, Understandable, and Useful Information **3a.1 Current Use:** In use **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*):	**3a** C☐ P☐ M☐ N☐

Florida (state)
Florida Health Finder
http://www.floridahealthfinder.gov/

Illinois (state hospital association)
Illinois Hospitals Caring for You
www.illinoishospitals.org

Kentucky (Norton Healthcare, a hospital system)
Norton Healthcare Quality Report
http://www.nortonhealthcare.com/body.cfm?id=157

Texas (state)
Reports on Hospital Performance
http://www.dshs.state.tx.us/thcic/

Vermont (state)
Dept of Banking, Insurance, Securities & Health Care Administration Comparison Report
http://www.bishca.state.vt.us/health-care/hospitals-health-care-practitioners/2009-vermont-hospital-report-card

The measure is also reported on HCUPnet:
http://hcupnet.ahrq.gov/HCUPnet.jsp?Id=EB57801381F71C41&Form=MAINSEL&JS=Y&Action=%3E%3ENext%3E%3E&_MAINSEL=AHRQ%20Quality%20Indicators

This measure will appear in the MONAHRQ system that is provided for public reporting and quality improvement throughout the United States: http://monahrq.ahrq.gov/

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***):**
University Healthcare Consortium (UHC) - An alliance of 103 academic medical centers and 219 of their affiliated hosptials. UCH reports the AHRQ QIs to their member hospitals. (See www.uhc.edu. Note that meaure results are reported to hospitals; not reported on the UHC site).

National Association of Children´s Hospitals and Related Institutions (NACHRI) reports all provider level PDIs to its approximately 85 member children´s hospitals. (See http://www.childrenshospitals.net. Note that meaure results are reported to hospitals; not reported on the NACHRI site).

Norton Healthcare - a multi-hospital system in Kentucky (see http://www.nortonhealthcare.com/about/Our_Performance/index.aspx)

Ministry Health Care - a multi-hospital system in Wisconsin (see http://ministryhealth.org/display/router.aspx. Note: measure results reported to hospitals; not reported on site).

Child Health Corporation of America (CHCA) reports all PDIs to its 42 member hospitals, which are large freestanding pediatric hospitals. (See http://www.chca.com/. Note that meaure results are reported to hospitals; not reported on the CHCA site).

This measure will be added to the MONAHRQ system that is provided for public reporting and quality improvement throughout the United States: http://monahrq.ahrq.gov/

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)***:** AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

3a.5 Methods *(e.g., focus group, survey, QI project)*: A research team from the School of Public Affairs, Baruch College, under contracts with the Department of Public Health, Weill Medical College and Battelle, Inc., has developed a pair of Hospital Quality Model Reports at the request of the Agency for Healthcare Research & Quality (AHRQ). These reports are designed specifically to report comparative information on hospital performance based on the AHRQ Quality Indicators (QIs). The work was done in close collaboration with AHRQ staff and the AHRQ Quality Indicators team. The Model Reports (discussed immediately above) are based on: • Extensive search and analysis of the literature on hospital quality measurement and reporting, as well as public reporting on health care quality more broadly; • Interviews with quality measurement and reporting experts, purchasers, staff of purchasing coalitions, and executives of integrated health care delivery systems who are responsible for quality in their facilities; • Two focus groups with chief medical officers of hospitals and/or systems and two focus groups with quality managers from a broad mix of hospitals; • Four focus groups with members of the public who had recently experienced a hospital admission; and • Four rounds of cognitive interviews (a total of 62 interviews) to test draft versions of the two Model Reports with members of the public with recent hospital experience, basic computer literacy but widely varying levels of education **3a.6 Results** *(qualitative and/or quantitative results and conclusions)*: Given the above review of the literature and original research that was conducted, a Model report was the result that could help sponsors use the best evidence on public reports so they are most likely to have the desired effects on quality	
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:** **5.1 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 measures found.	**3c** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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**

4a.1-2 How are the data elements that are needed to compute measure scores generated? Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	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*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	**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.** Coding professionals follow detail guidelines, are subject to training and credentialing requirements, peer review and audit	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** None **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** Administrative data are collected as part of the routine operations. Some staff time is required to download and execute the software from the AHRQ webs site, which is available at no cost. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm **4e.3 Evidence for costs:** All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm **4e.4 Business case documentation:** All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Co.2 Point of Contact
John, Bott, MSSW, MBA, John.Bott@AHRQ.hhs.gov, 301-427-1317-

Measure Developer If different from Measure Steward
Co.3 Organization
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Co.4 Point of Contact
John, Bott, MSSW, MBA, John.Bott@AHRQ.hhs.gov, 301-427-1317-

Co.5 Submitter If different from Measure Steward POC
John, Bott, MSSW, MBA, John.Bott@AHRQ.hhs.gov, 301-427-1317-, Agency for Healthcare Research and Quality

Co.6 Additional organizations that sponsored/participated in measure development
UC Davis,
Stanford University,
Battelle Memorial Institute

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.
None

Ad.2 If adapted, provide name of original measure: None
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: 2001
Ad.7 Month and Year of most recent revision: 10, 2010
Ad.8 What is your frequency for review/update of this measure? Annual
Ad.9 When is the next scheduled review/update for this measure? 05, 2011

Ad.10 Copyright statement/disclaimers: The AHRQ QI software is publicly available; no copyright disclaimers.

Ad.11 -13 Additional Information web page URL or attachment:

Date of Submission (*MM/DD/YY*)**:** 04/05/2011

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 0351 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Death among surgical inpatients with serious, treatable complications (PSI 4)
De.2 **Brief description of measure:** Percentage of cases having developed specified complications of care with an in-hospital death.
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** Not applicable
De.4 **National Priority Partners Priority Area:** Population health, Safety De.5 **IOM Quality Domain**: Effectiveness De.6 **Consumer Care Need:** Getting better

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary 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	**B** Y☐

every 3 years. Yes, information provided in contact section	N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	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:** Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Pending update. This indicator was originally proposed by Silber et al.31 as a more powerful tool than the risk adjusted mortality rate to detect true differences in patient outcomes across hospitals. The underlying premise was that better hospitals are distinguished not by having fewer adverse occurrences but by more successfully averting death among (i.e., rescuing) patients who experience such complications. Silber et al's original definition was based on key clinical findings abstracted from the medical records of 2,831 cholecystectomy patients and 3,141 transurethral prostatectomy patients admitted to 531 hospitals in 1985. The key postoperative diagnoses that defined the denominator at risk of "failure to rescue" included cardiac arrhythmias, congestive heart failure, cardiac arrest, pneumonia, pulmonary embolus, pneumothorax, renal dysfunction, stroke, wound infection, and unplanned return to surgery. More recently, Needleman and Buerhaus137 adapted failure to rescue to administrative data sets, hypothesizing that this outcome might be sensitive to nurse staffing. Their denominator definition included the ICD-9-CM codes for sepsis, pneumonia (including aspiration), acute upper gastrointestinal bleeding, shock, cardiac/respiratory arrest, deep vein thrombosis (DVT), and pulmonary embolus (PE). 1a.4 **Citations for Evidence of High Impact:** Updated citations will be presented in the May Steering Committee meeting	1a C☐ P☐ M☐ N☐

Measures of Patient Safety Based on Hospital Administrative Data -
The Patient Safety Indicators, August 2002
http://qualityindicators.ahrq.gov/downloads/technical/psi_technical_review.zip

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Silber and colleagues have published a series of studies establishing the construct validity of failure to rescue rates through their associations with
hospital characteristics and other measures of hospital performance. Among patients admitted for cholecystectomy and transurethral prostatectomy, failure to rescue was independent of severity of illness at admission, but was significantly associated with the presence of surgical housestaff and a lower percentage of board-certified anesthesiologists.31 The adverse occurrence rate was independent of this hospital characteristic. In a larger sample of 74,647 patients who underwent general surgical procedures in 1991-92, lower failure to rescue rates were found at hospitals with high ratios of registered nurses to beds.68 Failure rates were strongly associated with risk adjusted mortality rates, as expected, but not with complication rates.143 Finally, among
16,673 patients admitted for coronary artery bypass surgery, failure rates were lower (whereas complication rates were higher) at hospitals with magnetic resonance imaging facilities, bone marrow transplantation units, or approved residency training programs.32 More recently, Needleman and Buerhaus137 confirmed that higher registered nurse staffing (RN hours/adjusted patient day) and better nursing skill mix (RN hours/licensed nurse hours) were consistently associated with lower failure to rescue rates among major surgery patients from 799 hospitals in 11 states in 1997, even using administrative data to define complications. An increase from the 25th to the 75th percentile on these two
measures of staffing was associated with 5.9% (95% CI, 1.5% to 10.2%) and 3.9% (95% CI, -1.1% to 8.8%) decreases, respectively, in the rate of failure-to-rescue among major surgery patients.138 These associations were inconsistent among medical patients, in that nursing skill mix was associated with the failure-to-rescue rate (rate ratio 0.81, 95% CI 0.66-1.00) but aggregate registered nurse staffing was not (rate ratio 1.00, 95% CI 0.99-1.01). An increase from the 25th to the 75th percentile on nursing skill mix was associated with a 2.5% (95% CI, 0.0% to 5.0%) decrease in the failure-to-rescue rate among medical patients.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
1) Signal Variance 2) Signal Standard Deviation 3) Better Than Average 4) Worse than Average (95% probability interval)

1) 0.000996672391 2) 0.031570118641 3) 1.89% 4) 3.92%

1b.3 Citations for data on performance gap:
AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

1b.4 Summary of Data on disparities by population group:
1) Estimate 2) Standard error 3) P-value: Relative to marked group-c 4) P-value:
2007 relative to 2006
Median income of patient´s ZIP code:
First quartile (lowest income) 107.685 0.446 0.000 0.000
Second quartile 106.520 0.514 0.000 0.000
Third quartile 103.842 0.541 0.423 0.000
Fourth quartile (highest income)c 103.204 0.583 0.000

Expected payment source:
Private insurancec 101.823 0.497 0.000
Medicare 103.325 0.362 0.015 0.000
Medicaid 110.349 0.684 0.000 0.000
Other insurance 114.903 1.368 0.000 0.303
Uninsured / self-pay / no charge 126.797 1.093 0.000 0.000

1b.5 Citations for data on Disparities:

1b
C☐
P☐
M☐
N☐

AHRQ 2007 Nationwide Inpatient Sample (NIS) with 800 hospitals and 7 million discharges

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*)**:** Mortality is a frequent outcome among patients with serious treatable complications

1c.2-3. Type of Evidence: Expert opinion, Systematic synthesis of research

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Silber and colleagues have published a series of studies establishing the construct validity of failure to rescue rates through their associations with
hospital characteristics and other measures of hospital performance. Among patients admitted for cholecystectomy and transurethral prostatectomy, failure to rescue was independent of severity of illness at admission, but was significantly associated with the presence of surgical housestaff and a lower percentage of board-certified anesthesiologists.31 The adverse occurrence rate was independent of this hospital characteristic. In a larger sample of 74,647 patients who underwent general surgical procedures in 1991-92, lower failure to rescue rates were found at hospitals with high ratios of registered nurses to beds.68 Failure rates were strongly associated with risk adjusted mortality rates, as expected, but not with complication rates.143 Finally, among 16,673 patients admitted for coronary artery bypass surgery, failure rates were lower
(whereas complication rates were higher) at hospitals with magnetic resonance imaging facilities, bone marrow transplantation units, or approved residency training programs.32

More recently, Needleman and Buerhaus137 confirmed that higher registered nurse
staffing (RN hours/adjusted patient day) and better nursing skill mix (RN hours/licensed nurse hours) were consistently associated with lower failure to rescue rates among major surgery patients from 799 hospitals in 11 states in 1997, even using administrative data to define complications. An increase from the 25th to the 75th percentile on these two measures of staffing was associated with 5.9% (95% CI, 1.5% to 10.2%) and 3.9% (95% CI, -1.1% to 8.8%) decreases, respectively, in the rate of failure-to-rescue among major surgery patients.138 These associations were inconsistent among medical patients, in that nursing skill mix was associated with the failure-to-rescue rate (rate ratio 0.81, 95% CI 0.66-1.00) but aggregate registered nurse staffing was not (rate ratio 1.00, 95% CI 0.99-1.01). An increase from the 25th to the 75th percentile on nursing skill mix was associated with a 2.5% (95% CI, 0.0% to 5.0%) decrease in the failure-to-rescue rate among medical patients.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
Testing, rating, and review were conducted by the project team. A full report on the literature review and empirical evaluation can be found in Refinement of the HCUP Quality Indicators by the UCSF-Stanford EPC, Detailed coding information for each QI is provided in the document Prevention Quality Indicators Technical Specifications. Rating of performance on empirical evaluations, ranged from 0 to 26. The scores were intended as a guide for summarizing the performance of each indicator on four empirical tests of precision (signal variance, area-level share, signal ratio, and R-squared) and five tests of minimum bias (rank correlation, top and bottom decile movement, absolute change, and change over two deciles), as described in the previous section.

1c.6 Method for rating evidence: The project team conducted empirical analyses to explore the frequency and variation of the indicators, the potential bias, based on limited risk adjustment, and the relationship between indicators. The data sources used in the empirical analyses were the 1997 Florida State Inpatient Database (SID) for initial testing and development and the 1997 HCUP State Inpatient Database for 19 States (referred to in this guide as the HCUP SID) for the final empirical analyses.

All potential indicators were examined empirically by developing and conducting statistical tests for precision, bias, and relatedness of indicators. Three different estimates of hospital performance were calculated for each indicator:

1. The raw indicator rate was calculated using the number of adverse events in the numerator divided by the

1c
C☐
P☐
M☐
N☐

number of discharges in the population at risk by hospital.
2. The raw indicator was adjusted to account for differences among hospitals in age, gender, modified DRG, and comorbidities.
- Adjacent DRG categories that were separated by the presence or absence of comorbidities or complications were collapsed to avoid adjusting for the complication being measured. Most of the super-Major Diagnostic Category (MDC) DRG categories were excluded for the same reason.
- APR-DRG risk adjustment was not implemented because removing applicable complications from each indicator was beyond the scope of this project.
- The ICD-9-CM codes used to define comorbidity categories were modified to exclude conditions likely to represent potentially preventable complications in certain settings.
- "Acute on chronic" comorbidities were captured so that some patients with especially severe comorbidities would not be mislabeled as not having conditions of interest.
- Comorbidities in obstetric patients were added.
- 3. Multivariate signal extraction methods were applied to adjust for reliability by estimating the amount of "noise" (i.e., variation due to random error) relative to the amount of "signal" (i.e., systematic variation in hospital performance or reliability) for each indicator.

Similar reliability adjustment has been used in the literature for similar purposes.40 41 The project team constructed a set of statistical tests to examine precision, bias, and relatedness of indicators for all accepted Provider-level Indicators, and precision and bias for all accepted Area-level Indicators. It should be noted that rates based on fewer than 30 cases in the numerator or the denominator are not reported. This exclusion rule serves two purposes:
- It eliminates unstable estimates based on too few cases.
- It helps protect the identities of hospitals and patients.

1c.7 Summary of Controversy/Contradictory Evidence: Panelists expressed concern regarding patients with "do not resuscitate" (DNR) status. In cases where this DNR status is not a direct result of poor quality of care, it would be contrary to patient desire and poor quality of care to rescue a patient. In addition, very old patients?or patients with advanced cancer or HIV?may not desire or may be particularly difficult to rescue from these complications. As a result, this indicator definition was modified to exclude those patients age 75 years and older. In addition, panelists suggested the exclusion of patients admitted from long-term care facilities.
Panelists noted that several adverse incentives may be introduced by implementing this indicator. In particular, since some type of adjustment may be desirable, this indicator may encourage the upcoding of complications and comorbidities to inflate the denominator or manipulate risk adjustment. Others noted that this indicator could encourage irresponsible resource use and allocation, although this is likely to be a controversial idea. Finally, panelists emphasized that this indicator should be used internally by hospitals, as it is not validated for public reporting.

See the following for a complete treatment of the topic:
http://www.qualityindicators.ahrq.gov/downloads/psi/psi_guide_v31.pdf
Note: The Literature Review Findings column summarizes evidence specific to each potential concern on the link between the PQIs and quality of care, as described in step 3 above. A question mark (?) indicates that the concern is theoretical or suggested, but no specific evidence was found in the literature. A check mark indicates that the concern has been demonstrated in the literature.

1c.8 Citations for Evidence (*other than guidelines*): Updated citations will be presented in the May Steering Committee meeting

Silber JH, Williams SV, Krakauer H, Schwartz JS. Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue. Med Care 1992;30(7):615-29.
Silber J, Rosenbaum P, Ross R. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? J Am Stat Assoc 1995;90:7-18.
Silber JH, Rosenbaum PR, Williams SV, Ross RN, Schwartz JS. The relationship between choice of outcome measure and hospital rank in general surgical procedures: Implications for quality assessment. Int J Qual Health Care 1997;9(3):193-200.
Needleman J, Buerhaus PI, Mattke S, Stewart M, Zelevinsky K. Nurse Staffing and Patient Outcomes in Hospitals. Boston MA: Health Resources and Services Administration; 2001 February 28. Report No.:230-99-0021.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number*): Not applicable **1c.10 Clinical Practice Guideline Citation:** Not applicable **1c.11 National Guideline Clearinghouse or other URL:** Not applicable **1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom*):** Not applicable **1c.13 Method for rating strength of recommendation** (*If different from USPSTF system, also describe rating and how it relates to USPSTF)*: Not applicable **1c.14 Rationale for using this guideline over others:** Not applicable	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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*): All discharges with a disposition of "deceased" (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator. **2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator*):** Time window can be determined by user, but is generally a calendar year. **2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions*):** All discharges with a disposition of "deceased" (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator.	
2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured*): All surgical discharges age 18 years and older or MDC 14 (pregnancy, childbirth, and puerperium) defined by specific DRGs or MS-DRGs and an ICD-9-CM code for an operating room procedure, principal procedure within 2 days of admission OR admission type of elective (ATYPE=3) with potential complications of care listed in Death among Surgical definition (e.g., pneumonia, DVT/PE, sepsis, shock/cardiac arrest, or GI hemorrhage/acute ulcer). **2a.5 Target population gender:** Female **2a.6 Target population age range:** 18 and older	**2a-specs** **C☐** **P☐** **M☐** **N☐**

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
Time window can be determined by user, but is generally a calendar year.

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
All surgical discharges age 18 years and older or MDC 14 (pregnancy, childbirth, and puerperium) defined by specific DRGs or MS-DRGs and an ICD-9-CM code for an operating room procedure, principal procedure within 2 days of admission OR admission type of elective (ATYPE=3) with potential complications of care listed in Death among Surgical definition (pneumonia, DVT/PE, sepsis, shock/cardiac arrest, or GI hemorrhage/acute ulcer).

See Patient Safety Indicators Appendices:
• Appendix A – Operating Room Procedure Codes
• Appendix D – Surgical Discharge DRGs
• Appendix E – Surgical Discharge MS-DRGs
PSI appendices at:
http://www.qualityindicators.ahrq.gov/downloads/psi/TechSpecs42/PSI%20Appendices.pdf:

FTR 2 - DVT/PE: Denominator
A diagnosis of pulmonary embolism or deep vein thrombosis in any secondary diagnosis field

ICD-9-CM Pulmonary Embolism and Deep Vein Thrombosis diagnosis codes:
Pulmonary Embolism
4151
PULMONARY EMBOLISM AND INFARCTION
41511
IATROGENIC PULMONARY EMBOLISM AND INFARCTION
41519
PULMONARY EMBOLISM AND INFARCTION, OTHER
Deep Vein Thrombosis
45111
PHLEBITIS AND THROMBOSIS OF FEMORAL VEIN (DEEP) (SUPERFICIAL)
45119
PHLEBITIS AND THROMBOPHLEBITIS OF DEEP VESSEL OF LOWER EXTREMITIES – OTHER
4512
PHLEBITIS AND THROMBOPHLEBITIS OF LOWER EXTREMITIES UNSPECIFIED
45181
PHLEBITIS AND THROMBOPHLEBITIS OF ILIAC VEIN
4519
PHLEBITIS AND THROMBOPHLEBITIS OF OTHER SITES - OF UNSPECIFIED SITE
45340
DVT-EMBLSM LOWER EXT NOS (OCT 04)
45341
DVT-EMB PROX LOWER EXT (OCT 04)
45342
DVT-EMB DISTAL LOWER EXT (OCT 04)
4538
OTHER VENOUS EMBOLISM AND THROMBOSIS OF OTHER SPECIFIED VEINS
4539
OTHER VENOUS EMBOLISM AND THROMBOSIS OF UNSPECIFIED SITE

FTR 3 – Pneumonia: Denominator
A diagnosis of pneumonia in any secondary diagnosis field

ICD-9-CM Pneumonia diagnosis codes:
4820
PNEUMONIA DUE TO KLEBSIELLA PNEUMONIAE

4821
PNEUMONIA DUE TO PSEUDOMONAS
4822
PNEUMONIA DUE TO HEMOPHILUS INFLUENZAE [H. INFLUENZAE]
4823
PNEUMONIA DUE TO STREPTOCOCCUS
48230
PNEUMONIA DUE TO STREPTOCOCCUS – STREPTOCOCCUS, UNSPECIFIED
48231
PNEUMONIA DUE TO STREPTOCOCCUS – GROUP A
48232
PNEUMONIA DUE TO STREPTOCOCCUS – GROUP B
48239
PNEUMONIA DUE TO STREPTOCOCCUS – OTHER STREPTOCOCCUS
4824
PNEUMONIA DUE TO STAPHYLOCOCCUS
48240
PNEUMONIA DUE TO STAPHYLOCOCCUS – PNEUMONIA DUE TO STAPHYLOCOCCUS, UNSPECIFIED
48241
METHICILLIN SUSCEPTIBLE PNEUMONIA DUE TO STAPHYLOCOCCUS AUREUS OCT08-
48242
METHICILLIN RESISTANT PNEUMONIA DUE TO STAPHYLOCOCCUS AUREUS OCT08-
48249
PNEUMONIA DUE TO STAPHYLOCOCCUS – OTHER STAPHYLOCOCCUS PNEUMONIA
4828
PNEUMONIA DUE TO OTHER SPECIFIED BACTERIA
48281
PNEUMONIA DUE TO OTHER SPECIFIED BACTERIA – ANAEROBES
48282
PNEUMONIA DUE TO OTHER SPECIFIED BACTERIA – EXCHERICHIA COLI [E COLI]
48283
PNEUMONIA DUE TO OTHER SPECIFIED BACTERIA – OTHER GRAM-NEGATIVE BACTERIA
48284
PNEUMONIA DUE TO OTHER SPECIFIED BACTERIA – LEGIONNAIRES´ DISEASE
48289
PNEUMONIA DUE TO OTHER SPECIFIED BACTERIA – OTHER SPECIFIED BACTERIA
4829
BACTERIAL PNEUMONIA UNSPECIFIED
485
BRONCHOPNEUMONIA, ORGANISM UNSPECIFIED
486
PNEUMONIA, ORGANISM UNSPECIFIED
5070
DUE TO INHALATION OF FOOD OR VOMITUS
514
PULMONARY CONGESTION AND HYPOSTASIS

FTR 4 – Sepsis: Denominator
A diagnosis of sepsis in any secondary diagnosis field

Include ICD-9-CM Sepsis diagnosis codes:
0380
STREPTOCOCCAL SEPTICEMIA
0381
STAPHYLOCOCCAL SEPTICEMIA
03810
STAPHYLOCOCCAL SEPTICEMIA, UNSPECIFIED
03811

METHICILLIN SUSCEPTIBLE STAPHYLOCOCCUS AUREUS SEPTICEMIA OCT08-
03812
METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS SEPTICEMIA OCT08-
03819
OTHER STAPHYLOCOCCAL SEPTICEMIA
0382
PNEUMOCOCCAL SEPTICEMIA (STREPTOCOCCUS PNEUMONIAE SEPTICEMIA)
0383
SEPTICEMIA DUE TO ANAEROBES
03840
GRAM-NEGATIVE ORGANISM, UNSPECIFIED
03841
HEMOPHILUS INFLUENZAE
03842
ESCHERICHIA COLI
03843
PSEUDOMONAS
03844
SERRATIA
03849
SEPTICEMIA DUE TO OTHER GRAM-NEGATIVE ORGANISMS
0388
OTHER SPECIFIED SEPTICEMIAS
0389
UNSPECIFIED SEPTICEMIA
78552
SEPTIC SHOCK OCT03-
78559*
SHOCK W/O MENTION OF TRAUMA- OTHER
99591
SYSTEMIC INFLAMMATORY RESPONSE SYNDROME DUE TO INFECTIOUS PROCESS W/O ORGAN DYSFUNCTION
99592
SYSTEMIC INFLAMMATORY RESPONSE SYNDROME DUE TO INFECTIOUS PROCESS W/ ORGAN DYSFUNCTION
9980
POSTOPERATIVE SHOCK
*No longer valid in FY2005

FTR 5 - Shock or Cardiac Arrest: Denomniator
A diagnosis of shock or cardiac arrest in any secondary field or any procedure for shock or cardiac arrest

Include ICD-9-CM Shock or Cardiac Arrest diagnosis codes:
4275
CARDIAC ARREST
6395
COMPLICATIONS FOLLOWING ABORTION AND ECTOPIC AND MOLAR PREGNANCIES, SHOCK
66910
SHOCK DURING OR FOLLOWING LABOR AND DELIVERY – UNSPECIFIED AS TO EPISODE OF CARE OR NOT APPLICABLE
66911
SHOCK DURING OR FOLLOWING LABOR AND DELIVERY – DELIVERED, W/ OR W/O MENTION OF ANTEPARTUM CONDITION
66912
SHOCK DURING OR FOLLOWING LABOR AND DELIVERY – DELIVERED, W/ MENTION OF POSTPARTUM COMPLICATION
66913
SHOCK DURING OR FOLLOWING LABOR AND DELIVERY – ANTEPARTUM CONDITION OR COMPLICATION
66914
SHOCK DURING OR FOLLOWING LABOR AND DELIVERY – POSTPARTUM CONDITION OR COMPLICATION

7855
SHOCK NOS
78550
SHOCK, UNSPECIFIED
78551
CARDIOGENIC SHOCK
78552
SEPTIC SHOCK OCT03-
78559
SHOCK W/O MENTION OF TRAUMA- OTHER
7991
RESPIRATORY ARREST
9950
OTHER ANAPHYLACTIC SHOCK
9954
SHOCK DUE TO ANESTHESIA
9980
POSTOPERATIVE SHOCK
9994
ANAPHYLACTIC SHOCK DUE TO SERUM
ICD-9-CM Shock or Cardiac Arrest procedure codes:
9393
NONMECHANICAL METHODS OF RESUSCITATION
9960
CARDIOPULMONARY RESUSCITATION, NOS
9963
CLOSED CHEST CARDIAC MASSAGE

FTR 6 - GI Hemorrhage/Acute Ulcer: Denominator
A diagnosis of hemorrhage or acute ulcer in any secondary field

ICD-9-CM GI Hemorrhage/Acute Ulcer diagnosis codes:
4560
ESOPHAGEAL VARICES W/ BLEEDING
45620
ESOPHAGEAL VARICES IN DISEASES CLASSIFIED ELSEWHERE W/ BLEEDING
5307
GASTROESOPHAGEAL LACERATION-HEMORRHAGE SYNDROME
53082
ESOPHAGEAL HEMORRHAGE
Gastric ulcer:
53100
ACUTE W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53101
ACUTE W/ HEMORRHAGE – W/ OBSTRUCTION
53110
ACUTE W/ PERFORATION – W/O MENTION OF OBSTRUCTION
53111
ACUTE W/ PERFORATION – W/ OBSTRUCTION
53120
ACUTE W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53121
ACUTE W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53130
ACUTE W/O MENTION OF HEMORRHAGE OR PERFORATION – W/O MENTION OF OBSTRUCTION
53131
ACUTE W/O MENTION OF HEMORRHAGE OR PERFORATION – W/ OBSTRUCTION
53190

UNSPECIFIED AS ACUTE OR CHRONIC, W/O MENTION OF HEMORRHAGE OR PERFORATION – W/O MENTION OF OBSTRUCTION
53191
UNSPECIFIED AS ACUTE OR CHRONIC, W/O MENTION OF HEMORRHAGE OR PERFORATION – W/ OBSTRUCTION
Duodenal ulcer:
53200
ACUTE W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53201
ACUTE W/ HEMORRHAGE – W/ OBSTRUCTION
53210
ACUTE W/ PERFORATION – W/O MENTION OF OBSTRUCTION
53211
ACUTE W/ PERFORATION – W/ OBSTRUCTION
53220
ACUTE W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53221
ACUTE W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53230
ACUTE W/O MENTION OF HEMORRHAGE OR PERFORATION – W/O MENTION OF OBSTRUCTION
53231
ACUTE W/O MENTION OF HEMORRHAGE OR PERFORATION – W/ OBSTRUCTION
53290
UNSPECIFIED AS ACUTE OR CHRONIC, W/O MENTION OF HEMORRHAGE OR PERFORATION – W/O MENTION OF OBSTRUCTION
53291
UNSPECIFIED AS ACUTE OR CHRONIC, W/O MENTION OF HEMORRHAGE OR PERFORATION – W/ OBSTRUCTION
Peptic ulcer:
53300
SITE UNSPECIFIED ACUTE W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53301
SITE UNSPECIFIED ACUTE W/ HEMORRHAGE – W/ OBSTRUCTION
53310
SITE UNSPECIFIED ACUTE W/ PERFORATION – W/O MENTION OF OBSTRUCTION
53311
SITE UNSPECIFIED ACUTE W/ PERFORATION – W/ OBSTRUCTION
53320
SITE UNSPECIFIED ACUTE W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53321
SITE UNSPECIFIED ACUTE W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53330
SITE UNSPECIFIED ACUTE W/O MENTION OF HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53331
SITE UNSPECIFIED ACUTE W/O MENTION OF HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53390
SITE UNSPECIFIED AS ACUTE OR CHRONIC, W/O MENTION OF HEMORRHAGE OR PERFORATION – W/O MENTION OF OBSTRUCTION
53391
UNSPECIFIED AS ACUTE OR CHRONIC, W/O MENTION OF HEMORRHAGE OR PERFORATION – W/ OBSTRUCTION
Gastrojejunal ulcer:
53400
ACUTE W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53401
ACUTE W/ HEMORRHAGE – W/ OBSTRUCTION
53410
ACUTE W/ PERFORATION – W/O MENTION OF OBSTRUCTION
53411
ACUTE W/ PERFORATION – W/ OBSTRUCTION

53420
ACUTE W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53421
ACUTE W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53430
ACUTE W/O MENTION OF HEMORRHAGE OR PERFORATION – W/O MENTION OF OBSTRUCTION
53431
ACUTE W/O MENTION OF HEMORRHAGE OR PERFORATION – W/ OBSTRUCTION
53490
UNSPECIFIED AS ACUTE OR CHRONIC, W/O MENTION OF HEMORRHAGE OR PERFORATION – W/O MENTION OF OBSTRUCTION
53491
UNSPECIFIED AS ACUTE OR CHRONIC, W/O MENTION OF HEMORRHAGE OR PERFORATION – W/ OBSTRUCTION
Gastritis and duodenitis:
53501
ACUTE GASTRITIS – W/ HEMORRHAGE
53511
ATROPHIC GASTRITIS – W/ HEMORRHAGE
53521
GASTRIC MUCOSAL HYPERTROPHY – W/ HEMORRHAGE
53531
ALCOHOLIC GASTRITIS – W/ HEMORRHAGE
53541
OTHER SPECIFIED GASTRITIS – W/ HEMORRHAGE
53551
UNSPECIFIED GASTRITIS AND GASTRODUODENITIS – W/ HEMORRHAGE
53561
DUODENITIS – W/ HEMORRHAGE
53783
ANGIODYSPLASIA OF STOMACH AND DUODENUM – W/ HEMORRHAGE
53784
DIEULAFOY LESION (HEMORRHAGIC) OF STOMACH AND DUODENUM
56202
DIVERTICULOSIS OF SMALL INTESTINE – W/ HEMORRHAGE
56203
DIVERTICULITIS OF SMALL INTESTINE – W/ HEMORRHAGE
56212
DIVERTICULOSIS OF COLON – W/ HEMORRHAGE
56213
DIVERTICULITIS OF COLON – W/ HEMORRHAGE
5693
HEMORRHAGE OF RECTUM AND ANUS
56985
ANGIODYSPLASIA OF INTESTINE – W/ HEMORRHAGE
56986
DIEULAFOY LESION (HEMORRHAGIC) OF INTESTINE
5780
HEMATEMESIS
5781
BLOOD IN STOOL
5789
HEMORRHAGE OF GASTROINTESTINAL TRACT, UNSPECIFIED

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** Exclude cases:
• age 90 years and older
• transferred to an acute care facility (DISP = 2)
• missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1 =missing)

NOTE: Additional exclusion criteria is specific to each diagnosis (pneumonia, DVT/PE, sepsis, shock/cardiac arrest, or GI hemorrhage/acute ulcer). See 2a.10.

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
Exclude cases:
• age 90 years and older
• transferred to an acute care facility (DISP = 2)
• missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1 =missing)

NOTE: Additional exclusion criteria is specific to each diagnosis (pneumonia, DVT/PE, sepsis, shock/cardiac arrest, or GI hemorrhage/acute ulcer). See below for specifics.

FTR 2 - DVT/PE: Exclusions
• with a diagnosis of pulmonary embolism or deep vein thrombosis in the primary diagnosis field (Defined in 2a.8)
• with a diagnosis of abortion-related or postpartum obstetric pulmonary embolism in the primary diagnosis field

ICD-9-CM Abortion-related and Postpartum Obstetric Pulmonary Embolism diagnosis codes:
63460
SPONTANEOUS ABORTION W/ EMBOLISM - UNSPECIFIED
63461
SPONTANEOUS ABORTION W/ EMBOLISM - INCOMPLETE
63462
SPONTANEOUS ABORTION W/ EMBOLISM - COMPLETE
63560
LEGAL ABORTION W/ EMBOLISM - UNSPECIFIED
63561
LEGAL ABORTION W/ EMBOLISM - INCOMPLETE
63562
LEGAL ABORTION W/ EMBOLISM - COMPLETE
63660
ILLEGAL ABORTION W/ EMBOLISM - UNSPECIFIED
63661
ILLEGAL ABORTION W/ EMBOLISM - INCOMPLETE
63662
ILLEGAL ABORTION W/ EMBOLISM - COMPLETE
63760
ABORTION NOS W/ EMBOLISM - UNSPECIFIED
63761
ABORTION NOS W/ EMBOLISM - INCOMPLETE
63762
ABORTION NOS W/ EMBOLISM - COMPLETE
6386
ATTEMPTED ABORTION W/ EMBOLISM
6396
POSTABORTION EMBOLISM
67320
OBSTETRICAL BLOOD-CLOT EMBOLISM, UNSPECIFIED AS TO EPISODE OF CARE OR NOT APPLICABLE
67321
OBSTETRICAL BLOOD-CLOT EMBOLISM, DELIVERED, W/ OR W/O MENTION OF ANTEPARTUM CONDITION
67322
OBSTETRICAL BLOOD-CLOT EMBOLISM, DELIVERED, W/ MENTION OF POSTPARTUM COMPLICATION
67323
OBSTETRICAL BLOOD-CLOT EMBOLISM, ANTEPARTUM CONDITION OR COMPLICATION

67324
OBSTETRICAL BLOOD-CLOT EMBOLISM, POSTPARTUM CONDITION OR COMPLICATION

FTR 3 – Pneumonia: Exclusions
• with a diagnosis of pneumonia or respiratory complications in the primary diagnosis field (Defined in 2a.8)
• with any diagnosis code for viral pneumonia
• with any diagnosis of or procedure for immunocompromised state.
• MDC 4 (diseases/disorders of respiratory system)

See Patient Safety Indicators Appendices:
• Appendix I – Immunocompromised State Diagnosis and Procedure Codes
PSI appendices at:
http://www.qualityindicators.ahrq.gov/downloads/psi/TechSpecs42/PSI%20Appendices.pdf:

ICD-9-CM Respiratory Complications diagnosis code:
9973
RESPIRATORY COMPLICATIONS
ICD-9-CM Viral Pneumonia diagnosis codes:
4800
ADENOVIRAL PNEUMONIA
4801
RESPIRATORY SYNCYTIAL VIRAL PNEUMONIA
4802
PARAINFLUENZA VIRAL PNEUMONIA
4803
PNEUMONIA DUE TO SARS OCT03-
4808
VIRAL PNEUMONIA NOT ELSEWHERE CLASSIFIED
4809
VIRAL PNEUMONIA UNSPECIFIED
481
PNEUMOCOCCAL PNEUMONIA
4830
PNEUMONIA DUE TO MYCOPLASMA PNEUMONIAE
4831
PNEUMONIA DUE TO CHLAMYDIA
4838
PNEUMONIA DUE TO OTHER SPECIFIED ORGANISM
4841
PNEUMONIA IN CYTOMEGALIC INCLUSION DISEASE
4843
PNEUMONIA IN WHOOPING COUGH
4845
PNEUMONIA IN ANTHRAX
4846
PNEUMONIA IN ASPERGILLOSIS
4847
PNEUMONIA IN OTHER SYSTEMIC MYCOSES
4848
PNEUMONIA IN INFECTIOUS DISEASE NOT ELSEWHERE CLASSIFIED
4870
INFLUENZA W/ PNEUMONIA
4871
FLU W/ RESPIRATORY MANIFEST NOT ELSEWHERE CLASSIFIED
4878
FLU W/ MANIFESTATION NOT ELSEWHERE CLASSIFIED
488
FLU D/T AVIAN FLU VIRUS

4880
INFLUENZA DUE TO IDENTIFIED AVIAN INFLUENZA VIRUS OCT09-
4881
INFLUENZA DUE TO IDENTIFIED NOVEL H1N1 INFLUENZA VIRUS OCT09-

FTR 4 – Sepsis: Exclusions
• with a diagnosis of sepsis in the principal diagnosis field (Defined in 2a.8)
• with any diagnosis of infection
• with any diagnosis of or procedure for immunocompromised state
• with a length of stay of less than 4 days

See Patient Safety Indicators Appendices:
• Appendix F – Infection Diagnosis Codes
• Appendix I – Immunocompromised State Diagnosis and Procedure Codes
PSI appendices at:
http://www.qualityindicators.ahrq.gov/downloads/psi/TechSpecs42/PSI%20Appendices.pdf:

FTR 5 - Shock or Cardiac Arrest: Exclusions
• with a primary diagnosis of shock or cardiac arrest (Defined in 2a.8)
• with a primary diagnosis of trauma
• with a primary diagnosis of hemorrhage or GI hemorrhage
• with a primary diagnosis of abortion-related shock
• MDC 4 (diseases/disorders of respiratory system)
• MDC 5 (diseases/disorders of circulatory system)

See Patient Safety Indicators Appendices:
• Appendix G – Trauma Diagnosis Codes
PSI appendices at:
http://www.qualityindicators.ahrq.gov/downloads/psi/TechSpecs42/PSI%20Appendices.pdf:

ICD-9-CM Hemorrhage diagnosis codes:
2851
ACUTE POSTHEMORRHAGIC ANEMIA
4590
OTHER DISORDERS OF CIRCULATORY SYSTEM, HEMORRHAGE, UNSPECIFIED
56881
HEMOPERITONEUM (NONTRAUMATIC)
9582
CERTAIN EARLY COMPLICATIONS OF TRAUMA, SECONDARY AND RECURRENT HEMORRHAGE
99811
HEMORRHAGE COMPLICATING A PROCEDURE
ICD-9-CM Gastrointestinal (GI) Hemorrhage diagnosis codes:
4560
ESOPHAGEAL VARICES W/ BLEEDING
45620
ESOPHAGEAL VARICES IN DISEASES CLASSIFIED ELSEWHERE W/ BLEEDING
5307
GASTROESOPHAGEAL LACERATION – HEMORRHAGE SYNDROME
53082
ESOPHAGEAL HEMORRHAGE
53100
GASTRIC ULCER ACUTE W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53101
GASTRIC ULCER ACUTE W/ HEMORRHAGE – W/ OBSTRUCTION
53120
GASTRIC ULCER ACUTE W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53121
GASTRIC ULCER ACUTE W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION

53140
GASTRIC ULCER CHRONIC OR UNSPECIFIED W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53141
GASTRIC ULCER CHRONIC OR UNSPECIFIED W/ HEMORRHAGE – W/ OBSTRUCTION
53160
GASTRIC ULCER CHRONIC OR UNSPECIFIED W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53161
GASTRIC ULCER CHRONIC OR UNSPECIFIED W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53200
DUODENAL ULCER ACUTE W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53201
DUODENAL ULCER ACUTE W/ HEMORRHAGE – W/ OBSTRUCTION
53220
DUODENAL ULCER ACUTE W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53221
DUODENAL ULCER ACUTE W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53240
DUODENAL ULCER CHRONIC OR UNSPECIFIED W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53241
DUODENAL ULCER CHRONIC OR UNSPECIFIED W/ HEMORRHAGE – W/ OBSTRUCTION
53260
DUODENAL ULCER CHRONIC OR UNSPECIFIED W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53261
DUODENAL ULCER CHRONIC OR UNSPECIFIED W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53300
PEPTIC ULCER, SITE UNSPECIFIED, ACUTE W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53301
PEPTIC ULCER, SITE UNSPECIFIED, ACUTE W/ HEMORRHAGE – W/ OBSTRUCTION
53320
PEPTIC ULCER, SITE UNSPECIFIED, ACUTE W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53321
PEPTIC ULCER, SITE UNSPECIFIED, ACUTE W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53340
PEPTIC ULCER, SITE UNSPECIFIED, CHRONIC OR UNSPECIFIED W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53341
PEPTIC ULCER, SITE UNSPECIFIED, CHRONIC OR UNSPECIFIED W/ HEMORRHAGE – W/ OBSTRUCTION
53360
PEPTIC ULCER, SITE UNSPECIFIED, CHRONIC OR UNSPECIFIED W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53361
PEPTIC ULCER, SITE UNSPECIFIED, CHRONIC OR UNSPECIFIED W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53400
GASTROJEJUNAL ULCER, ACUTE W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53401
GASTROJEJUNAL ULCER, ACUTE W/ HEMORRHAGE – W/ OBSTRUCTION
53420
GASTROJEJUNAL ULCER, ACUTE W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53421
GASTROJEJUNAL ULCER, ACUTE W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53440
GASTROJEJUNAL ULCER, CHRONIC OR UNSPECIFIED W/ HEMORRHAGE – W/O MENTION OF OBSTRUCTION
53441
GASTROJEJUNAL ULCER, CHRONIC OR UNSPECIFIED W/ HEMORRHAGE – W/ OBSTRUCTION

53460
GASTROJEJUNAL ULCER, CHRONIC OR UNSPECIFIED W/ HEMORRHAGE AND PERFORATION – W/O MENTION OF OBSTRUCTION
53461
GASTROJEJUNAL ULCER, CHRONIC OR UNSPECIFIED W/ HEMORRHAGE AND PERFORATION – W/ OBSTRUCTION
53501
GASTRITIS AND DUODENITIS, ACUTE GASTRITIS W/ HEMORRHAGE
53511
GASTRITIS AND DUODENITIS, ATROPHIC GASTRITIS W/ HEMORRHAGE
53521
GASTRITIS AND DUODENITIS, GASTRIC MUCOSAL HYPERTROPHY, W/ HEMORRHAGE
53531
GASTRITIS AND DUODENITIS, ALCOHOLIC GASTRITIS, W/ HEMORRHAGE
53541
GASTRITIS AND DUODENITIS, OTHER SPECIFIED GASTRITIS – W/ HEMORRHAGE
53551
GASTRITIS AND DUODENITIS, UNSPECIFIED GASTRITIS AND GASTRODUODENITIS – W/ HEMORRHAGE
53561
GASTRITIS AND DUODENITIS, DUODENITIS – W/ HEMORRHAGE
53783
OTHER SPECIFIED DISORDERS OF STOMACH AND DUODENUM, ANGIODYSPLASIA OF STOMACH AND DUODENUM – W/ HEMORRHAGE
53784
DIEULAFOY LESION (HEMORRHAGIC) OF STOMACH AND DUODENUM
56202
DIVERTICULOSIS OF SMALL INTESTINE – W/ HEMORRHAGE
56203
DIVERTICULITIS OF SMALL INTESTINE – W/ HEMORRHAGE
56212
DIVERTICULOSIS OF COLON – W/ HEMORRHAGE
56213
DIVERTICULITIS OF COLON – W/ HEMORRHAGE
5693
HEMORRHAGE OF RECTUM AND ANUS
56985
ANGIODYSPLASIA OF INTESTINE - W/ HEMORRHAGE
56986
DIEULAFOY LESION (HEMORRHAGIC) OF INTESTINE
5780
GASTROINTESTINAL HEMORRHAGE, HEMATEMESIS
5781
GASTROINTESTINAL HEMORRHAGE, BLOOD IN STOOL
5789
GASTROINTESTINAL HEMORRHAGE, HEMORRHAGE OF GASTROINTESTINAL TRACT, UNSPECIFIED
ICD-9-CM Abortion-related Shock diagnosis codes:
63450
SPONTANEOUS ABORTION W/ SHOCK - UNSPECIFIED
63451
SPONTANEOUS ABORTION W/ SHOCK - INCOMPLETE
63452
SPONTANEOUS ABORTION W/ SHOCK - COMPLETE
63550
LEGAL ABORTION W/ SHOCK - UNSPECIFIED
63551
LEGAL ABORTION W/ SHOCK - INCOMPLETE
63552
LEGAL ABORTION W/ SHOCK - COMPLETE
63650

ILLEGAL ABORTION W/ SHOCK - UNSPECIFIED
63651
ILLEGAL ABORTION W/ SHOCK - INCOMPLETE
63652
ILLEGAL ABORTION W/ SHOCK - COMPLETE
63750
ABORTION NOS W/ SHOCK - UNSPECIFIED
63751
ABORTION NOS W/ SHOCK - INCOMPLETE
63752
ABORTION NOS W/ SHOCK - COMPLETE
6385
ATTEMPTED ABORTION W/ SHOCK

FTR 6 - GI Hemorrhage/Acute Ulcer: Exclusions
• with a primary diagnosis of hemorrhage or acute ulcer (Defined in 2a.8)
• with a primary diagnosis of trauma
• with a primary diagnosis of alcoholism
• with a primary diagnosis of anemia
• MDC 6 (diseases and disorders of the digestive system)
• MDC 7 (diseases and disorders of the hepatobiliary system and pancreas)

See Patient Safety Indicators Appendices:
• Appendix G – Trauma Diagnosis Codes
PSI appendices at:
http://www.qualityindicators.ahrq.gov/downloads/psi/TechSpecs42/PSI%20Appendices.pdf:

ICD-9-CM Alcoholism diagnosis codes:
2910
ALCOHOL WITHDRAWAL DELIRIUM
2911
ALCOHOL AMNESTIC SYNDROME
2912
OTHER ALCOHOLIC DEMENTIA
2913
ALCOHOL WITHDRAWAL HALLUCINOSIS
2914
IDIOSYNCRATIC ALCOHOL INTOXICATION
2915
ALCOHOLIC JEALOUSY
29181
OTHER SPECIFIED ALCOHOLIC PSYCHOSES, ALCOHOL WITHDRAWAL
29182
ALCOHOL INDUCED SLEEP DISORDERS OCT05-
29189
OTHER SPECIFIED ALCOHOLIC PSYCHOSES, OTHER
2919
UNSPECIFIED ALCOHOLIC PSYCHOSIS
30300
ACUTE ALCOHOLIC INTOXICATION - UNSPECIFIED
30301
ACUTE ALCOHOLIC INTOXICATION - CONTINUOUS
30302
ACUTE ALCOHOLIC INTOXICATION - EPISODIC
30303
ACUTE ALCOHOLIC INTOXICATION - IN REMISSION
30390
OTHER AND UNSPECIFIED ALCOHOL DEPENDENCE - UNSPECIFIED

30391
OTHER AND UNSPECIFIED ALCOHOL DEPENDENCE - CONTINUOUS
30392
OTHER AND UNSPECIFIED ALCOHOL DEPENDENCE - EPISODIC
30393
OTHER AND UNSPECIFIED ALCOHOL DEPENDENCE - IN REMISSION
30500
NONDEPENDENT ABUSE OF DRUGS, ALCOHOL ABUSE - UNSPECIFIED
30501
NONDEPENDENT ABUSE OF DRUGS, ALCOHOL ABUSE - CONTINUOUS
30502
NONDEPENDENT ABUSE OF DRUGS, ALCOHOL ABUSE - EPISODIC
30503
NONDEPENDENT ABUSE OF DRUGS, ALCOHOL ABUSE – IN REMISSION
4255
ALCOHOLIC CARDIOMYOPATHY
53530
ALCOHOLIC GASTRITIS, W/O MENTION OF HEMORRHAGE
53531
ALCOHOLIC GASTRITIS, W/ HEMORRHAGE
5710
ALCOHOLIC FATTY LIVER
5711
ACUTE ALCOHOLIC HEPATITIS
5712
ALCOHOLIC CIRRHOSIS OF LIVER
5713
ALCOHOLIC LIVER DAMAGE, UNSPECIFIED
9800
TOXIC EFFECT OF ALCOHOL, ETHYL ALCOHOL
9809
TOXIC EFFECT OF ALCOHOL, UNSPECIFIED ALCOHOL

ICD-9-CM Anemia diagnosis codes:
2800
SECONDARY TO BLOOD LOSS [CHRONIC]
2851
ACUTE POSTHEMORRHAGIC ANEMIA

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
User has an option to stratify by Gender, age (5-year age groups), race / ethnicity, primary payer, and custom stratifiers.

2a.12-13 Risk Adjustment Type: Risk adjustment method widely or commercially available

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age in years (in 5-year age groups), modified CMS DRG and AHRQ Comorbidities. The reference population used in the model is the universe of discharges for states that participate in the HCUP State Inpatient Databases (SID) for the year 2007 (updated annually), a database consisting of 43 states and approximately 30 million adult discharges. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., hospital, state, and region). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.

2a.15-17 Detailed risk model available Web page URL or attachment: URL None
http://qualityindicators.ahrq.gov/downloads/psi/PSI_Risk_Adjustment_Tables_(Version_4_2).pdf

2a.18-19 Type of Score: Rate/proportion
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***):**
Each indicator is expressed as a rate, is defined as outcome of interest / population at risk or numerator / denominator. The AHRQ Quality Indicators (AHRQ QI) software performs five steps to produce the rates. 1) Discharge-level data is used to mark inpatient records containing the outcome of interest and 2) the population at risk. For provider indicators, the population at risk is also derived from hospital discharge records; for area indicators, the population at risk is derived from U.S. Census data. 3) Calculate observed rates. Using output from steps 1 and 2, rates are calculated for user-specified combinations of stratifiers. 4) Calculate expected rates. Regression coefficients from a reference population database are applied to the discharge records and aggregated to the provider or area level. 5) Calculate risk-adjusted rate. Use the indirect standardization to account for case-mix. 6) Calculate smoothed rate. A Univariate shrinkage factor is applied to the risk-adjusted rates. The shrinkage estimate reflects a reliability adjustment unique to each indicator. Full information on calculation algorithms and specifications can be found at http://qualityindicators.ahrq.gov/PSI_download.htm

2a.22 Describe the method for discriminating performance *(e.g., significance testing)*:
Significance testing is not prescribed by the software. Users may calculate a confidence interval for the risk-adjusted rates and a posterior probability interval for the smoothed rates at a 95% or 99% level. Users may define the relevant benchmark and the methods of discriminating performance according to their application.

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)*:
Not applicable

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Electronic administrative data/claims

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.***):**
The data source is hospital discharge data such as the HCUP State Inpatient Databases (SID) or equivalent using UB-04 coding standards. The data collection instrument is public-use AHRQ QI software available in SAS or Windows versions.

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL None
http://www.qualityindicators.ahrq.gov/software.htm

2a.29-31 Data dictionary/code table web page URL or attachment: URL None
http://www.qualityindicators.ahrq.gov/downloads/winqi/AHRQ_QI_Windows_Software_Documentation_V41a.pdf

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency

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

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million discharges

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
Literature review, expert panels and empirical analysis

2b
C☐
P☐
M☐
N☐

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:

PSI 4 A higher risk-adjusted mortality rate for death among surgical inpatients with serious treatable complications is associated with significantly higher costs. The AHRQ QIs have the advantage of taking the multidimensional nature of hospital quality into account. As the coefficients on the AHRQ QIs show, measures of hospital quality can have conflicting effects on hospital costs. A single measure that combines these effects into one variable offers less insight into hospital performance than the outcomes for each measure. [1]

Patient Safety Events Are Common at U.S. Hospitals: Between 2005 and 2007 there were 913,215 total patient safety events among Medicare beneficiaries. Common Patient Safety Events are Very Costly: Between 2005 and 2007 these patient safety events were associated with over $6.9 billion of wasted healthcare cost. Less Improvement Seen Among Most Common Events: Eight patient safety indicators showed improvement while seven indicators worsened in 2007 compared to 2005. Some of the most common and most serious indicators worsened, including decubitus ulcer (bed sores), sepsis, respiratory failure, deep vein thrombosis (blood clots in the legs), and pulmonary embolism (potentially fatal blood clots forming in the lungs). Approximately One-in-Ten Medicare Patients with Patient Safety Events Died: Between 2005 and 2007 there were 97,755 actual inhospital deaths that occurred among patients who experienced one or more of the 15 patient safety events. [2]

PSI 4: death among surgical inpatients with serious treatable complications was not included because many procedure codes are required. [3]

The initial translation (electronic mapping, review and revision by expert coder, programming of codes and testing on data from 1996-1998 [ICD 9-CM] to 1998-2006 [ICD-10-AM, through 4 editions]) found that differences between ICD-9-CM and ICD-10-AM datasets presented some challenges. After this phase, which was faithful to AHRQ´s case definitions, the indicators were refined for use with the condition onset flag, resulting in the AusPSIs. [4]

Principal Findings. Excess 90-day expenditures likely attributable to PSIs ranged from $646 for technical problems (accidental laceration, pneumothorax, etc.) to $28,218 for acute respiratory failure, with up to 20 percent of these costs incurred postdischarge. With a third of all 90-day deaths occurring postdischarge, the excess death rate associated with PSIs ranged from 0 to 7 percent. The excess 90-day readmission rate associated with PSIs ranged from 0 to 8 percent. Overall, 11 percent of all deaths, 2 percent of readmissions, and 2 percent of expenditures were likely due to these 14 PSIs. Conclusions. The effects of medical errors continue long after the patient leaves the hospital. Medical error studies that focus only on the inpatient stay can underestimate the impact of patient safety events by up to 20-30 percent. [5]

References

[1] Laditka JN, Laditka SB, Cornman CB. Evaluating hospital care for individuals with Alzheimer´s disease using inpatient quality indicators. Am J Alzheimers Dis Other Demen. 2005 Jan-Feb;20(1):27-36. PMID: 15751451.
[2] HealthGrades. Every 1.7 Minutes a Medicare Beneficiary Experiences a Patient Safety Event. Business Wire. Available on-line: http://www.allbusiness.com/government/government-bodies-offices/12279340-1.html. Accessed 1/11/2011.
[3] Hude Quan, MD, PhD; Saskia Drösler, MD; Vijaya Sundararajan, et al. Adaptation of AHRQ Patient Safety Indicators for Use in ICD-10 Administrative Data by an International Consortium. In Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 1: Assessment). Henriksen K, Battles JB, Keyes MA, et al., editors. Rockville (MD): Agency for Healthcare Research and Quality; 2008 Aug. Bookshelf ID: NBK43634.
[4] McConchie S, Shepheard J, Waters S, McMillan AJ, Sundararajan V. The AusPSIs: the Australian version of the Agency of Healthcare Research and Quality patient safety indicators. Aust Health Rev. 2009 May;33(2):334-41. PMID: 19563325.
[5] Encinosa WE, Hellinger FJ. The impact of medical errors on ninety-day costs and outcomes: an examination of surgical patients. Health Serv Res. 2008 Dec;43(6):2067-85. Epub 2008 Jul 25. PMID: 18662169; DOI: 10.1111/j.1475-6773.2008.00882.x

2c. Validity testing — **2c** C☐ P☐

2c.1 Data/sample *(description of data/sample and size)*: We restricted our analysis to 20 states (4) for which

HCUP State Inpatient Databases (SID) were available. There were 1,601 nonfederal, urban, general hospitals in those 20 states. Over 300 hospitals were eliminated from the sample because of key missing variables in the American Hospital Association (AHA) Annual Survey of Hospital data, which was also used for this study, or because they had missing observations for some of the OIs that we used. Thus, our sample consisted of 1,290 urban, acute-care hospitals for which complete data were available for 2001. [1]

The Agency for Healthcare Research and Quality Patient Safety Indicators (PSIs) were used to identify 14 PSIs among 161,004 surgeries. [5]

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
A likelihood ratio test of the hypothesis that the coefficients on all of these variables were equal to 0 (lambda) = 35.3, p< .01). [1]

We used propensity score matching and multivariate regression analyses to predict expenditures and outcomes attributable to the 14 PSIs. [5]

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
PSI 4 A higher risk-adjusted mortality rate for death among surgical inpatients with serious treatable complications is associated with significantly higher costs. The AHRQ QIs have the advantage of taking the multidimensional nature of hospital quality into account. As the coefficients on the AHRQ QIs show, measures of hospital quality can have conflicting effects on hospital costs. A single measure that combines these effects into one variable offers less insight into hospital performance than the outcomes for each measure.[1]

Principal Findings. Excess 90-day expenditures likely attributable to PSIs ranged from $646 for technical problems (accidental laceration, pneumothorax, etc.) to $28,218 for acute respiratory failure, with up to 20 percent of these costs incurred postdischarge. With a third of all 90-day deaths occurring postdischarge, the excess death rate associated with PSIs ranged from 0 to 7 percent. The excess 90-day readmission rate associated with PSIs ranged from 0 to 8 percent. Overall, 11 percent of all deaths, 2 percent of readmissions, and 2 percent of expenditures were likely due to these 14 PSIs. Conclusions. The effects of medical errors continue long after the patient leaves the hospital. Medical error studies that focus only on the inpatient stay can underestimate the impact of patient safety events by up to 20-30 percent. [5]

References
[1] Laditka JN, Laditka SB, Cornman CB. Evaluating hospital care for individuals with Alzheimer´s disease using inpatient quality indicators. Am J Alzheimers Dis Other Demen. 2005 Jan-Feb;20(1):27-36. PMID: 15751451.
[5] Encinosa WE, Hellinger FJ. The impact of medical errors on ninety-day costs and outcomes: an examination of surgical patients. Health Serv Res. 2008 Dec;43(6):2067-85. Epub 2008 Jul 25. PMID: 18662169; DOI: 10.1111/j.1475-6773.2008.00882.

M☐ N☐

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
Exclusions remove cases where the outcome of interest is less likely to be preventable or more likely to be present on admission or with no or very low risk

2d.2 Citations for Evidence:
Updated citations will be presented in the May Steering Committee meeting

Measures of Patient Safety Based on Hospital Administrative Data -
The Patient Safety Indicators, August 2002
http://qualityindicators.ahrq.gov/downloads/technical/psi_technical_review.zip

2d.3 Data/sample *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2d.4 Analytic Method *(type analysis & rationale)*:
Expert panel and descriptive analyses stratified by exclusion categories

2d C☐ P☐ M☐ N☐ NA ☐

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
Measures of Patient Safety Based on Hospital Administrative Data -
The Patient Safety Indicators, August 2002
http://qualityindicators.ahrq.gov/downloads/technical/psi_technical_review.zip

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
Risk-adjustment models use a standard set of categories based on readily available classification systems for demographics, severity of illness and comorbidities. Within each category, covariates are initially selected based on a minimum of 30 cases in the outcome of interest. Then a stepwise regression process on a development sample is used to select a parsimonious set of covariates where p<.05. Model is then tested on a validation sample

2e.3 Testing Results *(risk model performance metrics)*:
c 0.738

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

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)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
Posterior probability distribution parameterized using the Gamma distribution

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)*:

5th	25th	Median	75th	95th
0.079961	0.104593	0.124460	0.146701	0.183056

2f
C☐
P☐
M☐
N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample *(description of data/sample and size)*: Not applicable

2g.2 Analytic Method *(type of analysis & rationale)*:
Not applicable

2g.3 Testing Results *(e.g., correlation statistics, comparison of rankings)*:
Not applicable

2g
C☐
P☐
M☐
N☐
NA ☐

2h. Disparities in Care

2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)*: [1]
Although we did find overall disparities in care, we found that indicators for blacks, Hispanics, and Asians were not statistically worse than corresponding quality indicators for whites in the same hospital. Only a few hospitals provide lower quality of care to minorities than to whites.

[1] Darrell J. Gaskin, Christine S. Spencer, Patrick Richard, Gerard F. Anderson, Neil R. Powe and Thomas A. LaVeist. Do Hospitals Provide Lower-Quality Care To Minorities Than To Whites? Health Affairs, 27, no. 2 (2008): 518-527 doi: 10.1377/hlthaff.27.2.518

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities,

2h
C☐
P☐
M☐
N☐
NA ☐

provide follow-up plans: Not applicable	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, met? **Rationale:**	**2** **C** ☐ **P** ☐ **M** ☐ **N** ☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Eval Rating**
3a. Meaningful, Understandable, and Useful Information **3a.1 Current Use:** In use **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** Arizona (NY QIO) Why Not the Best? http://www.http://whynotthebest.org/ Kentucky (Norton Healthcare, a hospital system) Norton Healthcare Quality Report http://www.nortonhealthcare.com/body.cfm?id=157 Kentucky (state hospital association) Kentucky Hospital Association Quality Data http://info.kyha.com/QualityData/IQISite/ Maine (state) Maine Health Data Organization http://gateway.maine.gov/mhdo2008Monahrq/home.html Minnesota (Minnesota Community Measurement) Minnesota Health Scores www.mnhealthscores.org Missouri (health care coalition) St Louis Area Business Health Coalition http://www.stlbhc.org/c_healthcare_4_3026553713.pdf Nevada (state hospital association) Nevada Hospital Association Hospital Performance http://www.nvhospitalquality.net/ New Hampshire (NY QIO) New York State Health Accountability Foundation http://nyshaf.org/juice/IPROSpikeChart.html New York (health care coalition)	**3a** **C** ☐ **P** ☐ **M** ☐ **N** ☐

New York State Hospital Report Card
http://www.myhealthfinder.com/

Rhode Island (NY QIO)
Why Not the Best?
http://www.http://whynotthebest.org/

Washington (health care coalition)
Washington State Hospital Report Card
http://www.myhealthfinder.com/wa09/index.php

The measure is also reported on HCUPnet:
http://hcupnet.ahrq.gov/HCUPnet.jsp?Id=EB57801381F71C41&Form=MAINSEL&JS=Y&Action=%3E%3ENext%3E%3E&_MAINSEL=AHRQ%20Quality%20Indicators

This measure is used in the MONAHRQ system that is provided for public reporting and quality improvement throughout the United States: http://monahrq.ahrq.gov/

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***):**
University Healthcare Consortium - An alliance of 103 academic medical centers and 219 of their affiliated hospitals. Reporting the AHRQ QIs to their member hospitals. (see www.uhc.edu. Note: measure results reported to hospitals; not reported on site).

Dallas Fort Worth Hospital Council – Reporting on measure results to over 70 hospitals in Texas (see www.dfwhc.ord. Note: measure results reported to hospitals; not reported on site).
Norton Healthcare - a multi-hospital system in Kentucky (see http://www.nortonhealthcare.com/about/Our_Performance/index.aspx)

Ministry Health Care - a multi-hospital system in Wisconsin (see http://ministryhealth.org/display/router.aspx. Note: measure results reported to hospitals; not reported on site).

Minnesota Hospital Association
http://www.mnhospitals.org/ Note: measure used in quality improvement. Not reported publicly by the association)

Premier - Premier´s "Quality Advisor" tool provides performance reports to approximately 650 hospitals for their use in monitoring and improving quality. Hospitals receive facility specific reports on this measure in Quality Advisor.

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)***:** AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

3a.5 **Methods** *(e.g., focus group, survey, QI project)***:**
A research team from the School of Public Affairs, Baruch College, under contracts with the Department of Public Health, Weill Medical College and Battelle, Inc., has developed a pair of Hospital Quality Model Reports at the request of the Agency for Healthcare Research & Quality (AHRQ). These reports are designed specifically to report comparative information on hospital performance based on the AHRQ Quality Indicators (QIs). The work was done in close collaboration with AHRQ staff and the AHRQ Quality Indicators team.
The Model Reports (discussed immediately above) are based on:
• Extensive search and analysis of the literature on hospital quality measurement and reporting, as well as public reporting on health care quality more broadly;
• Interviews with quality measurement and reporting experts, purchasers, staff of purchasing coalitions, and executives of integrated health care delivery systems who are responsible for quality in their facilities;

• Two focus groups with chief medical officers of hospitals and/or systems and two focus groups with quality managers from a broad mix of hospitals; • Four focus groups with members of the public who had recently experienced a hospital admission; and • Four rounds of cognitive interviews (a total of 62 interviews) to test draft versions of the two Model Reports with members of the public with recent hospital experience, basic computer literacy but widely varying levels of education. 3a.6 **Results** *(qualitative and/or quantitative results and conclusions)*: Given the above review of the literature and original research that was conducted, a Model report was the result that could help sponsors use the best evidence on public reports so they are most likely to have the desired effects on quality.	
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:** 5.1 **If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:**	**3c** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**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*) Yes 4b.2 **If not, specify the near-term path to achieve electronic capture by most providers.**	**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.** Coding professionals follow detail guidelines, are subject to training and credentialing requirements, peer review and audit.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** None **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** Administrative data are collected as part of the routine operations. Some staff time is required to download and execute the software from the AHRQ webs site, which is available at no cost. **4e.3 Evidence for costs:** Administrative data are collected as part of the routine operations. Some staff time is required to download and execute the software from the AHRQ webs site, which is available at no cost. **4e.4 Business case documentation:** Administrative data are collected as part of the routine operations. Some staff time is required to download and execute the software from the AHRQ webs site, which is available at no cost.	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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** Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850	

Co.2 **Point of Contact**
John, Bott, MSSW, MBA, John.Bott@AHRQ.hhs.gov, 301-427-1317-

Measure Developer If different from Measure Steward
Co.3 **Organization**
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Co.4 **Point of Contact**
John, Bott, MSSW, MBA, John.Bott@AHRQ.hhs.gov, 301-427-1317-

Co.5 **Submitter If different from Measure Steward POC**
John, Bott, MSSW, MBA, John.Bott@AHRQ.hhs.gov, 301-427-1317-, Agency for Healthcare Research and Quality

Co.6 **Additional organizations that sponsored/participated in measure development**
UC Davis,
Stanford University,
Battelle Memorial Institute'

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.**
None

Ad.2 **If adapted, provide name of original measure:** None
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:** 2003
Ad.7 **Month and Year of most recent revision:** 10, 2010
Ad.8 **What is your frequency for review/update of this measure?** Annual
Ad.9 **When is the next scheduled review/update for this measure?** 05, 2011

Ad.10 **Copyright statement/disclaimers:** The AHRQ QI software is publicly available; no copyright disclaimers

Ad.11 -13 **Additional Information web page URL or attachment:**

Date of Submission (*MM/DD/YY*)**:** 04/05/2011

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 0352 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Failure to Rescue In-Hospital Mortality (risk adjusted)
De.2 **Brief description of measure:** Percentage of patients who died with a complications in the hospital.
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** Failure to Rescue 30-day Mortality (risk adjusted)
De.4 **National Priority Partners Priority Area:** Safety De.5 **IOM Quality Domain**: Patient-centered 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 measure A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission 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	**B** Y☐

every 3 years. Yes, information provided in contact section	N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	**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, Frequently performed procedure, Severity of illness 1a.2 1a.3 **Summary of Evidence of High Impact:** Failure to Rescue measure has a very high impact because it is applicable to the majority of surgical procedures performed at acute care hospitals. Failure to Rescue affects large number of patients and applies to frequently performed procedures. Failure to Rescue, predicts death after an adverse event which accounts for severity of illness to properly adjust the death rate. The measure is less sensitive to errors in severity adjustment (because all patients in the analysis have complications) and more dependent on hospital characteristics relative to patient characteristics than the mortality rate, while having equivalent reliability. FTR has intuitive appeal as a quality marker, attempting to measure a hospital's ability to manage complications, while being less likely to confuse worse severity of illness with worse quality of care. 1a.4 **Citations for Evidence of High Impact:** 1. Silber JH, Williams SV, Krakauer H, et al. Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure-to-rescue. Med Care. 1992;30:615-629. 2. Silber JH, Romano PS, Rosen AK, et al. Failure-to-rescue: Comparing definitions to measure quality of care. Med Care. 2007;45:918-925. 3. Silber JH, Rosenbaum PR, Schwartz JS, et al. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. JAMA. 1995;274:317-323. 4. Silber JH, Rosenbaum PR, Williams SV, et al. The relationship between choice of outcome measure and	**1a** C☐ P☐ M☐ N☐

hospital rank in general surgical procedures: Implications for quality assessment. Int J Qual Health Care. 1997;9:193-200.
5. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist direction and patient outcomes. Anesthesiology. 2000;93:152-163.
6. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist board certification and patient outcomes. Anesthesiology. 2002;96:1044-1052.
7. Aiken LH, Clarke SP, Sloane DM, et al. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA. 2002;288:1987-1993.
8. Aiken LH, Clarke SP, Cheung RB, et al. Educational levels of hospital nurses and surgical patient mortality. JAMA. 2003;290:1617-1623.
9. Silber JH, Rosenbaum PR, Ross RN. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? J Am Stat Assoc. 1995;90:7-18.
10. Silber JH, Rosenbaum PR, Romano PS, Rosen AK, Wang Y, Teng Y, Halenar MJ, Even-Shoshan O, Volpp KG. Hospital Teaching Intensity, Patient Race,
and Surgical Outcomes. Arch Surg. 2009;144:113-120.
11. Friese CR, Earle CC, Silber JH, Aiken LH. Hospital characteristics, clinical severity, and outcomes for surgical oncology patients. Surgery 2010; 147:602-609.
12. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in Hospital Mortality Associated with Inpatient Surgery. N Engl J Med 2009; 361:1368-75.
13. Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational Levels of Hospital Nurses and Surgical Patient Mortality.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: The use of Failure to rescue, predicting death after an adverse occurrence, hospitals would be able to improve their quality of care. Hospitals and health care providers benefit from knowing not only their institution´s mortality rate, but also their institution´s ability to rescue patients after an adverse occurrence. Using failure to rescue measure is especially important if hospital resources needed for prevention were different from those needed for rescue. From a research and policy perspective knowing the failure to rescue rate in addition to the mortality rate will improve our understanding of mortality statistics.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
In Aiken et al. shows if the proportion of BSN nurses in all hospitals was 60% rather than 20% 14.2 fewer deaths per 1000 patients with complications (failure to rescue) would be expected. Moreover failure to rescue rates would be decidedly lower if both the workloads of nurses were lighter and the workforce were composed of higher percent-ages of BSN-prepared nurses. (see table 4 in Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational Levels of Hospital Nurses and Surgical Patient Mortality)

1b.3 Citations for data on performance gap:
Cross-sectional analyses of outcomes data for 232,342 general, orthopedic, and vascular surgery patinets discharged from 168 non-federal adult general Pennsylvania hospitals between April 1, 1998, and November 30, 1999, linked to administrative and survey data providing information on educational composition, staffing, and other chracteristics.

1b.4 Summary of Data on disparities by population group:
In Silber JH et al Hospital Teaching Intensity, Patient Race,
and Surgical Outcomes. Arch Surg. 2009, shows failure-to rescue rates were consistently lower in hospitals with higher resident-to-bed ratios. Hospitals of high teaching intensity (resident-to-bed ratio=0.6) compared with nonteaching hospitals (resident-to-bed ratio=0) were associated with 14%(95% CI, 12%-15%) lower odds of failure to rescue for combined surgery, with similar finding for subgroup analysis. (see table 3 in paper)

1b.5 Citations for data on Disparities:
For information reported in 1b4 the data sample was 2,021,214 patients with medicare claims on general, orthopedic, and vascular surgery admissions in the United States for 2000-2005.

1b
C☐ P☐ M☐ N☐

1c. Outcome or Evidence to Support Measure Focus

1c
C☐

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*)**:** Failure-to-rescue is defined as the probability of death following a complication. The measure will help improve both the management of the hospital and our understanding of hospital mortality rates.

1c.2-3. Type of Evidence: 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*)**:**
Failure to rescue is influenced by hospital characteristics. Rates differ based on different hospital characteristics such as number of hospital beds, anesthesiologists who are board certified, surgeons who are board certified, presence of house staff, and high technology hospitals, etc. Failure to rescue is an indicator of hospital quality of care. Patients in the age range of 18-90 are analyzed because patients under the age of 18 are considered a pediatric population and have a different set of complications. We use 90 years as a cut-point because of our concern regarding the increased use of do-not-resuscitate at higher ages [Wenger et al. Epidemiology of Do-Not Resuscitate Orders. Disparity by Age, Diagnosis, Gender, Race, and Functional Impairment. Arch Intern Med. 1995; 155(19):2056-62, Hakim et al. Factors Associated with Do-Not-Resuscitate Orders: Patients´, Preferences, Prognoses, and Physicians Judgments. Ann Intern Med.1996; 125:284-293.]. While we do adjust for admission severity when reporting FTR, and this includes age, we still thought it prudent to use an upper bound on age, since DNR status prior to the procedure is not well defined at hospitals [Tabak YP, Johannes RS, Silber JH, Kurtz SG, Gibber EM. Should do-not-resuscitate status be included as a mortality risk adjustor? The impact of DNR variations on performance reporting. Med Care 2005; 43:658-666]

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom*)**:**
Silber JH, Williams SV, Krakauer H, et al. Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure-to-rescue. Med Care. 1992;30:615-629. Silber JH, Rosenbaum PR, Schwartz JS, et al. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. JAMA. 1995;274:317-323 Silber JH, Romano PS, Rosen AK, et al. Failure-to-rescue: Comparing definitions to measure quality of care. Med Care. 2007;45:918-925

1c.6 Method for rating evidence: In Silber et al JAMA 1995, refers to the "power" of a measure as the ability of that measure to detect differences between hospitals or groups of hospitals, with respect to the outcome measure in question. Should the difference between two hospital failure rates achieve statistical significance, while the difference between those same hospitals´ death rates not achieve statistical significance, then we would consider the failure rate to be more powerful than the death rate. It can be shown that for equivalent adverse occurrence rates, the power to distinguish between two hospitals using the failure rate is always greater than or equal to the power using the death rate. Although somewhat counterintuitive, this result occurs because, although the failure rate and the death rate use the number of deaths as their numerators, the denominator of the failure rate is the number of patients with adverse occurrences, while the denominator of the death rate is the total number of patients. When adverse occurrence rates are not equal across hospitals, the power of the failure rate statistic may be greater than, equal to, or less than that of the death rate. When comparing two hospitals with failure rates F1 and F2 death rates Dl and D2 and adverse occurrence rates A1 and A2 it can be shown that whenever F1>= F2, Dl>= D2 and A1<=A2 then the power in distinguishing such hospitals using the failure rate is greater than or equal to the power when using the death rate. For situations where F1>=F2 and Dl < D2 the sufficient conditions for superior power using the failure rate instead of the death rate is given in the Appendix. Finally, these results are unchanged if one considers either hospital I or 2 in the above arguments to be a group of hospitals or the average of all hospitals (so that hospital 1 or 2 represents a very large sample size).
In summary, failure rate was a function of anesthesia board certification and the presence of surgical housestaff (hospital characteristics) but not a function of admission severity of illness score (patient characteristics). Since the death rate appears to be composed of two distinct rates, quality of care measurement may be improved if all three rates are reported instead of relying on the adjusted mortality rate alone. In so doing, we may better understand the reasons for variation in hospital mortality rates.

1c.7 Summary of Controversy/Contradictory Evidence: N/A

P☐
M☐
N☐

1c.8 Citations for Evidence (*other than guidelines*): 1. Silber JH, Williams SV, Krakauer H, et al. Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure-to-rescue. Med Care. 1992;30:615-629.
2. Silber JH, Romano PS, Rosen AK, et al. Failure-to-rescue: Comparing definitions to measure quality of care. Med Care. 2007;45:918-925.
3. Silber JH, Rosenbaum PR, Schwartz JS, et al. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. JAMA. 1995;274:317-323.
4. Silber JH, Rosenbaum PR, Williams SV, et al. The relationship between choice of outcome measure and hospital rank in general surgical procedures: Implications for quality assessment. Int J Qual Health Care. 1997;9:193-200.
5. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist direction and patient outcomes. Anesthesiology. 2000;93:152-163.
6. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist board certification and patient outcomes. Anesthesiology. 2002;96:1044-1052.
7. Aiken LH, Clarke SP, Sloane DM, et al. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA. 2002;288:1987-1993.
8. Aiken LH, Clarke SP, Cheung RB, et al. Educational levels of hospital nurses and surgical patient mortality. JAMA. 2003;290:1617-1623.
9. Silber JH, Rosenbaum PR, Ross RN. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? J Am Stat Assoc. 1995;90:7-18.
10. Silber JH, Rosenbaum PR, Romano PS, Rosen AK, Wang Y, Teng Y, Halenar MJ, Even-Shoshan O, Volpp KG. Hospital Teaching Intensity, Patient Race,
and Surgical Outcomes. Arch Surg. 2009;144:113-120.
11. Friese CR, Earle CC, Silber JH, Aiken LH. Hospital characteristics, clinical severity, and outcomes for surgical oncology patients. Surgery 2010; 147:602-609.
12. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in Hospital Mortality Associated with Inpatient Surgery. N Engl J Med 2009; 361:1368-75.
13. Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational Levels of Hospital Nurses and Surgical Patient Mortality.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number*):
N/A

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:
The motivation behind the development of FTR was based on 2 questions. The first was an empirical question—suppose hospitals were ranked by adjusted mortality and adjusted complication rates. Would these rankings be highly correlated? The answer is rather surprising—there is generally poor correlation or no correlation in most analyses. Second, suppose 2 hospitals had identical adjusted mortality rates but different adjusted complication rates. Would one prefer care at the hospital with the higher or lower complication rate? If one believes that complications are predominantly driven by patient characteristics, then one may decide to choose the hospital with the higher complication rate, as it achieved an equivalent mortality rate with a sicker population of patients. So there is an empirical question to ask—are adjusted complication rates more related to hospital or patient factors? This has been looked at in a number of ways—and the evidence to date suggests that complication measures are less sensitive to hospital characteristics, after adjusting for severity of illness, than mortality based measures. This is an underlying assumption of FTR theory—complications are undesirable outcome measures because they reflect underlying patient severity and diagnosis coding more than they reflect hospital care. Instead, a hospital's quality is

put to the test when a patient develops a complication, and whether a patient is salvaged after a complication will be a function of the care delivered by the hospital and its knowledge base, depth, and facilities. Thus, "good" hospitals will rescue patients by identifying complications quickly and treating them aggressively, resulting in lower FTR. Although many "failures," just like deaths, are often not preventable, we have argued that FTR may be a better measure for comparing hospital quality because of better severity adjustment properties, and because of its focus on hospital actions. By studying a population of patients who, by definition, have already developed a complication, the specifics of severity of illness adjustment becomes less important in failure rate analyses, because all patients have experienced complications and thus are more uniformly ill.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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***):** Patients who died with a complication plus patients who died without documented complications. Death is defined as death in the hospital. All patients in an FTR analysis have developed a complication (by definition). Complicated patient has at least one of the complications defined in Appendix B(see website http://www.research.chop.edu/programs/cor/outcomes.php). Complications are defined using the secondary ICD9 diagnosis and procedure codes and the DRG code of the current admission. Comorbidities are defined in Appendix C (see website http://www.research.chop.edu/programs/cor/outcomes.php) using secondary ICD9 diagnosis codes of the current admission and primary or secondary ICD9 diagnosis codes of previous admission within 90 days of the admission date of the current admission. *When physician part B is available, the definition of complications and comorbidities are augmented to include CPT codes. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** Index Hospitalization (Admission to Discharge) **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Patients who died with complication and patients who died without documented complications. Death is defined as death in the hospital.	**2a-specs** **C☐** **P☐** **M☐** **N☐**
2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):** General Surgery, Orthopedic and Vascular patients in specific DRGs with complications plus patients who died in the hospital without complications.	

Inclusions: adult patients admitted for one of the procedures in the General Surgery, Orthopedic or Vascular DRGs (see appendix A http://www.research.chop.edu/programs/cor/outcomes.php)

2a.5 **Target population gender:** Female, Male
2a.6 **Target population age range:** 18-90

2a.7 **Denominator Time Window (***The time period in which cases are eligible for inclusion in the denominator***):**
Index Hospitalization (Admission to Discharge)

2a.8 **Denominator Details (***All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Adult patients admitted for one of the procedures in the General Surgery, Orthopedic or Vascular DRGs (see Appendix A http://www.research.chop.edu/programs/cor/outcomes.php)who developed an in hospital complication and those who died without a complication.

2a.9 **Denominator Exclusions (***Brief text description of exclusions from the target population***):** Patients over age 90, under age 18.

2a.10 **Denominator Exclusion Details (***All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
N/A

2a.11 **Stratification Details/Variables (***All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
Complicated patient has at least one of the complications defined in Appendix B (http://www.research.chop.edu/programs/cor/outcomes.php) Complications are defined using the secondary ICD9 diagnosis and procedure codes and the DRG code of the current admission. When Physician Part B file is available, the definition of complications and comorbidities are augmented to include CPT codes.

2a.12-13 **Risk Adjustment Type:** Risk-adjustment devised specifically for this measure/condition

2a.14 **Risk Adjustment Methodology/Variables (***List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
Risk Adjustment: Model was developed using logistic regression analysis.

Associated data elements: age in years, sex, race, comorbidities, DRGs (combined with and without complications) and procedure codes within DRGs, transfer status.

Failure to rescue is adjusted using a logistic regression model where y is a failure and the total N is composed of patients who develop a complication and patients who died without a complication.

According to developer: The model adjustment variables can vary. We have found that FTR results are fairly stable, even with little adjustment, since all patients in an FTR analysis have developed a complication (by definition), they are a more homogeneous group of patients than the entire population. Hence severity adjustment plays somewhat less of a role than in other outcome measures.

2a.15-17 **Detailed risk model available Web page URL or attachment:** URL
http://www.research.chop.edu/programs/cor/outcomes.php

2a.18-19 **Type of Score:** Rate/proportion
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***):**
Refer to website (http://www.research.chop.edu/programs/cor/outcomes.php)

2a.22 **Describe the method for discriminating performance** *(e.g., significance testing)***:**
T-test for comparing rates

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): Measure not based on sample, all surgical patients between the ages of 18 and 90 admitted to an acute care hospital.	
2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)***</br>Electronic administrative data/claims **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.***):**</br>Linked patients hospitalizations claims records, augmented with Outpatient and Part B records; can also use unlinked data if linked files are not available to identify comorbidities and develop definitions of severity and other risk measure. **2a.26-28 Data source/data collection instrument reference web page URL or attachment:** URL</br>http://www.resdac.org/ **2a.29-31 Data dictionary/code table web page URL or attachment:** URL</br>http://www.research.chop.edu/programs/cor/outcomes.php **2a.32-35 Level of Measurement/Analysis** (*Check the level(s) for which the measure is specified and tested)*</br>Facility/Agency, Health Plan, Integrated delivery system, Population: national, Population: regional/network, Population: states, Population: counties or cities **2a.36-37 Care Settings (***Check the setting(s) for which the measure is specified and tested)*</br>Hospital **2a.38-41 Clinical Services** (*Healthcare services being measured, check all that apply*)</br>Clinicians: Physicians (MD/DO)	
TESTING/ANALYSIS	
2b. Reliability testing **2b.1 Data/sample** *(description of data/sample and size)***:** Medicare inpatient claims for general surgical admissions for the period July 1, 1999 to June 30, 2000. There were a total of 1467 hospitals and 403,679 patients. We included patients between 65 and 90 years of age. **2b.2 Analytic Method** *(type of reliability & rationale, method for testing)***:**</br>We defined reliability as described by Lord and Novick using split sample methodology. (Lord FM, Novick MR. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley; 1968) **2b.3 Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**</br>Using Spearman-Brown half split half sample reliability had a correlation of 0.31 and the upper bound on validity was 0.56.	**2b** C☐ P☐ M☐ N☐
2c. Validity testing **2c.1 Data/sample** *(description of data/sample and size)***:** Medicare inpatient claims for general surgical admissions for the period July 1, 1999 to June 30, 2000. There were a total of 1467 hospitals and 403,679 patients. We included patients between 65 and 90 years of age. **2c.2 Analytic Method** *(type of validity & rationale, method for testing)***:**</br>a) Rank correlation between various hospital outcomes (Death, Failure to Rescue, Complications, other measures of Failure to Rescue, Failure to Rescue Complement measures) b) Marginal and partial coefficients in logit models using detailed patient characteristics and hospital characteristics shown to be associated with better outcomes in previous studies.2, 7 The marginal results use one hospital characteristic at a time along with all patient characteristics. “Partial” regression results,	**2c** C☐ P☐ M☐ N☐

using all hospital and patient variables simultaneously have the disadvantage that correlation between hospital characteristics can cause difficulty in interpreting the effects of individual hospital variables. Hospital characteristics associated with better outcomes (1) teaching hospital status (member of the American Council of Teaching Hospitals); (2) high technology status (does the hospital perform open heart surgery or perform organ transplantation); (3) hospital size greater than 200 beds; (4) bed-to-nurse ratio (where nurses are the sum of RN plus LPN FTE positions); and (5) nursing skill mix (the ratio of RN/(RN+LPN)).2-8

c) The relative contribution of patient-to-hospital characteristics that predicted each outcome of interest, as provided by the omega statistic.2, 9 The omega statistic computes a ratio of the squared sum of the log odds for model patent variables divided by a similar quantity calculated for the model hospital variables. All else being equal, outcome measures that have lower omega ratios may be more desirable quality indicators, since the lower the omega, the greater the hospital's impact on outcome relative to the patient's impact. This is especially important if modeling patient severity is difficult (as with claims data) so that the lower the omega suggests the higher relative influence of hospital characteristics as compared to patient.

2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
FTR itself is highly correlated with death, with a Kendall's tau equal to 0.85, representing a probability of concordance equal to 0.92.

2d. Exclusions Justified

2d.1 **Summary of Evidence supporting exclusion(s):**
Patients younger than 18 are excluded because they are considered in the pediatric population and have a different set of complications. We use 90 years as a cut-point because of our concern regarding the increased use of do-not-resuscitate at higher ages [Wenger et al. Epidemiology of Do-Not Resuscitate Orders. Disparity by Age, Diagnosis, Gender, Race, and Functional Impairment. Arch Intern Med. 1995; 155(19):2056-62, Hakim et al. Factors Associated with Do-Not-Resuscitate Orders: Patients´, Preferences, Prognoses, and Physicians Judgments. Ann Intern Med.1996; 125:284-293.]. While we do adjust for admission severity when reporting FTR, and this includes age, we still thought it prudent to use an upper bound on age, since DNR status prior to the procedure is not well defined at hospitals [Tabak YP, Johannes RS, Silber JH, Kurtz SG, Gibber EM. Should do-not-resuscitate status be included as a mortality risk adjustor? The impact of DNR variations on performance reporting. Med Care 2005; 43:658-666]

2d.2 **Citations for Evidence:**
1. Wenger NS, Pearson ML, Desmond KA, Harrison ER, Rubenstein LV, Rogers WH, Kahn KL. Epidemiology of Do-Not Resuscitate Orders. Disparity by Age, Diagnosis, Gender, Race, and Functional Impairment. Arch Intern Med. 1995; 155(19):2056-62
2. Hakim RB, Teno JM, Harrell Jr. FE, Knaus WA, Wenger N, Phillips RS, Layde P, Califf R, Connors Jr. AF, Lynn J. Factors Associated with Do-Not-Resuscitate Orders: Patients´, Preferences, Prognoses, and Physicians Judgments. Ann Intern Med. 1996; 125:284-293.
3. Tabak YP, Johannes RS, Silber JH, Kurtz SG, Gibber EM. Should do-not-resuscitate status be included as a mortality risk adjustor? The impact of DNR variations on performance reporting. Med Care 2005; 43:658-666

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

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

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

2d
C☐
P☐
M☐
N☐
NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 **Data/sample** *(description of data/sample and size)*: Two different data samples were used to analyze risk adjustment. 1.) 5,972 Medicare patients undergoing elective cholecystectomy or transurethral prostatectomy (Silber et al. Hospital and Patient Characteristics Associated with Death After Surgery A study of Adverse Occrueenece and Failure to Rescue Med Care 1992).

2e
C☐
P☐
M☐
N☐
NA☐

2.) 2,021,214 patients with medicare claims on general, orthopedic, and vascular surgery admissions in the United States for 2000-2005. (Silber et al. Hospital Teaching Intensity, Patient Race, and Surgical Outcomes Arch Surg 2009)

2e.2 **Analytic Method** *(type of risk adjustment, analysis, & rationale)*:
Risk Adjustment: Model was developed using logistic regression analysis, where y is a failure and the total N is composed of patients who develop a complication and patients who died without a complication.

Associated data elements: age in years, sex, race, comorbidities, DRGs (combined with and without complications) and procedure codes within DRGs, transfer status.

The model adjustment variables can vary. We have found that FTR results are fairly stable, even with little adjustment since all patients in an FTR analysis have developed a complication, (by definition), they are a more homogenous group of patients than the entire population. Hence severity adjustment plays somewhat less of a role than in other outcome measures.

2e.3 **Testing Results** *(risk model performance metrics)*:
In earlier work we did report calibration as tested with the Hosmer-Lemeshow statistic, however the research community found that this calibration test fails its asymptotics, it overcalls with large sample size, we do not recommend its use. It is well known that the Hosmer-Lemeshow test is misleading with large data sets, and therefore we have not thought this to be a valid approach. C-statistic ranges 0.70 for the FTR 30 day risk adjustment model (Silber et. al Med Care 1992) to 0.792 (Silber et al. Arch Surg 2009). However c-statistics are also misleading when comparing across populations. Since FTR is a subset of the mortality and complication data set, one cannot compare, in a meaningful way, the c-statistic from FTR to that of mortality or complication.

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

2f. Identification of Meaningful Differences in Performance

2f.1 **Data/sample from Testing or Current Use** *(description of data/sample and size)*: Medicare inpatient claims for general surgical admissions for the period July 1, 1999 to June 30, 2000. There were a total of 1467 hospitals and 403,679 patients. We included patients between 65 and 90 years of age.

2f.2 **Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*:
T-test for comparing rates.

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)*:
75% Q3= 0.12, 50% Median=0.09, 25% Q1=0.06, Mean= 0.09, Std Deviation= 0.05

2f
C☐
P☐
M☐
N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 **Data/sample** *(description of data/sample and size)*: FTR was developed using standardized hospital discharge records, which are widely collected by states agencies and which hospitals are mandated to report to CMS. One of the big advantages of adopting FTR is that the data on which it is based is uniformely reported, checked for errors and edited. This is administrative data available for the entire population over 65 and for all patients admitted to acute care hospitals.

2g.2 **Analytic Method** *(type of analysis & rationale)*:
N/A

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

2g
C☐
P☐
M☐
N☐
NA☐

2h. Disparities in Care

2h.1 **If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*:

2h
C☐
P☐

Disparities in care are shown in Silber et al Arch Surg 2009 where the results show white patients displayed a reduction in failure-to-rescue rates in the teaching intensive hospitals vs non-teaching hospitals (OR, 0.94; 95% CI, 0.92-0.97), black patients displayed an increased failure-to-rescue rate (OR, 1.06; 95% CI, 1.00-1.12)(Results are based on 30 day mortality FTR however in-hospital showed similar results) 2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** N/A	**M☐ N☐ NA☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, met? **Rationale:**	***2*** **C☐ P☐ M☐ N☐**
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Eval Rating**
3a. Meaningful, Understandable, and Useful Information 3a.1 **Current Use:** In use 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** This measure has not yet been used in a public reporting initiative. This measure could be reported on a wide scale, the same way that mortality rates are reported. 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***):** Currently used to assess the impact of the change in the resident work hours regulations on patient outcomes in a recently NHLBI funded study (1R01HL094593-01) entitled "Work Hour Regulation for Physician Trainees: Educational and Clinical Outcomes" **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)***:** In Ghaferi et al "Variation in Hospital Mortality Associated with Inpatient Surgery" studied 84,730 patients who had undergone inpatient general and vascular surgery from 2005-2007 using data from the American College of Surgeons National Surgical Quality Improvement Program. 3a.5 **Methods** *(e.g., focus group, survey, QI project)***:** Ranked ranked hospitals according to their risk-adjusted overall rate of death and divided them into five groups. For hospitals in each overall mortality quintile, we then assessed the incidence of overall and major complications and the rate of death among patients with major complications 3a.6 **Results** *(qualitative and/or quantitative results and conclusions)***:** Rates of death varied widely across hospital quintiles, from 3.5% in very-low-mortality hospitals to 6.9% in very-high-mortality hospitals. Hospitals with either very high mortality or very low mortality had similar rates of overall complications (24.6% and 26.9%, respectively) and of major complications (18.2% and 16.2%, respectively). Rates of individual complications did not vary significantly across hospital mortality quintiles. In contrast, mortality in patients with major complications was almost twice as high in hospitals with very high overall mortality as in those with very low overall mortality (21.4% vs. 12.5%, P<0.001). Differences in rates of death among patients with major complications were also the primary determinant of variation in overall mortality with individual operations. In addition to efforts aimed at avoiding complications in the first place, reducing mortality associated with inpatient surgery will require greater attention to the timely	**3a** **C☐ P☐ M☐ N☐**

recognition and management of complications once they occur.	
3b/3c. Relation to other NQF-endorsed measures **3b.1 NQF # and Title of similar or related measures:** 0200 Death among surgical inpatients with treatable serious complications (failure to rescue)	
(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:** **5.1 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:** Needleman et al adapted the FTR measure to "nurse sensitive complications" by selecting a limited number of complications for the FTR measure. This change in definition, which we will call FTR-N, was developed to better focus on nursing quality of care. Because only deaths after nursing sensitive complications are studied, a large number of deaths are not used in the analysis. Subsequently, AHRQ again adapted the FTR-N definition to reflect quality from a "patient safety" perspective (ie, the identification of deaths that were especially likely to be preventable). Expert panels guided both of these adaptations through consensus development panels. The National Quality Forum, through its own process of selecting National Voluntary onsensus Standards for Nursing-Sensitive Care, endorsed Needleman et al's adaptation and assigned it to AHRQ for updating and support.FTR-N includes only 6 complications (pneumonia, shock, gastrointestinal bleeding, cardiac arrest, sepsis, and deep venous thrombosis) in its denominator definition, and it excludes deaths in patients without these complications. FTR-A adds renal failure to the FTR-N list of eligible complications, and modestly alters the definition of several others Table 1C and 1D display the impact of restricting the denominator of FTR to more limited sets of complications, as in the FTR-N and FTR-A definitions, respectively. Note first that the number of patients defined as having a complication fell from 189,031 (46.8%) in Table 1A to 43,500 (10.8%) in Table 1C and 39,101 (9.7%) in Table 1D. However, this smaller complication rate comes at an important cost—of all deaths, the proportion coded as having a complication (the precedence rate) fell from 95% in Table 1A to only 51% in Table 1C, and 58.5% in Table 1D. (Refer tp Silber et al. Med Care 2007)	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-	**4a** C☐ P☐ M☐

9 codes on claims, chart abstraction for quality measure or registry)	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*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	**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.** FTR is given to minimal susceptibility to inaccuracies or errors since it uses data collected uniformly across all hospitals and providers. The data is carefully checked by CMS before it is being released to researchers. However there may be unobserved differences among patients due to the lack of more detailed clinical information available only through chart abstraction.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** We have developed FTR measures based on restricted information, available only from the inpatient files. When possible, such as in the Medicare population, we improve the risk adjustment by using more patient level information available in the outpatient or Carrier file **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** CMS data is made available to researchers through ResDac, and its cost depends on the number of records requested, the number of years, and the type of file (inpatient, outpatient, or carrier) **4e.3 Evidence for costs:** N/A **4e.4 Business case documentation:** N/A	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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
The Children´s Hospital of Philadelphia, 3535 Market Street, Suite 1029, Philadelphia, Pennsylvania, 19104

Co.2 Point of Contact
Jeffrey H., Silber, MD, PhD, silber@email.chop.edu, 215-590-2540-

Measure Developer If different from Measure Steward
Co.3 Organization
The Children´s Hospital of Philadelphia, 3535 Market Street, Suite 1029, Philadelphia, Pennsylvania, 19104

Co.4 Point of Contact
Fabienne, Kyle, BA, kylef@email.chop.edu, 215-590-2484-

Co.5 Submitter If different from Measure Steward POC
Fabienne, Kyle, BA, kylef@email.chop.edu, 215-590-2484-, The Children´s Hospital of Philadelphia

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

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.
A group of clinicians and coding experts from the University of Pennsylvania reviewed the updated ICD, CPT, and DRG codes and updated the measure to reflect current coding.

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: URL
http://www.research.chop.edu/programs/cor/outcomes.php

Date of Submission (*MM/DD/YY*)**:** 03/29/2011

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 0353 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Failure to Rescue 30-Day Mortality (risk adjusted)
De.2 **Brief description of measure:** Percentage of patients who died with a complication within 30 days from admission.
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** Failure to Rescue In-Hospital Mortality (risk adjusted)
De.4 **National Priority Partners Priority Area:** Safety De.5 **IOM Quality Domain:** 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 measure A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission 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	**B**

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	Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	**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, Frequently performed procedure, Severity of illness 1a.2 1a.3 **Summary of Evidence of High Impact:** Failure to Rescue measure has a very high impact because it is applicable to the majority of surgical procedures performed at acute care hospitals. Failure to Rescue affects large number of patients and applies to frequently performed procedures. Failure to Rescue, predicts death after an adverse event which accounts for severity of illness to properly adjust the death rate. The measure is less sensitive to errors in severity adjustment (because all patients in the analysis have complications) and more dependent on hospital characteristics relative to patient characteristics than the mortality rate, while having equivalent reliability. FTR has intuitive appeal as a quality marker, attempting to measure a hospital's ability to manage complications, while being less likely to confuse worse severity of illness with worse quality of care. 1a.4 **Citations for Evidence of High Impact:** 1. Silber JH, Williams SV, Krakauer H, et al. Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure-to-rescue. Med Care. 1992;30:615-629. 2. Silber JH, Romano PS, Rosen AK, et al. Failure-to-rescue: Comparing definitions to measure quality of care. Med Care. 2007;45:918-925. 3. Silber JH, Rosenbaum PR, Schwartz JS, et al. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. JAMA. 1995;274:317-323.	**1a** C☐ P☐ M☐ N☐

4. Silber JH, Rosenbaum PR, Williams SV, et al. The relationship between choice of outcome measure and hospital rank in general surgical procedures: Implications for quality assessment. Int J Qual Health Care. 1997;9:193-200.
5. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist direction and patient outcomes. Anesthesiology. 2000;93:152-163.
6. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist board certification and patient outcomes. Anesthesiology. 2002;96:1044-1052.
7. Aiken LH, Clarke SP, Sloane DM, et al. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA. 2002;288:1987-1993.
8. Aiken LH, Clarke SP, Cheung RB, et al. Educational levels of hospital nurses and surgical patient mortality. JAMA. 2003;290:1617-1623.
9. Silber JH, Rosenbaum PR, Ross RN. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? J Am Stat Assoc. 1995;90:7-18.
10. Silber JH, Rosenbaum PR, Romano PS, Rosen AK, Wang Y, Teng Y, Halenar MJ, Even-Shoshan O, Volpp KG. Hospital Teaching Intensity, Patient Race,
and Surgical Outcomes. Arch Surg. 2009;144:113-120.
11. Friese CR, Earle CC, Silber JH, Aiken LH. Hospital characteristics, clinical severity, and outcomes for surgical oncology patients. Surgery 2010; 147:602-609.
12. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in Hospital Mortality Associated with Inpatient Surgery. N Engl J Med 2009; 361:1368-75.
13. Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational Levels of Hospital Nurses and Surgical Patient Mortality.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: The use of Failure to rescue, predicting death after an adverse occurrence, hospitals would be able to improve their quality of care. Hospitals and health care providers benefit from knowing not only their institution´s mortality rate, but also their institution´s ability to rescue patients after an adverse occurrence. Using failure to rescue measure is especially important if hospital resources needed for prevention were different from those needed for rescue. From a research and policy perspective knowing the failure to rescue rate in addition to the mortality rate will improve our understanding of mortality statistics.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
In Aiken et al. shows if the proportion of BSN nurses in all hospitals was 60% rather than 20% 14.2 fewer deaths per 1000 patients with complications (failure to rescue) would be expected. Moreover failure to rescue rates would be decidedly lower if both the workloads of nurses were lighter and the workforce were composed of higher percent-ages of BSN-prepared nurses. (see table 4 in Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational Levels of Hospital Nurses and Surgical Patient Mortality)

1b.3 Citations for data on performance gap:
In Silber JH et al Hospital Teaching Intensity, Patient Race,
Cross-sectional analyses of outcomes data for 232,342 general, orthopedic, and vascular surgery patients discharged from 168 non-federal adult general Pennsylvania hospitals between April 1, 1998, and November 30, 1999, linked to administrative and survey data providing information on educational composition, staffing, and other characteristics.

1b.4 Summary of Data on disparities by population group:
In Silber JH et al Hospital Teaching Intensity, Patient Race,
and Surgical Outcomes. Arch Surg. 2009, shows failure-to rescue rates were consistently lower in hospitals with higher resident-to-bed ratios. Hospitals of high teaching intensity (resident-to-bed ratio=0.6) compared with non-teaching hospitals (resident-to-bed ratio=0) were associated with 14%(95% CI, 12%-15%) lower odds of failure to rescue for combined surgery, with similar finding for subgroup analysis. (see table 3 in paper)

1b.5 Citations for data on Disparities:
For information reported in 1b4 the data sample was 2,021,214 patients with medicare claims on general, orthopedic, and vascular surgery admissions in the United States for 2000-2005.

1b
C ☐
P ☐
M ☐
N ☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Failure-to-rescue is defined as the probability of death following a complication. The measure will help improve both the management of the hospital and our understanding of hospital mortality rates.

1c.2-3. Type of Evidence: 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*)**:**
Failure to rescue is influenced by hospital characteristics. Rates differ based on different hospital characteristics such as number of hospital beds, anesthesiologists who are board certified, surgeons who are board certified, presence of house staff, and high technology hospitals, etc. Failure to rescue is an indicator of hospital quality of care. Patients in the age range of 18-90 are analyzed because patients under the age of 18 are considered a pediatric population and have a different set of complications. We use 90 years as a cut-point because of our concern regarding the increased use of do-not-resuscitate at higher ages [Wenger et al. Epidemiology of Do-Not Resuscitate Orders. Disparity by Age, Diagnosis, Gender, Race, and Functional Impairment. Arch Intern Med. 1995; 155(19):2056-62, Hakim et al. Factors Associated with Do-Not-Resuscitate Orders: Patients´, Preferences, Prognoses, and Physicians Judgments. Ann Intern Med.1996; 125:284-293.]. While we do adjust for admission severity when reporting FTR, and this includes age, we still thought it prudent to use an upper bound on age, since DNR status prior to the procedure is not well defined at hospitals [Tabak YP, Johannes RS, Silber JH, Kurtz SG, Gibber EM. Should do-not-resuscitate status be included as a mortality risk adjustor? The impact of DNR variations on performance reporting. Med Care 2005; 43:658-666]

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
Silber JH, Williams SV, Krakauer H, et al. Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure-to-rescue. Med Care. 1992;30:615-629. Silber JH, Rosenbaum PR, Schwartz JS, et al. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. JAMA. 1995;274:317-323 Silber JH, Romano PS, Rosen AK, et al. Failure-to-rescue: Comparing definitions to measure quality of care. Med Care. 2007;45:918-925

1c.6 Method for rating evidence: In Silber et al JAMA 1995, refers to the "power" of a measure as the ability of that measure to detect differences between hospitals or groups of hospitals, with respect to the outcome measure in question. Should the difference between two hospital failure rates achieve statistical significance, while the difference between those same hospitals´ death rates not achieve statistical significance, then we would consider the failure rate to be more powerful than the death rate. It can be shown that for equivalent adverse occurrence rates, the power to distinguish between two hospitals using the failure rate is always greater than or equal to the power using the death rate. Although somewhat counterintuitive, this result occurs because, although the failure rate and the death rate use the number of deaths as their numerators, the denominator of the failure rate is the number of patients with adverse occurrences, while the denominator of the death rate is the total number of patients. When adverse occurrence rates are not equal across hospitals, the power of the failure rate statistic may be greater than, equal to, or less than that of the death rate. When comparing two hospitals with failure rates F1 and F2 death rates Dl and D2 and adverse occurrence rates A1 and A2 it can be shown that whenever F1>= F2, Dl>= D2 and A1<=A2 then the power in distinguishing such hospitals using the failure rate is greater than or equal to the power when using the death rate. For situations where F1>=F2 and Dl < D2 the sufficient conditions for superior power using the failure rate instead of the death rate is given in the Appendix. Finally, these results are unchanged if one considers either hospital I or 2 in the above arguments to be a group of hospitals or the average of all hospitals (so that hospital 1 or 2 represents a very large sample size).
In summary, failure rate was a function of anesthesia board certification and the presence of surgical housestaff (hospital characteristics) but not a function of admission severity of illness score (patient characteristics). Since the death rate appears to be composed of two distinct rates, quality of care measurement may be improved if all three rates are reported instead of relying on the adjusted mortality rate alone. In so doing, we may better understand the reasons for variation in hospital mortality rates.

1c
C ☐
P ☐
M ☐
N ☐

1c.7 Summary of Controversy/Contradictory Evidence: N/A

1c.8 Citations for Evidence (*other than guidelines*): 1. Silber JH, Williams SV, Krakauer H, et al. Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure-to-rescue. Med Care. 1992;30:615-629.
2. Silber JH, Romano PS, Rosen AK, et al. Failure-to-rescue: Comparing definitions to measure quality of care. Med Care. 2007;45:918-925.
3. Silber JH, Rosenbaum PR, Schwartz JS, et al. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. JAMA. 1995;274:317-323.
4. Silber JH, Rosenbaum PR, Williams SV, et al. The relationship between choice of outcome measure and hospital rank in general surgical procedures: Implications for quality assessment. Int J Qual Health Care. 1997;9:193-200.
5. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist direction and patient outcomes. Anesthesiology. 2000;93:152-163.
6. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist board certification and patient outcomes. Anesthesiology. 2002;96:1044-1052.
7. Aiken LH, Clarke SP, Sloane DM, et al. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA. 2002;288:1987-1993.
8. Aiken LH, Clarke SP, Cheung RB, et al. Educational levels of hospital nurses and surgical patient mortality. JAMA. 2003;290:1617-1623.
9. Silber JH, Rosenbaum PR, Ross RN. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? J Am Stat Assoc. 1995;90:7-18.
10. Silber JH, Rosenbaum PR, Romano PS, Rosen AK, Wang Y, Teng Y, Halenar MJ, Even-Shoshan O, Volpp KG. Hospital Teaching Intensity, Patient Race,
and Surgical Outcomes. Arch Surg. 2009;144:113-120.
11. Friese CR, Earle CC, Silber JH, Aiken LH. Hospital characteristics, clinical severity, and outcomes for surgical oncology patients. Surgery 2010; 147:602-609.
12. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in Hospital Mortality Associated with Inpatient Surgery. N Engl J Med 2009; 361:1368-75.
13. Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational Levels of Hospital Nurses and Surgical Patient Mortality.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number*):
N/A

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:
The motivation behind the development of traditional FTR was based on 2 questions. The first was an empirical question—suppose hospitals were ranked by adjusted mortality and adjusted complication rates. Would these rankings be highly correlated? The answer is rather surprising—there is generally poor correlation or no correlation in most analyses. Second, suppose 2 hospitals had identical adjusted mortality rates but different adjusted complication rates. Would one prefer care at the hospital with the higher or lower complication rate? If one believes that complications are predominantly driven by patient characteristics, then one may decide to choose the hospital with the higher complication rate, as it achieved an equivalent mortality rate with a sicker population of patients. So there is an empirical question to ask—are adjusted complication rates more related to hospital or patient factors? This has been looked at in a number of ways—and the evidence to date suggests that complication measures are less sensitive to hospital characteristics, after adjusting for severity of illness, than mortality based measures. This is an

underlying assumption of FTR theory—complications are undesirable outcome measures because they reflect underlying patient severity and diagnosis coding more than they reflect hospital care. Instead, a hospital's quality is put to the test when a patient develops a complication, and whether a patient is salvaged after a complication will be a function of the care delivered by the hospital and its knowledge base, depth, and facilities. Thus, "good" hospitals will rescue patients by identifying complications quickly and treating them aggressively, resulting in lower FTR. Although many "failures," just like deaths, are often not preventable, we have argued that FTR may be a better measure for comparing hospital quality because of better severity adjustment properties, and because of its focus on hospital actions. By studying a population of patients who, by definition, have already developed a complication, the specifics of severity of illness adjustment becomes less important in failure rate analyses, because all patients have experienced complications and thus are more uniformly ill.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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***):** Patients who died with a complication plus patients who died without documented complications. Death is defined as death within 30 days from admission. All patients in an FTR analysis have developed a complication (by definition). Complicated patient has at least one of the complications defined in Appendix B(see website http://www.research.chop.edu/programs/cor/outcomes.php). Complications are defined using the secondary ICD9 diagnosis and procedure codes and the DRG code of the current admission. Comorbidities are defined in Appendix C(see website http://www.research.chop.edu/programs/cor/outcomes.php) using secondary ICD9 diagnosis codes of the current admission and primary or secondary ICD9 diagnosis codes of previous admission within 90 days of the admission date of the current admission. *When physician part B is available, the definition of complications and comorbidities are augmented to include CPT codes. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** Within 30 days from admission. **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Patients who died with complication and patients who died without documented complications. Death is defined as death within 30 days from admission.	**2a-specs** **C☐** **P☐** **M☐** **N☐**
2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**	

General Surgery, Orthopedic and Vascular patients in specific DRGs with complications plus patients who died in the hospital without complications.

Inclusions: adult patients admitted for one of the procedures in the General Surgery, Orthopedic or Vascular DRGs (see appendix A http://www.research.chop.edu/programs/cor/outcomes.php)
Inclusions: adult patients admitted for one of the procedures in the General Surgery, Orthopedic or Vascular DRGs (see appendix A)

2a.5 **Target population gender:** Female, Male
2a.6 **Target population age range:** 18-90

2a.7 **Denominator Time Window (***The time period in which cases are eligible for inclusion in the denominator***):**
Within 30 days from admission

2a.8 **Denominator Details (***All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Adult patients admitted for one of the procedures in the General Surgery, Orthopedic or Vascular DRGs (see Appendix A http://www.research.chop.edu/programs/cor/outcomes.php)who developed an in hospital complication and those who died without a complication.

2a.9 **Denominator Exclusions (***Brief text description of exclusions from the target population***):** Patients over age 90, under age 18.

2a.10 **Denominator Exclusion Details (***All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
N/A

2a.11 **Stratification Details/Variables (***All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
Complicated patient has at least one of the complications defined in Appendix B (http://www.research.chop.edu/programs/cor/outcomes.php) Complications are defined using the secondary ICD9 diagnosis and procedure codes and the DRG code of the current admission. When Physician Part B file is available, the definition of complications and comorbidities are augmented to include CPT codes.

2a.12-13 **Risk Adjustment Type:** Risk-adjustment devised specifically for this measure/condition

2a.14 **Risk Adjustment Methodology/Variables (***List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
Risk Adjustment: Model was developed using logistic regression analysis.

Associated data elements: age in years, sex, race, comorbidities, DRGs (combined with and without complications) and procedure codes within DRGs, transfer status.

Failure to rescue is adjusted using a logistic regression model where y is a failure and the total N is composed of patients who develop a complication and patients who died without a complication.

According to developer: The model adjustment variables can vary. We have found that FTR results are fairly stable, even with little adjustment, since all patients in an FTR analysis have developed a complication (by definition), they are a more homogeneous group of patients than the entire population. Hence severity adjustment plays somewhat less of a role than in other outcome measures.

2a.15-17 **Detailed risk model available Web page URL or attachment:** URL
http://www.research.chop.edu/programs/cor/outcomes.php

2a.18-19 **Type of Score:** Rate/proportion
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***):**
Refer to website (http://www.research.chop.edu/programs/cor/outcomes.php)

2a.22 Describe the method for discriminating performance *(e.g., significance testing)*:
T-test for comparing rates

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)*:
Measure not based on sample, all surgical patients between the ages of 18 and 90 admitted to an acute care hospital.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Electronic administrative data/claims

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.***):**
Linked patients hospitalizations claims records, augmented with Outpatient and Part B records; can also use unlinked data if linked files are not available to identify comorbidities and develop definitions of severity and other risk measure.

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL
http://www.resdac.org/

2a.29-31 Data dictionary/code table web page URL or attachment: URL
http://www.research.chop.edu/programs/cor/outcomes.php

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency, Health Plan, Integrated delivery system, Population: national, Population: regional/network, Population: states, Population: counties or cities

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

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: Medicare inpatient claims for general surgical admissions for the period July 1, 1999 to June 30, 2000. There were a total of 1467 hospitals and 403,679 patients. We included patients between 65 and 90 years of age.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
We defined reliability as described by Lord and Novick using split sample methodology. (Lord FM, Novick MR. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley; 1968)

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Using Spearman-Brown half split half sample reliability had a correlation of 0.32 and the upper bound on validity was 0.56.

2b
C☐
P☐
M☐
N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: Medicare inpatient claims for general surgical admissions for the period July 1, 1999 to June 30, 2000. There were a total of 1467 hospitals and 403,679 patients. We included patients between 65 and 90 years of age.

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
a) Rank correlation between various hospital outcomes (Death, Failure to Rescue, Complications, other measures of Failure to Rescue, Failure to Rescue Complement measures)

2c
C☐
P☐
M☐
N☐

b) Marginal and partial coefficients in logit models using detailed patient characteristics and hospital characteristics shown to be associated with better outcomes in previous studies.2, 7 The marginal results use one hospital characteristic at a time along with all patient characteristics. "Partial" regression results, using all hospital and patient variables simultaneously have the disadvantage that correlation between hospital characteristics can cause difficulty in interpreting the effects of individual hospital variables. Hospital characteristics associated with better outcomes (1) teaching hospital status (member of the American Council of Teaching Hospitals); (2) high technology status (does the hospital perform open heart surgery or perform organ transplantation); (3) hospital size greater than 200 beds; (4) bed-to-nurse ratio (where nurses are the sum of RN plus LPN FTE positions); and (5) nursing skill mix (the ratio of RN/(RN+LPN)).2-8

c) The relative contribution of patient-to-hospital characteristics that predicted each outcome of interest, as provided by the omega statistic.2, 9 The omega statistic computes a ratio of the squared sum of the log odds for model patent variables divided by a similar quantity calculated for the model hospital variables. All else being equal, outcome measures that have lower omega ratios may be more desirable quality indicators, since the lower the omega, the greater the hospital's impact on outcome relative to the patient's impact. This is especially important if modeling patient severity is difficult (as with claims data) so that the lower the omega suggests the higher relative influence of hospital characteristics as compared to patient.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
FTR itself is highly correlated with death, with a Kendall's tau equal to 0.83, representing a probability of concordance equal to 0.91.

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
Patients younger than 18 are excluded because they are considered in the pediatric population and have a different set of complications. We use 90 years as a cut-point because of our concern regarding the increased use of do-not-resuscitate at higher ages [Wenger et al. Epidemiology of Do-Not Resuscitate Orders. Disparity by Age, Diagnosis, Gender, Race, and Functional Impairment. Arch Intern Med. 1995; 155(19):2056-62, Hakim et al. Factors Associated with Do-Not-Resuscitate Orders: Patients´, Preferences, Prognoses, and Physicians Judgments. Ann Intern Med.1996; 125:284-293.]. While we do adjust for admission severity when reporting FTR, and this includes age, we still thought it prudent to use an upper bound on age, since DNR status prior to the procedure is not well defined at hospitals [Tabak YP, Johannes RS, Silber JH, Kurtz SG, Gibber EM. Should do-not-resuscitate status be included as a mortality risk adjustor? The impact of DNR variations on performance reporting. Med Care 2005; 43:658-666]

2d.2 Citations for Evidence:
1. Wenger NS, Pearson ML, Desmond KA, Harrison ER, Rubenstein LV, Rogers WH, Kahn KL. Epidemiology of Do-Not Resuscitate Orders. Disparity by Age, Diagnosis, Gender, Race, and Functional Impairment. Arch Intern Med. 1995; 155(19):2056-62
2. Hakim RB, Teno JM, Harrell Jr. FE, Knaus WA, Wenger N, Phillips RS, Layde P, Califf R, Connors Jr. AF, Lynn J. Factors Associated with Do-Not-Resuscitate Orders: Patients´, Preferences, Prognoses, and Physicians Judgments. Ann Intern Med. 1996; 125:284-293.
3. Tabak YP, Johannes RS, Silber JH, Kurtz SG, Gibber EM. Should do-not-resuscitate status be included as a mortality risk adjustor? The impact of DNR variations on performance reporting. Med Care 2005; 43:658-666

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

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

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

2d
C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e
C☐

2e.1 Data/sample *(description of data/sample and size)*: Two different data samples were used to analyze risk adjustment. 1.) 5,972 Medicare patients undergoing elective cholecystectomy or transurethral prostatectomy (Silber et al. Hospital and Patient Characteristics Associated with Death After Surgery A study of Adverse Occrueenece and Failure to Rescue Med Care 1992).
2.) 2,021,214 patients with medicare claims on general, orthopedic, and vascular surgery admissions in the United States for 2000-2005. (Silber et al. Hospital Teaching Intensity, Patient Race, and Surgical Outcomes Arch Surg 2009)

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
Risk Adjustment: Model was developed using logistic regression analysis, where y is a failure and the total N is composed of patients who develop a complication and patients who died without a complication.

2e.3 Testing Results *(risk model performance metrics)*:
In earlier work we did report calibration as tested with the Hosmer-Lemeshow statistic, however the research community found that this calibration test fails its asymptotics, it overcalls with large sample size, we do not recommend its use. It is well known that the Hosmer-Lemeshow test is misleading with large data sets, and therefore we have not thought this to be a valid approach. C-statistic ranges 0.70 for the FTR 30 day risk adjustment model (Silber et. al Med Care 1992) to 0.792 (Silber et al. Arch Surg 2009). However c-statistics are also misleading when comparing across populations. Since FTR is a subset of the mortality and complication data set, one cannot compare, in a meaningful way, the c-statistic from FTR to that of mortality or complication.

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

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)*: Medicare inpatient claims for general surgical admissions for the period July 1, 1999 to June 30, 2000. There were a total of 1467 hospitals and 403,679 patients. We included patients between 65 and 90 years of age.

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
T-test for comparing rates.

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)*:
75% Q3 = 0.16, Median= 0.12, 25% Q1 =0.09, Mean= 0.13, Std Deviation =0.05.

2f C☐ P☐ M☐ N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample *(description of data/sample and size)*: FTR was developed using standardized hospital discharge records, which are widely collected by states agencies and which hospitals are mandated to report to CMS. One of the big advantages of adopting FTR is that the data on which it is based is uniformely reported, checked for errors and edited. This is administrative data available for the entire population over 65 and for all patients admitted to acute care hospitals.

2g.2 Analytic Method *(type of analysis & rationale)*:
N/A

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

2g C☐ P☐ M☐ N☐ NA☐

2h. Disparities in Care

2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)*:
Disparities in care are shown in Silber et al Arch Surg 2009 where the results show white patients displayed a reduction in failure-to-rescue rates in the teaching intensive hospitals vs non-teaching hospitals (OR, 0.94; 95% CI, 0.92-0.97), black patients displayed an increased failure-to-rescue rate (OR, 1.06; 95% CI, 1.00-1.12)

2h C☐ P☐ M☐ N☐ NA☐

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: N/A	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, met? **Rationale:**	***2*** **C☐** **P☐** **M☐** **N☐**
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Eval Rating**
3a. Meaningful, Understandable, and Useful Information **3a.1 Current Use:** In use **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** This measure has not yet been used in a public reporting initiative. This measure could be reported on a wide scale, the same way that mortality rates are reported. **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***):** Currently used to assess the impact of the change in the resident work hours regulations on patient outcomes in a recently NHLBI funded study (1R01HL094593-01) entitled "Work Hour Regulation for Physician Trainees: Educational and Clinical Outcomes" **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)***:** In Ghaferi et al "Variation in Hospital Mortality Associated with Inpatient Surgery" studied 84,730 patients who had undergone inpatient general and vascular surgery from 2005-2007 using data from the American College of Surgeons National Surgical Quality Improvement Program. **3a.5 Methods** *(e.g., focus group, survey, QI project)***:** Ranked ranked hospitals according to their risk-adjusted overall rate of death and divided them into five groups. For hospitals in each overall mortality quintile, we then assessed the incidence of overall and major complications and the rate of death among patients with major complications (failure to rescue rate). **3a.6 Results** *(qualitative and/or quantitative results and conclusions)***:** Rates of death varied widely across hospital quintiles, from 3.5% in very-low-mortality hospitals to 6.9% in very-high-mortality hospitals. Hospitals with either very high mortality or very low mortality had similar rates of overall complications (24.6% and 26.9%, respectively) and of major complications (18.2% and 16.2%, respectively). Rates of individual complications did not vary significantly across hospital mortality quintiles. In contrast, mortality in patients with major complications was almost twice as high in hospitals with very high overall mortality as in those with very low overall mortality (21.4% vs. 12.5%, P<0.001). Differences in rates of death among patients with major complications were also the primary determinant of variation in overall mortality with individual operations. In addition to efforts aimed at avoiding complications in the first place, reducing mortality associated with inpatient surgery will require greater attention to the timely recognition and management of complications once they occur.	**3a** **C☐** **P☐** **M☐** **N☐**
3b/3c. Relation to other NQF-endorsed measures **3b.1 NQF # and Title of similar or related measures:**	

0200 Death among surgical inpatients with treatable serious complications (failure to rescue)	
(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:** **5.1 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:** Needleman et al adapted the FTR measure to "nurse sensitive complications" by selecting a limited number of complications for the FTR measure. This change in definition, which we will call FTR-N, was developed to better focus on nursing quality of care. Because only deaths after nursing sensitive complications are studied, a large number of deaths are not used in the analysis. Subsequently, AHRQ again adapted the FTR-N definition to reflect quality from a "patient safety" perspective (ie, the identification of deaths that were especially likely to be preventable). Expert panels guided both of these adaptations through consensus development panels. The National Quality Forum, through its own process of selecting National Voluntary onsensus Standards for Nursing-Sensitive Care, endorsed Needleman et al's adaptation and assigned it to AHRQ for updating and support.FTR-N includes only 6 complications (pneumonia, shock, gastrointestinal bleeding, cardiac arrest, sepsis, and deep venous thrombosis) in its denominator definition, and it excludes deaths in patients without these complications. FTR-A adds renal failure to the FTR-N list of eligible complications, and modestly alters the definition of several others Table 1C and 1D display the impact of restricting the denominator of FTR to more limited sets of complications, as in the FTR-N and FTR-A definitions, respectively. Note first that the number of patients defined as having a complication fell from 189,031 (46.8%) in Table 1A to 43,500 (10.8%) in Table 1C and 39,101 (9.7%) in Table 1D. However, this smaller complication rate comes at an important cost—of all deaths, the proportion coded as having a complication (the precedence rate) fell from 95% in Table 1A to only 51% in Table 1C, and 58.5% in Table 1D. (Refer tp Silber et al. Med Care 2007)	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure*	**4b** C☐ P☐

scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims) Yes 4b.2 **If not, specify the near-term path to achieve electronic capture by most providers.**	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.** FTR is given to minimal susceptibility to inaccuracies or errors since it uses data collected uniformly across all hospitals and providers. The data is carefully checked by CMS before it is being released to researchers. However there may be unobserved differences among patients due to the lack of more detailed clinical information available only through chart abstraction.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation 4e.1 **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** We have developed FTR measures based on restricted information, available only from the inpatient files. When possible, such as in the Medicare population, we improve the risk adjustment by using more patient level information available in the outpatient or Carrier file 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** CMS data is made available to researchers through ResDac, and its cost depends on the number of records requested, the number of years, and the type of file (inpatient, outpatient, or carrier) 4e.3 **Evidence for costs:** N/A 4e.4 **Business case documentation:** N/A	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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** The Children´s Hospital of Philadelphia, 34th St. and Civic Center Blvd., Philadelphia, Pennsylvania, 19104 **Co.2 Point of Contact** Jeffrey, Silber, PhD, MD, silber@email.chop.edu, 215-590-2540-
Measure Developer If different from Measure Steward **Co.3 Organization** The Children´s Hospital of Philadelphia, 34th St. and Civic Center Blvd., Philadelphia, Pennsylvania, 19104 **Co.4 Point of Contact** Fabienne, Kyle, BA, kylef@email.chop.edu, 215-590-2484-
Co.5 Submitter If different from Measure Steward POC Fabienne, Kyle, BA, kylef@email.chop.edu, 215-590-2484-, Center for Outcomes Research, Children´s Hospital of Philadelphia
Co.6 Additional organizations that sponsored/participated in measure development N/A
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.** A group of clinicians and coding experts from the University of Pennsylvania reviewed the updated ICD, CPT, and DRG codes and updated the measure to reflect current coding.
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: URL http://www.research.chop.edu/programs/cor/outcomes.php
Date of Submission (*MM/DD/YY*)**:** 03/29/2011

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 0515 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Ambulatory surgery patients with appropriate method of hair removal
De.2 **Brief description of measure:** Percentage of ASC admissions with appropriate surgical site hair removal.
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** Not included in a composite or paired with another measure
De.4 **National Priority Partners Priority Area:** Safety De.5 **IOM Quality Domain:** Effectiveness De.6 **Consumer Care Need:** Staying healthy

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 measure A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission 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	**B** Y☐

every 3 years. Yes, information provided in contact section	N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	**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, Frequently performed procedure, Leading cause of morbidity/mortality, High resource use, Severity of illness, Patient/societal consequences of poor quality **1a.2** **1a.3 Summary of Evidence of High Impact:** As a result of advances in surgery and anesthesia, approximately 80 percent of surgeries in the United States are now performed on an outpatient basis. Ambulatory surgical centers perform approximately 40%, or more than 22 million, of those outpatient surgeries.1 Appropriate surgical site hair removal is measured for surgical patients in the hospital inpatient setting, and given the high volume of outpatient surgical procedures, should also be measured in the outpatient setting. Accumulated evidence suggests that shaving the surgical site is associated with an increased incidence of surgical site infections. Razors are thought to cause microabrasions that may subsequently become infected. Hair removal with clippers has been demonstrated to reduce the rate of surgical site infections and associated healthcare expenditures. 2-12 Surgical site infection rates in ambulatory surgery are not well understood. However, in other settings, surgical site infections occur in 2 to 5 percent of clean extra-abdominal surgeries. Evidence suggests each infection increases a hospital stay by 7 to 10 days and adds from $3,000 to $29,000 in charges. Patients who develop surgical site infections are thought to have at least twice the incidence of mortality when compared to surgical patients without a surgical site infection. 13-19	**1a** C☐ P☐ M☐ N☐

1a.4 Citations for Evidence of High Impact:

1. U.S. Department of Health and Human Services. Centers for Medicare & Medicaid Services. http://www.cms.gov/.

2. Alexander JW, Fischer JE, Boyajian M, Palmquist J, Morris MJ. The influence of hair-removal methods on wound infections. Arch Surg. 1983 Mar;118(3):347-52.

3. Balthazar ER, Colt JD, Nichols RL. Preoperative hair removal: a random prospective study of shaving versus clipping. South Med J. 1982 Jul;75(7):799-801.

4. Court-Brown CM. Preoperative skin depilation and its effect on postoperative wound infections. J R Coll Surg Edinb. 1981 Jul;26(4):238-41.

5. Kjonniksen I, Andersen BM, Sondenaa VG, Segadal L. Preoperative hair removal--a systematic literature review. AORN J. 2002 May;75(5):928-38, 940.

6. Ko W, Lazenby WD, Zelano JA, Isom OW, Krieger KH. Effects of shaving methods and intraoperative irrigation on suppurative mediastinitis after bypass operations. Ann Thorac Surg. 1992 Feb;53(2):301-5.

7. Powis SJ, Waterworth TA, Arkell DG. Preoperative skin preparation: clinical evaluation of depilatory cream. Br Med J. 1976 Nov 13;2(6045):1166-8.

8. Seropian R, Reynolds BM. Wound infections after preoperative depilatory versus razor preparation. Am J Surg. 1971 Mar;121(3):251-4.

9. Tanner J, Moncaster K, Woodings D. Preoperative hair removal to reduce surgical site infection. Cochrane Database Syst Rev. 2006 Jul 19;3:CD004122.

10. Thur de Koos P, McComas B. Shaving versus skin depilatory cream for preoperative skin preparation. A prospective study of wound infection rates. Am J Surg. 1983 Mar;145(3):377-8.

11. Gurkan I, Wenz Sr, JF. Perioperative infection control: an update for patient safety in orthopedic surgery. Orthopedics. 2006 Apr;29(4):329.

12. Fletcher N, Sofianos D, Berkes MB, Obremskey WT. Prevention of perioperative infection. J Bone Joint Surg Am. 2007;89:1605-18.

13. Cruse P. Wound infection surveillance. Rev Infect Dis 1981; 3:734-737.

14. Cruse PJ, Foord R. The epidemiology of wound infection: a 10-year prospective study of 62,939 wounds. Surg Clin North Am 1980; 60:27-40.

15. Engemann JJ, Carmeli Y, Cosgrove SE, et al. Adverse clinical and economic outcomes attributable to methicillin resistance among patients with Staphylococcus aureus surgical site infection. Clin Infect Dis 2003; 36:592-598.

16. Kirkland K, Briggs J, Trivette S, Wilkinson W, and Sexton D. 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):725-30.

17. Coello R, Glenister H, Fereres J, et al. The cost of infection in surgical patients: a case-control study. J Hosp Infect 1993; 25:239-250.

18. Vegas AA, Jodra VM, Garcia ML. Nosocomial infection in surgery wards: a controlled study of increased duration of hospital stays and direct cost of hospitalization. Eur J Epidemiol 1993; 9:504-510.

19. Whitehouse JD, Friedman ND, Kirkland KB, Richardson WJ, Sexton DJ. The impact of surgical-site infections following orthopedic surgery at a community hospital and a university hospital: adverse quality of

life, excess length of stay, and extra cost. Infect Control Hosp Epidemiol 2002; 23:183-189.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Improving the rate of appropriate surgical site hair removal is expected to reduce the risk of surgical site infection.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
The rates for this measure were collected for 192 ambulatory surgery centers throughout the US for services provided during July to September 2010. The rate for appropriate surgical site hair with removal clippers or depilatory cream ranged from a minimum of 0.0% to a maximum of 100%. The mean rate was 96% (SD: 18%), while the median rate was 100%. The minimum rate of 0% and the fact that 7.3% of the centers reported a rate of lower than 100% demonstrate that there is an opportunity for improvement in this measure.

1b.3 Citations for data on performance gap:
A convenience sample of 192 ambulatory surgery centers was selected to assess the opportunity for improvement for this measure. The centers were located throughout the US. Services from the third calendar quarter of 2010 were included in this portion of the study.

1b.4 Summary of Data on disparities by population group:
This measure is not intended to evaluate disparities by population group.

1b.5 Citations for data on Disparities:
No data available for disparities by population group. Please see 1b.4. above, this measure is not intended to evaluate disparities by population group.

Regarding 1b.2. above, a convenience sample of 192 ambulatory surgery centers was selected to assess the opportunity for improvement for this measure. The centers were located throughout the US. Services from the third calendar quarter of 2010 were included in this portion of the study.

1b
C☐
P☐
M☐
N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Evidence suggests improving the rate of appropriate surgical site hair removal can be expected to reduce the risk of surgical site infection.

1c.2-3. Type of Evidence: Observational study, Evidence-based guideline, Randomized controlled trial, Expert opinion, Systematic synthesis of research

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
The literature regarding preoperative hair removal has been systematically reviewed twice, once by Kjonniksen et al in 2002 and again by Tanner et al in 2007. Three randomized controlled trials (Alexander et al 1983, Balthazar et al 1983, Ko et al 1992) compared the rates of infection at the surgical site when hair removal at the site was performed with clippers or with razors. A statistically significant difference in infection rates in the pooled results (Tanner et al 2007) was seen, with 2.8% of the patients who were shaved developing a surgical site infection compared with 1.4% rate of surgical site infection in the patients who were clipped. Additional randomized controlled trials (Court-Brown 1981, Powis et al 1976, Seropian 1971, Thur de Koos 1983) have demonstrated that patients were more likely to develop a surgical site infection when shaved as compared to having hair removal with a depilatory. Observational studies have suggested that no hair removal is less likely to result in surgical site infection, but this has not been confirmed in randomized controlled trials.

The HICPAC/CDC Guideline for Prevention of Surgical Site Infection (Mangram at al 1999), the Association of Operating Room Nurses Recommended Practices for Preoperative Patient Skin Antisepsis (AORN 2002) and the SHEA/IDSA Strategies to Prevent Surgical Site Infections in Acute Care Hospitals (Anderson et al 2008) are consistent with the intent of this measure.

1c
C☐
P☐
M☐
N☐

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom*): Most recently rated II-A by SHEA/IDSA in 2008. II: Evidence from >1 well-designed clinical trial without randomization, from cohort or case-controlled analytic studies (preferably from >1 center), from multiple time-series studies, or from dramatic results of uncontrolled experiments ; A: Good evidence to support a recommendation for use

1c.6 Method for rating evidence: The Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA) Standards and Practice Guidelines Committee convened experts in the prevention and monitoring of healthcare-associated infections (HAIs). In evaluating the evidence regarding the prevention and monitoring of HAIs, the HAI Allied Task Force followed a process used in the development of other IDSA guidelines, including a systematic weighting of the quality of the evidence and the grade of recommendation.

The weighting methodology was adapted from the Canadian Task Force on the Periodic Health Examination.
Strength of recommendation:
A Good evidence to support a recommendation for use
B Moderate evidence to support a recommendation for use
C Poor evidence to support a recommendation
Quality of evidence:
I Evidence from > or = 1 properly randomized, controlled trial
II Evidence from > or = 1 well-designed clinical trial, without randomization; from cohort or case-control analytic studies (preferably from >1 center); from multiple time series; or from dramatic results from uncontrolled experiments
III Evidence from opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees

1c.7 Summary of Controversy/Contradictory Evidence: In July 2009, the Surgical Quality Alliance requested review of another hair removal measure endorsed by the National Quality Forum (NQF), entitled "Surgery Patients with Appropriate Hair Removal" (#0301). This measure was considered under ad hoc review because of concerns that the evidence did not support hair removal for specific surgeries. The NQF Board voted to continue endorsement of this measure with minor modifications to reflect evidence changes regarding hair removal for neurosurgical and scrotum procedures.

Some recent studies suggest that better adherence to individual infection-related process measures is not significantly associated with better outcomes. See:

Ingraham AM, Cohen ME, Bilimoria KY, Dimick JB, Richards KE, Raval MV, Fleisher LA, Hall BL, Ko CY. Association of surgical care improvement project infection-related process measure compliance with risk-adjusted outcomes: implications for quality measurement. J Am Coll Surg. 2010 Dec;211(6):705-14.

Stulberg JJ, Delaney CP, Neuhauser DV, Aron DC, Fu P, Koroukian SM. Adherence to surgical care improvement project measures and the association with postoperative infections. JAMA. 2010 Jun 23;303(24):2479-85.

1c.8 Citations for Evidence (*other than guidelines*): Alexander JW, Fischer JE, Boyajian M, Palmquist J, Morris MJ. The influence of hair-removal methods on wound infections. Arch Surg. 1983 Mar;118(3):347-52.

Anderson DJ, Kaye KS, Classen D, Arias KM, Podgorny K, Burstin H, Calfee DP, Coffin SE, Dubberke ER, Fraser V, Gerding DN, Griffin FA, Gross P, Klompas M, Lo E, Marschall J, Mermel LA, Nicolle L, Pegues DA, Perl TM, Saint S, Salgado CD, Weinstein RA, Wise R, Yokoe DS. Strategies to prevent surgical site infections in acute care hospitals. Infect Control Hosp Epidemiol 2008 Oct;29 Suppl 1:S51-61.

Association of Operating Room Nurses. Recommended practices for skin preparation of patients. AORN J. 2002 Jan;75(1):184-7.

Balthazar ER, Colt JD, Nichols RL. Preoperative hair removal: a random prospective study of shaving versus clipping. South Med J. 1982 Jul;75(7):799-801.

Court-Brown CM. Preoperative skin depilation and its effect on postoperative wound infections. J R Coll Surg Edinb. 1981 Jul;26(4):238-41.

Kjonniksen I, Andersen BM, Sondenaa VG, Segadal L. Preoperative hair removal--a systematic literature review. AORN J. 2002 May;75(5):928-38, 940.

Ko W, Lazenby WD, Zelano JA, Isom OW, Krieger KH. Effects of shaving methods and intraoperative irrigation on suppurative mediastinitis after bypass operations. Ann Thorac Surg. 1992 Feb;53(2):301-5.

Powis SJ, Waterworth TA, Arkell DG. Preoperative skin preparation: clinical evaluation of depilatory cream. Br Med J. 1976 Nov 13;2(6045):1166-8.

Seropian R, Reynolds BM. Wound infections after preoperative depilatory versus razor preparation. Am J Surg. 1971 Mar;121(3):251-4.

Tanner J, Moncaster K, Woodings D. Preoperative hair removal to reduce surgical site infection. Cochrane Database Syst Rev. 2006 Jul 19;3:CD004122.

Thur de Koos P, McComas B. Shaving versus skin depilatory cream for preoperative skin preparation. A prospective study of wound infection rates. Am J Surg. 1983 Mar;145(3):377-8.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
Guideline for prevention of surgical site infection, page 266:

2. Do not remove hair preoperatively unless the hair at or around the incision site will interfere with the operation. Category IA
3. If hair is removed, remove immediately before the operation, preferably with electric clippers. Category IA

1c.10 Clinical Practice Guideline Citation: Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR. Guideline for prevention of surgical site infection, 1999. Hospital Infection Control Practices Advisory Committee. Infect Control Hosp Epidemiol. 1999;20:250 -78.
1c.11 National Guideline Clearinghouse or other URL:
http://www.cdc.gov/ncidod/dhqp/pdf/guidelines/SSI.pdf

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
Category IA: Strongly recommended for implementation and supported by well-designed experimental, clinical, or epidemiological studies. Rating given by HICPAC.

1c.13 Method for rating strength of recommendation (*If different from USPSTF system, also describe rating and how it relates to USPSTF)*:
Category I recommendations, including IA and IB, are those recommendations that are viewed as effective by HICPAC and experts in the fields of surgery, infectious diseases, and infection control.

1c.14 Rationale for using this guideline over others:
The HICPAC/CDC guideline provides guidance for surgical care in all settings, whereas the SHEA/IDSA guideline has a acute care hospital focus.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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*)**:** ASC admissions with surgical site hair removal with a razor or clippers from the scrotal area, or with clippers or depilatory cream from all other surgical sites **2a.2 Numerator Time Window** (*The time period in which cases are eligible for inclusion in the numerator*)**:** In-facility, prior to discharge **2a.3 Numerator Details** (*All information required to collect/calculate the numerator, including all codes, logic, and definitions*)**:** DEFINITIONS: Admission: completion of registration upon entry into the facility	
2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured*)**:** All ASC admissions with surgical site hair removal **2a.5 Target population gender:** Female, Male **2a.6 Target population age range:** All ages **2a.7 Denominator Time Window** (*The time period in which cases are eligible for inclusion in the denominator*)**:** In-facility, prior to discharge **2a.8 Denominator Details** (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions*)**:** DEFINITIONS: Admission: completion of registration upon entry into the facility	
2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population*)**:** ASC admissions who perform their own hair removal **2a.10 Denominator Exclusion Details** (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions*)**:** To collect data for the denominator exclusion, centers must track patients who perform their own hair removal	
2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions*)**:** The measure is not stratified	
2a.12-13 Risk Adjustment Type: No risk adjustment necessary **2a.14 Risk Adjustment Methodology/Variables** (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method*)**:** Not applicable **2a.15-17 Detailed risk model available Web page URL or attachment:**	**2a-specs** **C☐** **P☐** **M☐** **N☐**

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
1a. The number of admissions with surgical site hair removal is determined.
1b. The number of admissions who performed their own surgical site hair removal is determined.
1c. The value determined in step 1b is subtracted from the value determined in step 1a to yield the measure denominator.

2. The number of admissions with appropriate surgical site hair removal (hair removal with razor or clippers from the scrotal area, or hair removal with clippers or depilatory cream from all other surgical sites) is determined. This value is the measure numerator.

3. The number of ASC admissions with appropriate surgical site hair removal (step 2) is divided by the number of ASC admissions with surgical site hair removal (steps 1a through 1c) during the reporting period, yielding the rate of appropriate surgical site hair removal for the reporting period.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)*:
Facilities reporting data may compare their performance to the average performance. Alternatively, facilities may compare their performance to a percentile ranking (such as the 50th percentile (median)) to determine their relative performance.

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)*:
The measure is not based on a sample

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Paper medical record/flow-sheet

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.***):**
Facilities may review records such as a pre-surgical checklist, nursing notes, operating room record, and operative report as needed for documentation of method of hair removal. Clinical logs designed to capture information relevant to preoperative hair removal may also be used.

No specific collection instrument is required, although the ASC Quality Collaboration has developed a sample data collection instrument that may be used as desired. Facilities may use any collection instrument that allows tracking of the method of hair removal for all admissions with surgical site hair removal.

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL Not required http://ascquality.org/documents/ASCQualityCollaborationImplementationGuide.pdf

2a.29-31 Data dictionary/code table web page URL or attachment: URL Not required
http://ascquality.org/documents/ASCQualityCollaborationImplementationGuide.pdf

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency

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

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Other Ambulatory surgical center

TESTING/ANALYSIS

2b. Reliability testing

2b
C☐
P☐
M☐
N☐

2b.1 Data/sample *(description of data/sample and size)*: Reliability testing was performed using a previous numerator definition (ASC admissions with surgical site hair removal with clippers or depilatory cream). The revised numerator statement (ASC admissions with surgical site hair removal with a razor or clippers from

the scrotal area, or with clippers or depilatory cream from all other surgical sites) was formulated to harmonize with the current SCIP measure for appropriate hair removal. This change impacts a very small proportion of the ASC admissions and therefore it is unlikely that this change will have a material impact on the reliability statistics.

A convenience sample of 19 ambulatory surgery centers was selected for a retrospective chart audit comparing the reported values for the measure versus the values identified from the medical record. The centers were located in twelve different states throughout the US. Services from second and third calendar quarter of 2010 were reviewed in the course of the reliability testing. One center was eliminated from the sample due to obvious data collection errors involving the numerator and denominator of the rate. This was a very small ASC (8 patients requiring hair removal) that erroneously excluded seven patients from their reported denominator and reported those same cases as having hair removal in the numerator. The errors were attributed to data entry/transcription errors.

2b.2 **Analytic Method** *(type of reliability & rationale, method for testing)*:
The numerator (number of ASC admissions during the period who received hair with removal clippers or depilatory cream) and denominator (number of ASC admissions requiring hair removal during the period) values were compared for all 18 centers in the validated sample.

2b.3 **Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
The error rates at 16 of the 18 (88.9%) ASCs are zero for both the numerator and denominator. The overall error rate for the numerator and denominator were 0.2% and 0.9% respectively. The median error rates by center were zero for both the numerator and denominator. The results show an excellent level of reliability with an overall median center accuracy rate of 100%.

2c. Validity testing

2c.1 **Data/sample** *(description of data/sample and size)*: Validity testing was performed using a previous numerator definition (ASC admissions with surgical site hair removal with clippers or depilatory cream). The revised numerator statement (ASC admissions with surgical site hair removal with a razor or clippers from the scrotal area, or with clippers or depilatory cream from all other surgical sites) was formulated to harmonize with the current SCIP measure for appropriate hair removal. This change impacts a very small proportion of the ASC admissions and therefore it is unlikely that this change will have a material impact on the validity statistics.

Validity was measured via a formal consensus process. A questionnaire that included ratings of the various characteristics of the measure was distributed to 6 clinicians (RNs) who currently work in ambulatory surgery centers or have responsibility for multiple surgery centers. Two have credentials in quality and the others are involved in quality in their current positions. Responses were received from all 6 of the panel members. Respondents were asked to rate agreement with a series of statements regarding the validity of the measure on a scale from 1 to 5 (5 being the highest level of agreement).

2c.2 **Analytic Method** *(type of validity & rationale, method for testing)*:
Validity was measured via a formal consensus process. Five of the six respondents responded with a 5/5 rating for the question most related to content validity for this measure. Due to the high level of consensus on the primary validity question, multiple rounds of Delphi-type evaluations were not necessary. These results demonstrate a high level of agreement around the validity of the measure.

2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
Each attribute was measured on a 5 point Likert Scale. The attributes related to validity and average scores are listed below:
1. The measure appears to measure what it is intended to. (Median: 5.0/5.0; Mean: 4.7/5.0)
2. The measure is defined in a way that will allow for consistent interpretation of the inclusion and exclusion criteria from center to center. (Median: 5.0/5.0; Mean 4.7/5.0)
3. The data required for the measure are likely to be obtained with reasonable effort. (Median: 4.0/5.0; Mean: 4.0/5.0)
4. The data required for the measure are likely to be obtained with reasonable cost. (Median: 5.0/5.0; Mean:

2c
C☐
P☐
M☐
N☐

4.7/5.0)
5. The data required for the measure can be generated during care delivery. (Median: 5.0/5.0; Mean: 4.8/5.0)

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
The exclusion for this measure (ASC admissions who perform their own hair removal) was developed by expert consensus and reflects the need to focus the measure on the ASC facility´s hair removal processes and practices.

2d.2 Citations for Evidence:
No citations. Please see 2d.1. above, the exclusion for this measure (ASC admissions who perform their own hair removal) was developed by expert consensus and reflects the need to focus the measure on the ASC facility´s hair removal processes and practices.

2d.3 Data/sample *(description of data/sample and size)*: For validity testing, a questionnaire that included ratings of the various characteristics of the measure was distributed to 6 clinicians (RNs) who currently work in ambulatory surgery centers or have responsibility for multiple surgery centers. Two have credentials in quality and the others are involved in quality in their current positions. Responses were received from all 6 of the panel members.

For reliability testing of the exclusion criteria, a convenience sample of 19 ambulatory surgery centers was selected for a retrospective chart audit comparing the reported values for the measure versus the values identified from the medical record. The centers were located in twelve different states throughout the US. Services from second and third calendar quarter of 2010 were reviewed in the course of the reliability testing. One center was eliminated from the sample due to obvious data collection errors involving the numerator and denominator of the rate. This was a very small ASC (8 patients requiring hair removal) that erroneously excluded seven patients from their reported denominator and reported those same cases as having hair removal in the numerator. The errors were attributed to data entry/transcription errors.

Reliability testing was performed using a previous numerator definition (ASC admissions with surgical site hair removal with clippers or depilatory cream). The revised numerator statement (ASC admissions with surgical site hair removal with a razor or clippers from the scrotal area, or with clippers or depilatory cream from all other surgical sites) was formulated to harmonize with the current SCIP measure for appropriate hair removal. This change impacts a very small proportion of the ASC admissions and therefore it is unlikely that this change will have a material impact on the reliability statistics.

2d.4 Analytic Method *(type analysis & rationale)*:
Validity was measured via a formal consensus process. Respondents were asked to rate agreement with the following statement from 1 to 5 (5 being the highest level of agreement): The measure is defined in a way that will allow for consistent interpretation of the inclusion and exclusion criteria from center to center.

For study of reliability, the denominator exclusion (number of ASC admissions performing their own hair removal during the period) values were compared for all 18 centers in the validated sample.

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
Five of the six respondents responded with a 5/5 rating for the question: The measure is defined in a way that will allow for consistent interpretation of the inclusion and exclusion criteria from center to center. The average rating for this question was 4.7 with a median of 5.0. Thus the exclusion criteria were rated highly for validity.

Only two of the 18 ASCs with valid data for the reliability study reported errors in application of the exclusion criteria. The error rates for those sites were 6.7% and 6.9%. The overall error rate in application of the exclusion criteria was 0.3%.

2d
C☐
P☐
M☐
N☐
NA
☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: This measure is not risk adjusted

2e
C☐
P☐

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*: Not applicable **2e.3 Testing Results** *(risk model performance metrics)*: Not applicable **2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:** This process measure does not require risk adjustment. Surgical site hair removal occurs frequently in the ASC setting. The likelihood of appropriate hair removal is not dependent on risk factors based on patient characteristics. Thus we believe this measure should not be risk adjusted.	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)*: The rates for this measure were collected for 192 ambulatory surgery centers throughout the US for services provided during the third calendar quarter of 2010. **2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*: Using the ASC as the unit of analysis, a 95% confidence interval around the mean appropriate surgical site hair with removal clippers or depilatory cream rate of 96% is (94%, 99%). Appropriate hair removal rates below 94% of all patients requiring hair removal would be considered statistically different from the population rate and represents a meaningful difference from the mean compliance rate as well as the gold standard of 100%. **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)*: The rate for appropriate surgical site hair removal ranged from a minimum of 0% to a maximum of 100%. The mean rate was 96% (SD: 18%), while the median rate was 100%. The minimum appropriate hair removal rate of 0% as well as the fact that 7.3% of all centers reported rates lower than 100% demonstrate that there is an opportunity for improvement in this measure.	2f C☐ P☐ M☐ N☐
2g. Comparability of Multiple Data Sources/Methods **2g.1 Data/sample** *(description of data/sample and size)*: This measure is specified for a single data source (paper medical record/flowsheet) as noted in 2a.24. above. The data collection methodology and data source is consistent across ASCs for this measure. **2g.2 Analytic Method** *(type of analysis & rationale)*: The data collection methodology and data source is consistent across ASCs for this measure. **2g.3 Testing Results** *(e.g., correlation statistics, comparison of rankings)*: The data collection methodology and data source is consistent across ASCs for this measure.	2g C☐ P☐ M☐ N☐ NA ☐
2h. Disparities in Care **2h.1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: This measure is not stratified. **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** At the present time, a federal quality reporting system has not yet been proposed or implemented for ambulatory surgical centers. We anticipate that CMS will issue its proposals for an ASC quality reporting system in the near future. When the system is implemented, we anticipate patient level demographic data will be collected in association with ASC data on hair removal practices, allowing for the detection of any	2h C☐ P☐ M☐ N☐ NA ☐

disparities.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, met? **Rationale:**	**2** **C☐** **P☐** **M☐** **N☐**
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Eval Rating**
3a. Meaningful, Understandable, and Useful Information 3a.1 **Current Use:** In use 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** The ASC Quality Collaboration posts a public report of quality data on six ASC quality measures endorsed by the NQF on a quarterly basis. This quarterly report includes aggregated performance data on the Appropriate Surgical Site Hair Removal measure. The report for the third quarter of 2010 is available at: http://www.ascquality.org/qualityreport.cfm. Six hundred seventy-five (675) ASCs submitted data on appropriate surgical site hair removal for the third quarter 2010 report. 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***):** This measure is in use in several other initiatives. For example, the ASC Association includes this metric in its Outcomes Monitoring Project, which is described at http://www.ascassociation.org/outcomes/. It is also in use in various state association quality data collection and reporting projects, including the Texas Ambulatory Surgery Center Association, located at http://tascs.org/. In addition, the measure has been adopted by the Minnesota Department of Health (MDH) for state reporting by ASCs beginning July 2011. This is described at the MDH website at: http://www.health.state.mn.us/healthreform/measurement/adoptedrule/QualityMeasurementAppendices_101129.pdf **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)***:** Interpretability testing was performed using a previous numerator definition (ASC admissions with surgical site hair removal with clippers or depilatory cream). The revised numerator statement (ASC admissions with surgical site hair removal with a razor or clippers from the scrotal area, or with clippers or depilatory cream from all other surgical sites) was formulated to harmonize with the current SCIP measure for appropriate hair removal. This change impacts a very small proportion of the ASC admissions and therefore it is unlikely that this change will have a material impact on the interpretability statistics. Interpretability was measured via a formal consensus process. A questionnaire that included ratings of the various characteristics of the measure was distributed to 6 clinicians (RNs) who currently work in ambulatory surgery centers or have responsibility for multiple surgery centers. Two have credentials in quality and the others are involved in quality in their current positions. Responses were received from all 6 of the panel members. 3a.5 **Methods** *(e.g., focus group, survey, QI project)***:** The survey was summarized to assess the panel's level of agreement with statements that measured the	**3a** **C☐** **P☐** **M☐** **N☐**

interpretability of the measure.

3a.6 **Results** *(qualitative and/or quantitative results and conclusions)*:
Each attribute was measured on a 5 point Likert Scale. The attributes related to usability and average scores are listed below:
1. A provider can understand the results of the measure. (Median: 5.0/5.0; Mean: 5.0/5.0)
2. If necessary, a provider can use the results of the measure to take action. (Median: 5.0/5.0; Mean: 5.0/5.0)
3. This measure has a direct link to improving the outcome and/or process of care. (Median: 4.5/5.0; Mean: 4.3/5.0)

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

3b.1 **NQF # and Title of similar or related measures:**
#0301 Surgery patients with appropriate hair removal

(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?**
#0301, Surgery patients with appropriate hair removal, is designed for hospital use. Certain, but not all, of the measure specifications have been harmonized. The most significant differences and the rationale for these differences are as follows:

The measure specifications do not include patients with no surgical site hair removal. In the aggregate, the most common procedures performed for ASC patients do not involve hair removal. For example, the most commonly performed procedures for Medicare patients in the ASC setting are cataract and after-cataract procedures, endoscopic procedures, and pain management injections. These represent over 75% of the volume of ASC procedures and none require hair removal. Knowing this, we have not included patients with no surgical site hair removal in order to minimize data collection burden for ASC providers.

Identification of the denominator population is not based on ICD-9 procedure codes, as is the case with the hospital-based measure, but rather on patient criteria that can be identified concurrent with the process of care. This was done for two reasons. First, ICD-9 procedural codes are not valid code set in the outpatient setting. Secondly, we seek to minimize provider data collection burden. Procedure codes are assigned after care has been rendered. By avoiding the use of code sets to determine the denominator, we allow providers to determine the target population during the process of care. This is a much more efficient approach, minimizing the amount of personnel and time required for data collection. This efficiency is essential for providers whose non-clinical personnel resources are limited, especially in comparison to hospitals.

Similarly, measure exclusions have been made ASC appropriate and are designed to allow concurrent data collection.

The ASC Quality Collaboration has recently updated the measure specifications to harmonize with the related SCIP hair removal measure by identifying scrotal hair removal with a razor as an appropriate hair removal methodology for that specific site. Neurosurgical cases were not addressed for this ASC measure, as patients with a current traumatic head injury are not treated in ASCs.

3b
C ☐
P ☐
M ☐
N ☐
NA ☐

3c. Distinctive or Additive Value
3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:**
This hair removal measure allows outpatient surgical service providers to measure an important process of care. The measure specifications have been designed to ensure usability and feasibility in the outpatient setting.

5.1 **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:**
As noted above, this measure offers improved efficiency of data collection for ASC providers. Patients in

3c
C ☐
P ☐
M ☐
N ☐
NA ☐

both the numerator and denominator populations can be identified concurrent with the process of care, avoiding the additional cost, resource use and inefficiency that results when these determinations are made retrospectively.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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 care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition)	**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.** Widespread adoption of electronic health records in ambulatory surgical centers would be needed to achieve electronic capture of data elements.	**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.** Experience with this measure and feedback from users indicates that reliability is high. Most errors appear to be the result of human factors, such as data entry errors. The ASC Quality Collaboration is not aware of any unintended consequences as a result of the use of this measure.	**4d** **C☐** **P☐** **M☐** **N☐**
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** The ASC Quality Collaboration has included "Frequently Asked Questions" in the Implementation Guide for the measure to assist users in their implementation of data collection. **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** The measure is designed to allow the possibility of concurrent data collection, which minimizes staff time, effort and cost.	**4e** **C☐** **P☐** **M☐** **N☐**

There are no fees associated with the use of this measure and benchmarking data is publicly available on the ASC Quality Collaboration´s website. 4e.3 **Evidence for costs:** Evidence for costs was gathered using a previous numerator definition (ASC admissions with surgical site hair removal with clippers or depilatory cream). The revised numerator statement (ASC admissions with surgical site hair removal with a razor or clippers from the scrotal area, or with clippers or depilatory cream from all other surgical sites) was formulated to harmonize with the current SCIP measure for appropriate hair removal. This change impacts a very small proportion of the ASC admissions and therefore it is unlikely that this change will have a material impact on the costs of data collection. The survey used for validity and interpretability also asked respondents about the feasibility and cost of collecting data. The following two questions support the premise that the cost to collect this information is reasonable for the ASC. Ques #3. The data required for the measure are likely to be obtained with reasonable effort. (Median: 4.0/5.0; Mean: 4.0/5.0) Ques #4. The data required for the measure are likely to be obtained with reasonable cost. (Median: 5.0/5.0; Mean: 4.7/5.0) 4e.4 **Business case documentation:** Not applicable	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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** ASC Quality Collaboration, 5686 Escondida Blvd S, St. Petersburg, Florida, 33715 Co.2 **Point of Contact** Donna, Slosburg, BSN, LHRM, CASC, donnaslosburg@ascquality.org, 727-867-0072-	
Measure Developer If different from Measure Steward Co.3 **Organization** ASC Quality Collaboration, 5686 Escondida Blvd S, St. Petersburg, Florida, 33715 Co.4 **Point of Contact** Donna, Slosburg, BSN, LHRM, CASC, donnaslosburg@ascquality.org, 727-867-0072-	
Co.5 **Submitter If different from Measure Steward POC** Donna, Slosburg, BSN, LHRM, CASC, donnaslosburg@ascquality.org, 727-867-0072-, ASC Quality Collaboration	
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.**

The ASC Quality Collaboration workgroup members meet via teleconference to develop, critique, and modify candidate measures; to maintain existing measures; and to offer sites willing to participate in testing. No contractors are used.

The following is a list of the individuals (and their affiliation at the time of their participation) serving on the workgroup and contributing to this measure:

AAAHC: Naomi Kuznets, PhD
Ambulatory Surgery Foundation: Debra Stinchcomb, BSN, CASC, David Shapiro, MD,
Sarah Martin, RN, BS, CASC and Marian Lowe
AMSURG: Deby Samuels, Lorri Smith RN, BSN, Linda Brooks-Belli and Kathy Wilson
AOA/HFAP: Monda Shaver, RN, BSN, CPHIT and Susan Lautner, RN, BSN, MSHL
AORN: Bev Kirchner BSN, CNOR, CASC and Bonnie Denholm, RN, MS, CNOR
ASCOA: Ann Geier RN, MS, CNOR, CASC
ASC Quality Collaboration: Donna Slosburg, BSN, LHRM, CASC
HCA: Carol Harbin, RN, BSN, MBA
The Joint Commission: Michael Kulczycki and Kathleen Domzalski
NATIONAL: Rhonda Arnwine and Terry Hawes, RN, BHA
Novamed: Cassandra Speier
NUETERRA: Rachelle Babin RN, BSN and Mary Hibdon, RN
Surgical Care Affiliates: Kim Wood, MD
Symbion: Steve Whitmore and Gina Throneberry RN, MBA, CASC
USPI: David Zarin, MD, Julie Gunderson RN, MM, CPHQ, Clint Chain, RN, BSN and Ann Shimek RN, BSN, CASC

Ad.2 **If adapted, provide name of original measure:** Not adapted
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:** 2008
Ad.7 **Month and Year of most recent revision:** 03, 2011
Ad.8 **What is your frequency for review/update of this measure?** Annually, or more frequently if indicated
Ad.9 **When is the next scheduled review/update for this measure?** 03, 2012

Ad.10 **Copyright statement/disclaimers:** None

Ad.11 -13 **Additional Information web page URL or attachment:**

Date of Submission (*MM/DD/YY*)**:** 03/28/2011

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1536 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Cataracts: Improvement in Patient's Visual Function within 90 Days Following Cataract Surgery
De.2 **Brief description of measure:** Percentage of patients aged 18 years and older who had cataract surgery and had improvement in visual function achieved within 90 days following the cataract 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** Composite measure including existing PQRI measures Measures 191 – 20/40 or better visual acuity within 90 days following cataract surgery and 192 – complications within 30 days of cataract surgery requiring additional surgical procedures, and another new measure: Cataracts: Patient Satisfaction within 90 Days Following Cataract Surgery
De.4 **National Priority Partners Priority Area:** Patient and family engagement De.5 **IOM Quality Domain**: Patient-centered De.6 **Consumer Care Need:** Getting better

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	NQF Staff
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission	**A** Y☐ N☐

A.4 Measure Steward Agreement attached: txNQFMeasureStewardAgreement_020309_Final.pdf	
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, Internal quality improvement Accountability, Payment incentive	**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.1Testing:** 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, Frequently performed procedure, Leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality **1a.2** **1a.3 Summary of Evidence of High Impact:** Cataracts are the leading cause of blindness worldwide and remain an important cause of blindness and visual impairment in the United States, accounting for approximately 50% of visual impairment in adults over the age of 40. Cataracts are the leading cause of treatable blindness among Americans of African descent age 40 and older and are the leading cause of visual impairment among Americans of African, Hispanic/Latino, and European descent. Cataract surgery with IOL implantation was the most frequently performed operation and the single largest expenditure for any Part B surgical procedure in the Medicare program, calculated by Part B procedure codes based on allowed charges. In 2008 (latest year available), payment for cataract was $2.1 billion, which is 1.8% of total allowed charges. **1a.4 Citations for Evidence of High Impact:** 1. Congdon N, O´Colmain B, Klaver CC, et al. Causes and prevalence of visual impairment among adults in the United States. Arch Ophthalmol 2004;122:477-85. 2. Cotter SA, Varma R, Ying-Lai M, et al. Causes of low vision and blindness in adult Latinos: the Los	**1a** C☐ P☐ M☐ N☐

Angeles Latino Eye Study. Ophthalmology 2006;113:1574-82.
3. Centers for Medicare and Medicaid Services. Medicare leading Part B procedure codes based on allowed charges: calendar year 2010. Available at: www.cms.hhs.gov/datacompendium/. Accessed December 10, 2010.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: The benefits are to enhance improvement of visual function of patients receiving cataract surgery. The primary indication for surgery is visual function that no longer meets the patient's needs and for which cataract surgery provides a reasonable likelihood of improved vision.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
This is an outcome of surgery indicator of direct relevance and import to patients, their families and referring providers. The available evidence suggests that cataract surgery achieves this in about 90% of patients. While the potential for improvement is seemingly small, the volume of cataract surgery in the U.S. of over 2.8 million surgeries means that the impact could affect more than 280,000 patients per year. Ideally, performance on this indicator would be as high as possible, with lower rates suggestive of opportunities for improvement.

1b.3 Citations for data on performance gap:
1. Monestam E, Wachtmeister L. Impact of cataract surgery on visual acuity and subjective functional outcomes: a population-based study in Sweden. Eye 1999; 13:711-19.
2. Steinberg EP, Tielsch JM, Schein OD, et al. National study of cataract surgery outcomes. Variation in 4-month postoperative outcomes as reflected in multiple outcome measures. Ophthalmology 1994; 101:1131-40; discussion 1140-1.
3. Lundström M, Brege KG, Florén I, et al. Impaired visual function after cataract surgery assessed using the Catquest questionnaire. J Cataract Refract Surg 2000; 26:101–8.
4. Lum F, Schein O, Schachat AP, et al. Initial two years of experience with the AAO National Eyecare Outcomes Network (NEON) cataract surgery database. Ophthalmology 2000; 107:691-7.
5. Lum F, Schachat AP, Jampel HD. The development and demise of a cataract surgery database. The Joint Commission Journal on Quality Improvement 2202; 28:108-114.
6. Mozaffarieh M, Krepler K, Heinzl H et al. Visual function, quality of life and patient satisfaction after ophthalmic surgery: a comparative study. Ophthalmologica 2004; 218:26-30.

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

1b.5 Citations for data on Disparities:

1b
C☐
P☐
M☐
N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** The multiple components of visual function include central near, intermediate, and distance visual acuity; peripheral vision; visual search; binocular vision; depth perception; contrast sensitivity; perception of color; adaptation; and visual processing speed. Visual function also can be measured in terms of functional disability caused by visual impairment. Many activities of daily living require function of more than one of these visual components. Improved function and quality of life are the treatment outcomes that are most critical and applicable to the patient.

1c.2-3. Type of Evidence: Evidence-based guideline

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
n well-designed observational studies, cataract surgery consistently has been shown to have a significant impact on vision-dependent function; up to 90% of patients undergoing first-eye cataract surgery note

1c
C☐
P☐
M☐
N☐

improvement in functional status and satisfaction with vision. Several studies have reported an association between improved visual function after cataract surgery and an improved health-related quality of life. Visual function plays an important role in physical function and well-being, particularly in terms of mobility. The loss of visual function in the elderly is associated with a decline in physical and mental functioning as well as in independence in activities of daily living, including night-time driving, daytime driving, community activities, and home activities. A long-term (10-year) evaluation of patients in the Blue Mountain Study found that cataract surgery patients had a significant improvement in the mental health domain scores with SF-36 evaluation. Cataract surgery may also improve insomnia.

Visual impairment is an important risk factor for falls and for hip fracture; poor depth perception and decreased contrast sensitivity has been found to increase independently the risk of hip fracture. In a randomized controlled trial, first-eye cataract surgery was found to reduce the rate of falling and fracture over a 12-month period. Similar improvement following second eye surgery has also been confirmed. Visual impairment, in particular a decrease of visual acuity and contrast sensitivity, has been shown to be associated with difficulties in driving. Drivers with visually significant cataracts were 2.5 times more likely to have had an at-fault involvement in a motor vehicle crash over a 5-year period compared with drivers without cataracts. When older adults with cataracts who have undergone surgery are compared with those who did not undergo surgery, motor vehicle crash rates in the 4 to 6 years of follow-up were halved in the surgery group.

One large study found that in visual function assessment pre- and postoperatively, the largest improvements were noted for “driving during the day,” “self-care activities,” and “driving during the night.”

In summary, there are numerous studies showing that physical function, emotional well-being, safety and overall quality of life can be enhanced when visual function is restored by cataract extraction

Improved visual function as a result of cataract surgery includes the following:

The multiple components of visual function include central near, intermediate, and distance visual acuity; peripheral vision; visual search; binocular vision; depth perception; contrast sensitivity; perception of color; adaptation; and visual processing speed.93-95 Visual function also can be measured in terms of functional disability caused by visual impairment. Many activities of daily living require function of more than one of these visual components.

Improved function and quality of life are the treatment outcomes that are most critical and applicable to the patient. In well-designed observational studies, cataract surgery consistently has been shown to have a significant impact on vision-dependent function; up to 90% of patients undergoing first-eye cataract surgery note improvement in functional status and satisfaction with vision. Several studies have reported an association between improved visual function after cataract surgery and an improved health-related quality of life. Visual function plays an important role in physical function and well-being, particularly in terms of mobility. The loss of visual function in the elderly is associated with a decline in physical and mental functioning as well as in independence in activities of daily living, including night-time driving, daytime driving, community activities, and home activities. A long-term (10-year) evaluation of patients in the Blue Mountain Study found that cataract surgery patients had a significant improvement in the mental health domain scores with SF-36 evaluation. Cataract surgery may also improve insomnia.

Visual impairment is an important risk factor for falls and for hip fracture122; poor depth perception and decreased contrast sensitivity has been found to increase independently the risk of hip fracture. In a randomized controlled trial, first-eye cataract surgery was found to reduce the rate of falling and fracture over a 12-month period. Similar improvement following second eye surgery has also been confirmed. Visual impairment, in particular a decrease of visual acuity and contrast sensitivity, has been shown to be associated with difficulties in driving. Drivers with visually significant cataracts were 2.5 times more likely to have had an at-fault involvement in a motor vehicle crash over a 5-year period compared with drivers without cataracts. When older adults with cataracts who have undergone surgery are compared with those who did not undergo surgery, motor vehicle crash rates in the 4 to 6 years of follow-up were halved in the surgery group.

One large study found that in visual function assessment pre- and postoperatively, the largest improvements were noted for “driving during the day,” “self-care activities,” and “driving during the night.”

In summary, there are numerous studies showing that physical function, emotional well-being, safety and overall quality of life can be enhanced when visual function is restored by cataract extraction

Improved visual function as a result of cataract surgery includes the following:

- Better optically corrected vision
- Better uncorrected vision with reduced spectacle dependence
- Increased ability to read or do near work
- Reduced glare

- Improved ability to function in dim levels of light
- Improved depth perception and binocular vision by elimination of anisometropia and achievement of good functional acuity in both eyes
- Improved color vision

Improved physical function as a critical outcome of cataract surgery includes the following:

- Increased ability to perform activities of daily living
- Increased ability to continue or resume an occupation
- Increased mobility (walking, driving)
- Reduced mortality

Improved mental health and emotional well-being as a second critical outcome of cataract surgery includes the following benefits:

- Improved self-esteem and independence
- Increased ability to avoid injury
- Increased social contact and ability to participate in social activities
- Relief from fear of blindness

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
Not rated in guideline because it does not serve as a treatment recommendation

1c.6 Method for rating evidence: The panel rated each recommendation on the strength of evidence in the available literature to support the recommendation made. The "ratings of strength of evidence" also are divided into three levels.
Level I includes evidence obtained from at least one properly conducted, well-designed, randomized controlled trial. It could include meta-analyses of randomized controlled trials.
Level II includes evidence obtained from the following:

- Well-designed controlled trials without randomization
- Well-designed cohort or case-control analytic studies, preferably from more than one center
- Multiple-time series with or without the intervention

Level III includes evidence obtained from one of the following:

- Descriptive studies
- Case reports
- Reports of expert committees/organizations (e.g., PPP panel consensus with peer review)

The I, II and III can also be correlated with the USPSTF system of high, moderate and low.

1c.7 Summary of Controversy/Contradictory Evidence:

1c.8 Citations for Evidence (*other than guidelines***):** 1.Brenner MH, Curbow B, Javitt JC, et al. Vision change and quality of life in the elderly. Response to cataract surgery and treatment of other chronic ocular conditions. Arch Ophthalmol 1993;111:680-5.
2. Sloane ME, Ball K, Owsley C, et al. The Visual Activities Questionnaire: developing an instrument for assessing problems in everyday visual tasks. Technical Digest, Noninvasive Assessment of the Visual System 1992;1:26-9.
3. Datta S, Foss AJ, Grainge MJ, et al. The importance of acuity, stereopsis, and contrast sensitivity for health-related quality of life in elderly women with cataracts. Invest Ophthalmol Vis Sci 2008;49:1-6.
4 Steinberg EP, Tielsch JM, Schein OD, et al. The VF-14. An index of functional impairment in patients with cataract. Arch Ophthalmol 1994;112:630-8.
5. Bilbao A, Quintana JM, Escobar A, et al. Responsiveness and clinically important differences for the VF-14 index, SF-36, and visual acuity in patients undergoing cataract surgery. Ophthalmology 2009;116:418-24.
6. Ishii K, Kabata T, Oshika T. The impact of cataract surgery on cognitive impairment and depressive mental status in elderly patients. Am J Ophthalmol 2008;146:404-9.
7. Lundstrom M, Pesudovs K. Catquest-9SF patient outcomes questionnaire: nine-item short-form Rasch-scaled revision of the Catquest questionnaire. J Cataract Refract Surg 2009;35:504-13.
8. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Visual Activities Questionnaire: assessment of subscale validity for cataract surgery outcomes. J Cataract Refract Surg 2009;35:1961-9.
9. Schein OD, Steinberg EP, Javitt JC, et al. Variation in cataract surgery practice and clinical

outcomes. Ophthalmology 1994;101:1142-52.
10. Mangione CM, Phillips RS, Lawrence MG, et al. Improved visual function and attenuation of declines in health-related quality of life after cataract extraction. Arch Ophthalmol 1994;112:1419-25.
11. Desai P, Minassian DC, Reidy A. National cataract surgery survey 1997-8: a report of the results of the clinical outcomes. Br J Ophthalmol 1999;83:1336-40.
12. McGwin G, Jr, Scilley K, Brown J, Owsley C. Impact of cataract surgery on self-reported visual difficulties: comparison with a no-surgery reference group. J Cataract Refract Surg 2003;29:941-8.
13. Monestam E, Wachtmeister L. Impact of cataract surgery on visual acuity and subjective functional outcomes: a population-based study in Sweden. Eye 1999;13 (Pt 6):711-9.
14. Steinberg EP, Tielsch JM, Schein OD, et al. National study of cataract surgery outcomes. Variation in 4-month postoperative outcomes as reflected in multiple outcome measures. Ophthalmology 1994;101:1131-40; discussion 40-1.
15. Harwood RH, Foss AJ, Osborn F, et al. Falls and health status in elderly women following first eye cataract surgery: a randomised controlled trial. Br J Ophthalmol 2005;89:53-9.
16. Gray CS, Karimova G, Hildreth AJ, et al. Recovery of visual and functional disability following cataract surgery in older people: Sunderland Cataract Study. J Cataract Refract Surg 2006;32:60-6.
17. Lee P, Smith JP, Kington R. The relationship of self-rated vision and hearing to functional status and well-being among seniors 70 years and older. Am J Ophthalmol 1999;127:447-52.
18. Lee PP, Spritzer K, Hays RD. The impact of blurred vision on functioning and well-being. Ophthalmology 1997;104:390-6.
19. Lundstrom M, Fregell G, Sjoblom A. Vision related daily life problems in patients waiting for a cataract extraction. Br J Ophthalmol 1994;78:608-11.
20. Broman AT, Munoz B, Rodriguez J, et al. The impact of visual impairment and eye disease on vision-related quality of life in a Mexican-American population: proyecto VER. Invest Ophthalmol Vis Sci 2002;43:3393-8.
21. Salive ME, Guralnik J, Glynn RJ, et al. Association of visual impairment with mobility and physical function. J Am Geriatr Soc 1994;42:287-92.
22. Foss AJ, Harwood RH, Osborn F, et al. Falls and health status in elderly women following second eye cataract surgery: a randomised controlled trial. Age Ageing 2006;35:66-71.
23. Laforge RG, Spector WD, Sternberg J. The relationship of vision and hearing impairment to one-year mortality and functional decline. J Aging Health 1992;4:126-48.
24. Klein BE, Klein R, Knudtson MD. Lens opacities associated with performance-based and self-assessed visual functions. Ophthalmology 2006;113:1257-63.
25. Chandrasekaran S, Wang JJ, Rochtchina E, Mitchell P. Change in health-related quality of life after cataract surgery in a population-based sample. Eye (Lond) 2008;22:479-84.
26. Asplund R, Ejdervik Lindblad B. The development of sleep in persons undergoing cataract surgery. Arch Gerontol Geriatr 2002;35:179-87.
27. Asplund R, Lindblad BE. Sleep and sleepiness 1 and 9 months after cataract surgery. Arch Gerontol Geriatr 2004;38:69-75.
28. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. N Engl J Med 1988;319:1701-7.
29. De Coster C, Dik N, Bellan L. Health care utilization for injury in cataract surgery patients. Can J Ophthalmol 2007;42:567-72.
30. Felson DT, Anderson JJ, Hannan MT, et al. Impaired vision and hip fracture. The Framingham Study. J Am Geriatr Soc 1989;37:495-500.
31. Cummings SR, Nevitt MC, Browner WS, et al. Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. N Engl J Med 1995;332:767-73.
32. McGwin G, Jr, Chapman V, Owsley C. Visual risk factors for driving difficulty among older drivers. Accid Anal Prev 2000;32:735-44.
33. Owsley C, Stalvey BT, Wells J, et al. Visual risk factors for crash involvement in older drivers with cataract. Arch Ophthalmol 2001;119:881-7.
34. Subzwari S, Desapriya E, Scime G, et al. Effectiveness of cataract surgery in reducing driving-related difficulties: a systematic review and meta-analysis. Inj Prev 2008;14:324-8.
35. Wood JM, Carberry TP. Bilateral cataract surgery and driving performance. Br J Ophthalmol 2006;90:1277-80.
36. Owsley C, Stalvey B, Wells J, Sloane ME. Older drivers and cataract: driving habits and crash risk. J Gerontol A Biol Sci Med Sci 1999;54:M203-11.
37. Owsley C, McGwin G, Jr, Sloane M, et al. Impact of cataract surgery on motor vehicle crash

involvement by older adults. JAMA 2002;288:841-9.
38. Bassett K, Noertjojo K, Nirmalan P, et al. RESIO revisited: visual function assessment and cataract surgery in British Columbia. Can J Ophthalmol 2005;40:27-33.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
Cataract in the Adult Eye, 2005, American Academy of Ophthalmology
Page 9
Function and quality of life are the outcomes of treatment that are most critical and applicable to the patient.
In summary, these studies show that physical function, emotional well-being, safety, and overall quality of life can be enhanced when visual function is restored by cataract extraction.

1c.10 Clinical Practice Guideline Citation: American Academy of Ophthalmology. Cataract in the Adult Eye, Preferred Practice Pattern. San Francisco: American Academy of Ophthalmology, 2006. Available at: www.aao.org/ppp.
1c.11 National Guideline Clearinghouse or other URL:
http://www.guideline.gov/content.aspx?id=10173&search=cataract+and+cataract+2005+and+cataract+2006

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

1c.13 Method for rating strength of recommendation (*If different from USPSTF system, also describe rating and how it relates to USPSTF)*:
The panel rated each recommendation according to its importance to the care process. This "importance to the care process" rating represents care that the panel thought would improve the quality of the patient's care in a meaningful way. The ratings of importance are divided into three levels.
- Level A, defined as most important
- Level B, defined as moderately important
- Level C, defined as relevant but not critical

The A, B, C ratings can be correlated with the USPSTF system of A, B, C for strength of recommendation.

1c.14 Rationale for using this guideline over others:
This guideline is the only United States guideline on cataract surgery contained in the National Guideline Clearinghouse.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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***):**
Patients who had improvement in visual function achieved within 90 days following cataract surgery

2a-specs
C☐
P☐
M☐
N☐

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator*):
One year

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions*):
Reporting Numerator includes each of the following instances:
A. Patients who had an improvement in their visual function achieved within 90 days following cataract surgery
C. Patients who did not complete their visual function assessment within 90 days following cataract surgery but for whom there is a documented medical or patient reason for not doing so
D. Patients who did not have an improvement in their visual function achieved within 90 days following cataract surgery and there is no documented medical or patient reason for not doing so

For the reporting calculation, documented medical and patient reasons for not doing so include the following:
Medical reasons:
When cataract surgery was performed for these indications:
- Clinically significant anisometropia in the presence of a cataract
- The lens opacity interferes with optimal diagnosis or management of posterior segment conditions
- The lens causes inflammation (phacolysis, phacoanaphylaxis)
- The lens induces angle closure (phacomorphic or phacotopic)

Patient reasons:
- The patient refuses to participate
- The patient is unable to complete the questionnaire

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured*):
All patients aged 18 years and older who had cataract surgery

2a.5 Target population gender: Female, Male
2a.6 Target population age range: 18 years and older

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator*):
One year

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions*):
Denominator (Eligible Population): All patients aged 18 years and older who had cataract surgery
- CPT Procedure Codes (with or without modifiers): 66840, 66850, 66852, 66920, 66930, 66940, 66982, 66983, 66984

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population*): A patient is excluded if the following condition(s) exist:
Medical reasons:
When cataract surgery was performed for these indications:
- Clinically significant anisometropia in the presence of a cataract
- The lens opacity interferes with optimal diagnosis or management of posterior segment conditions
- The lens causes inflammation (phacolysis, phacoanaphylaxis)
- The lens induces angle closure (phacomorphic or phacotopic)

The rationale for these medical exclusions is that cataract surgery is being performed for a medical reason other than improvement of visual function impaired due to cataract; either for visualization of posterior structures of the eye or to provide relief of aggravation of other conditions such as inflammation or angle closure. In these situations, therefore, an improvement of visual acuity and visual function would not be expected, and cataract surgery is not undertaken with this purpose in mind.

Patient reasons:
- The patient refuses to participate
- The patient is unable to complete the questionnaire

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
Denominator Exclusions: Documentation of medical reason for not improving visual function within 90 days of cataract surgery
- Append modifier to CPT Category II Code: -1P

Documentation of patient reason for not improving visual function within 90 days of cataract surgery
- Append modifier to CPT Category II Code: -2P

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
This measure can be stratified into two major groups: those patients with ocular co-morbidities and those patients without ocular co-morbidities. An improvement in visual function after cataract surgery would be expected in both groups, however the magnitude of the difference would vary by group. The Cataract Patient Outcomes Research Team found that an important preoperative patient characteristic that was independently associated with failure to improve on one of the outcomes measured (including the VF-14) was ocular comorbidity. The authors explained that this was expected, because it is reasonable to assume that other diseases that impair visual function would be correlated with a reduced improvement in functional status. The National Eye Care Outcomes Network also found that there were differences in the mean postooperative VF-14 scores across groups of patients with and without ocular co-morbidities, as seen in the table below. The study involving the Rasch-scaled short version of the VF-14 also found differences between the preoperative and postoperative visual function test scores and differences between preoperative and postoperative visual function tests, as seen below.

National Eyecare Outcomes Network

Mean VF-14 (postoperative)
- Total 92.7
- With ocular comorbidity 89.9
- Without ocular comorbidity 94.6

Rasch-Scaled Short Version of the VF-14

Patients without Ocular Comorbidity - Preop VF-8R - 68.87
Postop VF-8R - 86.22
Mean Diff = 17.35
Patients with Ocular Comorbidity - Preop VF-8R - 67.71
Postop VF-8R - 81.58
Mean Diff = 13.87

A list of codes for comorbidities can be found in the AMA PCPI measure for 20/40 visual acuity after cataract surgery:

Acute and subacute iridocyclitis 364.00
Acute and subacute iridocyclitis 364.01
Acute and subacute iridocyclitis 362.02
Acute and subacute iridocyclitis 364.03
Acute and subacute iridocyclitis 364.04
Acute and subacute iridocyclitis 364.05
Amblyopia 368.01
Amblyopia 368.02
Amblyopia 368.03
Burn confined to eye and adnexa 940.0
Burn confined to eye and adnexa 940.1
Burn confined to eye and adnexa 940.2
Burn confined to eye and adnexa 940.3

Burn confined to eye and adnexa	940.4
Burn confined to eye and adnexa	940.5
Burn confined to eye and adnexa	940.9
Cataract secondary to ocular disorders	366.32
Cataract secondary to ocular disorders	366.33
Certain types of iridocyclitis	364.21
Certain types of iridocyclitis	364.22
Certain types of iridocyclitis	364.23
Certain types of iridocyclitis	364.24
Certain types of iridocyclitis	364.3
Choroidal degenerations	363.43
Choroidal detachment	363.72
Choroidal hemorrhage and rupture	363.61
Choroidal hemorrhage and rupture	363.62
Choroidal hemorrhage and rupture	363.63
Chorioretinal scars	363.30
Chorioretinal scars	363.31
Chorioretinal scars	363.32
Chorioretinal scars	363.33
Chorioretinal scars	363.35
Chronic iridocyclitis	364.10
Chronic iridocyclitis	364.11
Cloudy cornea	371.01
Cloudy cornea	371.02
Cloudy cornea	371.03
Cloudy cornea	371.04
Corneal edema	371.20
Corneal edema	371.21
Corneal edema	371.22
Corneal edema	371.23
Corneal edema	371.43
Corneal edema	371.44
Corneal opacity and other disorders of cornea	371.00
Corneal opacity and other disorders of cornea	371.03
Corneal opacity and other disorders of cornea	371.04
Degenerative disorders of globe	360.20
Degenerative disorders of globe	360.21
Degenerative disorders of globe	360.23
Degenerative disorders of globe	360.24
Degenerative disorders of globe	360.29
Degeneration of macula and posterior pole	362.50
Degeneration of macula and posterior pole	362.51
Degeneration of macula and posterior pole	362.52
Degeneration of macula and posterior pole	362.53
Degeneration of macula and posterior pole	362.54
Degeneration of macula and posterior pole	362.55
Degeneration of macula and posterior pole	362.56
Degeneration of macula and posterior pole	362.57
Disseminated chorioretinitis and disseminated retinochoroiditis	363.10
Disseminated chorioretinitis and disseminated retinochoroiditis	363.11
Disseminated chorioretinitis and disseminated retinochoroiditis	363.12
Disseminated chorioretinitis and disseminated retinochoroiditis	363.13
Disseminated chorioretinitis and disseminated retinochoroiditis	363.14
Disseminated chorioretinitis and disseminated retinochoroiditis	363.15
Diabetic retinopathy	362.01
Diabetic retinopathy	362.02
Diabetic retinopathy	362.03
Diabetic retinopathy	362.04

Diabetic retinopathy	362.05
Diabetic retinopathy	362.06
Diabetic macular edema	362.07
Disorders of optic chiasm	377.51
Disorders of optic chiasm	377.52
Disorders of optic chiasm	377.53
Disorders of optic chiasm	377.54
Disorders of visual cortex	377.75
Focal chorioretinitis and focal retinochoroiditis	363.00
Focal chorioretinitis and focal retinochoroiditis	363.01
Focal chorioretinitis and focal retinochoroiditis	363.03
Focal chorioretinitis and focal retinochoroiditis	363.04
Focal chorioretinitis and focal retinochoroiditis	363.05
Focal chorioretinitis and focal retinochoroiditis	363.06
Focal chorioretinitis and focal retinochoroiditis	363.07
Focal chorioretinitis and focal retinochoroiditis	363.08
Glaucoma	365.10
Glaucoma	365.11
Glaucoma	365.12
Glaucoma	365.13
Glaucoma	365.14
Glaucoma	365.15
Glaucoma	365.20
Glaucoma	365.21
Glaucoma	365.22
Glaucoma	365.23
Glaucoma	365.24
Glaucoma	365.31
Glaucoma	365.32
Glaucoma	365.51
Glaucoma	365.52
Glaucoma	365.59
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.41
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.42
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.43
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.44
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.60
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.61
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.62
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.63
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.64
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.65
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.81
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.82
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.83
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.89
Glaucoma associated with congenital anomalies, dystrophies, and systemic syndromes	365.9
Hereditary corneal dystrophies	371.50
Hereditary corneal dystrophies	371.51
Hereditary corneal dystrophies	371.52
Hereditary corneal dystrophies	371.53
Hereditary corneal dystrophies	371.54
Hereditary corneal dystrophies	371.55
Hereditary corneal dystrophies	371.56
Hereditary corneal dystrophies	371.57
Hereditary corneal dystrophies	371.58
Hereditary choroidal dystrophies	363.50
Hereditary choroidal dystrophies	363.51

Hereditary choroidal dystrophies	363.52
Hereditary choroidal dystrophies	363.53
Hereditary choroidal dystrophies	363.54
Hereditary choroidal dystrophies	363.55
Hereditary choroidal dystrophies	363.56
Hereditary choroidal dystrophies	363.57
Hereditary retinal dystrophies	362.70
Hereditary retinal dystrophies	362.71
Hereditary retinal dystrophies	362.72
Hereditary retinal dystrophies	362.73
Hereditary retinal dystrophies	362.74
Hereditary retinal dystrophies	362.75
Hereditary retinal dystrophies	362.76
High myopia	360.20
High myopia	360.21
Injury to optic nerve and pathways	950.0
Injury to optic nerve and pathways	950.1
Injury to optic nerve and pathways	950.2
Injury to optic nerve and pathways	950.3
Injury to optic nerve and pathways	950.9
Keratitis	370.03
Moderate or severe impairment, better eye, profound impairment lesser eye	369.10
Moderate or severe impairment, better eye, profound impairment lesser eye	369.11
Moderate or severe impairment, better eye, profound impairment lesser eye	369.12
Moderate or severe impairment, better eye, profound impairment lesser eye	369.13
Moderate or severe impairment, better eye, profound impairment lesser eye	369.14
Moderate or severe impairment, better eye, profound impairment lesser eye	369.15
Moderate or severe impairment, better eye, profound impairment lesser eye	369.16
Moderate or severe impairment, better eye, profound impairment lesser eye	369.17
Moderate or severe impairment, better eye, profound impairment lesser eye	369.18
Nystagmus and iother irregular eye movements	379.51
Open wound of eyeball	871.0
Open wound of eyeball	871.1
Open wound of eyeball	871.2
Open wound of eyeball	871.3
Open wound of eyeball	871.4
Open wound of eyeball	871.5
Open wound of eyeball	871.6
Open wound of eyeball	871.7
Open wound of eyeball	871.9
Optic atrophy	377.10
Optic atrophy	377.11
Optic atrophy	377.12
Optic atrophy	377.13
Optic atrophy	377.14
Optic atrophy	377.15
Optic atrophy	377.16
Optic neuritis	377.30
Optic neuritis	377.31
Optic neuritis	377.32
Optic neuritis	377.33
Optic neuritis	377.34
Optic neuritis	377.39
Other background retinopathy and retinal vascular changes	362.12
Other background retinopathy and retinal vascular changes	362.16
Other background retinopathy and retinal vascular changes	362.18
Other corneal deformities	371.70
Other corneal deformities	371.71

Other corneal deformities	371.72
Other corneal deformities	371.73
Other disorders of optic nerve	377.41
Other disorders of sclera	379.11
Other disorders of sclera	379.12
Other endophthalmitis	360.11
Other endophthalmitis	360.12
Other endophthalmitis	360.13
Other endophthalmitis	360.14
Other endophthalmitis	360.19
Other retinal disorders	362.81
Other retinal disorders	362.82
Other retinal disorders	362.83
Other retinal disorders	362.84
Other retinal disorders	362.85
Other retinal disorders	362.89
Other and unspecified forms of chorioretinitis and retinochoroiditis	363.20
Other and unspecified forms of chorioretinitis and retinochoroiditis	363.21
Other and unspecified forms of chorioretinitis and retinochoroiditis	363.22
Prior penetrating keratoplasty	371.60
Prior penetrating keratoplasty	371.61
Prior penetrating keratoplasty	371.62
Profound impairment, both eyes	369.00
Profound impairment, both eyes	369.01
Profound impairment, both eyes	369.02
Profound impairment, both eyes	369.03
Profound impairment, both eyes	369.04
Profound impairment, both eyes	369.05
Profound impairment, both eyes	369.06
Profound impairment, both eyes	369.07
Profound impairment, both eyes	369.08
Purulent endophthalmitis	360.00
Purulent endophthalmitis	360.01
Purulent endophthalmitis	360.02
Purulent endophthalmitis	360.03
Purulent endophthalmitis	360.04
Retinal detachment with retinal defect	361.00
Retinal detachment with retinal defect	361.01
Retinal detachment with retinal defect	361.02
Retinal detachment with retinal defect	361.03
Retinal detachment with retinal defect	361.04
Retinal detachment with retinal defect	361.05
Retinal detachment with retinal defect	361.06
Retinal detachment with retinal defect	361.07
Retinal vascular occlusion	362.31
Retinal vascular occlusion	362.32
Retinal vascular occlusion	362.35
Retinal vascular occlusion	362.36
Retinopathy of prematurity	362.21
Scleritis and episcleritis	379.04
Scleritis and episcleritis	379.05
Scleritis and episcleritis	379.06
Scleritis and episcleritis	379.07
Scleritis and episcleritis	379.09
Separation of retinal layers	362.41
Separation of retinal layers	362.42
Separation of retinal layers	362.43
Uveitis	360.11

Uveitis 360.12
Visual field defects 368.41

References:
1. Schein OD, Steinberg EP, Cassard SD et al. Predictors of outcome in patients who underwent cataract surgery. Ophthalmology 1995; 102:817-23.
2. Lum F, Schachat AP, Jampel HD.The development and demise of a cataract surgery database. Jt Comm J Qual Improv. 2002 Mar;28(3):108-14.
3. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Measuring outcomes of cataract surgery using the Visual Function Index-14. J Cataract Refract Surg 2010; 36:1181-8.

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
A risk adjustment methodology is not necessary if the stratification schema is utilized, as described above.

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
Calculation for Reporting:

For reporting purposes, this measure is calculated by creating a fraction with the following components: Reporting Numerator and Reporting Denominator.

Reporting Numerator includes each of the following instances:

A. Patients who had an improvement in their visual function achieved within 90 days following cataract surgery

C. Patients who did not complete their visual function assessment within 90 days following cataract surgery but for whom there is a documented medical or patient reason for not doing so

D. Patients who did not have an improvement in their visual function achieved within 90 days following cataract surgery and there is no documented medical or patient reason for not doing so

Reporting Denominator (RD) includes:
- Patients aged 18 years and older AND
- Had cataract surgery

Reporting Calculation
A (# of patients meeting measure criteria) + C (# of patients with valid exclusions) + D (# of patients NOT meeting numerator criteria)
_____

RD (# of patients in denominator)

A (# of patients meeting measure criteria)
A (A
PD (# of patients in denominator)

Components for this measure are defined as:
A # of patients who had an improvement in their visual function achieved within 90 days following cataract surgery

C # of patients who did not complete their visual function assessment within 90 days following cataract surgery but for whom there is a documented medical or patient reason for not doing so
D # of patients who did not have an improvement in their visual function achieved within 90 days following cataract surgery and there is no documented medical or patient reason for not doing so
RD # of patients aged 18 years and older who had cataract surgery

2a.22 Describe the method for discriminating performance *(e.g., significance testing)*:
Methods would include comparison of means and percentiles, and analysis of variance against established benchmarks in the literature.

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 this physician-level measure, it is anticipated to be used as a group or composite measure. Utilizing a sample, work in the field has indicated that a sample size of 30 patients would be adequate for typical practice sizes.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Survey: Patient

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.***):**
The data collection instrument is specified as an assessment tool that has been appropriately validated for the population for which it being used. Examples of tools for visual function assessment include, but are not limited to: National Eye Institute-Visual Function Questionnaire (VFQ), the Visual Function (VF)-14, the modified VF-8, the Activities of Daily Vision Scale (ADVS), the Catquest and the modified Catquest-9. For this measure, we are proposing the Rasch-scaled short version of the VF-14, otherwise referred to as the VF-8R hereafter.

2a.26-28 Data source/data collection instrument reference web page URL or attachment: Attachment VF8 Pesudovs.pdf

2a.29-31 Data dictionary/code table web page URL or attachment:

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Clinicians: Individual

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

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: There are several validated instruments to measure visual function that are available for use. We are proposing use of one such instrument, the Rasch-scaled Short Version of the VF-14 is described here for which reliability and validity testing have been performed. The VF-14 is a health status measurement listed by the Agency for Healthcare Research and Quality (www.ahrq.gov/clinic/out2res/outcom5.htm#/) as an instrument tested for reliability and validity in their Patient Outcomes Research Team studies and identified as a discriminative and evaluative health status measurement instrument. If there is greater detail needed on the reliability and validity testing of the VF-14 itself, please let us know. References are listed below.

In the following, we describe the testing performed on the Rasch-scaled Short Version, otherwise referred to as the VF-8R. In this study, the purpose was to determine which version of the Visual Function Index-14 most precisely measured cataract surgery outcomes, to rescale the VF-14 using Rasch analysis and to create a short-form version. Participants were selected from the cataract surgery waiting list at the Flinders

2b
C ☐
P ☐
M ☐
N ☐

Medical Centre, Adelaide, Australia. All patients had cataract surgery performed using phacoemulsification with intraocular lens placement. The eligibility criteria were age 18 years or older, ability to provide written informed consent, and English-speaking. There were two patient populations. The first cohort were preoperative cataract patients, whose data were used for the Rasch analysis to refine the VF-14, called the development group. The second cohort were patients whose results were used to measure the outcomes of cataract surgery, called the outcomes group. The instrument was mailed to 414 patients, of whom 210 returned the completed questionnaire preoperatively (development group), and 51 of the 81 patients postoperatively returned the questionnaire (outcomes group). In the development group (n= 210), the mean age was 74.3 years, 42% were male, and 58% were female, 48% had a ocular comorbidity and 84% had a systemic comorbidity. In the outcomes group (n = 51), the mean age was 73.0 years, 57% were male and 43% were female, 59% had ocular comorbidity, and 78% had a systemic comorbidity.

The reference for the visual function instrument described here (VF-8R)is:

1. Gothwal VK, Wright TA, Lamoureux EL, and Pesudovs K. Measuring outcomes of cataract surgery using the Visual Function Index-14. J Cataract Refract Surg 2010; 36:1181-1188.

A reference describing more of the Rasch analysis is:

1. Lamoureux EL, Pesudovs K, Thumboo J, Saw S-M, and Wong T.Y. An evaluation of the reliability and validity of the Visual Functioning Questionnaire (VF-11) Using Rasch Analysis in an Asian population. Invest Ophthalmol Vis Sci 2009; 50:2607-13.

Original references for the VF-14 include:

1. Steinberg EP, Tielsch JM, Schein OD, Javitt JC, Sharkey P, Cassard SD, Legro MW, Diener-West M, Bass EB, Damiano AM, et al. The VF-14. An index of functional impairment in patients with cataract. Arch Ophthalmol. 1994 May;112(5):630-8.1.

2. Cassard SD, Patrick DL, Damiano AM, Legro MW, Tielsch JM, Diener-West M, Schein OD, Javitt JC, Bass EB, Steinberg EP. Reproducibility and responsiveness of the VF-14. An index of functional impairment in patients with cataracts. Arch Ophthalmol. 1995 Dec;113(12):1508-13.

3. Schein OD, Steinberg EP, Cassard SD, Tielsch JM, Javitt JC, Sommer A. Predictors of outcome in patients who underwent cataract surgery. Ophthalmology. 1995 May;102(5):817-23.

4. Damiano AM, Steinberg EP, Cassard SD, Bass EB, Diener-West M, Legro MW, Tielsch J, Schein OD, Javitt J, Kolb M. Comparison of generic versus disease-specific measures of functional impairment in patients with cataract. Med Care. 1995 Apr;33(4 Suppl):AS120-30.

5. Steinberg EP, Tielsch JM, Schein OD, Javitt JC, Sharkey P, Cassard SD, Legro MW, Diener-West M, Bass EB, Damiano AM, et al. National study of cataract surgery outcomes. Variation in 4-month postoperative outcomes as reflected in multiple outcome measures. Ophthalmology. 1994 Jun;101(6):1131-40; discussion 1140-1.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
In summary, Rasch analysis was used to re-define the VF-14 into two valid forms, the VF-11R and VF-8R form. Then, the ability of the different versions of the VF-14 to discriminate outcomes of cataract surgery was compared with the standard VF-14, using the relative precision method.

Rasch analysis: The Rasch model, where the total score summarizes completely a person´s standing on a variable, arises from a more fundamental requirement: that the comparison of two people is independent of which items may be used within the set of items assessing the same variable. Thus the Rasch model is taken as a criterion for the structure of the responses, rather than a mere statistical description of the responses. For example, the comparison of the performance of two students´ work marked by different graders should be independent of the graders.

In this case it is considered that the researcher is deliberately developing items that are valid for the

purpose and that meet the Rasch requirements of invariance of comparisons.

Analyzing data according to the Rasch model, that is, conducting a Rasch analysis, gives a range of details for checking whether or not adding the scores is justified in the data. This is called the test of fit between the data and the model. If the invariance of responses across different groups of people does not hold, then taking the total score to characterize a person is not justified. Of course, data never fit the model perfectly, and it is important to consider the fit of data to the model with respect to the uses to be made of the total scores. If the data do fit the model adequately for the purpose, then the Rasch analysis also linearises the total score, which is bounded by 0 and the maximum score on the items, into measurements. The linearised value is the location of the person on the unidimensional continuum - the value is called a parameter in the model and there can be only one number in a unidimensional framework. This parameter can then be used in analysis of variance and regression more readily than the raw total score which has floor and ceiling effects. Relative precision is a ratio of pairwise F statistics. The extent to which the relative precision ratio differs from 1.0 indicates the extent to which scoring methods differed in their ability to detect change in scores; values greater than 1.0 indicate an increase in precision.

2b.3 **Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Results for the VF-8R: Mean item location = 0; mean person location = 1.97 and principal components analysis (eigenvalue) = 1.6; relative precision to the VF-14 = 2.25;

Results for the VF-14: (based on 552 patients who underwent cataract surgery in one eye and completed a 4 month postoperative survey) Highly reproducible,w ith an intraclass correlation coefficient of 0.79 when patient-rated criteria were used to define stable patients.

2c. Validity testing

2c.1 **Data/sample** *(description of data/sample and size)*: The VF-14 was mailed to 414 patients, of whom 210 returned the completed questionnaire, and 51 returned the VF-15 postoperatively. The mean age of the patients submitting preoperative VF-14 scores was 74.3 years. In this group, 42% were male, and 58% were female, 48% had a ocular comorbidity and 84% had a systemic comorbidity.

2c.2 **Analytic Method** *(type of validity & rationale, method for testing)*:
Content validity was evaluated by using person and item fit residual statistics. It is expected that the mean and SD values approximate 0 and 1, respectively. An estimate of overall scale functioning is the person separation reliability (PSR) index. This is linked to the targeting of the scale, because it differentiates the number of statistically distinct groups of respondents that can be identified by this trait. In other words, this can demonstrate if an instrument can discriminate among different levels of the patient´s visual functioning.

Also, ANOVA was used to see if the change in preoperative to postoperative score for the original VF-14 and the shortened version differed significantly from zero. The F statistic with a $P < 0.05$ was then considered significant. Then relative precision as described above was used to evaluate how well the different versions of VF-14 discriminated between visual functioning in the preoperative period compared with the postoperative period.

2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
Person separation = 2.29 (the minimum acceptable value is 2.0) ; Misfitting items = 0; (ideal value = 0)

Overall, the VF-8R showed the following results for cataract surgery patients:

Mean preoperative score and standard error - 67.75, SE = 2.36
Mean postoperative score and standard error - 83.15, SE = 2.43
Mean difference preop vs. postop and standard error - 15.39, SE = 2.66
F statistic 20.67
Relative precision 2.25

The overall results of the testing found these benefits of using the VF-8R over the original VF-14: 1) all

2c
C ☐
P ☐
M ☐
N ☐

items measure a single construct of visual functioning, which is a desirable measurement property and unlike the original VF-14 which has more than a single construct; 2) it has better measurement precision for distinguishing outcomes (125% gain in relative precision) than the original VF-14; 3) it has other similar psychometric properties to the original VF-14. Testing Results for the VF-14 (from the original VF-14 publications): (based on 552 patients who underwent cataract surgery in one eye and completed a 4 month postoperative survey): high internal consistency with a Cronbach´s a = 0.85, with item-to-total correlations ranging from 0.32 to 0.61. It was also found to be three times more responsive to a change in vision than a generic health status measure (Sickness Impact Profile) with an impact size of approximately 1.00 to 0.30, respectively. The criterion validity was assessed by examining the correlation between the VF-14 scores and several other measures of vision. The correlation between the VF-14 score and self-reported trouble with vision and overall satisfaction with vision (0.45 and 0.34, respectively) were higher than correlations between several measures of visual acuity and trouble or satisfaction with vision.	
2d. Exclusions Justified **2d.1 Summary of Evidence supporting exclusion(s):** Other indications for a cataract removal include the following: - Clinically significant anisometropia in the presence of a cataract.[A:III] - The lens opacity interferes with optimal diagnosis or management of posterior segment conditions.[A:III] - The lens causes inflammation (phacolysis, phacoanaphylaxis).[A:III] - The lens induces angle closure (phacomorphic or phacotopic).[A:III] The rationale for these medical exclusions is that cataract surgery is being performed for a medical reason other than improvement of visual function impaired due to cataract; either for visualization of posterior structures of the eye or to provide relief of aggravation of other conditions such as inflammation or angle closure. In these situations, therefore, an improvement of visual acuity and visual function would not be expected, and cataract surgery is not undertaken with this purpose in mind. **2d.2 Citations for Evidence:** American Academy of Ophthalmology. Cataract in the Adult Eye, Preferred Practice Pattern. San Francisco: American Academy of Ophthalmology, 2006. Available at: www.aao.org/ppp. **2d.3 Data/sample** *(description of data/sample and size)*: **2d.4 Analytic Method** *(type analysis & rationale)*: **2d.5 Testing Results** *(e.g., frequency, variability, sensitivity analyses)*:	**2d** C☐ P☐ M☐ N☐ NA☐
2e. Risk Adjustment for Outcomes/ Resource Use Measures **2e.1 Data/sample** *(description of data/sample and size)*: There is no risk adjustment strategy necessary given that a stratification of results is proposed. **2e.2 Analytic Method** *(type of risk adjustment, analysis, & rationale)*: **2e.3 Testing Results** *(risk model performance metrics)*: **2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:**	**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)*: The VF-14 was mailed to 414 patients, of whom 210 returned the completed questionnaire, and 51 returned the VF-15	**2f** C☐ P☐ M☐

postoperatively. The mean age of the patients submitting preoperative VF-14 scores was 74.3 years. In this group, 42% were male, and 58% were female, 48% had a ocular comorbidity and 84% had a systemic comorbidity.

2f.2 **Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*:
The VF-8 preoperative and postoperative scores for patients with ocular comorbidity (30) and for patients without ocular comorbidity (20) were compared in terms of mean scores and standard errors.

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)*:
The group with ocular comorbidity had a mean preoperative and postoperative + SE score of 67.71 + 3.29 and 81.58 + 3.57, respectively. The mean difference preop vs. postop was 13.87 + 3.81. The F Statistic was 8.15. The group without ocular comorbidity had a mean preoperative and postoperative + SE score of 68.87 + 3.36 and 86.22 + 3.03, respectively. The mean difference preop vs. postop was 17.35 + 3.72 and the F Statistic was 14.70.

N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 **Data/sample** *(description of data/sample and size)*: The VF-14 can be interviewed-administered, and self-administered. There don´t appear to be peer-reviewed reports comparing the interviewed-administered and the self-administered versions of the VF-14. However, there are at least two peer-reviewed reports demonstrating the validity and responsiveness of the self-administered VF-14 in the literature.

One study evaluated the validity and responsiveness of two self-administered instruments, the VF-14 and the Quality of Well-Being Scale. This was performed in 233 adults who had small-incision phacoemulsification cataract surgery in a Southern California Health Maintenance Organization. The mean age of patients was 72.5 years old, and 60.5% were men. Approximately 50% of the patients had ocular morbidities and 82% had at least one chronic illness.

A second study tested the validity of the self-administered VF-14 in a group of patients with retinal disease. The patient population were 547 patients attending the Vancouver General Hospital Eye Care Centre. 48% were female and 52% were male. The mean age of the group was 55 years, ranging from 16 to 95 years old.

References
1. Rosen PN, Kaplan Rn, David K. Measuring outcomes of cataract surgery using the Quality of Well-Being Scale and VF-14 Visual Function Index. J Cataract Refract Surg 2005; 31:369-78.
2. Linder M, Chang TS, Scott IU et al. Validity of the Visual Function Index (VF-14) in Patients with Retinal Disease. Arch Ophthalmol 1999; 117:1611-16.

2g.2 **Analytic Method** *(type of analysis & rationale)*:
One study evaluated the validity and responsiveness of two self-administered instruments, the VF-14 and the Quality of Well-Being Scale. Bivariate analysis was performed on the effect of cataract surgery on the VF-14 score using Pearson correlations and independent and paired t tests. One-way analysis of variance was used to test the VF-14 in discriminating between categories of satisfaction and toruble with vision.

A second study tested the validity of the self-administered VF-14 in a group of patients with retinal disease. Criterion validity was evaluated through measurement of the Spearman correlation coefficients between VF-14 score and the global self-assessments scales within the VF-14: amount of trouble with vision, level of satisfaction with vision and overall quality of vision. Also, the Spearman correlations between the VF-14 score and the global scores were compared with the correlation of visual acuity scores and the global scales.

2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)*:
One study evaluated the validity and responsiveness of two self-administered instruments, the VF-14 and the Quality of Well-Being Scale. The VF-14 was found to correlate significantly with self-reported

2g
C☐
P☐
M☐
N☐
NA☐

satisfaction and trouble with vision, and responsive to improvements in quality of life after cataract surgery. The postoperative correlations of the VF-14 were as follows: Trouble with vision r = .520 (p<.01) Self vision rating r = .497 (p<.01) Satisfaction with vision r = .462 (p<.01) Satisfaction with surgery result r = .460 (p<.01) Visual symptoms r = .465 (p<.01) Visual acuity of operated eye r = .157 (p<.05) A second study tested the validity of the self-administered VF-14 in a group of patients with retinal disease. The Cronbach alpha coefficient for the sample was 0.91, indicating high internal consistency. The results showed that the VF-14 had a moderately strong association with patient self-rating of the amount of trouble with vision, satisfaction with vision and overall quality of vision. This was stronger than the associations found with a more general health status instrument, the Short-Form Health Survey. The VF-14 was also correlated with visual acuity. The correlations were as follows: VF-14 score - Visual acuity better eye -0.34 (p= .001) Visual acuity worse eye -0.43 (p= .001) Average visual acuity -0.45 (p= .001) WMAR (weighted average logMar) visual acuity -0.45 (p = .001) Overall quality of vision scale 0.50 (p = .001) Satisfaction with vision scale 0.43 (p = .001) Trouble with vision scale -0.63 (p = .001)	
2h. Disparities in Care **2h.1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: The stratified results are as follows: Rasch-Scaled Short Version of the VF-14 Results by Stratification Group with Ocular Comorbidity: The group with ocular comorbidity had a mean preoperative and postoperative + SE score of 67.71 + 3.29 and 81.58 + 3.57, respectively. The mean difference preop vs. postop was 13.87 + 3.81. The F Statistic was 8.15. Group without Ocular Comorbidity: The group without ocular comorbidity had a mean preoperative and postoperative + SE score of 68.87 + 3.36 and 86.22 + 3.03, respectively. The mean difference preop vs. postop was 17.35 + 3.72 and the F Statistic was 14.70. **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:**	**2h** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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:** Not in use but testing completed **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** **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*)**:** **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)***:** **3a.5 Methods** *(e.g., focus group, survey, QI project)***:** **3a.6 Results** *(qualitative and/or quantitative results and conclusions)***:**	**3a** C☐ P☐ M☐ N☐
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:** **5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:**	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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** C☐

4a.1-2 **How are the data elements that are needed to compute measure scores generated?** Survey	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 web-based survey instrument could be used and results uploaded into a data registry. Paper survey instruments could be scanned and incorporated into a data registry.	**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.**	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation 4e.1 **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** There is a burden upon the office practice to survey patients pre and post cataract surgery. The majority of these patients are elderly, and they may require assistance/prompting in responding to the surveys. This then will entail time taken out by the practice staff. The follow-up survey also requires close attention. 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** There are costs of data collection and follow up of patients who haven´t filled out the surveys. There are no fees associated with proprietary measures. 4e.3 **Evidence for costs:** 4e.4 **Business case documentation:**	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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 Academy of Ophthalmology and Hoskins Center for Quality Eye Care, 655 Beach Street, San Francisco, California, 94109-1336

Co.2 Point of Contact
Flora, Lum, M.D., flum@aao.org, 415-561-8592-

Measure Developer If different from Measure Steward
Co.3 Organization
American Academy of Ophthalmology and Hoskins Center for Quality Eye Care, 655 Beach Street, San Francisco, California, 94109-1336

Co.4 Point of Contact
Flora, Lum, M.D., flum@aao.org, 415-561-8592-

Co.5 Submitter If different from Measure Steward POC
Flora, Lum, M.D., flum@aao.org, 415-561-8592-, American Academy of Ophthalmology and Hoskins Center for Quality Eye Care

Co.6 Additional organizations that sponsored/participated in measure development
American Society of Cataract and Refractive Surgery

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.
Priscilla Arnold, MD; David Chang, MD, Kevin Miller, MD, John Thompson, MD, Leon Herndon, MD

The group developed and reviewed the measure specifications

Ad.2 If adapted, provide name of original measure:
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: 2010
Ad.7 Month and Year of most recent revision: 12, 2010
Ad.8 What is your frequency for review/update of this measure? Every 3 years
Ad.9 When is the next scheduled review/update for this measure? 12, 2013

Ad.10 Copyright statement/disclaimers: Copyright by the American Academy of Ophthalmology 2010

Ad.11 -13 Additional Information web page URL or attachment: Attachment visual functionand patient satisfaction measure Nov 2010.doc

Date of Submission (*MM/DD/YY*)**:** 03/21/2011

American Academy of Ophthalmology

Eye Care III
Physician Performance Measurement Set

Eye Care Work Group *(specialty)*
Priscilla P. Arnold, MD (Co-Chair) (ophthalmologist)

Surgical Management Subgroup:
David Chang, MD (ophthalmologist**)**
Leon W. Herndon, MD (ophthalmologist)
Kevin Miller MD (ophthalmologist)
John T. Thompson, MD (ophthalmologist)

Staff:

American Academy of Ophthalmology
Flora Lum, MD

Purpose of Measures:
These clinical performance measures, developed by the American Academy of Ophthalmology, are designed for individual quality improvement. Unless otherwise indicated, the measures are also appropriate for accountability if appropriate methodological, statistical, and implementation rules are achieved.

The proposed measures seek to advance performance measures for eye care by including explicit measures of patient visual function and patient satisfaction so as to more directly connect process measures to issues of patient interest, satisfaction, and empowerment.

Accountability Measures:
Measure #1 Cataracts: Improvement in Patient's Visual Function within 90 Days Following Cataract Surgery
Measure #2 Cataracts: Patient Satisfaction within 90 Days Following Cataract Surgery

Intended Audience and Patient Population:
Ophthalmologists may implement these measures if and when they provide the cataract surgery addressed in the measures. The measures are designed for calculating reporting or performance measurement at the individual level.

Measure Specifications
Draft specifications to report on these measures for eye care using administrative (claims) data are included in this document. We have identified codes for these measures, including ICD-9 and CPT (Evaluation & Management Codes, Category I and where Category II codes would apply). Specifications for additional data sources, including EHRs, will be fully developed at a later date.

Measure Exclusions:
For ***process measures***, there exist three categories of reasons for which a patient may be excluded from the denominator of an individual measure:

- **Medical reasons**

Includes:
- not indicated (absence of organ/limb, already received/performed, other)
- contraindicated (patient allergic history, potential adverse drug interaction, other)

- **Patient reasons**

Includes:
- patient declined
- economic, social, or religious reasons
- other patient reasons

- **System reasons**

Includes:
- resources to perform the services not available
- insurance coverage/payor-related limitations
- other reasons attributable to health care delivery system

These measure exclusion categories are not available uniformly across all measures; for each measure, there must be a clear rationale to permit an exclusion for a medical, patient, or system reason. The exclusion of a patient may be reported by appending the appropriate modifier to the CPT Category II code designated for the measure:

- **Medical reasons**: modifier 1P

- **Patient reasons**: modifier 2P

- **System reasons**: modifier 3P

Although this methodology does not require the external reporting of more detailed exclusion data, physicians should document the *specific* reasons for exclusion in patients' medical records for purposes of optimal patient management and audit-readiness. Also, each physician's exclusions data could be self-assessed to identify practice patterns and opportunities for quality improvement.

For example, it is possible for implementers to calculate the percentage of patients that physicians have identified as meeting the criteria for exclusion.

Please refer to documentation for each individual measure for information on the acceptable exclusion categories and the codes and modifiers to be used for reporting.

For ***outcome measures***, there are acceptable reasons for which a patient may be excluded from the denominator. Each specified reason is reportable with a CPT Category II code or CPT Category I code designated for that purpose.

Data Capture and Measure Calculation
This is intended for physicians to collect data on each patient eligible for a measure. Feedback on measures should be available to physicians by patient to facilitate patient management and in aggregate to identify opportunities for improvement across a physician's patient population.

Measure calculations will differ depending on whether a rate is being calculated for performance or reporting purposes.

The method of calculation for performance follows these steps: first, identify the patients who meet the eligibility criteria for the denominator (PD); second, identify which of those patients meet the numerator criteria (A); and third, for those patients who do not meet the numerator criteria, determine whether an appropriate exclusion applies and subtract those patients from the denominator (C). (see examples below)

The methodology also enables implementers to calculate the rates of patient exclusions and to further analyze both low and high rates, as appropriate (see examples below).

The method of calculation for reporting differs. One program which currently focuses on reporting rates is the Centers for Medicare and Medicaid Services (CMS) Physician Quality Reporting Initiative (PQRI). Currently, under that program design, there will be a reporting denominator determined solely from claims data (CPT and ICD-9), which in some cases result in a reporting denominator that is much larger than the eligible population for the performance denominator. Additional components of the reporting denominator are explained below.

The components that make up the numerator for reporting include all patients from the eligible population for which the physician has reported, including: the number of patients who meet the numerator criteria (A), the number of patients for whom valid exclusions apply (C) and also the number of patients who do not meet the numerator criteria (D). These components, where applicable, are summed together to make up the inclusive reporting numerator. The calculation for reporting will be the reporting numerator divided by the reporting denominator. (see examples below).

Examples of calculations for reporting and performance are provided for each measure.

Calculation for Performance

For performance purposes, this measure is calculated by creating a fraction with the following components: Numerator, Denominator, and Denominator Exclusions.

Numerator (A) Includes:
Number of patients meeting numerator criteria

Performance Denominator (PD) Includes:
Number of patients meeting criteria for denominator inclusion

Denominator Exclusions (C) Include:
Number of patients with valid medical, patient or system exclusions (where applicable; will differ by measure)

Performance Calculation

A (# of patients meeting numerator criteria) PD (# patients in denominator) – C (# patients with valid denominator exclusions)

It is also possible to calculate the percentage of patients excluded overall, or excluded by medical, patient, or system reason where applicable**:**

Overall Exclusion Calculation

C (# of patients with any valid exclusion) PD (# patients in denominator)

OR

Exclusion Calculation by Type

C_1 (# patients with medical reason) **PD** (# patients in denominator) C_2 (# patients with patient reason) **PD** (# patients in denominator) C_3 (# patients with system reason) **PD** (# patients in denominator)

Calculation for Reporting

For reporting purposes, this measure is calculated by creating a fraction with the following components: Reporting Numerator and Reporting Denominator

Reporting Numerator includes each of the following components, where applicable. (There may be instances where there are no patients to include in A, C, D, or E).

A. Number of patients meeting additional denominator criteria (for measures where true denominator cannot be determined through ICD-9 and CPT Category I coding alone) AND numerator criteria
C. Number of patients with valid medical, patient or system exclusions (where applicable; will differ by measure)
D. Number of patients not meeting numerator criteria and without a valid exclusion
E. All other patients not meeting additional denominator criteria (for measures where true denominator cannot be determined through ICD-9 and CPT Category I coding alone)

Reporting Denominator (RD) Includes:
RD. Denominator criteria (identifiable through ICD-9 and CPT Category I coding)

Reporting Calculation

A(# of patients meeting additional denominator criteria AND numerator criteria) + **C**(# of patients with valid exclusions) + **D**(# of patients NOT meeting numerator criteria) + **E**(# of patients not meeting additional denominator criteria) **RD** (# of patients in denominator)

Eye Care

Measure #1 Cataracts: Improvement in Patient's Visual Function within 90 Days Following Cataract Surgery

This measure may be used as an Accountability measure.

Clinical Performance Measure
Numerator: Patients who had improvement in visual function achieved within 90 days following cataract surgery **Denominator:** All patients aged 18 years and older who had cataract surgery **Denominator Exclusions:** The patient refuses to participate or the patient is unable to complete the questionnaire, or there is a medical reason **Measure:** Percentage of patients aged 18 years and older who had cataract surgery and had improvement in visual function achieved within 90 days following the cataract surgery
The following clinical recommendation statements are quoted <u>verbatim</u> from the referenced clinical guidelines and represent the evidence base for the measure: **This is an outcomes measure.** As such, no statements in the guideline are specific to this measurement topic.
Rationale for the measure: 1. Scientific basis for measuring visual function outcomes after cataract surgery. Visual function has been described as having multiple components, including central near, intermediate, and distance visual acuity; peripheral vision;[1] visual search; binocular vision; depth perception; contrast sensitivity; perception of color; adaptation; and visual processing speed.[2] Visual function also can be measured in terms of functional disability caused by visual impairment.[3] Many activities are affected by more than one of these visual components. Health services researchers have increasingly emphasized function and quality of life as the outcomes of treatment that are most critical and applicable to the patient. As previously stated, the primary purpose in managing a patient with cataract is to improve functional vision and the quality of life. In well-designed observational studies, cataract surgery consistently has been shown to have a significant impact on vision-dependent function. The Cataract Patient Outcomes Research Team (PORT) reported that 90% of patients under-going first-eye cataract surgery noted improvement in functional status and satisfaction with vision.[4] The Activities of Daily Vision Study of elderly patients with a high prevalence of coexisting ocular and medical diseases reported improved visual function in 80% of patients at 12 months after surgery.[5] A National Cataract Study conducted in England of 1,139 patients who had cataract surgery found that preoperative functional impairment varied in relation to gender, age, and visual acuity. Men were more likely to have trouble with driving, glare, and employment, and women

were more likely to have difficulties with activities of daily living and recreational activities.[6] Studies have found that regardless of the preoperative visual acuity in the better eye, most patients reported improvement in their ability to perform visually dependent tasks after undergoing cataract surgery.[4-6]

Several studies have reported an association between improved visual function after cataract surgery and improved health-related quality of life.[1,5,7-8] Visual function plays an important role in physical function, particularly in terms of mobility.[9] The loss of visual function in the elderly is associated with a decline in physical and mental functioning as well as in independence in activities of daily living,[10] including night-time driving, daytime driving, community activities, and home activities. Elderly patients with visual impairment only (and no other physical or mental impairments) were 2.5 times as likely to experience functional decline than elderly patients without visual impairment.

Improved visual function following cataract surgery can ameliorate the progressive deterioration of quality of life seen in elderly patients.[1,5] In a cohort of 464 patients 65 years old and older, cataract extraction improved visual function and health-related quality of life. Patients with an improvement in their Activities of Daily Vision Scale (ADVS), a brief measure of vision-specific functional status,[11] had from 10% to 59% less decline in nearly all Short Form (SF)-36 dimensions.[5] The SF-36 is a generic global measure of multidimensional health-related quality of life.[12] A nationally representative population of 7,114 persons who were 70 years old and older showed that limitations in vision correlated with decreased functional status.[13] The unadjusted functional score of a person with reported poor vision was four times worse than the score for a person with excellent vision.[13] This difference was comparable with the differences found in other chronic conditions such as arthritis. This relationship with vision persisted, even after adjustment for health, demographics, and economic status. Individuals who rated their vision as other than excellent reported worse functional status, even when controlled for the presence of other medical conditions, education, income, general health status, and other symptoms. By improving visual function, cataract surgery may play an important role in preserving overall functional status, reducing associated injuries and accidents, and preventing disability in at-risk elderly patients.[10]

An analysis of the Medical Outcomes Study found that having blurred vision more than once or twice a month has a significant impact on functional status and well-being, particularly on problems with work or other daily activities as a result of physical health.[14] This impact was found to be greater than the impact of several other chronic conditions, such as hypertension, history of myocardial infarction, type 2 diabetes mellitus, indigestion, trouble urinating, and headache. In one study, patients planning to undergo cataract surgery assigned a mean preoperative preference value of 0.68 on a scale ranging from 0 to 1 (where 0 is death and 1 is excellent health), indicating that the visual impairment from cataracts had a substantial impact on their quality of life.[11] Visual impairment is an important risk factor for falls[15] and for hip fracture.[16] Specifically, the Study for Osteoporotic Fractures Research Group found that poor depth perception and decreased contrast sensitivity independently increased the risk of hip fracture.[17]

Visual impairment, in particular a decrease of visual acuity and contrast sensitivity, has been shown to be associated with difficulties in driving.[18] In one study, older drivers with visually significant cataract were twice as likely as older drivers without cataract to report reduction in days driven and four times as likely to report difficulties in challenging driving situations.[19] Drivers with visually significant cataract were 2.5 times more likely to have had an at-fault involvement in a motor vehicle crash in the past 5 years compared with drivers without cataract.[19] This association was significant, even after accounting for other factors such as impaired general health, age, mental status deficit or depression. In this study, visually significant cataract was determined by reviewing the participant's medical record and most recent eye examination by an eye care specialist. The study required that cataract in both eyes was the cause of the visual impairment, based on the medical record; an additional inclusion criterion was best-corrected visual acuity in one eye of 20/40 or worse. A further study in the same group demonstrated that drivers with a history of crash involvement were eight times more likely to have a serious contrast

sensitivity deficit (defined as a Pelli-Robson score of 1.25 or less) in the worse eye than those who had no history of crash involvement.[20] A severe contrast sensitivity deficit in only one eye was still significantly associated with crash involvement.[20]

Binocular vision is better than the vision of a single eye. The simultaneous use of the two eyes is complex and requires the integration of disparate images from each eye. A study demonstrated that binocular vision resulted in better perception of form, color, and the relationship of the body to the environment, which facilitated manipulation, reaching, and balance, particularly under dim illumination.[21] However, if the vision of one eye is reduced due to cataract, visual performance can fall below the level of monocular vision by a mechanism known as binocular inhibition,[22] which reduces patients' visual acuity and contrast sensitivity.[23] A study of the Framingham Study Cohort found that poor vision in one or both eyes was associated with an increased risk of hip fracture. It also found that patients with good vision in one eye and moderately impaired vision in the other eye had a higher risk of fracture than those with similar visual impairment in both eyes.[24] A study of 150 patients before and after cataract surgery found that poor binocular visual acuity was related to more problems in activities of daily living.[25] Another study, based on patients who reported no beneficial outcomes after first-eye cataract surgery in the National Swedish Cataract Outcome register, found that anisometropia was the reason for the poor outcome in one-third of cases.[26] These studies have shown that second-eye surgery is important to visual and physical function.

In summary, these studies demonstrate that physical function, emotional well-being, and overall quality of life can be enhanced when visual function is restored by cataract extraction.[27]

Improved visual function as a result of cataract surgery includes the following:

- Better optically corrected vision.
- Better uncorrected vision with reduced spectacle dependence.
- Increased ability to read or do near work.
- Reduced glare.
- Improved ability to function in dim levels of light.
- Improved depth perception and binocular vision.
- Improved color vision.

Improved physical function as a critical outcome of cataract surgery includes the following:

- Increased ability to perform activities of daily living.
- Increased opportunity to continue or resume an occupation.
- Increased mobility (walking, driving).

Improved mental health and emotional well-being as a second critical outcome of cataract surgery includes the following benefits:

- Improved self-esteem and independence.
- Increased ability to avoid injury.
- Increased social contact and ability to participate in social activities.
- Relief from fear of blindness.

Most patients achieve improved visual function after cataract surgery. This outcome is achieved consistently through careful attention through the patient selection process, accurate measurement of axial length and corneal power, appropriate selection of an IOL power calculation formula, etc. As such, it reflects the care and diligence with which the surgery is assessed, planned and executed. Failure to achieve this after surgery would reflect patterns of patient selection or treatment that should be

assessed for opportunities for improvement.

Sometimes cataract surgery is performed for other medical reasons other than to improve impaired visual function caused by cataract. These circumstances include the following: clinically significant anisometropia in the presence of a cataract; when the lens opacity interferes with optimal diagnosis or management of posterior segment conditions, when the lens causes inflammation (phacolysis, phacoanaphylaxis) and when the lens induces angle closure (phacomorphic or phacotopic). In these situations, improved visual function as a result of the removal of the cataract is not expected, because of the pre-existing comorbid conditions.

2. Evidence of a gap in care

This is an outcome of surgery indicator of direct relevance and import to patients, their families and referring providers. The available evidence suggests that cataract surgery achieves this in about 90% of patients. While the potential for improvement is seemingly small, the volume of cataract surgery in the U.S. of over 2.8 million surgeries means that the impact could affect more than 100,000 patients per year. Ideally, performance on this indicator would be as high as possible, with lower rates suggestive of opportunities for improvement.

Definitions:

Standardized Tool – An assessment tool that has been appropriately validated for the population for which it being used. Examples of tools for visual function assessment include, but are not limited to: National Eye Institute-Visual Function Questionnaire (VFQ), the Visual Function (VF)-14, the modified VF-8, the Activities of Daily Vision Scale (ADVS), the Catquest and the modified Catquest-9.

Vision Function Assessment – Questionnaires designed to measure a patient's ability to perform the everyday tasks requiring vision.

Data Capture and Calculations:

Calculation for *Performance*

For performance purposes, this measure is calculated by creating a fraction with the following components: Numerator, Denominator, and Denominator Exclusions.

Performance Numerator (A) Includes:

- Patients who had an improvement in their visual function achieved within 90 days following cataract surgery

Performance Denominator (PD) Includes:

- All patients aged 18 years and older

AND

- Had cataract surgery

Performance Denominator Exclusions (C) Includes:

A patient is excluded if the following condition(s) exist:

Medical reasons:

When cataract surgery was performed for these indications:

- Clinically significant anisometropia in the presence of a cataract
- The lens opacity interferes with optimal diagnosis or management of posterior segment conditions
- The lens causes inflammation (phacolysis, phacoanaphylaxis)
- The lens induces angle closure (phacomorphic or phacotopic)

Patient reasons:

- The patient refuses to participate
- The patient is unable to complete the questionnaire

Performance Calculation

$$\frac{\text{A (\# of patients meeting measure criteria)}}{\text{PD (\# of patients in denominator)} - \text{C (\# of patients with valid denominator exclusions)}}$$

Components for this measure are defined as:

A	# of patients who had an improvement in their visual function achieved within 90 days following cataract surgery
PD	# of patients aged 18 years and older who had cataract surgery

C	# of patients with documented patient reason for not completing their visual function assessment within 90 days following cataract surgery

Calculation for Reporting:
For reporting purposes, this measure is calculated by creating a fraction with the following components: Reporting Numerator and Reporting Denominator.

Reporting Numerator includes each of the following instances:

A. Patients who had an improvement in their visual function achieved within 90 days following cataract surgery

C. Patients who did not complete their visual function assessment within 90 days following cataract surgery but for whom there is a documented medical or patient reason for not doing so

D. Patients who did not have an improvement in their visual function achieved within 90 days following cataract surgery and there is no documented medical or patient reason for not doing so

Reporting Denominator (RD) includes:

- Patients aged 18 years and older AND
- Had cataract surgery

Reporting Calculation

$$\frac{\text{A (\# of patients meeting measure criteria) + C (\# of patients with valid exclusions) + D (\# of patients NOT meeting numerator criteria)}}{\text{RD (\# of patients in denominator)}}$$

Components for this measure are defined as:

A	# of patients who had an improvement in their visual function achieved within 90 days following cataract surgery
C	# of patients who did not complete their visual function assessment within 90 days following cataract surgery but for whom there is a documented medical or patient reason for not doing so
D	# of patients who did not have an improvement in their visual function achieved within 90 days following cataract surgery and there is no documented medical or patient reason for not doing so
RD	# of patients aged 18 years and older who had cataract surgery

Measure Specifications - Measure #1 Cataracts: Improvement in Patient's Visual Function within 90 Days Following Cataract Surgery

Measure specifications will be provided for multiple data sources.

A. Administrative claims data

Administrative claims data collection requires users to identify the eligible population (denominator) and numerator using codes recorded on claims or billing forms (electronic or paper). Users report a rate based on all patients in a given practice for whom data are available and who meet the eligible population/denominator criteria.

(Note: The specifications listed below are those needed for performance calculation)

Denominator (Eligible Population): All patients aged 18 years and older who had cataract surgery

- CPT Procedure Codes (with or without modifiers): 66840, 66850, 66852, 66920, 66930, 66940, 66982, 66983, 66984

Numerator: Patients who had an improvement in their visual function achieved within 90 days following cataract surgery

Report the following CPT Category II code:

_____ - Improved visual function achieved within the 90 days following cataract Surgery

Denominator Exclusions: Documentation of medical reason for not improving visual function within 90 days of cataract surgery

- Append modifier to CPT Category II Code: -1P

Documentation of patient reason for not improving visual function within 90 days of cataract surgery

- Append modifier to CPT Category II Code: -2P

B. Registry

Registry reporting requires users to identify the eligible population (denominator) using CPT codes and patient demographics. The numerator options as described in the CPT Category II codes are used to report the numerator of the measure. The CPT Category II codes listed do not need to be submitted for registry-based submissions, however these codes may be submitted for those registries that utilize claims data.

C. Electronic Health Record System (in development)

D. Paper Medical Record (in development)

Eye Care

Measure #2 Cataracts: Patient Satisfaction within 90 Days Following Cataract Surgery

This measure may be used as an Accountability measure.

Clinical Performance Measure
Numerator: Patients who were satisfied with their care within 90 days following cataract surgery **Denominator:** All patients aged 18 years and older who had cataract surgery **Denominator Exclusions:** The patient refuses to participate or the patient is unable to complete the questionnaire **Measure:** Percentage of patients aged 18 years and older who had cataract surgery and were satisfied with their care within 90 days following the cataract surgery
The following clinical recommendation statements are quoted <u>verbatim</u> from the referenced clinical guidelines and represent the evidence base for the measure: **This is an outcomes measure.** As such, no statements in the guideline are specific to this measurement topic.
Rationale for the measure: 1. Scientific basis for measuring patient satisfaction after cataract surgery. Patient satisfaction is a valuable performance indicator for measuring the quality of care delivered by ophthalmologists providing cataract surgery. In the broadest sense, patient satisfaction is an assessment of the patient's experience with the care process delivered by health plans, clinicians, health systems, hospitals, etc. This experience can cover domains as diverse as information/education, interpersonal manner, emotional support, accessibility, convenience, outcomes or results, environment, personalization, involvement in care, finances, etc. In 1996, The American Academy of Ophthalmology launched the National Eyecare Outcomes Network (NEON) database. [28,29] From January 1, 1996 through March 30, 2001, 249 ophthalmologists at 114 different practice sites submitted data to the NEON cataract surgery database. Post-operative patient satisfaction responses were collected for 6,154 patients, or about 34.5% of all patients who had pre-operative forms submitted. This assessment was performed at a median of 4.1 weeks postoperatively for all patients enrolled in the database. A 12-item questionnaire was used to assess patient satisfaction. Patient satisfaction was associated with younger age and absence of ocular comorbidity. Other studies of patient satisfaction after cataract surgery in Austria and in Spain. One study found that patients with pre-existing eye disease, including those patients with improved visual acuity after surgery, were the least satisfied with the results of surgery. [30] In these cases, improved patient

education prior to surgery could be helpful in improving patient satisfaction. Another study found that patient satisfaction was associated with expectations prior to surgery.[31]

Most patients are satisfied with their care and results after cataract surgery. This outcome is achieved consistently through careful attention through the patient selection process, accurate measurement of axial length and corneal power, appropriate selection of an IOL power calculation formula, etc. As such, it reflects the care and diligence with which the surgery is assessed, planned and executed. Failure to achieve this satisfaction after surgery would reflect patterns of patient selection or treatment that should be assessed for opportunities for improvement.

Use of this indicator in the PQRI program in the claims reporting method would require some modification to the current reporting of post-operative care for patients undergoing cataract surgery, since this indicator would be operative during the 90 day global period. However, there is a strong and practical precedent for such modifications in that reporting arrangements have previously been made to accommodate co-management of care by different providers during the post-operative period. A similar adjustment to allow for filing of a claim of meeting this goal at one point in the 90 day global period would be sufficient, potentially drawing upon the methods to demarcate the onset of co-management transfer of post-operative care.

Various patient satisfaction instruments exist, but an instrument developed by the program, Consumer Assessment of Healthcare Providers and Systems (CAHPS), Agency for Healthcare Research and Quality develops and supports the use of a comprehensive and evolving family of standardized surveys that ask consumers and patients to report on and evaluate their experiences with health care. These surveys cover topics that are important to consumers, such as the communication skills of providers and the accessibility of services. AHRQ first launched the CAHPS program in October 1995 in response to concerns about the lack of good information about the quality of health plans from the enrollees' perspective. At that time, numerous public and private organizations collected information on enrollee and patient satisfaction, but the surveys varied from sponsor to sponsor and often changed from year to year.

The CAPHS Surgical Care Survey asks adult patients to report on surgical care, surgeons, their staff, and anesthesiologists. It was developed by the American College of Surgeons and the Surgical Quality Alliance to assess patients' experiences before, during, and after surgery. In early 2010, the CAHPS Consortium voted to adopt the Surgical Care Survey as an official CAHPS survey. The Surgical Care Survey expands on the current CAHPS Clinician & Group Survey, which focuses on primary and specialty care, by incorporating domains that are relevant to surgical care, such as informed consent, anesthesia care, and post-operative follow-up. The survey is unique in that it assesses patients' experiences with surgical care in both the inpatient and outpatient settings by asking respondents about their care before, during, and after surgery

The main purpose of the CAHPS Surgical Care Survey is to address the need to assess and improve the experiences of surgical patients. Like other CAHPS surveys, this questionnaire focuses on aspects of surgical quality that are important to patients and for which patients are the best source of information. The survey results are expected to be useful to everyone with a need for information on the quality of surgeons and surgical care, including patients, practice groups, health plans, insurers, and specialty boards. Patients can use the information to help make better and more informed choices about their surgical care. Practices, health plans, and insurers can use the survey results for quality improvement initiatives and incentives. Specialty boards may use the survey for maintenance of certification.

https://www.cahps.ahrq.gov/content/products/sc/PROD_SC_Surgical_Care.asp?p=1021&s=213

2. Evidence of a gap in care

This is an outcome of surgery indicator of direct relevance and import to patients , their families and referring providers. The available evidence suggests that cataract surgery achieves this in about 90% of patients. While the potential for improvement appears seemingly small, the volume of cataract surgery in the U.S. of over 2.8 million surgeries means that the impact could affect more than 100,000 patients per year. Ideally, performance on this indicator to be as high as possible, with rates lower than 95-100% suggestive of opportunities for improvement.

Definitions:

Standardized Tool – An assessment tool that has been appropriately validated for the population for which it being used. Examples of tools for patient satisfaction include, but are not limited to: Surgical Consumer Assessment of Health Plans and Systems, which is also approved by the Agency for Health Care Research and Quality.

Patient Satisfaction Assessment – Questionnaires designed to measure a patient's satisfaction with the care that they received from their surgeon.

Data Capture and Calculations:

Calculation for *Performance*

For performance purposes, this measure is calculated by creating a fraction with the following components: Numerator, Denominator, and Denominator Exclusions.

Performance Numerator (A) Includes:

- Patients who were satisfied with their care within 90 days following cataract surgery

Performance Denominator (PD) Includes:

- All patients aged 18 years and older

AND

- Had cataract surgery

Performance Denominator Exclusions (C) Includes:

A patient is excluded if the following condition(s) exist:

- The patient refuses to participate
- The patient is unable to complete the questionnaire

Performance Calculation

$$\frac{\text{A (\# of patients meeting measure criteria)}}{\text{PD (\# of patients in denominator)} - \text{C (\# of patients with valid denominator exclusions)}}$$

Components for this measure are defined as:

A	# of patients who were satisfied with their care within 90 days following cataract surgery
PD	# of patients aged 18 years and older who had cataract surgery
C	# of patients with documented patient reason for not performing the patient satisfaction assessment within 90 days following cataract surgery

Calculation for Reporting:

For reporting purposes, this measure is calculated by creating a fraction with the following components: Reporting Numerator and Reporting Denominator.

Reporting Numerator includes each of the following instances:

A. Patients who were satisfied with their care within 90 days following cataract surgery

C.Patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery but for whom there is a documented patient reason for not doing so

D. Patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery and there is no documented patient reason for not doing so

Reporting Denominator (RD) includes:

- Patients aged 18 years and older AND
- Had cataract surgery

Reporting Calculation

$$\frac{\text{A (\# of patients meeting measure criteria) + C (\# of patients with valid exclusions) + D (\# of patients NOT meeting numerator criteria)}}{\text{RD (\# of patients in denominator)}}$$

Components for this measure are defined as:

A	# of patients who were satisfied with their care within 90 days following cataract surgery
C	# of patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery but for whom there is a documented patient reason for not doing so
D	# of patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery and there is no documented patient reason for not doing so
RD	# of patients aged 18 years and older who had cataract surgery

Measure Specifications - Measure #2 Cataracts: Patient Satisfaction within 90 Days Following Cataract Surgery

Measure specifications will be provided for multiple data sources.

A. Administrative claims data

Administrative claims data collection requires users to identify the eligible population (denominator) and numerator using codes recorded on claims or billing forms (electronic or paper). Users report a rate based on all patients in a given practice for whom data are available and who meet the eligible population/denominator criteria.

(Note: The specifications listed below are those needed for performance calculation)

Denominator (Eligible Population): All patients aged 18 years and older who had cataract surgery

- CPT Procedure Codes (with or without modifiers): 66840, 66850, 66852, 66920, 66930, 66940, 66982, 66983, 66984

Numerator: Patients who were satisfied with their care within 90 days following cataract surgery

Report the following CPT Category II code:
_____ - Patient satisfaction achieved within the 90 days following cataract Surgery

Denominator Exclusions: Documentation of patient reason for not completing patient satisfaction assessment within 90 days of cataract surgery

- Append modifier to CPT Category II Code: -2P

B. Registry

Registry reporting requires users to identify the eligible population (denominator) using CPT codes and patient demographics. The numerator options as described in the CPT Category II codes are used to report the numerator of the measure. The CPT Category II codes listed do not need to be submitted for registry-based submissions, however these codes may be submitted for those registries that utilize claims data.

C. Electronic Health Record System (in development)

D. Paper Medical Record (in development)

References:

1. Brenner MH, Curbow B, Javitt JC, et al. Vision change and quality of life in the elderly. Response to cataract surgery and treatment of other chronic ocular conditions. Arch Ophthalmol 1993; 111:680-5.
2. Sloane ME, Ball K, Owsley C, et al. The Visual Activities Questionnaire: developing an instrument for assessing problems in everyday visual tasks. Technical Digest, Noninvasive Assessment of the Visual System 1992; 1:26-29.
3. Steinberg EP, Tielsch JM, Schein OD, et al. The VF-14. An index of functional impairment in patients with cataract. Arch Ophthalmol 1994; 112:630-8.
4. Schein OD, Steinberg EP, Javitt JC, et al. Variation in cataract surgery practice and clinical outcomes. Ophthalmology 1994; 101:1142-52.
5. Mangione CM, Phillips RS, Lawrence MG, et al. Improved visual function and attenuation of declines in health-related quality of life after cataract extraction. Arch Ophthalmol 1994; 112:1419-25.
6. Desai P, Minassian DC, Reidy A. National cataract surgery survey 1997-8: a report of the results of the clinical outcomes. Br J Ophthalmol 1999; 83:1336-40.
7. Monestam E, Wachtmeister L. Impact of cataract surgery on visual acuity and subjective functional outcomes: a population-based study in Sweden. Eye 1999; 13:711-19.
8. Steinberg EP, Tielsch JM, Schein OD, et al. National study of cataract surgery outcomes. Variation in 4-month postoperative outcomes as reflected in multiple outcome measures. Ophthalmology 1994; 101:1131-40; discussion 1140-1.
9. Damiano AM, Steinberg EP, Cassard SD, et al. Comparison of generic versus disease-specific measures of functional impairment in patients with cataract. Med Care 1995; 33 (4 Suppl):AS120-30.
10. Salive ME, Guralnik J, Glynn RJ, et al. Association of visual impairment with mobility and physical function. J Am Geriatr Soc 1994; 42:287-92.
11. Laforge RG, Spector WD, Sternberg J. The relationship of vision and hearing impairment to one-year mortality and functional decline. J Aging Health 1992; 4:126–48.
12. Mangione CM, Phillips RS, Seddon JM, et al. Development of the 'Activities of Daily Vision Scale'. A measure of visual functional status. Med Care 1992; 30:1111-26.
13. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30:473-83.
14. Lee P, Smith JP, Kington R. The relationship of self-rated vision and hearing to functional status and well-being among seniors 70 years and older. Am J Ophthalmol 1999; 127:447-52.
15. Lee PP, Spritzer K, Hays RD. The impact of blurred vision on functioning and well-being. Ophthalmology 1997; 104:390-6.
16. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. N Engl J Med 1988; 319:1701-7.
17. Felson DT, Anderson JJ, Hannan MT, et al. Impaired vision and hip fracture. The Framingham Study. J Am Geriatr Soc 1989; 37:495-500.

18. Cummings SR, Nevitt MC, Browner WS, et al. Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. N Engl J Med 1995; 332:767-73.
19. McGwin G Jr, Chapman V, Owsley C. Visual risk factors for driving difficulty among older drivers. Accid Anal Prev 2000; 32:735-44.
20. Owsley C, Stalvey B, Wells J, Sloane ME. Older drivers and cataract: driving habits and crash risk. J Gerontol A Biol Sci Med Sci 1999; 54:M203-11.
21. Owsley C, Stalvey BT, Wells J, et al. Visual risk factors for crash involvement in older drivers with cataract. Arch Ophthalmol 2001; 119:881-7.
22. Jones RK, Lee DN. Why two eyes are better than one: the two views of binocular vision. J Exp Psychol 1981; 7:30–40.
23. Pardhan S, Gilchrist J. The importance of measuring binocular contrast sensitivity in unilateral cataract. Eye 1991; 5:31-5.
24. Taylor RH, Misson GP, Moseley MJ. Visual acuity and contrast sensitivity in cataract: summation and inhibition of visual performance. Eye 1991; 5:704-7.
25. Lundström M, Fregell G, Sjöblom A. Vision related daily life problems in patients waiting for a cataract extraction. Br J Ophthalmol 1994; 78:608–11.
26. Lundström M, Brege KG, Florén I, et al. Impaired visual function after cataract surgery assessed using the Catquest questionnaire. J Cataract Refract Surg 2000; 26:101–8.
27. Cataract Management Guideline Panel. Cataract in Adults: Management of Functional Impairment. Clinical Practice Guideline, Number 4. Rockville, MD: USDHHS, AHCPR Publ. No. (PHS) 93-0542; 1993.
28. Lum F, Schein O, Schachat AP, et al. Initial two years of experience with the AAO National Eyecare Outcomes Network (NEON) cataract surgery database. Ophthalmology 2000; 107:691-7.
29. Lum F, Schachat AP, Jampel HD. The development and demise of a cataract surgery database. The Joint Commission Journal on Quality Improvement 2202; 28:108-114.
30. Mozaffarieh M, Krepler K, Heinzl H et al. Visual function, quality of life and patient satisfaction after ophthalmic surgery: a comparative study. Ophthalmologica 2004; 218:26-30.
31. Lledo R, Rodriguez T, Fontenia JR et al. Cataract surgery: An analysis of patient satisfaction with medical care. International Ophthalmology 22:227-32.

Measuring outcomes of cataract surgery using the Visual Function Index-14

Vijaya K. Gothwal, PhD, Thomas A. Wright, Ecosse L. Lamoureux, PhD, Konrad Pesudovs, PhD

PURPOSE: To determine which version of the Visual Function Index-14 (VF-14) most precisely measured cataract surgery outcomes, rescale the VF-14 using Rasch analysis, and create a short-form version for comparison.

SETTING: Flinders Medical Centre, Adelaide, South Australia, Australia.

METHODS: In this cohort study incorporating questionnaire development, participants were drawn from the cataract surgery waiting list at Flinders Medical Centre. There were 2 cohorts: a preoperative cohort used for questionnaire development and an outcomes cohort. All patients had cataract surgery by phacoemulsification with intraocular lens implantation. Rasch analysis was used to refine the VF-14 into valid long-form (VF-11R) and short-form (VF-8R) versions. The ability of 8 versions (original; 2 proposed versions; 5 previously proposed versions) of the VF-14 to discriminate cataract surgery outcomes was compared with that of the standard VF-14 using the relative precision method.

RESULTS: The preoperative cohort comprised 210 patients and the outcomes cohort, 51 patients. Large gains in visual functioning occurred with cataract surgery, and these were detectable with all versions of the VF-14. The largest gain in precision, 125% (relative precision. 2.25), occurred for VF-8R. Short forms that were not Rasch scaled showed gains in precision, from 23% to 80%. The VF-8R also showed the largest gains in precision in 2 subgroups: with ocular comorbidity (relative precision, 2.14) and without ocular comorbidity (relative precision, 2.48).

CONCLUSIONS: Results show an unequivocal advantage to using Rasch-scaled scores for assessing cataract surgery outcomes. The 8-item, Rasch-scaled VF-8R appears ideally suited for measuring cataract surgery outcomes given its high precision and short test time.

Financial Disclosure: No author has a financial or proprietary interest in any material or method mentioned.

J Cataract Refract Surg 2010; 36:1181–1188 © 2010 ASCRS and ESCRS

A patient's perspective is critical in evaluating the need for, and outcomes of, cataract surgery.[1–3] Questionnaires are increasingly being required for these evaluations. One such questionnaire is the Visual Function Index-14 (VF-14), which was developed to assess functional impairment in cataract patients.[4] The VF-14 is a popular questionnaire. It possesses adequate traditional psychometric properties,[5,6] has a concise format, is easy to administer, and has been validated internationally.[3,7] However, researchers have suggested it is too time consuming for routine use and therefore have proposed shortened versions.[8–10] Uusitalo et al.[8] proposed a VF-7, derived by selecting items that best correlated with patient satisfaction. Pager[9] also advocated a VF-7, which included items (different from Uusitalo et al.) that closely correlated with the overall preoperative VF-14 score. Moghimi et al.[10] advocated a VF-9 for use in specific conditions, including cataract surgery outcomes in traumatic aniridia.

The most recent short-form of the VF-14 is the VF-9, a Rasch-scaled version proposed by Lamoureux et al.[11] for use in a population-based study. Before this, Mallinson et al.[12] had used the VF-14 as an illustrative example to show the benefits of using Rasch analysis to shorten questionnaires. In contrast, Friedman et al.[13] proposed a shortened VF-11 but questioned the advantages of shortening the original VF-14.

Given there are many short forms of the VF-14, each varying in item content and number, which version best measures cataract surgery outcomes is unclear. To bring clarity to this problem, we aimed to compare the precision (ie, usefulness in making comparisons between preoperative and postoperative participants)[14] of current short-form versions of the VF-14

doi:10.1016/j.jcrs.2010.01.029

in assessing cataract surgery outcomes to determine the preferred version for future use.

Furthermore, questionnaires reexamined using Rasch analysis have shown more sensitivity to change postoperatively[2]; therefore, we hypothesized that Rasch-scaled versions of the VF-14 may improve the precision of outcomes measurement. Although this has been done in a population-based setting, the high rate of normal visual functioning may make such a population unsuitable for refining the instrument. Therefore, we evaluated a cataract population to revise the VF-14 using Rasch analysis and included this version in our comparison.

PATIENTS AND METHODS

Study Group and Protocol

Since 2005, as part of a long-term Cataract Outcomes Assessment Study, data on a number of cataract-specific questionnaires (including the VF-14) were collected. This assessment was implemented by routinely mailing packs of questionnaires (10) to consecutive patients on the waiting list for cataract extraction surgery at Flinders Medical Centre, Adelaide, South Australia. Inclusion criteria were English speaking, aged 18 years or older, and ability to provide written informed consent. Patients self-administered the questionnaires and returned them in a prepaid envelope. Patients chose to complete as many questionnaires as they wished. A demographic data form was included in the pack to obtain information regarding ocular and systemic status, which was subsequently confirmed from the patient's medical record at the time of data entry.

During a single 6-month data-collection window, the same pack was mailed 6 months after cataract surgery. Patients had coexisting systemic and ocular conditions, which is typical of an elderly cataract patient cohort in Australia.[15] Ethics approval for this research was obtained from the Flinders Clinical Ethics Committee. This research adhered to the tenets of the Declaration of Helsinki.

Submitted: November 23, 2009.
Final revision submitted: January 2, 2010.
Accepted: January 21, 2010.

From NH&MRC Centre for Clinical Eye Research (Gothwal, Wright, Pesudovs), Department of Optometry and Vision Science, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia; Centre for Eye Research Australia (Lamoureux), Department of Ophthalmology, University of Melbourne, Victoria; Vision CRC (Lamoureux), Sydney, New South Wales, Australia; Meera and LB Deshpande Centre for Sight Enhancement (Gothwal), Vision Rehabilitation Centres, LV Prasad Eye Institute, Hyderabad, India; Singapore Eye Research Institute (Lamoureux), Singapore National Eye Centre, Singapore.

Corresponding author: Konrad Pesudovs, PhD, NH&MRC Centre for Clinical Eye Research, Department of Optometry and Vision Science, Flinders Medical Centre, Bedford Park, South Australia, 5042, Australia. E-mail: konrad.pesudovs@flinders.edu.au.

There were 2 patient populations. The first cohort comprised preoperative cataract patients whose data were used to refine the VF-14 with Rasch analysis (development group). The second cohort comprised patients whose data were used to measure cataract surgery outcomes (outcomes group).

Standardized eye examinations were performed before and after (minimum 21 days) cataract surgery. Habitual monocular and binocular visual acuity assessments were performed using computerized testing based on the logMAR principles with screen illumination of 150 candelas/m^2.[16] The visual acuity in the operated and fellow eyes of patients who had cataract surgery is presented here.

Questionnaires

Visual Function Index-14 The VF-14 contains questions (items) related to the degree of difficulty in performing 14 vision-dependent activities (eg, reading, watching television).[4] Table 1 shows the activities the VF-14 addresses and the response categories. Responses were coded as recommended by the developers. "Not applicable" responses were treated as missing data in the analysis. Higher scores represent better visual functioning (ie, less difficulty) and, therefore, greater ability in performing the activity.

Short-Form Versions of the Visual Function Index-14 Five studies that proposed short-form versions of the VF-14 were identified. They were Friedman et al.'s VF-11 (11 items),[13] Uusitalo et al.'s VF-7 (VF-7U; 7 items),[8] Pager's VF-7 (VF-7P; 7 items), [9] Moghimi et al.'s VF-9 for women (VF-9MF; 9 items) and for men (VF-9MM; 10 items),[10] and Lamoureux et al.'s Rasch-analyzed VF-9 (VF-9L).[11] Each shortened version contains a different set of the original VF-14 items.

The response options used in all short-form versions were similar to the original VF-14. Although Lamoureux et al.[11] proposed a reduction in categories from 5 to 4 for their VF-9L, in this study the original 5 categories for data collection were retained as that was how Lamoureux et al. collected their data.

Outcome Measures

Change in overall visual functioning with cataract surgery was the primary outcome measure. This outcome was also tested for 2 subgroups: with ocular comorbidity and without ocular comorbidity. Change in visual acuity was the secondary outcome measure.

Assessment of the Psychometric Properties of Visual Function Index-14 by Rasch Analysis

The native scoring system of the VF-14 is an ordinal (Likert) scale (ie, numerical values in an increasing order are assigned to categories of increasing difficulty) that uses summary scoring. This approach falsely assumes the spacing between response categories is equal and that all the items have the same level of difficulty. Ordinal scores are not a measurement; thus, they are inappropriate for measuring the degree of difference between patients or between preoperative and postoperative periods.[17] Therefore, before using the VF-14, it was imperative to assess its psychometric properties using Rasch analysis. A series of analyses was performed that included assessment of the following: (1) behavior of response categories (ie, whether higher

Table 1. Item content for VF-14 and the 2 Rasch-scaled versions of the VF-14 (VF-11R and VF-8R).*

Item	Item Description in VF-14	Items in VF-11R	Items in VF-8R
1	Reading small print, such as labels on medicine bottles, a telephone book, food labels	Retained	Retained
2	Reading a newspaper or a book	Retained	Retained
3	Reading a large-print book or large-print newspaper or numbers on a telephone	Retained	Eliminated
4	Recognizing people when they are close to you	Retained	Eliminated
5	Seeing steps, stairs, or curbs	Retained	Retained
6	Reading traffic signs, street signs, or store signs	Retained	Retained
7	Doing fine handwork, such as sewing, knitting, crocheting, carpentry	Retained	Retained
8	Writing checks or filling out forms	Retained	Retained
9	Playing games, such as bingo, dominos, card games, mahjong	Retained	Retained
10	Taking part in sports, such as bowling, handball, tennis, golf	Eliminated	Eliminated
11	Cooking	Retained	Eliminated
12	Watching television	Retained	Retained

*For items 1 through 12, the frame question was, "Do you have any difficulty, even with glasses?"; there were 5 scoring response options (no = 4; a little = 3; a moderate amount = 2; a great deal = 1; unable to do the activity = 0). Items 13 through 18 are driving items. Two are scoring items with 5 response options, and there are different frame question for these items; these were eliminated from the Rasch-scaled versions (VF-11R and VF-8R).

categories represented better visual functioning), (2) measurement precision (represented by person separation; minimum acceptable value of 2.0[18]), (3) unidimensionality (ie, whether all the items contribute and measure a single underlying latent trait of visual functioning measured by infit mean square statistic with acceptable range of 0.7 to 1.3 and also by principal components analysis,), and (4) whether items match the patient's visual functioning (represented by targeting; ideal <0.5 logits). If all the items did not measure visual functioning (representing lack of unidimensionality), the goal was to provide remedial measures. As in other studies, this one considered shortening the VF-14 without compromising its original properties. Details about applying Rasch analysis to the questionnaires for this purpose have been described[2,19,20] and are reported in brief here. In the context of Rasch analysis, an item (activity) is considered difficult if a high level of visual functioning is required to complete it. In Rasch analysis, item difficulty and patient ability are calibrated on the same scale and are expressed in logit units.[18,21]

Using the data from all preoperative cataract patients, Rasch analysis was performed using the Andrich rating scale model for polytomous data (ie, multiple response options for an item) in the Winsteps software (version 3.68).[22,23] In contrast to the need to combine categories, as reported by Lamoureux et al.[11] for the VF-9L, the patients in this study used the response options as they were intended to and, therefore, the original 5 response categories were retained. The VF-14 showed adequate stratification of visual functioning evidenced by a person separation of 2.45 (minimum acceptable value, 2.0) indicating that it was able to discriminate between 3 strata of patient's visual functioning (Table 2). Targeting was suboptimum (1.86 logits), indicating that the items were mismatched to the patient's visual functioning. This result indicated that, overall, the items were too easy for patients.[12]

Two items did not fit. This indicated a lack of unidimensionality (ie, these 2 items measured a construct different than the remaining 12 items [not visual functioning]). Principal component analysis further confirmed the lack of unidimensionality by revealing the presence of a secondary dimension, which could be described as relating to driving.

Taken together, the above findings suggested that the VF-14 required revision. Specifically, unidimensionality had to be restored and item misfit minimized. Unidimensionality was restored by deleting the 2 driving items. However, after deletion of the items, a further item (playing games) showed misfit and therefore was also deleted. The remaining 11 items then fit the Rasch model. That is, these items formed a unidimensional measure of visual functioning that could be used in the comparisons along with previously proposed short-form versions. This new version is referred to here as the VF-11R (R for Rasch) (Table 2).

In the VF-11R, certain items possessed the same difficulty level as others. This suggested redundancy in the measure and that further items in the VF-11R could be removed. The following criteria were used to drive the selection of items to be retained in the short-form: (1) maintain a minimum person separation value of 2.0 and (2) maintain targeting.

Two further items were removed from the VF-11R. In this process, an additional item also misfit and was deleted. Thus, 8 items remained in this unidimensional short-form version, which is referred to here as the VF-8R (Table 1). In terms of being a unidimensional measure of visual functioning, the VF-8R was superior to the VF-14, although person separation and targeting were marginally lower than for the VF-11R (Table 2). Nevertheless the VF-8R was shorter than the original scale by 6 items. The reliability of these short-form versions was not tested.

To fulfill the study's main aim of determining the best version of VF-14 for assessing the change in visual functioning after cataract surgery, the VF-11R and VF-8R were appended to the existing list of the 5 shortened versions of the VF-14.[8–11, 13]

Statistical Analysis

For the Rasch analysis of the outcomes, the data obtained from the preoperative patients and postoperative patients were combined; that is, all data were assembled in a single data set, with the postoperative data treated as "new patients".[24] Preoperative and postoperative visual functioning scores (in logits) were then estimated for each patient. This

Table 2. Overall performance of the VF-14 and the included short-form versions of the VF-14.

Parameter	VF-14	VF-11R*	VF-8R*	VF-11†	VF-7U†	VF-7P†	VF-9MF†	VF-9MM†	VF-9L*
Misfitting items (n)	2	0	0	1	1	1	0	2	1
Person separation	2.45	2.46	2.29	2.29	1.86	2.07	2.31	2.18	2.73
Mean item location	0	0	0	0	0	0	0	0	0
Mean person location	1.86	2.57	1.97	1.39	1.53	1.75	2.64	1.67	2.26
Principal components analysis (eigenvalue)	2.3	1.6	1.6	2.2	1.6	1.6	1.7	2.3	1.7

VF-14 = Visual Functioning Index 14 (14 items[4]; VF-11R = 11 items (Rasch scaled version from present study); VF-8R = 8 items (Rasch scaled version from present study); VF-11 = 11 items (Friedman et al.[13]); VF-7U = 7 items (Uusitalo et al.[8]); VF-7P = 7 items (Pager[9]); VF-9MF = 9 items for females (Moghimi et al.[10]); VF-9MM = 9 items for males (Moghimi et al.[10]); VF-9L = 9 items (Lamoureux et al.[11])
*Rasch-scaled versions
†Non Rasch–scaled versions

was done so that the preoperative and postoperative scores were derived on the same scale and would therefore provide an accurate measure of outcomes.

A 1-way analysis of variance (ANOVA) was used to determine whether the change in preoperative to postoperative score for the original VF-14 and each shortened version differed significantly from zero. The F statistic with a *P* value less than 0.05 was considered significant. Relative precision was then used to examine how well each version of the VF-14 distinguished visual functioning between preoperative and postoperative periods, relative to the Likert scoring of the original VF-14.[25] Relative precision is a ratio of pairwise F statistics (F for each version versus F for the Likert scoring of VF-14). The extent to which the relative precision ratio differed from 1.0 indicated the degree to which the 2 scoring methods differed in their ability to detect the change in scores; values greater than 1.0 indicated increased precision.

To maximize comparability, the ordinal raw scores (from VF-14, VF-11, VF-7U, VF-7P, VF-9MF, and VF-9MM) and Rasch measures (from VF-11R, VF-8R, and VF-9L) were transformed from their original scale to a 0 to 100 metric; minimum visual functioning (maximum difficulty) was set at 0 and maximum visual functioning (minimum difficulty), at 100.[26]

SPSS for Windows software (version 15.0, SPSS, Inc.) was used for all general descriptive statistics. A paired *t* test was used to compare improvements in visual acuity within the group for those with ocular comorbidity and without ocular comorbidity. Independent-samples *t* tests were used to compare the improvement in visual acuity between these groups. A *P* value less than 0.05 was considered statistically significant.

RESULTS

Response and Patient Characteristics

The VF-14 was mailed to 414 patients, of whom 210 (50.7% response rate) returned the completed questionnaire. Postoperatively, 51 of the 81 patients who were mailed the VF-14 returned it (62.9% response rate). Table 3 shows the baseline characteristics of the patients by group.

Clinical Outcomes

Combining the data of the preoperative patients and postoperative patients for Rasch analysis of the outcomes yielded 102 patient records. Table 4 shows the mean preoperative and postoperative visual acuity values in the operated eyes and fellow eyes. Visual acuity improved significantly from preoperatively to postoperatively overall ($P < .0001$) and in the comorbidity subgroup ($P < .0001$) and no-comorbidity subgroup ($P = .02$). The final postoperative visual acuity was not significantly different between the 3 groups ($F = 2.69$ and $P = .08$, ANOVA).

Relative Precision: Clinical Discrimination

Tables 5, 6, and 7 show the mean preoperative and postoperative scores (and mean change) for the VF-14 and the various short-form versions in the overall group, the ocular comorbidity subgroup, and the no-ocular comorbidity subgroup, respectively. Overall, regardless of the scoring method used, the mean postoperative scores were consistently higher than the preoperative scores across all versions (Table 5). The

Table 3. Baseline sociodemographic and clinical characteristics of the cataract patients who completed the VF-14.

Characteristic	Group	
	Development	Outcomes
Patients (n)	210	51
Mean age (y) ± SD	74.3 ± 9.3	73.0 ± 7.5
Sex, n (%)		
Male	88 (42)	29 (57)
Female	122 (58)	22 (43)
Ocular comorbidity,* n (%)		
Present	98 (48)	30 (59)
Absent	106 (52)	21 (41)
Systemic comorbidity,† n (%)		
Present	142 (84)	40 (78)
Absent	27 (16)	11 (22)

*Includes age-related macular degeneration, glaucoma, diabetic retinopathy, etc. Data were missing for 6 cases in the development group.
†Includes hypertension, diabetes, angina, etc. Data were missing for 41 cases in the development group.

Table 4. Preoperative and postoperative visual acuity of 51 patients who had cataract surgery and completed the VF-14.

Group/Exam Time	Visual Acuity
Operated eyes	
All (n = 51)	
Preoperative*	
Mean logMAR ± SD	0.52 ± 0.40
Range	0.00 to 2.00
Snellen	$6/19^{-1}$
Postoperative*	
Mean logMAR ± SD	0.18 ± 0.21
Range	−0.12 to 0.80
Snellen	$6/7.5^{-1}$
With comorbidity (n = 30)	
Preoperative*	
Mean logMAR ± SD	0.41 ± 0.32
Range	0.00 to 1.30
Snellen	6/15
Postoperative*	
Mean logMAR ± SD	0.23 ± 0.21
Range	−0.10 to 0.80
Snellen	$6/9.5^{-1}$
With no comorbidity (n = 20)	
Preoperative (better eye)	
Mean logMAR ± SD	0.69 ± 0.45
Range	0.10 to 2.00
Snellen	$6/30^{-1}$
Postoperative (better eye) †	
Mean logMAR ± SD	0.07 ± 0.17
Range	−0.12 to 0.44
Snellen	$6/7.5^{+1}$
Fellow eyes	
All	
Mean logMAR ± SD	0.20 ± 0.20
Range	−0.30 to 0.80
Snellen	6/9.5
With comorbidity	
Mean logMAR ± SD	0.18 ± 0.18
Range	−0.30 to 0.50
Snellen	$6/9.5^{+1}$
Without comorbidity	
Mean logMAR ± SD	0.22 ± 0.24
Range	−0.1 to 0.80
Snellen	$6/9.5^{-1}$

Notes on logMAR values: 1.3 represents visual acuity of 3/60 or 6/120; 2.00 represents light perception, 0 represents 6/6, negative logMAR values indicate visual acuity of better than 6/6.
Snellen notation: Minus sign in the superscript indicates patient could not read the line completely and missed letters, for example, $6/19^{-2}$ indicates patient missed 2 letters from this line. Plus sign indicates patient read this line completely correctly and read 2 more letters correctly in the subsequent smaller line
*$P < 0.0001$ (Paired t test)
†$P = 0.02$ (Paired t test)

largest improvement in scores occurred for the VF-8R. Figure 1 shows the relative distribution of the VF-14 and VF-8R scores preoperatively and postoperatively in the overall group.

As hypothesized, all Rasch-scaled versions achieved significantly greater gains in precision in discriminating between visual functioning of preoperative and postoperative patients (Table 5). The gain in precision compared with the original Likert scored VF-14 was 98% for the VF-11R, 125% for the VF-8R, and 98% for the VF-9L.

Similar to the overall group, the mean postoperative scores were significantly higher than the preoperative scores in both subgroups (Tables 6 and 7). The gain in precision was consistently the largest for VF-8R with ocular comorbidity (114%) and without ocular comorbidity (148%). In the subgroup without ocular comorbidity, 2 Likert-scored versions (VF-11 and the VF-9MM) had less precision (12% and 22%, respectively) than the VF-14.

DISCUSSION

After cataract surgery, visual acuity improved significantly overall (by a mean of 3.4 lines) and in both subgroups, with the largest gains in eyes without ocular comorbidity (6.2 lines). Visual acuity is, of course, a surrogate for visual functioning, albeit limited to the high contrast acuity spectrum of function. More important, visual functioning also improved significantly overall and in both subgroups. For example, postoperatively, patients in the overall group had a mean VF-8R Rasch-score of 83.15 logits (15.39-logit improvement from preoperative assessment), while the ocular comorbidity subgroup had a mean VF-8R Rasch score gain of 13.87 logits, and the no comorbidity subgroup gained 17.35 logits. Similar improvements, albeit smaller in magnitude by comparison, were observed for the VF-14 and the other 7 short-form versions.

The main objective of our study was to determine the best short-form version of the VF-14 by comparing the relative precision of 8 short-form versions against the original VF-14 in measuring the outcomes of cataract surgery. We found larger gains in precision for Rasch-scoring (range of relative precision 98% to 125% increase) in discriminating the visual functioning in the overall group; the largest gain of 125% was for the VF-8R (relative precision, 2.25). Similar large gains were observed for Rasch-scoring across both subgroups. In fact, the largest gain in precision (relative precision = 2.48) was for the VF-8R in discriminating the visual functioning for those who did not have ocular comorbidity. That is, the precision of VF-8R in this subgroup was 2.48 times better than that of the original VF-14. Thus, the results in our study provide strong evidence of the benefits of Rasch-scaling questionnaires. These results are consistent with those of other researchers, who have also showed the benefits

Table 5. Mean preoperative and postoperative scores for cataract surgery patients (overall, n = 51) and relative precision for the VF-14 and the included short-form versions of the VF-14.

Version	Mean ± SE		Mean Differences* ± SE: Preop Vs Postop	F Statistic†	Relative Precision‡
	Preoperative	Postoperative			
VF-14/Likert	82.49 ± 1.99	90.61 ± 1.79	8.12 ± 1.87	9.18	1.00
VF-11R/Rasch	79.59 ± 1.50	88.92 ± 1.59	9.33 ± 1.61	18.14	1.98
VF-8R/Rasch	67.75 ± 2.36	83.15 ± 2.43	15.39 ± 2.66	20.67	2.25
VF-11/Likert	78.68 ± 2.20	89.43 ± 2.07	10.75 ± 2.31	12.66	1.38
VF-7U/Likert	78.17 ± 2.10	88.37 ± 1.96	10.20 ± 2.00	12.57	1.37
VF-7P/Likert	77.26 ± 2.38	90.17 ± 2.10	12.91 ± 2.43	16.53	1.80
VF-9MF/Likert	83.18 ± 1.95	92.14 ± 1.73	8.95 ± 1.68	11.77	1.28
VF-9MM/Likert	81.34 ± 2.03	90.50 ± 1.82	9.16 ± 1.85	11.27	1.23
VF-9L/Rasch	79.49 ± 1.55	89.17 ± 1.66	9.68 ± 1.68	18.14	1.98

SE = standard error
*The follow-up time for self-administration of the VF-14 postoperatively was a minimum of 6 months from the date of surgery. The mean difference was calculated by subtracting the postoperative score from the preoperative score, with a positive result indicating a gain postoperatively.
†$P < .05$
‡Relative precision was calculated by dividing the F statistic for each version by that of the VF-14 (as baseline).

of Rasch-scaled versions over Likert scores for ophthalmic and nonophthalmic questionnaires.[25–28]

The main reason the Rasch-scaled versions had relatively greater precision in measuring outcomes is the reduction in error in estimating the measurement of visual disability, as evidenced by reduced standard errors of the measures.[25,29] Smaller standard errors, typical of Rasch scaling, were noted in the present study for the VF-11R and VF-9L, but not for the VF-8R.[28] Second, as a result of logistic transformation, Rasch-scaling increases measurement precision by expanding the range of measurement. It is the larger range of measurement for the VF-8R that probably caused its increased standard errors, although further reliability testing of this version could be informative. In contrast, Likert-scaled scores are constrained at each end of the scale. The larger range of measurement in the Rasch-scaled versions implies reduced ceiling and floor effects (ie, patients with extreme scores), as was evidenced with the use of VF-8R. Patients with

Table 6. Mean preoperative and postoperative scores for cataract surgery patients who had ocular comorbidity (n = 30) and relative precision for the VF-14 and the included short-form versions of the VF-14.

Version	Mean ± SE		Mean Differences* ± SE: Preop Vs Postop	F Statistic	Relative Precision¶
	Preoperative	Postoperative			
VF-14/Likert	81.54 ± 2.78	89.21 ± 2.76	7.66 ± 2.36	3.81†	1.00
VF-11R/Rasch	79.45 ± 2.13	87.69 ± 2.35	8.24 ± 2.26	6.73‡	1.77
VF-8R/Rasch	67.71 ± 3.29	81.58 ± 3.57	13.87 ± 3.81	8.15‡	2.14
VF-11/Likert	77.87 ± 2.94	88.60 ± 2.99	10.73 ± 2.77	6.54‡	1.72
VF-7U/Likert	76.11 ± 2.91	86.73 ± 2.99	10.62 ± 2.83	6.48‡	1.70
VF- 7P/Likert	77.54 ± 3.17	88.21 ± 3.33	10.66 ± 3.33	5.37‡	1.41
VF-9MF/Likert	82.51 ± 2.79	90.18 ± 2.74	7.67 ± 2.22	3.84‡	1.01
VF-9MM/Likert	80.01 ± 2.74	89.39 ± 2.74	9.31 ± 2.14	5.77‡	1.51
VF-9L/Rasch	79.59 ± 2.18	88.30 ± 2.46	8.71 ± 2.35	7.02‡	1.84

SE = standard error
*The follow-up time for self-administration of the VF-14 postoperatively was a minimum of 6 months from the date of surgery. The mean difference was calculated by subtracting the postoperative score from the preoperative score, with a positive result indicating a gain postoperatively.
†$P > .05$ for VF-14 only
‡$P < .05$
¶Relative precision was calculated by dividing the F statistic for each version by that of the VF-14 (as baseline).

Table 7. Mean preoperative and postoperative scores for cataract surgery patients who did not have ocular comorbidity (n = 20) and relative precision for the VF-14 and the included short-form versions of the VF-14.

Version	Mean ± SE		Mean Differences* ± SE: Preop Vs Postop	F statistic†	Relative Precision‡
	Preoperative	Postoperative			
VF-14/Likert	84.97 ± 2.71	92.97 ± 1.86	8.00 ± 3.20	5.92	1.00
VF-11R/Rasch	80.50 ± 2.03	91.20 ± 1.94	10.70 ± 2.31	14.44	2.44
VF-8R/Rasch	68.87 ± 3.36	86.22 ± 3.03	17.35 ± 3.72	14.70	2.48
VF-11/Likert	81.09 ± 3.29	91.02 ± 2.82	9.93 ± 4.17	5.26	0.89
VF-7U/Likert	82.49 ± 2.69	91.70 ± 1.97	9.21 ± 2.92	7.62	1.29
VF- 7P/Likert	78.21 ± 3.59	93.46 ± 1.79	15.25 ± 3.54	14.42	2.43
VF-9MF/Likert	85.29 ± 2.49	95.15 ± 1.46	9.86 ± 2.53	11.63	1.96
VF-9MM/Likert	84.43 ± 2.92	92.39 ± 2.14	7.96 ± 3.39	4.84	0.82
VF-9L/Rasch	80.04 ± 2.16	91.04 ± 2.02	11.03 ± 2.47	13.90	2.35

SE = standard error
*The follow-up time for self-administration of the VF-14 postoperatively was a minimum of 6 months from the date of surgery. The mean difference was calculated by subtracting the postoperative score from the preoperative score, with a positive result indicating a gain postoperatively.
†$P < .05$
‡Relative precision was calculated by dividing the F statistic for each version by that of the VF-14 (as baseline).

high visual functioning scored at the upper end of the VF-8R, while those with low visual functioning scored at the lower end. Although it appears as though there was some truncation of measurement in the postoperative samples, the truncation seemed to be less with the VF-8R (Figure 1).

Nevertheless, the overarching question is which version(s) of VF-14 should be used for assessing outcomes of cataract surgery? Our results clearly indicate that the Rasch-scaled VF-8R is the most appropriate. There are many potential benefits to using it. First, it provides interval-level measurement, making comparison between patients meaningful. Second, all items measure a single construct of visual functioning (implying unidimensionality, which is an essential measurement property); this is unlike the original VF-14, which is confounded by more than 1 construct. Third, it has better measurement precision for discriminating outcomes, indicating a smaller sample size will be required to find significant differences. Finally, with only 8 items, respondent burden and administration time are minimal.

The proposed VF-8R version is not without limitations. It has suboptimum targeting, marginally lower than the original VF-14. Except for the Catquest-9SF,[2] problems with targeting (ie, items being too easy) have been evident for cataract patients with all other questionnaires.[1,19,20] There may also be marginal differences in patient response if the questionnaire were administered in an 8-item format instead of a 14-item format[30]; however, this has not been tested.

In conclusion, our results show that Rasch-scaled versions of VF-14 perform better than Likert-scored versions. In particular, the VF-8R measures cataract surgery outcomes with high precision, possesses psychometric properties comparable to those of the original VF-14, and performs even better than VF-14 in terms of measuring a single construct. Given these benefits, we believe the VF-8R would prove to be a superior tool in cataract outcomes assessment.

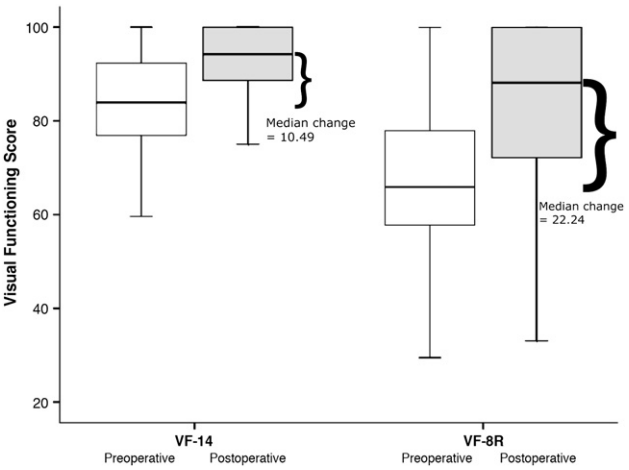


Figure 1. Box-and-whisker plot of preoperative scores (*empty boxes*) and postoperative scores (*solid boxes*) of visual functioning in the overall group of patients (n = 51) using the VF-14 and the Rasch-scaled version, the VF-8R. The boxes contain the interquartile range, and the line running across the center of each box represents the median. The change in the median score was statistically significantly larger for the VF-8R than for the VF-14 (both $P < .0001$, paired *t* test).

REFERENCES

1. Pesudovs K, Garamendi E, Keeves JP, Elliott DB. The Activities of Daily Vision Scale for cataract surgery outcomes: re-evaluating validity with Rasch analysis. Invest Ophthalmol Vis Sci 2003; 44:2892–2899. Available at: http://www.iovs.org/cgi/reprint/44/7/2892. Accessed March 29, 2010
2. Lundström M, Pesudovs K. Catquest-9SF patient outcomes questionnaire; nine-item short-form Rasch-scaled revision of the Catquest questionnaire. J Cataract Refract Surg 2009; 35:504–513
3. Steinberg EP, Tielsch JM, Schein OD, Javitt JC, Sharkey P, Cassard SD, Legro MW, Diener-West M, Bass EB, Damiano AM, Steinwachs DM, Sommer A. National study of cataract surgery outcomes; variation in 4-month postoperative outcomes as reflected in multiple outcome measures. Ophthalmology 1994; 101:1131–1140; discussion by DM O'Day, 1140–1131
4. Steinberg EP, Tielsch JM, Schein OD, Javitt JC, Sharkey P, Cassard SD, Legro MW, Diener-West M, Bass EB, Damiano AM, Steinwachs DM, Sommer A. The VF-14. An index of functional impairment in patients with cataract. Arch Ophthalmol 1994; 112:630–638. Available at: http://archopht.ama-assn.org/cgi/reprint/112/5/630. Accessed March 29, 2010
5. Cassard SD, Patrick DL, Damiano AM, Legro MW, Tielsch JM, Diener-West M, Schein OD, Javitt JC, Bass EB, Steinberg EP. Reproducibility and responsiveness of the VF-14. An index of functional impairment in patients with cataracts. Arch Ophthalmol 1995; 113:1508–1513. Available at: http://archopht.ama-assn.org/cgi/reprint/113/12/1508. Accessed March 29, 2010
6. Bilbao A, Quintana JM, Escobar A, García S, Andradas E, Baré M, Elizalde B. Responsiveness and clinically important differences for the VF-14 index, SF-36, and visual acuity in patients undergoing cataract surgery; for the IRYSS-Cataract Group. Ophthalmology 2009; 116:418–424; e1
7. Alonso J, Espallargues M, Andersen TF, Cassard SD, Dunn E, Bernth-Petersen P, Norregaard JC, Black C, Steinberg EP, Anderson GF. International applicability of the VF-14; an index of visual function in patients with cataracts. Ophthalmology 1997; 104:799–807
8. Uusitalo RJ, Brans T, Pessi T, Tarkkanen A. Evaluating cataract surgery gains by assessing patients' quality of life using the VF-7. J Cataract Refract Surg 1999; 25:989–994
9. Pager CK. Assessment of visual satisfaction and function after cataract surgery. J Cataract Refract Surg 2004; 30:2510–2516
10. Moghimi S, Riazi Esfahani M, Maghsoudipour M. Visual function after implantation of aniridia intraocular lens for traumatic aniridia in vitrectomized eye. Eur J Ophthalmol 2007; 17:660–665
11. Lamoureux EL, Pesudovs K, Thumboo J, Saw S-M, Wong TY. An evaluation of the reliability and validity of the visual functioning questionnaire (VF-11) using Rasch analysis in an Asian population. Invest Ophthalmol Vis Sci 2009; 50:2607–2613
12. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient α in the creation of minimum item sets. Med Care 2004; 42(suppl 1):I-17–I-24
13. Friedman DS, Tielsch JM, Vitale S, Bass EB, Schein OD, Steinberg EP. VF-14 item specific responses in patients undergoing first eye cataract surgery: can the length of the VF-14 be reduced? Br J Ophthalmol 2002; 86:885–891. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1771229/pdf/bjo08600885.pdf. Accessed March 29, 2010
14. McHorney CA, Ware JE Jr, Rogers W, Raczek AE, Lu JFR. The validity and relative precision of MOS short- and long-form health status scales and Dartmouth COOP charts; results from the Medical Outcomes Study. Med Care 1992; 30(suppl 5):MS253–MS265
15. Kirkwood BJ, Pesudovs K, Latimer P, Coster DJ. The efficacy of a nurse-led preoperative cataract assessment and postoperative care clinic. Med J Aust 2006; 184:278–281. Available at: http://www.mja.com.au/public/issues/184_06_200306/kir10534_fm.pdf. Accessed March 29, 2010
16. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. Invest Ophthalmol Vis Sci 1991; 32:422–432. Available at: http://www.iovs.org/cgi/reprint/32/2/422.pdf. Accessed March 29, 2010
17. Mallinson T. Why measurement matters for measuring patient vision outcomes. Optom Vis Sci 2007; 84:E675–E682. Available at: http://journals.lww.com/optvissci/Fulltext/2007/08000/Why_Measurement_Matters_for_Measuring_Patient.6.aspx. Accessed March 29, 2010
18. Wright BD, Masters GN. Rating Scale Analysis (Rasch Measurement Book). Chicago, IL, MESA Press, 1982
19. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Rasch Analysis of Visual Function and Quality of Life Questionnaires. Optom Vis Sci 2009; 86:1160–1168
20. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Rasch analysis of the Quality of Life and Vision Function Questionnaire. Optom Vis Sci 2009; 86:E836–E844
21. Wright BD, Stone MH. Best Test Design; Rausch Measurement. Chicago, IL, MESA Press, 1979
22. Linacre JM. WINSTEPS® Rasch measurement computer program. Beaverton, OR, Winsteps.com, 2009
23. Andrich D. A rating scale formulation for ordered response categories. Psychometrika 1978; 43:561–573
24. Wolfe EW, Chiu CW. Measuring change across multiple occasions using the Rasch Rating Scale Model. J Outcome Meas 1999; 3:360–381
25. McHorney CA, Haley SM, Ware JE Jr. Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. J Clin Epidemiol 1997; 50:451–461
26. Garamendi E, Pesudovs K, Stevens MJ, Elliott DB. The Refractive Status and Vision Profile: evaluation of psychometric properties and comparison of Rasch and summated Likert-scaling. Vision Res 2006; 46:1375–1383
27. Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C. Comparing alternative Rasch-based methods vs raw scores in measuring change in health. Med Care 2004; 42(suppl 1):I-25–I-36
28. Fitzpatrick R, Norquist JM, Dawson J, Jenkinson C. Rasch scoring of outcomes of total hip replacement. J Clin Epidemiol 2003; 56:68–74
29. Raczek AE, Ware JE, Bjorner JB, Gandek B, Haley SM, Aaronson NK, Apolone G, Bech P, Brazier JE, Bullinger M, Sullivan M. Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: results from the IQOLA Project. J Clin Epidemiol 1998; 51:1203–1214
30. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. Optom Vis Sci 2007; 84:663–674. Available at: http://www.optvissci.com/pt/re/ovs/pdfhandler.00006324-200708000-00005.pdf;jsessionid=KZcJrBxMKHzzcx4VGhCDgGwV2vDSD5HknjcG729LKxDlX1XPvfm1!-701738752!181195629!8091!-1. Accessed March 29, 2010

First author
Vijaya K Gothwal, PhD

Department of Ophthalmology, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1549 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Cataracts: Patient Satisfaction within 90 Days Following Cataract Surgery
De.2 **Brief description of measure:** Percentage of patients aged 18 years and older who had cataract surgery and were satisfied with their care within 90 days following the cataract surgery
1.1-2 **Type of Measure:** Patient experience De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** This is intended to be included in a composite measure for cataract surgery to provide a comprehensive evaluation of both the clinical and patient-centered outcomes. This group includes approved NQF measures and PQRI measures Measures 191 – 20/40 or better visual acuity within 90 days following cataract surgery and 192 – complications within 30 days of cataract surgery requiring additional surgical procedures, and a newly submitted measure: Improvement in Patient's Visual Function within 90 Days Following Cataract Surgery
De.4 **National Priority Partners Priority Area:** Patient and family engagement De.5 **IOM Quality Domain**: Patient-centered De.6 **Consumer Care Need:** Getting better

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	NQF Staff
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of	**A** Y☐ N☐

measure submission A.4 **Measure Steward Agreement attached:** txNQFMeasureStewardAgreement_020309_Final-634278446871486346.pdf	
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, Internal quality improvement Payment incentive	**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, Frequently performed procedure, Leading cause of morbidity/mortality, High resource use 1a.2 1a.3 **Summary of Evidence of High Impact:** Cataracts are the leading cause of blindness worldwide and remain an important cause of blindness and visual impairment in the United States, accounting for approximately 50% of visual impairment in adults over the age of 40. Cataracts are the leading cause of treatable blindness among Americans of African descent age 40 and older and are the leading cause of visual impairment among Americans of African, Hispanic/Latino, and European descent. Cataract surgery with IOL implantation was the most frequently performed operation and the single largest expenditure for any Part B surgical procedure in the Medicare program, calculated by Part B procedure codes based on allowed charges. In 2008 (latest year available), payment for cataract was $2.1 billion, which is 1.8% of total allowed charges. 1a.4 **Citations for Evidence of High Impact:** 1. Congdon N, O´Colmain B, Klaver CC, et al. Causes and prevalence of visual impairment among adults in the United States. Arch Ophthalmol 2004;122:477-85.	**1a** C☐ P☐ M☐ N☐

2. Cotter SA, Varma R, Ying-Lai M, et al. Causes of low vision and blindness in adult Latinos: the Los Angeles Latino Eye Study. Ophthalmology 2006;113:1574-82.
3. Centers for Medicare and Medicaid Services. Medicare leading Part B procedure codes based on allowed charges: calendar year 2010. Available at: www.cms.hhs.gov/datacompendium/. Accessed December 10, 2010.

1b. Opportunity for Improvement

1b.1 **Benefits (improvements in quality) envisioned by use of this measure:** The benefits are to enhance satisfaction of patients receiving cataract surgery. The primary indication of surgery is visual function that no longer meets the patient´s needs and for which cataract surgery provides a reasonable likelihood of improved vision, leading to satisfaction.

1b.2 **Summary of data demonstrating performance gap (variation or overall poor performance) across providers:**
This is an outcome of surgery indicator of direct relevance and importance to patients, their families and referring providers. The available evidence suggests that satisfaction with cataract surgery is found in about 90% of patients surveyed. While the potential for improvement appears seemingly small, the volume of cataract surgery in the U.S. of over 2.8 million surgeries means that the impact could affect more than 280,000 patients per year. Ideally, performance on this indicator would be as high as possible,with lower rates suggestive of opportunities for improvement.

1b.3 **Citations for data on performance gap:**
1. Mozaffarieh M, Krepler K, Heinzl H et al. Visual function, quality of life and patient satisfaction after ophthalmic surgery: a comparative study. Ophthalmologica 2004; 218:26-30.
2. Lledo R, Rodriguez T, Fontenia JR et al. Cataract surgery: An analysis of patient satisfaction with medical care. International Ophthalmology 22:227-32.
3. Lum F, Schein O, Schachat AP, et al. Initial two years of experience with the AAO National Eyecare Outcomes Network (NEON) cataract surgery database. Ophthalmology 2000; 107:691-7.
4. Lum F, Schachat AP, Jampel HD. The development and demise of a cataract surgery database. The Joint Commission Journal on Quality Improvement 2202; 28:108-114.

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

1b.5 **Citations for data on Disparities:**

1b
C☐
P☐
M☐
N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 **Relationship to Outcomes** (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Patient satisfaction is a relevant, patient-centered patient experience type outcome for cataract surgery.

1c.2-3. **Type of Evidence:** Evidence-based guideline

1c.4 **Summary of Evidence** (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Several constructs have been found to be associated with patient satisfaction, with the physician having control over several of these. Some of these constructs include: physician-patient communication, information, accessibility, quality of medical care and outcomes, premises, professional care, length of communication, caring/trust, interpersonal skills, affordability of care, etc. Physician-patient communications and patient´s understanding of expectations and outcomes is a critical construct.

In the focus groups conducted for the S-CAHPS instrument, the following three constructs were identified as drivers of surgical care experience (good or bad):
1. surgeon´s interpersonal skills and behaviors
2. surgeon´s expertise/technical competence
3. surgeon´s skill in communicating and providing health information and patient education

1c
C☐
P☐
M☐
N☐

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
Not rated in guideline because it does not serve as a treatment recommendation

1c.6 Method for rating evidence: The panel rated each recommendation on the strength of evidence in the available literature to support the recommendation made. The "ratings of strength of evidence" also are divided into three levels.
Level I includes evidence obtained from at least one properly conducted, well-designed, randomized controlled trial. It could include meta-analyses of randomized controlled trials.
Level II includes evidence obtained from the following:
- Well-designed controlled trials without randomization
- Well-designed cohort or case-control analytic studies, preferably from more than one center
- Multiple-time series with or without the intervention

Level III includes evidence obtained from one of the following:
- Descriptive studies
- Case reports
- Reports of expert committees/organizations (e.g., PPP panel consensus with peer review)

The I, II, and III can also be correlated with the USPSTF system of high, moderate and low.

1c.7 Summary of Controversy/Contradictory Evidence:

1c.8 Citations for Evidence (*other than guidelines***):** 1. Schein OD, Steinberg EP, Javitt JC, et al. Variation in cataract surgery practice and clinical outcomes. Ophthalmology 1994;101:1142-52.
2. Mangione CM, Phillips RS, Lawrence MG, et al. Improved visual function and attenuation of declines in health-related quality of life after cataract extraction. Arch Ophthalmol 1994;112:1419-25.
3. Desai P, Minassian DC, Reidy A. National cataract surgery survey 1997-8: a report of the results of the clinical outcomes. Br J Ophthalmol 1999;83:1336-40.
4. McGwin G, Jr, Scilley K, Brown J, Owsley C. Impact of cataract surgery on self-reported visual difficulties: comparison with a no-surgery reference group. J Cataract Refract Surg 2003;29:941-8.
5. Colin J, El Kebir S, Eydoux E, Hoang-Xuan T, Rozot P, Weiser M. Assessment of patient satisfaction with outcomes of and ophthalmic care of cataract surgery. J Cataract Refract Surg. 2010 Aug;36(8):1373-9.
6. Nijkamp MD, Nuijts RM, Borne B, Webers CA, van der Horst F, Hendrikse F. Determinants of patient satisfaction after cataract surgery in 3 settings. J Cataract Refract Surg 2000 Sep;26(9):1379-88.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
In well-designed observational studies, cataract surgery consistently has been shown to have a significant impact on vision-dependent function; up to 90% of patients undergoing first-eye cataract surgery note improvement in functional status and satisfaction with vision.

Also, the guideline outlines the ophthalmologist´s responsibility for communication to the patient:

The ophthalmologist who is to perform the cataract surgery has the following responsibilities:
- To examine the patient preoperatively (see Ophthalmic Evaluation).[A:III]
- To ensure that the evaluation accurately documents the symptoms, findings, and indications for treatment.[A:III]
- To obtain informed consent from the patient or the patient's surrogate decision maker after discussing the risks, benefits, and expected outcomes of surgery, including anticipated refractive outcome and the surgical experience.[A:III]
- To review the results of presurgical and diagnostic evaluations with the patient or the patient's surrogate decision maker.[A:III]
- To formulate a surgical plan, including selection of an appropriate IOL.[A:III]
- To formulate postoperative care plans and inform the patient or the patient's surrogate decision maker of these arrangements (setting of care, individuals who will provide care).[A:III]
- To afford the patient or the patient's surrogate decision maker the opportunity to discuss the costs

associated with surgery.[B:III] 1c.10 **Clinical Practice Guideline Citation:** American Academy of Ophthalmology. Cataract in the Adult Eye, Preferred Practice Pattern. San Francisco: American Academy of Ophthalmology, 2006. Available at: www.aao.org/ppp. 1c.11 **National Guideline Clearinghouse or other URL:** http://www.guideline.gov/content.aspx?id=10173&search=cataract+and+cataract+2005+and+cataract+2006 1c.12 **Rating of strength of recommendation** (*also provide narrative description of the rating and by whom*)**:** The ratings for communications to the patients are rated [A:III] which indicates the highest importance to care rating, based on expert opinion/consensus evidence. 1c.13 **Method for rating strength of recommendation** (*If different from USPSTF system, also describe rating and how it relates to USPSTF)***:** The panel rated each recommendation according to its importance to the care process. This "importance to the care process" rating represents care that the panel thought would improve the quality of the patient's care in a meaningful way. The ratings of importance are divided into three levels. - Level A, defined as most important - Level B, defined as moderately important - Level C, defined as relevant but not critical The A, B, C ratings can be correlated with the USPSTF system of A, B, C for strength of recommendation. 1c.14 **Rationale for using this guideline over others:** This guideline is the only United States guideline on cataract surgery contained in the National Guideline Clearinghouse.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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*)**:** Patients who were satisfied with their care within 90 days following cataract surgery. Valid exclusions for not performing the measure for the reporting calculation include: • The patient refuses to participate • The patient is unable to complete the questionnaire 2a.2 **Numerator Time Window** (*The time period in which cases are eligible for inclusion in the numerator*)**:** One year 2a.3 **Numerator Details** (*All information required to collect/calculate the numerator, including all codes, logic, and definitions*)**:** A. Patients who were satisfied with their care within 90 days following cataract surgery	**2a-specs** **C☐** **P☐** **M☐** **N☐**

C.Patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery but for whom there is a documented patient reason for not doing so

D. Patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery and there is no documented patient reason for not doing so

The calculation for Reporting Numerator is as follows:

(A (# of patients meeting measure criteria) + C (# of patients with valid exclusions) + D (# of patients NOT meeting numerator criteria))

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
All patients aged 18 years and older who had cataract surgery

2a.5 Target population gender: Female, Male
2a.6 Target population age range: 18 years and older

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
One year

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
All patients aged 18 years and older who had cataract surgery
- CPT Procedure Codes (with or without modifiers): 66840, 66850, 66852, 66920, 66930, 66940, 66982, 66983, 66984

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** No exclusions for the reporting calculation. If for performance, the denominator exclusions are as follows:
- The patient refuses to participate
- The patient is unable to complete the questionnaire

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
Documentation of patient reason for not completing patient satisfaction assessment within 90 days of cataract surgery
- Append modifier to CPT Category II Code: -2P

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

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
Reporting Calculation
A (# of patients meeting measure criteria) + C (# of patients with valid exclusions) + D (# of patients NOT meeting numerator criteria)
_____
RD (# of patients in denominator)

Components for this measure are defined as:

A # of patients who were satisfied with their care within 90 days following cataract surgery
C # of patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery but for whom there is a documented patient reason for not doing so
D # of patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery and there is no documented patient reason for not doing so
RD # of patients aged 18 years and older who had cataract surgery

2a.22 **Describe the method for discriminating performance** *(e.g., significance testing)*:
Methods would include comparison of means and percentiles and analysis of variance against established benchmarks in the literature.

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 this physician-level measure, it is anticipated to be used as a group or composite measure. Utilizing a sample, work in the field has indicated that a sample size of 30 patients would be adequate for typical practice sizes.

2a.24 **Data Source (***Check the source(s) for which the measure is specified and tested)*
Survey: Patient

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.***):**
Surgical Consumer Assessment of Healthcare Providers and Systems (S-CAPHS)

2a.26-28 **Data source/data collection instrument reference web page URL or attachment:** URL
https://www.cahps.ahrq.gov/content/products/sc/PROD_SC_Surgical_Care.asp?p=1021&s=213

2a.29-31 **Data dictionary/code table web page URL or attachment:**

2a.32-35 **Level of Measurement/Analysis** (*Check the level(s) for which the measure is specified and tested)*
Clinicians: Individual

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

2a.38-41 **Clinical Services** (*Healthcare services being measured, check all that apply*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 **Data/sample** *(description of data/sample and size)*: The field test involved 96 surgeons in 33 different practices, representing a range of surgical specialties. A total of 5,627 adult patients were sent questionnaires, a total of 2,285 completed the questionnaire by mail. The major criteria for patient selection was having had a major surgery as defined by CPT codes with a 90 day global within 3 to 6 months prior to the start of the survey.

2b.2 **Analytic Method** *(type of reliability & rationale, method for testing)*:
Surgeon-level reliability (that is, inter-rater reliability) is based on the theory that consumers who use the same surgeon should generally agree in their assessments of that surgeon. The reliability of aggregate surgeon scores increases with the ratio of between-to-within-surgeon variation in consumer assessments and with the number of respondents (which causes the within-surgeon-variance to shrink). This relationship of between- to within- surgeon variability was examined using analysis of variance with surgeon as the class variable and the consumer assessments as the dependent variable. Standard practice with CAHPS surveys is that surgeon-level reliabilities should be at least 0.25 and ideally greater than 0.40, corresponding to moderate and large effect sizes, respectively.
Internal consistency reliabilities were calculated using Cronbach´s coefficient alpha.

2b
C ☐
P ☐
M ☐
N ☐

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
The testing results for surgeon-level reliability showed that for 3 out of 4 composites, the surgeon-level reliabilities were ideal. The results were as follows for the mail mode group: pre-surgical = 0.50; peri-operative = 0.67; post-surgical = 0.43 and office staff = 0.00. The reliability coefficient of 0 for the fourth composite means that this cannot be used to detect differences among surgeons in the quality of their office staff.
The internal consistency reliabilities were high for three of the four composites and compares favorably to those found for other CAHPS surveys.
The results were as follows for the mail mode group: pre-surgical = 0.82; peri-operative = 0.69; post-surgical = 0.90; and office staff = 0.88. The lower score for the peri-operative composite reflects the heterogeneity of the sample.

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: The field test involved 96 surgeons in 33 different practices, representing a range of surgical specialties. A total of 5,627 adult patients were sent questionnaires, a total of 2,285 completed the questionnaire by mail. The major criteria for patient selection was having had a major surgery as defined by CPT codes with a 90 day global within 3 to 6 months prior to the start of the survey.

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
Structural equation modeling as implemented by PROC CALIS to evaluate the fit of the data to the structure around which the questionnaire was designed. The maximum likelihood estimation method was used, taking into account that simulation studies suggest that the ML method is likely to result in conservative estimates of model fit. These data were also treated as continuous, consistent with the observed imputed values that comprised a portion of the data. The goodness of fit of the model to the data was evaluated using chi-square, the comparative fit index (CFI), the non-normed fit index (NNFI) and the average root mean square residual approximation (RMSEA). Current practice with regard to these indicators of model fit is to: 1) report chi-square and p-values but not to reject models where the p-value is <0.05 in data sets greater than 250 observations; 2) require RMSEA to be less than 0.10 and ideally less than 0.06 and 3) require the CFI and NNFI to be greater than 0.90.
Exploratory factor analysis on the correlation matrix was used with the principle factor method with squared multiple correlations as initial communality estimates and oblique rotation (promax) with Kaiser normalization. In determining the number of factors, the following information was considered: 1) the number of eigen values greater than one; 2) the point at which additional factors explained a trivial amount of variance in the data as evidence by the scree plot; and 3) the interpretability of the rotated vector, based on simple structure. Simple structure was determined by the pattern fo factor loadings after rotation. An item was considered to be conforming to simple structure if it had comparatively larger loadings on one factor and smaller loadings on all others. Large loadings were considered to be those greater than 0.40 and small loadings to be no larger than half the size of the larger loading and less than 0.25.
The investigators reviewed the exploratory factor analysis and used the formative research to select among the candidate composite models. The hypothetical model to be evaluated by the confirmatory factor analysis included 15 items and specified 4 composites concerning the following: Presurgical care; perioperative care, post-surgical followup and quality of interactions with the surgeon´s office staff.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
The results show that the model fit the observed correlation matrix of the mail mode responses reasonably well. The results were X2 = 463, df = 74, CFI = 0.95, NNFI = 0.94 and RMSEA = 0.07. With the combined set of mail and web responses, the results also showed a good fit, with X2 = 513, df = 74, CFI = 0.95, NNFI = 0.93 and RMSEA = 0.06.

The results for the confirmatory factor analysis for the final model found that all t-tests for beta-weights describing the relationship of items to their hypothesized composites were highly significant (p<0.0001), ranging from 0.38 to 0.91.

2c
C☐
P☐
M☐
N☐

2d. Exclusions Justified — **2d**

2d.1 Summary of Evidence supporting exclusion(s):
No measure exclusions except if the patient refuses or is unable to take the survey.

2d.2 Citations for Evidence:

2d.3 Data/sample *(description of data/sample and size)*:

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

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

C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: No risk adjustment strategy was used.

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:

2e.3 Testing Results *(risk model performance metrics)*:

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

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)*: The field test involved 96 surgeons in 33 different practices, representing a range of surgical specialties. A total of 5,627 adult patients were sent questionnaires, a total of 2,285 completed the questionnaire by mail. The major criteria for patient selection was having had a major surgery as defined by CPT codes with a 90 day global within 3 to 6 months prior to the start of the survey.

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
The variability of assessments was evaluated by evaluating the percentage of consumers for whom the highest (i.e., the ceiling effect) and the lowest (i.e., the floor effect) possible scores were tabulated.

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)*:
The percent at the highest score in the mail mode group were as follows: pre-surgical: 70%; peri-operative: 56%; post-surgical: 64%; and office staff: 87%. The results on the office staff indicates that there is little information about differences in the quality of office staff across surgeons. The relatively high ceiling effects on composites is believed to be due to a restricted range of performance in the field test sample, since participating surgeons were volunteers and were not randomly selected. Thus, high performers are likely to have been over-represented in the sample. A random sample of surgeons would probably provide a more accurate picture of the distribution of the composite scores.

2f
C☐ P☐ M☐ N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample *(description of data/sample and size)*: The survey was also administered in a web-based version. The web-based version was completed by 465 of the respondents, who were about 17% of the respondents. This was field tested in the summer of 2008. In terms of modality of questionnaire (mail vs. web-based), this was investigated as a potential case mix adjuster and was not found to have any significant impact.

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

2g
C☐ P☐ M☐ N☐ NA☐

Structural equation modeling as implemented by PROC CALIS to evaluate the fit of the data to the structure around which the questionnaire was designed. The maximum likelihood estimation method was used, taking into account that simulation studies suggest that the ML method is likely to result in conservative estimates of model fit. These data were also treated as continuous, consistent with the observed imputed values that comprised a portion of the data. The goodness of fit of the model to the data was evaluated using chi-square, the comparative fit index (CFI), the non-normed fit index (NNFI) and the average root mean square residual approximation (RMSEA). Current practice with regard to these indicators of model fit is to: 1) report chi-square and p-values but not to reject models where the p-value is <0.05 in data sets greater than 250 observations; 2) require RMSEA to be less than 0.10 and ideally less than 0.06 and 3) require the CFI and NNFI to be greater than 0.90. 2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)*: The web-administered questionnaire is comparable to the mailed questionnaire in terms of reliability and validity estimates. These are the statistics for the internal consistency reliability for the web only version: pre-surgical 0.77; peri-operative = 0.70; post-surgical = 0.87; and office staff = 0.79. The correlation with rating of surgeon was as follows: pre-surgical = 0.69; peri-operative = 0.29; post-surgical = 0.78; and office staff = 0.46. The mean composite scores were also identical to the first decimal point of those in the mail mode: pre-surgical = 3.83; peri-operative = 2.27; post-surgical = 3.79 and office staff = 3.82.	
2h. Disparities in Care 2h.1 **If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: The measure is not stratified 2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:**	**2h** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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:** Not in use but testing completed 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*): 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***):** **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)*: Thirty patients were interviewed for the cognitive testing phase of the project. There were 20 English and 10 Spanish speaking patients who had surgery within the past 12 months. There were two rounds of cognitive testing, one with the initial survey and one with a revised survey.	**3a** C☐ P☐ M☐ N☐

3a.5 Methods *(e.g., focus group, survey, QI project)*: Interviews **3a.6 Results** *(qualitative and/or quantitative results and conclusions)*: The cognitive testing revealed problems in the initial survey drafts. These problems were then addressed and implemented and tested in Round 2 with improvements to facilitate respondent comprehension and to result in more valid and readily interpretable results.	
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:** **5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:**	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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?** Survey	**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 web-based survey could be used and results uploaded into a data registry. Paper survey instruments could be scanned and incorporated into a data registry.	**4b** C☐ P☐ M☐ N☐
4c. Exclusions	**4c** C☐

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.**	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.**	4d C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** There is a burden upon the office practice to survey patients post cataract surgery. The vast majority of patients are elderly and they may require assistance/prompting in responding to the surveys. This then will entail time taken out by the office staff. To ensure compliance with the follow-up service will also require attention. **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** There are costs of data collection and follow up of patients who haven´t filled out the surveys. There are no fees associated with proprietary measures. **4e.3 Evidence for costs:** **4e.4 Business case documentation:**	4e C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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 Academy of Ophthalmology and the Hoskins Center for Quality Eye Care, 655 Beach Street, San Francisco, California, 94109-1336 **Co.2 Point of Contact** Flora, Lum, MD, flum@aao.org, 415-561-8592-	

Measure Developer If different from Measure Steward
Co.3 Organization
American Academy of Ophthalmology and the Hoskins Center for Quality Eye Care, 655 Beach Street, San Francisco, California, 94109-1336

Co.4 Point of Contact
Flora, Lum, MD, flum@aao.org, 415-561-8592-

Co.5 Submitter If different from Measure Steward POC
Flora, Lum, MD, flum@aao.org, 415-561-8592-, American Academy of Ophthalmology and the Hoskins Center for Quality Eye Care

Co.6 Additional organizations that sponsored/participated in measure development
American Society of Cataract and Refractive Surgery

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.
Priscilla Arnold, MD; David Chang, MD; John Thompson, MD, Kevin Miller, MD, Leon Herndon, MD

Ad.2 If adapted, provide name of original measure:
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: 2010
Ad.7 Month and Year of most recent revision: 12, 2010
Ad.8 What is your frequency for review/update of this measure? Every 3 years
Ad.9 When is the next scheduled review/update for this measure? 12, 2013

Ad.10 Copyright statement/disclaimers: Copyright by the American Academy of Ophthalmology 2010

Ad.11 -13 Additional Information web page URL or attachment: Attachment visual functionand patient satisfaction measure Nov 2010-634279328820242414.doc

Date of Submission (*MM/DD/YY*)**:** 03/22/2011

American Academy of Ophthalmology

Eye Care III
Physician Performance Measurement Set

Eye Care Work Group *(specialty)*
Priscilla P. Arnold, MD (Co-Chair) (ophthalmologist)

Surgical Management Subgroup:
David Chang, MD (ophthalmologist**)**
Leon W. Herndon, MD (ophthalmologist)
Kevin Miller MD (ophthalmologist)
John T. Thompson, MD (ophthalmologist)

Staff:

American Academy of Ophthalmology
Flora Lum, MD

Purpose of Measures:
These clinical performance measures, developed by the American Academy of Ophthalmology, are designed for individual quality improvement. Unless otherwise indicated, the measures are also appropriate for accountability if appropriate methodological, statistical, and implementation rules are achieved.

The proposed measures seek to advance performance measures for eye care by including explicit measures of patient visual function and patient satisfaction so as to more directly connect process measures to issues of patient interest, satisfaction, and empowerment.

Accountability Measures:
Measure #1 Cataracts: Improvement in Patient's Visual Function within 90 Days Following Cataract Surgery
Measure #2 Cataracts: Patient Satisfaction within 90 Days Following Cataract Surgery

Intended Audience and Patient Population:
Ophthalmologists may implement these measures if and when they provide the cataract surgery addressed in the measures. The measures are designed for calculating reporting or performance measurement at the individual level.

Measure Specifications
Draft specifications to report on these measures for eye care using administrative (claims) data are included in this document. We have identified codes for these measures, including ICD-9 and CPT (Evaluation & Management Codes, Category I and where Category II codes would apply). Specifications for additional data sources, including EHRs, will be fully developed at a later date.

Measure Exclusions:
For ***process measures***, there exist three categories of reasons for which a patient may be excluded from the denominator of an individual measure:

- **Medical reasons**

Includes:
- not indicated (absence of organ/limb, already received/performed, other)
- contraindicated (patient allergic history, potential adverse drug interaction, other)

- **Patient reasons**

Includes:
- patient declined
- economic, social, or religious reasons
- other patient reasons

- **System reasons**

Includes:
- resources to perform the services not available
- insurance coverage/payor-related limitations
- other reasons attributable to health care delivery system

These measure exclusion categories are not available uniformly across all measures; for each measure, there must be a clear rationale to permit an exclusion for a medical, patient, or system reason. The exclusion of a patient may be reported by appending the appropriate modifier to the CPT Category II code designated for the measure:

- **Medical reasons**: modifier 1P

- **Patient reasons**: modifier 2P

- **System reasons**: modifier 3P

Although this methodology does not require the external reporting of more detailed exclusion data, physicians should document the *specific* reasons for exclusion in patients' medical records for purposes of optimal patient management and audit-readiness. Also, each physician's exclusions data could be self-assessed to identify practice patterns and opportunities for quality improvement.

For example, it is possible for implementers to calculate the percentage of patients that physicians have identified as meeting the criteria for exclusion.

Please refer to documentation for each individual measure for information on the acceptable exclusion categories and the codes and modifiers to be used for reporting.

For ***outcome measures***, there are acceptable reasons for which a patient may be excluded from the denominator. Each specified reason is reportable with a CPT Category II code or CPT Category I code designated for that purpose.

Data Capture and Measure Calculation
This is intended for physicians to collect data on each patient eligible for a measure. Feedback on measures should be available to physicians by patient to facilitate patient management and in aggregate to identify opportunities for improvement across a physician's patient population.

Measure calculations will differ depending on whether a rate is being calculated for performance or reporting purposes.

The method of calculation for performance follows these steps: first, identify the patients who meet the eligibility criteria for the denominator (PD); second, identify which of those patients meet the numerator criteria (A); and third, for those patients who do not meet the numerator criteria, determine whether an appropriate exclusion applies and subtract those patients from the denominator (C). (see examples below)

The methodology also enables implementers to calculate the rates of patient exclusions and to further analyze both low and high rates, as appropriate (see examples below).

The method of calculation for reporting differs. One program which currently focuses on reporting rates is the Centers for Medicare and Medicaid Services (CMS) Physician Quality Reporting Initiative (PQRI). Currently, under that program design, there will be a reporting denominator determined solely from claims data (CPT and ICD-9), which in some cases result in a reporting denominator that is much larger than the eligible population for the performance denominator. Additional components of the reporting denominator are explained below.

The components that make up the numerator for reporting include all patients from the eligible population for which the physician has reported, including: the number of patients who meet the numerator criteria (A), the number of patients for whom valid exclusions apply (C) and also the number of patients who do not meet the numerator criteria (D). These components, where applicable, are summed together to make up the inclusive reporting numerator. The calculation for reporting will be the reporting numerator divided by the reporting denominator. (see examples below).

Examples of calculations for reporting and performance are provided for each measure.

Calculation for Performance

For performance purposes, this measure is calculated by creating a fraction with the following components: Numerator, Denominator, and Denominator Exclusions.

Numerator (A) Includes:
Number of patients meeting numerator criteria

Performance Denominator (PD) Includes:
Number of patients meeting criteria for denominator inclusion

Denominator Exclusions (C) Include:
Number of patients with valid medical, patient or system exclusions (where applicable; will differ by measure)

Performance Calculation

A (# of patients meeting numerator criteria) PD (# patients in denominator) – C (# patients with valid denominator exclusions)

It is also possible to calculate the percentage of patients excluded overall, or excluded by medical, patient, or system reason where applicable**:**

Overall Exclusion Calculation

C (# of patients with any valid exclusion) PD (# patients in denominator)

OR

Exclusion Calculation by Type

C_1 (# patients with medical reason) **PD** (# patients in denominator) C_2 (# patients with patient reason) **PD** (# patients in denominator) C_3 (# patients with system reason) **PD** (# patients in denominator)

Calculation for Reporting

For reporting purposes, this measure is calculated by creating a fraction with the following components: Reporting Numerator and Reporting Denominator

Reporting Numerator includes each of the following components, where applicable. (There may be instances where there are no patients to include in A, C, D, or E).

A. Number of patients meeting additional denominator criteria (for measures where true denominator cannot be determined through ICD-9 and CPT Category I coding alone) AND numerator criteria
C. Number of patients with valid medical, patient or system exclusions (where applicable; will differ by measure)
D. Number of patients not meeting numerator criteria and without a valid exclusion
E. All other patients not meeting additional denominator criteria (for measures where true denominator cannot be determined through ICD-9 and CPT Category I coding alone)

Reporting Denominator (RD) Includes:
RD. Denominator criteria (identifiable through ICD-9 and CPT Category I coding)

Reporting Calculation

A(# of patients meeting additional denominator criteria AND numerator criteria) + **C**(# of patients with valid exclusions) + **D**(# of patients NOT meeting numerator criteria) + **E**(# of patients not meeting additional denominator criteria) **RD** (# of patients in denominator)

Eye Care

Measure #1 Cataracts: Improvement in Patient's Visual Function within 90 Days Following Cataract Surgery

This measure may be used as an Accountability measure.

Clinical Performance Measure
Numerator: Patients who had improvement in visual function achieved within 90 days following cataract surgery **Denominator:** All patients aged 18 years and older who had cataract surgery **Denominator Exclusions:** The patient refuses to participate or the patient is unable to complete the questionnaire, or there is a medical reason **Measure:** Percentage of patients aged 18 years and older who had cataract surgery and had improvement in visual function achieved within 90 days following the cataract surgery
The following clinical recommendation statements are quoted <u>verbatim</u> from the referenced clinical guidelines and represent the evidence base for the measure: **This is an outcomes measure.** As such, no statements in the guideline are specific to this measurement topic.
Rationale for the measure: 1. Scientific basis for measuring visual function outcomes after cataract surgery. Visual function has been described as having multiple components, including central near, intermediate, and distance visual acuity; peripheral vision;[1] visual search; binocular vision; depth perception; contrast sensitivity; perception of color; adaptation; and visual processing speed.[2] Visual function also can be measured in terms of functional disability caused by visual impairment.[3] Many activities are affected by more than one of these visual components. Health services researchers have increasingly emphasized function and quality of life as the outcomes of treatment that are most critical and applicable to the patient. As previously stated, the primary purpose in managing a patient with cataract is to improve functional vision and the quality of life. In well-designed observational studies, cataract surgery consistently has been shown to have a significant impact on vision-dependent function. The Cataract Patient Outcomes Research Team (PORT) reported that 90% of patients under-going first-eye cataract surgery noted improvement in functional status and satisfaction with vision.[4] The Activities of Daily Vision Study of elderly patients with a high prevalence of coexisting ocular and medical diseases reported improved visual function in 80% of patients at 12 months after surgery.[5] A National Cataract Study conducted in England of 1,139 patients who had cataract surgery found that preoperative functional impairment varied in relation to gender, age, and visual acuity. Men were more likely to have trouble with driving, glare, and employment, and women

were more likely to have difficulties with activities of daily living and recreational activities.[6] Studies have found that regardless of the preoperative visual acuity in the better eye, most patients reported improvement in their ability to perform visually dependent tasks after undergoing cataract surgery.[4-6]

Several studies have reported an association between improved visual function after cataract surgery and improved health-related quality of life.[1,5,7-8] Visual function plays an important role in physical function, particularly in terms of mobility.[9] The loss of visual function in the elderly is associated with a decline in physical and mental functioning as well as in independence in activities of daily living,[10] including night-time driving, daytime driving, community activities, and home activities. Elderly patients with visual impairment only (and no other physical or mental impairments) were 2.5 times as likely to experience functional decline than elderly patients without visual impairment.

Improved visual function following cataract surgery can ameliorate the progressive deterioration of quality of life seen in elderly patients.[1,5] In a cohort of 464 patients 65 years old and older, cataract extraction improved visual function and health-related quality of life. Patients with an improvement in their Activities of Daily Vision Scale (ADVS), a brief measure of vision-specific functional status,[11] had from 10% to 59% less decline in nearly all Short Form (SF)-36 dimensions.[5] The SF-36 is a generic global measure of multidimensional health-related quality of life.[12] A nationally representative population of 7,114 persons who were 70 years old and older showed that limitations in vision correlated with decreased functional status.[13] The unadjusted functional score of a person with reported poor vision was four times worse than the score for a person with excellent vision.[13] This difference was comparable with the differences found in other chronic conditions such as arthritis. This relationship with vision persisted, even after adjustment for health, demographics, and economic status. Individuals who rated their vision as other than excellent reported worse functional status, even when controlled for the presence of other medical conditions, education, income, general health status, and other symptoms. By improving visual function, cataract surgery may play an important role in preserving overall functional status, reducing associated injuries and accidents, and preventing disability in at-risk elderly patients.[10]

An analysis of the Medical Outcomes Study found that having blurred vision more than once or twice a month has a significant impact on functional status and well-being, particularly on problems with work or other daily activities as a result of physical health.[14] This impact was found to be greater than the impact of several other chronic conditions, such as hypertension, history of myocardial infarction, type 2 diabetes mellitus, indigestion, trouble urinating, and headache. In one study, patients planning to undergo cataract surgery assigned a mean preoperative preference value of 0.68 on a scale ranging from 0 to 1 (where 0 is death and 1 is excellent health), indicating that the visual impairment from cataracts had a substantial impact on their quality of life.[11] Visual impairment is an important risk factor for falls[15] and for hip fracture.[16] Specifically, the Study for Osteoporotic Fractures Research Group found that poor depth perception and decreased contrast sensitivity independently increased the risk of hip fracture.[17]

Visual impairment, in particular a decrease of visual acuity and contrast sensitivity, has been shown to be associated with difficulties in driving.[18] In one study, older drivers with visually significant cataract were twice as likely as older drivers without cataract to report reduction in days driven and four times as likely to report difficulties in challenging driving situations.[19] Drivers with visually significant cataract were 2.5 times more likely to have had an at-fault involvement in a motor vehicle crash in the past 5 years compared with drivers without cataract.[19] This association was significant, even after accounting for other factors such as impaired general health, age, mental status deficit or depression. In this study, visually significant cataract was determined by reviewing the participant's medical record and most recent eye examination by an eye care specialist. The study required that cataract in both eyes was the cause of the visual impairment, based on the medical record; an additional inclusion criterion was best-corrected visual acuity in one eye of 20/40 or worse. A further study in the same group demonstrated that drivers with a history of crash involvement were eight times more likely to have a serious contrast

sensitivity deficit (defined as a Pelli-Robson score of 1.25 or less) in the worse eye than those who had no history of crash involvement.[20] A severe contrast sensitivity deficit in only one eye was still significantly associated with crash involvement.[20]

Binocular vision is better than the vision of a single eye. The simultaneous use of the two eyes is complex and requires the integration of disparate images from each eye. A study demonstrated that binocular vision resulted in better perception of form, color, and the relationship of the body to the environment, which facilitated manipulation, reaching, and balance, particularly under dim illumination.[21] However, if the vision of one eye is reduced due to cataract, visual performance can fall below the level of monocular vision by a mechanism known as binocular inhibition,[22] which reduces patients' visual acuity and contrast sensitivity.[23] A study of the Framingham Study Cohort found that poor vision in one or both eyes was associated with an increased risk of hip fracture. It also found that patients with good vision in one eye and moderately impaired vision in the other eye had a higher risk of fracture than those with similar visual impairment in both eyes.[24] A study of 150 patients before and after cataract surgery found that poor binocular visual acuity was related to more problems in activities of daily living.[25] Another study, based on patients who reported no beneficial outcomes after first-eye cataract surgery in the National Swedish Cataract Outcome register, found that anisometropia was the reason for the poor outcome in one-third of cases.[26] These studies have shown that second-eye surgery is important to visual and physical function.

In summary, these studies demonstrate that physical function, emotional well-being, and overall quality of life can be enhanced when visual function is restored by cataract extraction.[27]

Improved visual function as a result of cataract surgery includes the following:

- Better optically corrected vision.
- Better uncorrected vision with reduced spectacle dependence.
- Increased ability to read or do near work.
- Reduced glare.
- Improved ability to function in dim levels of light.
- Improved depth perception and binocular vision.
- Improved color vision.

Improved physical function as a critical outcome of cataract surgery includes the following:

- Increased ability to perform activities of daily living.
- Increased opportunity to continue or resume an occupation.
- Increased mobility (walking, driving).

Improved mental health and emotional well-being as a second critical outcome of cataract surgery includes the following benefits:

- Improved self-esteem and independence.
- Increased ability to avoid injury.
- Increased social contact and ability to participate in social activities.
- Relief from fear of blindness.

Most patients achieve improved visual function after cataract surgery. This outcome is achieved consistently through careful attention through the patient selection process, accurate measurement of axial length and corneal power, appropriate selection of an IOL power calculation formula, etc. As such, it reflects the care and diligence with which the surgery is assessed, planned and executed. Failure to achieve this after surgery would reflect patterns of patient selection or treatment that should be

assessed for opportunities for improvement.

Sometimes cataract surgery is performed for other medical reasons other than to improve impaired visual function caused by cataract. These circumstances include the following: clinically significant anisometropia in the presence of a cataract; when the lens opacity interferes with optimal diagnosis or management of posterior segment conditions, when the lens causes inflammation (phacolysis, phacoanaphylaxis) and when the lens induces angle closure (phacomorphic or phacotopic). In these situations, improved visual function as a result of the removal of the cataract is not expected, because of the pre-existing comorbid conditions.

2. Evidence of a gap in care

This is an outcome of surgery indicator of direct relevance and import to patients, their families and referring providers. The available evidence suggests that cataract surgery achieves this in about 90% of patients. While the potential for improvement is seemingly small, the volume of cataract surgery in the U.S. of over 2.8 million surgeries means that the impact could affect more than 100,000 patients per year. Ideally, performance on this indicator would be as high as possible, with lower rates suggestive of opportunities for improvement.

Definitions:

Standardized Tool – An assessment tool that has been appropriately validated for the population for which it being used. Examples of tools for visual function assessment include, but are not limited to: National Eye Institute-Visual Function Questionnaire (VFQ), the Visual Function (VF)-14, the modified VF-8, the Activities of Daily Vision Scale (ADVS), the Catquest and the modified Catquest-9.

Vision Function Assessment – Questionnaires designed to measure a patient's ability to perform the everyday tasks requiring vision.

Data Capture and Calculations:

Calculation for *Performance*

For performance purposes, this measure is calculated by creating a fraction with the following components: Numerator, Denominator, and Denominator Exclusions.

Performance Numerator (A) Includes:

- Patients who had an improvement in their visual function achieved within 90 days following cataract surgery

Performance Denominator (PD) Includes:

- All patients aged 18 years and older

AND

- Had cataract surgery

Performance Denominator Exclusions (C) Includes:

A patient is excluded if the following condition(s) exist:

Medical reasons:

When cataract surgery was performed for these indications:

- Clinically significant anisometropia in the presence of a cataract
- The lens opacity interferes with optimal diagnosis or management of posterior segment conditions
- The lens causes inflammation (phacolysis, phacoanaphylaxis)
- The lens induces angle closure (phacomorphic or phacotopic)

Patient reasons:

- The patient refuses to participate
- The patient is unable to complete the questionnaire

Performance Calculation

$$\frac{\text{A (\# of patients meeting measure criteria)}}{\text{PD (\# of patients in denominator)} - \text{C (\# of patients with valid denominator exclusions)}}$$

Components for this measure are defined as:

A	# of patients who had an improvement in their visual function achieved within 90 days following cataract surgery
PD	# of patients aged 18 years and older who had cataract surgery

C	# of patients with documented patient reason for not completing their visual function assessment within 90 days following cataract surgery

Calculation for Reporting:
For reporting purposes, this measure is calculated by creating a fraction with the following components: Reporting Numerator and Reporting Denominator.

Reporting Numerator includes each of the following instances:

A. Patients who had an improvement in their visual function achieved within 90 days following cataract surgery

C. Patients who did not complete their visual function assessment within 90 days following cataract surgery but for whom there is a documented medical or patient reason for not doing so

D. Patients who did not have an improvement in their visual function achieved within 90 days following cataract surgery and there is no documented medical or patient reason for not doing so

Reporting Denominator (RD) includes:

- Patients aged 18 years and older AND
- Had cataract surgery

Reporting Calculation

$$\frac{\text{A (\# of patients meeting measure criteria) + C (\# of patients with valid exclusions) + D (\# of patients NOT meeting numerator criteria)}}{\text{RD (\# of patients in denominator)}}$$

Components for this measure are defined as:

A	# of patients who had an improvement in their visual function achieved within 90 days following cataract surgery
C	# of patients who did not complete their visual function assessment within 90 days following cataract surgery but for whom there is a documented medical or patient reason for not doing so
D	# of patients who did not have an improvement in their visual function achieved within 90 days following cataract surgery and there is no documented medical or patient reason for not doing so
RD	# of patients aged 18 years and older who had cataract surgery

Measure Specifications - Measure #1 Cataracts: Improvement in Patient's Visual Function within 90 Days Following Cataract Surgery

Measure specifications will be provided for multiple data sources.

A. Administrative claims data

Administrative claims data collection requires users to identify the eligible population (denominator) and numerator using codes recorded on claims or billing forms (electronic or paper). Users report a rate based on all patients in a given practice for whom data are available and who meet the eligible population/denominator criteria.

(Note: The specifications listed below are those needed for performance calculation)

Denominator (Eligible Population): All patients aged 18 years and older who had cataract surgery

- CPT Procedure Codes (with or without modifiers): 66840, 66850, 66852, 66920, 66930, 66940, 66982, 66983, 66984

Numerator: Patients who had an improvement in their visual function achieved within 90 days following cataract surgery

Report the following CPT Category II code:

_____ - Improved visual function achieved within the 90 days following cataract Surgery

Denominator Exclusions: Documentation of medical reason for not improving visual function within 90 days of cataract surgery

- Append modifier to CPT Category II Code: -1P

Documentation of patient reason for not improving visual function within 90 days of cataract surgery

- Append modifier to CPT Category II Code: -2P

B. Registry

Registry reporting requires users to identify the eligible population (denominator) using CPT codes and patient demographics. The numerator options as described in the CPT Category II codes are used to report the numerator of the measure. The CPT Category II codes listed do not need to be submitted for registry-based submissions, however these codes may be submitted for those registries that utilize claims data.

C. Electronic Health Record System (in development)

D. Paper Medical Record (in development)

Eye Care

Measure #2 Cataracts: Patient Satisfaction within 90 Days Following Cataract Surgery

This measure may be used as an Accountability measure.

Clinical Performance Measure
Numerator: Patients who were satisfied with their care within 90 days following cataract surgery **Denominator:** All patients aged 18 years and older who had cataract surgery **Denominator Exclusions:** The patient refuses to participate or the patient is unable to complete the questionnaire **Measure:** Percentage of patients aged 18 years and older who had cataract surgery and were satisfied with their care within 90 days following the cataract surgery
The following clinical recommendation statements are quoted <u>verbatim</u> from the referenced clinical guidelines and represent the evidence base for the measure: **This is an outcomes measure.** As such, no statements in the guideline are specific to this measurement topic.
Rationale for the measure: 1. Scientific basis for measuring patient satisfaction after cataract surgery. Patient satisfaction is a valuable performance indicator for measuring the quality of care delivered by ophthalmologists providing cataract surgery. In the broadest sense, patient satisfaction is an assessment of the patient's experience with the care process delivered by health plans, clinicians, health systems, hospitals, etc. This experience can cover domains as diverse as information/education, interpersonal manner, emotional support, accessibility, convenience, outcomes or results, environment, personalization, involvement in care, finances, etc. In 1996, The American Academy of Ophthalmology launched the National Eyecare Outcomes Network (NEON) database. [28,29] From January 1, 1996 through March 30, 2001, 249 ophthalmologists at 114 different practice sites submitted data to the NEON cataract surgery database. Post-operative patient satisfaction responses were collected for 6,154 patients, or about 34.5% of all patients who had pre-operative forms submitted. This assessment was performed at a median of 4.1 weeks postoperatively for all patients enrolled in the database. A 12-item questionnaire was used to assess patient satisfaction. Patient satisfaction was associated with younger age and absence of ocular comorbidity. Other studies of patient satisfaction after cataract surgery in Austria and in Spain. One study found that patients with pre-existing eye disease, including those patients with improved visual acuity after surgery, were the least satisfied with the results of surgery. [30] In these cases, improved patient

education prior to surgery could be helpful in improving patient satisfaction. Another study found that patient satisfaction was associated with expectations prior to surgery.[31]

Most patients are satisfied with their care and results after cataract surgery. This outcome is achieved consistently through careful attention through the patient selection process, accurate measurement of axial length and corneal power, appropriate selection of an IOL power calculation formula, etc. As such, it reflects the care and diligence with which the surgery is assessed, planned and executed. Failure to achieve this satisfaction after surgery would reflect patterns of patient selection or treatment that should be assessed for opportunities for improvement.

Use of this indicator in the PQRI program in the claims reporting method would require some modification to the current reporting of post-operative care for patients undergoing cataract surgery, since this indicator would be operative during the 90 day global period. However, there is a strong and practical precedent for such modifications in that reporting arrangements have previously been made to accommodate co-management of care by different providers during the post-operative period. A similar adjustment to allow for filing of a claim of meeting this goal at one point in the 90 day global period would be sufficient, potentially drawing upon the methods to demarcate the onset of co-management transfer of post-operative care.

Various patient satisfaction instruments exist, but an instrument developed by the program, Consumer Assessment of Healthcare Providers and Systems (CAHPS), Agency for Healthcare Research and Quality develops and supports the use of a comprehensive and evolving family of standardized surveys that ask consumers and patients to report on and evaluate their experiences with health care. These surveys cover topics that are important to consumers, such as the communication skills of providers and the accessibility of services. AHRQ first launched the CAHPS program in October 1995 in response to concerns about the lack of good information about the quality of health plans from the enrollees' perspective. At that time, numerous public and private organizations collected information on enrollee and patient satisfaction, but the surveys varied from sponsor to sponsor and often changed from year to year.

The CAPHS Surgical Care Survey asks adult patients to report on surgical care, surgeons, their staff, and anesthesiologists. It was developed by the American College of Surgeons and the Surgical Quality Alliance to assess patients' experiences before, during, and after surgery. In early 2010, the CAHPS Consortium voted to adopt the Surgical Care Survey as an official CAHPS survey. The Surgical Care Survey expands on the current CAHPS Clinician & Group Survey, which focuses on primary and specialty care, by incorporating domains that are relevant to surgical care, such as informed consent, anesthesia care, and post-operative follow-up. The survey is unique in that it assesses patients' experiences with surgical care in both the inpatient and outpatient settings by asking respondents about their care before, during, and after surgery

The main purpose of the CAHPS Surgical Care Survey is to address the need to assess and improve the experiences of surgical patients. Like other CAHPS surveys, this questionnaire focuses on aspects of surgical quality that are important to patients and for which patients are the best source of information. The survey results are expected to be useful to everyone with a need for information on the quality of surgeons and surgical care, including patients, practice groups, health plans, insurers, and specialty boards. Patients can use the information to help make better and more informed choices about their surgical care. Practices, health plans, and insurers can use the survey results for quality improvement initiatives and incentives. Specialty boards may use the survey for maintenance of certification.

https://www.cahps.ahrq.gov/content/products/sc/PROD_SC_Surgical_Care.asp?p=1021&s=213

2. Evidence of a gap in care

This is an outcome of surgery indicator of direct relevance and import to patients , their families and referring providers. The available evidence suggests that cataract surgery achieves this in about 90% of patients. While the potential for improvement appears seemingly small, the volume of cataract surgery in the U.S. of over 2.8 million surgeries means that the impact could affect more than 100,000 patients per year. Ideally, performance on this indicator to be as high as possible, with rates lower than 95-100% suggestive of opportunities for improvement.

Definitions:

Standardized Tool – An assessment tool that has been appropriately validated for the population for which it being used. Examples of tools for patient satisfaction include, but are not limited to: Surgical Consumer Assessment of Health Plans and Systems, which is also approved by the Agency for Health Care Research and Quality.

Patient Satisfaction Assessment – Questionnaires designed to measure a patient's satisfaction with the care that they received from their surgeon.

Data Capture and Calculations:

Calculation for *Performance*

For performance purposes, this measure is calculated by creating a fraction with the following components: Numerator, Denominator, and Denominator Exclusions.

Performance Numerator (A) Includes:

- Patients who were satisfied with their care within 90 days following cataract surgery

Performance Denominator (PD) Includes:

- All patients aged 18 years and older

AND

- Had cataract surgery

Performance Denominator Exclusions (C) Includes:

A patient is excluded if the following condition(s) exist:

- The patient refuses to participate
- The patient is unable to complete the questionnaire

Performance Calculation

$$\frac{\text{A (\# of patients meeting measure criteria)}}{\text{PD (\# of patients in denominator)} - \text{C (\# of patients with valid denominator exclusions)}}$$

Components for this measure are defined as:

A	# of patients who were satisfied with their care within 90 days following cataract surgery
PD	# of patients aged 18 years and older who had cataract surgery
C	# of patients with documented patient reason for not performing the patient satisfaction assessment within 90 days following cataract surgery

Calculation for Reporting:

For reporting purposes, this measure is calculated by creating a fraction with the following components: Reporting Numerator and Reporting Denominator.

Reporting Numerator includes each of the following instances:

A. Patients who were satisfied with their care within 90 days following cataract surgery

C.Patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery but for whom there is a documented patient reason for not doing so

D. Patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery and there is no documented patient reason for not doing so

Reporting Denominator (RD) includes:

- Patients aged 18 years and older AND
- Had cataract surgery

Reporting Calculation

$$\frac{\text{A (\# of patients meeting measure criteria) + C (\# of patients with valid exclusions) + D (\# of patients NOT meeting numerator criteria)}}{\text{RD (\# of patients in denominator)}}$$

Components for this measure are defined as:

A	# of patients who were satisfied with their care within 90 days following cataract surgery
C	# of patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery but for whom there is a documented patient reason for not doing so
D	# of patients who did not complete a patient satisfaction assessment within 90 days following cataract surgery and there is no documented patient reason for not doing so
RD	# of patients aged 18 years and older who had cataract surgery

Measure Specifications - Measure #2 Cataracts: Patient Satisfaction within 90 Days Following Cataract Surgery

Measure specifications will be provided for multiple data sources.

A. Administrative claims data

Administrative claims data collection requires users to identify the eligible population (denominator) and numerator using codes recorded on claims or billing forms (electronic or paper). Users report a rate based on all patients in a given practice for whom data are available and who meet the eligible population/denominator criteria.

(Note: The specifications listed below are those needed for performance calculation)

Denominator (Eligible Population): All patients aged 18 years and older who had cataract surgery

- CPT Procedure Codes (with or without modifiers): 66840, 66850, 66852, 66920, 66930, 66940, 66982, 66983, 66984

Numerator: Patients who were satisfied with their care within 90 days following cataract surgery

Report the following CPT Category II code:
_____ - Patient satisfaction achieved within the 90 days following cataract Surgery

Denominator Exclusions: Documentation of patient reason for not completing patient satisfaction assessment within 90 days of cataract surgery

- Append modifier to CPT Category II Code: -2P

B. Registry

Registry reporting requires users to identify the eligible population (denominator) using CPT codes and patient demographics. The numerator options as described in the CPT Category II codes are used to report the numerator of the measure. The CPT Category II codes listed do not need to be submitted for registry-based submissions, however these codes may be submitted for those registries that utilize claims data.

C. Electronic Health Record System (in development)

D. Paper Medical Record (in development)

References:

1. Brenner MH, Curbow B, Javitt JC, et al. Vision change and quality of life in the elderly. Response to cataract surgery and treatment of other chronic ocular conditions. Arch Ophthalmol 1993; 111:680-5.
2. Sloane ME, Ball K, Owsley C, et al. The Visual Activities Questionnaire: developing an instrument for assessing problems in everyday visual tasks. Technical Digest, Noninvasive Assessment of the Visual System 1992; 1:26-29.
3. Steinberg EP, Tielsch JM, Schein OD, et al. The VF-14. An index of functional impairment in patients with cataract. Arch Ophthalmol 1994; 112:630-8.
4. Schein OD, Steinberg EP, Javitt JC, et al. Variation in cataract surgery practice and clinical outcomes. Ophthalmology 1994; 101:1142-52.
5. Mangione CM, Phillips RS, Lawrence MG, et al. Improved visual function and attenuation of declines in health-related quality of life after cataract extraction. Arch Ophthalmol 1994; 112:1419-25.
6. Desai P, Minassian DC, Reidy A. National cataract surgery survey 1997-8: a report of the results of the clinical outcomes. Br J Ophthalmol 1999; 83:1336-40.
7. Monestam E, Wachtmeister L. Impact of cataract surgery on visual acuity and subjective functional outcomes: a population-based study in Sweden. Eye 1999; 13:711-19.
8. Steinberg EP, Tielsch JM, Schein OD, et al. National study of cataract surgery outcomes. Variation in 4-month postoperative outcomes as reflected in multiple outcome measures. Ophthalmology 1994; 101:1131-40; discussion 1140-1.
9. Damiano AM, Steinberg EP, Cassard SD, et al. Comparison of generic versus disease-specific measures of functional impairment in patients with cataract. Med Care 1995; 33 (4 Suppl):AS120-30.
10. Salive ME, Guralnik J, Glynn RJ, et al. Association of visual impairment with mobility and physical function. J Am Geriatr Soc 1994; 42:287-92.
11. Laforge RG, Spector WD, Sternberg J. The relationship of vision and hearing impairment to one-year mortality and functional decline. J Aging Health 1992; 4:126–48.
12. Mangione CM, Phillips RS, Seddon JM, et al. Development of the 'Activities of Daily Vision Scale'. A measure of visual functional status. Med Care 1992; 30:1111-26.
13. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30:473-83.
14. Lee P, Smith JP, Kington R. The relationship of self-rated vision and hearing to functional status and well-being among seniors 70 years and older. Am J Ophthalmol 1999; 127:447-52.
15. Lee PP, Spritzer K, Hays RD. The impact of blurred vision on functioning and well-being. Ophthalmology 1997; 104:390-6.
16. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. N Engl J Med 1988; 319:1701-7.
17. Felson DT, Anderson JJ, Hannan MT, et al. Impaired vision and hip fracture. The Framingham Study. J Am Geriatr Soc 1989; 37:495-500.

18. Cummings SR, Nevitt MC, Browner WS, et al. Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. N Engl J Med 1995; 332:767-73.
19. McGwin G Jr, Chapman V, Owsley C. Visual risk factors for driving difficulty among older drivers. Accid Anal Prev 2000; 32:735-44.
20. Owsley C, Stalvey B, Wells J, Sloane ME. Older drivers and cataract: driving habits and crash risk. J Gerontol A Biol Sci Med Sci 1999; 54:M203-11.
21. Owsley C, Stalvey BT, Wells J, et al. Visual risk factors for crash involvement in older drivers with cataract. Arch Ophthalmol 2001; 119:881-7.
22. Jones RK, Lee DN. Why two eyes are better than one: the two views of binocular vision. J Exp Psychol 1981; 7:30–40.
23. Pardhan S, Gilchrist J. The importance of measuring binocular contrast sensitivity in unilateral cataract. Eye 1991; 5:31-5.
24. Taylor RH, Misson GP, Moseley MJ. Visual acuity and contrast sensitivity in cataract: summation and inhibition of visual performance. Eye 1991; 5:704-7.
25. Lundström M, Fregell G, Sjöblom A. Vision related daily life problems in patients waiting for a cataract extraction. Br J Ophthalmol 1994; 78:608–11.
26. Lundström M, Brege KG, Florén I, et al. Impaired visual function after cataract surgery assessed using the Catquest questionnaire. J Cataract Refract Surg 2000; 26:101–8.
27. Cataract Management Guideline Panel. Cataract in Adults: Management of Functional Impairment. Clinical Practice Guideline, Number 4. Rockville, MD: USDHHS, AHCPR Publ. No. (PHS) 93-0542; 1993.
28. Lum F, Schein O, Schachat AP, et al. Initial two years of experience with the AAO National Eyecare Outcomes Network (NEON) cataract surgery database. Ophthalmology 2000; 107:691-7.
29. Lum F, Schachat AP, Jampel HD. The development and demise of a cataract surgery database. The Joint Commission Journal on Quality Improvement 2202; 28:108-114.
30. Mozaffarieh M, Krepler K, Heinzl H et al. Visual function, quality of life and patient satisfaction after ophthalmic surgery: a comparative study. Ophthalmologica 2004; 218:26-30.
31. Lledo R, Rodriguez T, Fontenia JR et al. Cataract surgery: An analysis of patient satisfaction with medical care. International Ophthalmology 22:227-32.

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1550 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA)
De.2 **Brief description of measure***:* This measure estimates hospital risk-standardized complication rates (RSCRs) associated with primary elective THA and TKA in patients 65 years and older. The measure uses Medicare claims data to identify complications occurring from the date of index admission to 90 days post date of the index admission.
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** This measure is paired with a readmission measure for THA and TKA.
De.4 **National Priority Partners Priority Area:** Care coordination, Safety De.5 **IOM Quality Domain**: Effectiveness, Patient-centered, Efficiency, Safety De.6 **Consumer Care Need:** Getting better, Living with illness

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	NQF Staff
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary 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, Internal quality improvement	**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, Frequently performed procedure, High resource use, Other 1a.2 High cost 1a.3 **Summary of Evidence of High Impact:** High complication rate We conducted analyses using 2008 Medicare Part A inpatient claims data and found a median 30-day unadjusted hospital complication rate of 6.7%. This rate is high considering these are elective procedures typically performed on younger, healthier patients, compared to other Medicare patients. Complication rates following THA and TKA warrant investigation as these procedures are elective, costly, and projected to increase over the coming years (Kurtz et al., 2007). Complication rates have been shown to vary across hospitals, suggesting care can be improved. Prospective studies show risk adjusted rates for periprosthetic joint infection, a rare but devastating complication, vary between 2.3 to 1.6 percent after 1 and 2 years of follow-up respectively (Kurtz et al., 2010; Bongartz et al., 2008). Ninety-day death rates following THA also range from 0.7 to 2.7 percent and are high for an elective procedure (Cram et al., 2007; Soohoo et al., 2010). Rates for pulmonary embolism following TKA range from 0.5 and 0.9 percent (Cram et al., 2007; Mahomed et al., 2003; Khatod et al., 2008; Solomon et al., 2006;). Rates for wound infection in Medicare population based studies vary between 0.3 and 1.0 percent (Cram et al., 2007; Mahomed et al., 2003; Solomon et al., 2006). Rates for septicemia range from 0.1%, during the index admission (Browne et al., 2010) to 0.3%, 90 days following discharge for primary TKA (Cram et al.,	**1a** C☐ P☐ M☐ N☐

2007). Rates for bleeding and hematoma following TKA range from 0.94 (Browne et al., 2010) to 1.7% (Huddleston et al., 2009).

The variation in complication rates across hospitals indicates there is room for quality improvement and targeted efforts to reduce these complications could result in better patient care and potential cost savings.

High volume
THA and TKA are priority areas for outcomes measure development, as they are commonly performed procedures in the US. In 2003 there were 202,500 primary hip arthroplasties and 402,100 primary total knee arthroplasties performed (Kurtz et al., 2007). The number of procedures performed has increased steadily over the past decade (Kurtz et al., 2007; Ong et al., 2006) and complications may increase the risk of revision procedures which are even more costly and associated with higher resource utilization (Ong et al., 2006).

High cost
Although these procedures can dramatically improve health-related quality-of-life, they are costly. In 2005 annual hospital charges totaled $3.95 billion and $7.42 billion for primary THA and TKA, respectively (Kurtz et al., 2007). These costs are projected to increase by 340% to 17.4 billion for THA and by 450% to 40.8 billion for TKA by 2015 (Kurtz et al., 2007). Medicare is the single largest payer for these procedures, covering approximately two-thirds of all THAs and TKAs performed in the US (Ong et al., 2006). THA and TKA procedures combined account for the largest procedural cost in the Medicare budget (Bozic et al., 2008).

1a.4 **Citations for Evidence of High Impact:** Bongartz, T, Halligan CS, Osmon D, et al. Incidence and risk factors of prosthetic joint infection after total hip or knee replacement in patients with rheumatoid arthritis. Arthritis Rheum. 2008; 59(12): 1713-1720.

Bozic KJ, Rubash HE, Sculco TP, Berry DJ. An analysis of medicare payment policy for total joint arthroplasty. Journal of Arthroplasty. 2008;23(6 Suppl 1):133-138.

Browne, JA, Cook C, Hofmann A, Bolognesi MP. Postoperative morbidity and mortality following total knee arthroplasty with computer navigation. Knee. 2010;17(2): 152-156.

Cram P,Vaughan-Sarrazin MS,Wolf B,Katz JN,Rosenthal GE. A comparison of total hip and knee replacement in specialty and general hospitals. J Bone Joint Surg Am. Aug 2007;89(8):1675-1684.

Huddleston JI,Maloney WJ,Wang Y,Verzier N,Hunt DR,Herndon JH. Adverse Events After Total Knee Arthroplasty: A National Medicare Study. The Journal of Arthroplasty. 2009;24(6, Supplement 1):95-100.

Khatod M,Inacio M,Paxton EW, et al. Knee replacement: epidemiology, outcomes, and trends in Southern California: 17,080 replacements from 1995 through 2004. Acta Orthop. Dec 2008;79(6):812-819.
Kurtz S,Ong K,Lau E,Bozic K,Berry D,Parvizi J. Prosthetic joint infection risk after TKA in the Medicare population. Clin Orthop Relat Res. 2010;468:5.
Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am. Apr 2007;89(4):780-785.
Kurtz SM, Ong KL, Schmier J, et al. Future clinical and economic impact of revision total hip and knee arthroplasty. J Bone Joint Surg Am. Oct 2007;89 Suppl 3:144-151.
Mahomed NN,Barrett JA,Katz JN, et al. Rates and outcomes of primary and revision total hip replacement in the United States medicare population. J Bone Joint Surg Am. Jan 2003;85-A(1):27-32.
Ong KL, Mowat FS, Chan N, Lau E, Halpern MT, Kurtz SM. Economic burden of revision hip and knee arthroplasty in Medicare enrollees. Clin Orthop Relat Res. May 2006;446:22-28.
Solomon DH,Chibnik LB,Losina E, et al. Development of a preliminary index that predicts adverse events after total knee replacement. Arthritis & Rheumatism. 2006;54(5):1536-1542.
Soohoo NF,Farng E,Lieberman JR,Chambers L,Zingmond DS. Factors That Predict Short-term Complication Rates After Total Hip Arthroplasty. Clin Orthop Relat Res. Sep 2010;468(9):2363-2371.

1b. Opportunity for Improvement

1b
C☐
P☐
M☐

1b.1 **Benefits (improvements in quality) envisioned by use of this measure:** Measuring and reporting complication rates will inform health care providers about opportunities to improve care, strengthen

incentives for quality improvement, and ultimately improve the quality of care for Medicare patients. The measure will also provide patients with information that could guide their choices. In addition, it has the potential to lower health care costs associated with complications. The measure will increase transparency for consumers.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
There is considerable variation in practice patterns, patient outcomes, and adherence to payer-defined practice guidelines for both THA and/or TKA (Bozic et al., 2008). The unadjusted mean complication rate was 4.98% and ranged from 0% to 100% across 3,311 hospitals in 2008. After adjustment for patient and clinical characteristics, the mean hospital-level complication rate was 4.23% ranging from 2.20-8.88%. The variation observed for complications suggested there are differences in the quality of care delivered across hospitals that result in variation in outcomes.

Primary elective THA and TKA are beneficial procedures that greatly improve the quality of life for patients who choose to undergo these procedures (Hawker et al., 1998). Understanding and addressing causes of complications in this elective group of patients may improve the quality of care and reduce costs associated with THA and TKA.

1b.3 Citations for data on performance gap:
Bozic KJ, Chiu V. Quality Measurement and Public Reporting in Total Joint Replacement. The Journal of Replacement. 2008; 23:146-149.

Hawker GJ, Wright J, Coyte P, Paul J, Dittus R, Croxford B, et al. Health-related quality of life after knee replacement. J Bone Joint Surg Am. 1998; 80:163-73.

1b.4 Summary of Data on disparities by population group:
We conducted analyses to explore disparities by SES. We used Medicaid eligibility status identified in the Medicare claims enrollment database (EDB) as a proxy for SES. This approach is consistent with prior research as well as NQF recommendations (http://www.nysna.org/images/pdfs/practice/nqf_ana_outcomes_draft10.pdf). Patients were categorized into two groups, based on their eligibility status for Medicaid (yes/no). The Medicaid eligible population represents lower SES status. Analyses demonstrated that although SES is a significant predictor of readmission at the patient level, it does not affect overall hospital performance in the risk-adjusted readmission model. Consistent with NQF guidelines, this measure does not risk-adjust for SES factors.

1b.5 Citations for data on Disparities:
N/A

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 measure will calculate hospital-level complication rates following elective primary THA and/or TKA with the goal to reduce complication rates. It addresses a priority condition (osteoarthritis) and will lead to reduced morbidity and mortality post THA and TKA.

1c.2-3. Type of Evidence: Expert opinion, Systematic synthesis of research

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Complications following primary elective THA and/or TKA are important patient outcomes that may reflect quality of care delivered to patients undergoing these procedures. However, the evidence available on the relationship between healthcare processes and complication outcomes from primary elective THA and/or TKA is sparse. Most of the research into complications of primary elective THA and/or TKA estimate rates and patient level characteristics that predict outcomes. Few studies examine hospital and provider level characteristics associated with complications from THA and/or TKA. However, a working group and technical expert panel (TEP) of orthopedists, rheumatologists, consumer and purchaser perspective, disparities experts, and quality improvement experts were consulted in determining which complications are likely

N☐

1c
C☐
P☐
M☐
N☐

attributable to care processes (see section 2c for details) and can be reduced.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):** N/A (outcome measure)

1c.6 Method for rating evidence: N/A (outcome measure)

1c.7 Summary of Controversy/Contradictory Evidence: Defining Complications
After conducting a comprehensive literature review and in consultation with the working group, YNHHSC/CORE identified complications for potential inclusion in a complications measure. To be considered as candidates for inclusion in the outcome, the complications had to:
? Represent meaningful complications attributable to the THA/TKA procedures
? Be identifiable in administrative claims data
? Be fair to hospitals and physicians

Based on these criteria and in consultation with the working group, we identified several candidate complications for inclusion in a composite complications measure:
? Death
? Mechanical complications
? Periprosthetic joint infection
? Surgical site bleeding
? Wound infection
? Pulmonary embolism
? Acute myocardial infarction (AMI)
? Pneumonia
? Sepsis/septicemia
? Deep vein thrombosis (DVT)
? Urinary tract infection (UTI)

DVT and UTI were excluded based on working group feedback and the literature. We excluded DVT because there is wide variability across hospitals in screening (Geerts et al., 2004; Pierce et al., 2008) and readmission practices for this complication. We excluded UTI because there is wide variability in diagnosing UTI, and the rates are likely inflated due to overdiagnosis in patients post THA/TKA (Woodford et al., 2009).

Based on these considerations, we included the following complications in the measure:
? Death
? Mechanical complications
? Periprosthetic joint infection
? Surgical site bleeding
? Wound infection
? Pulmonary embolism
? AMI
? Pneumonia
? Sepsis/septicemia

A potential area of controversy may be the varying degrees of severity for some of the complications. Degrees of severity are not conveyed in the ICD-9 diagnosis codes, specifically, wound infection, periprosthetic joint infection, and surgical site bleeding. For example, the diagnosis codes used to identify wound infection may reflect redness and swelling around the incision site, or a true wound infection, requiring incision and drainage. Thus, to capture clinically important complications and to reduce the likelihood of capturing miscoded complications, working group and TEP members recommended only counting these complications in the outcome if they are associated with accompanying ICD-9 procedure codes indicating that they were severe enough to require specific interventions. We therefore imposed additional coding requirements for these complications to set an appropriate threshold for severity.

Complication-specific follow-up periods
We identified the follow-up period for each complication based on preliminary data analyses and expert clinical input. Our empirical analyses indicated that the rates for all complications were elevated during the

index admission and returned to baseline within 30 to 90 days post the index date of admission, depending on the complication. We confirmed the follow-up periods with an expert panel that included orthopaedic surgeons, a rheumatologist, and experts in quality measurement.
The inclusion of medical complications (acute myocardial infarction, pneumonia, and sepsis) may be controversial because some clinicians may feel these medical conditions are not attributable to the procedure. Our data indicated, however, that the rates for these medical complications are elevated during the index admission period and decrease sharply 7 days from admission, returning to baseline within 30 days of the index admission date. Therefore, the follow-up period for these medical complications was limited to 7 days post index admission date, as they are more likely to be attributable to the procedure if they occur within 7 days of the index date of admission. Restricting the follow-up period to 7 days also limits overlap with the 30-day all-cause readmission measure.

Use of Hierarchical Generalized Linear Modeling
Hierarchical modeling for hospital outcomes measurement is the appropriate statistical approach for hospital outcomes measures given the structure of the data and the underlying assumption of such measures, which is that hospital quality of care influences complication rates. However, CMS frequently receives comments and questions about this approach, so we are concisely reiterating the rationale for and merits of using hierarchical logistic regression. Patients are clustered within hospitals and, as such, have a shared exposure to the hospital quality and processes. The use of hierarchical modeling accounts for the clustering of patients within hospitals. Second, hierarchical models distinguish within-hospital variation and between-hospital variation to estimate the hospital's contribution to the risk of complications. This allows for an estimation of the hospital's influence on patient outcomes. Finally, within hierarchical models we can account for both differences in case mix and sample size to fairly profile hospital performance. If we did not use hierarchical modeling we could overestimate variation and potentially misclassify hospitals' performance. Accurately estimating variation is an important objective for models used in public reporting and potentially used in value-based purchasing programs.

1c.8 Citations for Evidence (*other than guidelines***):** Geerts WH,Pineo GF,Heit JA, et al. Prevention of Venous Thromboembolism. Chest. September 1, 2004 2004;126(3 suppl):338S-400S.

Pierce C,Haut E,Kardooni S, et al. Surveillance bias and deep vein thrombosis in the national trauma data bank: the more we look, the more we find. J Trauma. 2008;64:6.

Woodford HJ,George J. Diagnosis and Management of Urinary Tract Infection in Hospitalized Older People. Journal of the American Geriatrics Society. 2009;57(1):107-114.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
N/A - We did not set any clinical practice guidelines as this is an outcomes measure, not a process of care measure.

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 subcriteria 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***):** This outcome measure does not have a traditional numerator and denominator like a core process measure (e.g., percentage of adult patients with diabetes aged 18-75 years receiving one or more hemoglobin A1c tests per year); thus, we are using this field to define the outcome (i.e. adverse events) following THA and/or TKA procedures. The composite complication is a binary outcome (yes for any complication(s); no for no complications). Therefore, if a patient experiences 1 or more complications, the outcome variable will get coded as a "yes." Complications are counted in the measure only if they occur during the index hospital admission or during a readmission. The complications captured in the numerator are identified during the index admission or associated with a readmission up to 90 days post date of index admission, depending on the complication. The follow-up period for complications from date of index admission is as follows: 1) Mechanical complications - 90 days 2) Periprosthetic joint infection (PJI) - 90 days 3) Wound infection - 90 days 4) Surgical site bleeding - 30 days 5) Pulmonary embolism - 30 days 6) Death - 30 days 7) AMI - 7 days 8) Pneumonia - 7 days 9) Sepsis/septicemia - 7days **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** The specific time frame for the complication varies (depending on the complication) from 7 to 90 days post date of the index admission (see “Numerator Details”). **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Complications are identified using the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) diagnosis and procedure codes. The complications listed below are counted in the measure if coded in the primary or secondary diagnosis fields during either the index admission or a readmission. Multiple complications count only once toward the numerator. For example, if a patient experiences a mechanical complication and also has an acute myocardial infarction, the combined events will be counted only once in the measure. ICD-9 diagnosis and procedure codes used to identify complications are listed below: Complications identified from the date of index admission to 7 days post date of index admission: 1. Acute Myocardial Infarction Presence of one of the following diagnosis codes: 410.xx excluding 410.x2 2. Pneumonia	**2a-specs** C☐ P☐ M☐ N☐

Presence of one of the following diagnosis codes: 480, 480.0, 480.1, 480.2, 480.3, 480.8, 480.9, 481, 482, 482.0, 482.1, 482.2, 482.3, 482.30,482.31, 482.32, 482.39, 482.4, 482.40, 482.41, 482.42, 482.49, 482.81, 482.82, 482.83, 482.84, 482.89, 482.9, 483, 483.0, 483.1, 483.8, 485, 486, 487.0, 507.0

3. Sepsis/Septicemia
Presence of one of the following diagnosis codes: 038, 038.0, 038.1, 038.10, 038.11, 038.12, 038.19, 038.2, 038.3, 038.4, 038.40, 038.41, 038.42, 038.43, 038.44, 038.49, 038.8, 038.9, 785.52, 785.59, 790.7, 995.91, 995.92, 998.0, 998.59, 790.7, 998.59

Complications identified from date of index admission to 30 days post date of index admission:
4. Pulmonary Embolism
Presence of one of the following diagnosis codes: 415.1, 415.11, 415.19

5. Surgical Site Bleeding
Presence of one of the following diagnosis codes: 998.1,998.11, 998.12, 998.13, 286.5, 719.10, 719.16, 719.17

AND the following procedure code:
- Incision and Drainage: 86.04

6. Death (Source: Medicare Enrollment Database)

Complications identified from date of index admission to 90 days post date of index admission:
7. Wound Infection
Presence of one of the following diagnosis codes: 998.6, 998.83, 998.3, 998.30, 998.31, 998.32, 998.33, 998.5, 998.51, 998.59, 996.67

AND at least one of the following procedure codes:
- Incision and Drainage: 86.22, 86.28, 86.04
- Revision: 81.53, 81.55, 81.59, 00.70, 00.71, 00.72, 00.73, 00.80, 00.81, 00.82, 00.83, 00.84
- Removal: 80.05, 80.06, 80.09

8. Periprosthetic Joint Infection
Presence of the following diagnosis code: 996.66

AND at least one of the following procedure codes:
- Incision and Drainage: 86.22, 86.28, 86.04
- Revision: 81.53, 81.55, 81.59, 00.70, 00.71, 00.72, 00.73, 00.80, 00.81, 00.82, 00.83, 00.84
- Removal: 80.05, 80.06, 80.09

9. Mechanical Complication
Presence of one of the following diagnosis codes: 996.4, 996.40, 996.41, 996.42, 996.44, 996.47, 996.49

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
The target population for this measure includes admissions for patients at least 65 years of age undergoing elective primary THA and/or TKA procedures.

2a.5 Target population gender: Female, Male
2a.6 Target population age range: 65 years of age and older

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
This measure was developed using claims data from calendar year 2007 and 2008. The time period for public reporting has not been determined.

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**

The denominator includes patients aged 65 and older admitted to non-federal acute care hospitals for an elective, primary THA and/or TKA in 2007 and 2008. Patients are eligible for inclusion in the denominator if they had a THA and/or a TKA AND had continuous enrollment in Medicare FFS one year prior to the date of index admission.

This cohort is defined using the following ICD-9-CM procedure codes identified in Medicare Part A Inpatient claims data:
81.51 Total Hip Arthroplasty
81.54 Total Knee Arthroplasty

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** Patients will be excluded from the cohort if they meet any of the followed criteria:

1. Patients with hip fractures
   Presence of one of the following diagnosis codes: 733.1, 733.10, 733.14, 733.15, 733.19, 733.8, 733.81, 733.82, 733.95, 733.96, 733.97, 808.0, 808.1, 820.00, 820.01, 820.02, 820.03, 820.09, 820.10, 820.11, 820.12, 820.13, 820.19, 820.20, 820.21, 820.22, 820.30, 820.31, 820.32, 820.8, 820.9, 821, 821.0, 821.00, 821.01, 821.1, 821.10, 821.11
   Rationale: Patients with hip fractures have higher mortality, complication and readmission rates and the procedure (THA) is not elective.

2. Patients undergoing revision procedures (with or without a concurrent THA/TKA)
   Presence of one of the following diagnosis codes: 81.53, 81.55, 81.59, 00.70, 00.71, 00.72, 00.73, 00.80, 00.81, 00.82, 00.83, 00.84
   Rationale: Revision procedures may be performed at a disproportionately small number of hospitals and are associated with higher mortality, complication and readmission rates.

3. Patients undergoing partial hip arthroplasty procedures (with or without a concurrent THA/TKA)
   Presence of the following diagnosis code: 81.52
   Rationale: Partial arthroplasties are primarily done for hip fractures and are typically performed on patients who are older, more frail, and with more comorbid conditions.

4. Patients undergoing resurfacing procedures (with or without a concurrent THA/TKA)
   Presence of one of the following diagnosis codes: 00.85, 00.86, 00.87
   Rationale: Resurfacing procedures are a different type of procedure which are typically performed on younger, healthier patients.

5. Patients who are transferred in to the index hospital
   Rationale: If the patient is transferred from another acute care facility to the hospital where the index procedure occurs, it is likely that the procedure is not elective.

6. Patients who leave the hospital against medical advice (AMA)
   Rationale: Hospitals and physicians do not have the opportunity to provide the highest quality care.

7. Patients with more than two THA/TKA procedure codes during the index hospitalization
   Rationale: Patients with more than two procedure codes for THA/TKA are excluded because it is rare that a patient would have 3 arthroplasty procedures done at one time. This is likely to be a coding error.

8. Patients with multiple admissions for THA/TKA in the 12 months studied; one hospitalization per patient was randomly selected for inclusion after applying the other exclusion criteria
   Rationale: Admissions for the same patient are statistically dependent and it is preferable to include one admission per year in the measure.

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
See "Denominator Exclusion" section

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**

This measure is not stratified.

2a.12-13 Risk Adjustment Type: Risk-adjustment devised specifically for this measure/condition

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
The measure estimates hospital-level RSCRs using hierarchical logistic regression models. In brief, the approach simultaneously models outcomes at two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand et al., 2007). At the patient level, the model adjusts the log-odds of a complication for age, sex, and selected clinical covariates. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of complication at the hospital, after accounting for case mix. If there were no differences among hospitals, then after adjusting for case mix, the hospital intercepts should be identical across all hospitals.

The measure adjusts for key variables that were clinically relevant and had strong relationships with the outcome (e.g. demographic factors, disease severity indicators, and indicators of frailty). For each patient, covariates are obtained from Medicare claims extending 12 months prior to and including the index admission. The model adjusts for case mix differences based on the clinical status of the patient at the time of admission. We use condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis and procedure codes. Conditions that may represent adverse outcomes due to care received during the index admission are not considered for inclusion in the risk adjusted model. Although they may increase the risk of mortality and complications, including them as covariates in a risk-adjusted model could attenuate the measure's ability to characterize the quality of care delivered by hospitals. Hence, these conditions are not adjusted for if they only appear in the index admission and not in the 12 months prior to admission.
The risk adjustment model included 33 variables which are listed below:
Demographic
1. Age-65 (years above 65, continuous)
2. Sex

THA/TKA Procedure
3. THA procedure
4. Number of procedures performed

Clinical Risk Factors
5. Skeletal deformities (ICD-9 code 755.63)
6. Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16)
7. Morbid obesity (ICD-9 code 278.01)
8. Metastatic cancer and acute leukemia (CC 7)
9. Cancer (CC 8-10)
10. Respiratory/Heart/Digestive/Urinary/Other Neoplasms (CC 11-13)
11. Diabetes and DM complications (CC 15-20,119,120)
12. Protein-calorie malnutrition (CC 21)
13. Bone/Joint/Muscle Infections/Necrosis (CC 37)
14. Rheumatoid Arthritis and Inflammatory Connective Tissue Disease (CC 38)
15. Osteoarthritis of hip and knee (CC 40)
16. Osteoporosis and Other Bone/Cartilage Disorders (CC 41)
17. Dementia and senility (CC 49, 50)
18. Major psychiatric disorders (CC 54-56)
19. Hemiplegia, paraplegia, paralysis, function disability (CC 67-69, 100-102, 177-178)
20. Cardio-respiratory failure and shock (CC 79)
21. Chronic atherosclerosis (CC 83-84)
22. Stroke (CC 95, 96)
23. Vascular or circulatory disease (CC 104-106)
24. COPD (CC 108)
25. Pneumonia (CC 111-113)
26. Pleural effusion/pneumothorax (CC 114)
27. End-stage renal disease or dialysis (CC 129, 130)
28. Renal Failure (CC 131)
29. Decubitus ulcer or chronic skin ulcer (CC 148, 149)

30. Trauma (CC 154-156,158-161)
31. Vertebral Fractures (CC 157)
32. Other injuries (CC 162)
33. Major complications of medical care and trauma (CC 164)

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22 (2): 206-226.

2a.15-17 Detailed risk model available Web page URL or attachment: Attachment THA-TKA Complications Technical Report.pdf

2a.18-19 Type of Score: Rate/proportion
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***):**
The RSCR is calculated as the ratio of the number of "predicted" to the number of "expected" complications, multiplied by the national unadjusted complication rate. For each hospital, the "numerator" of the ratio is the number of complications predicted on the basis of the hospital's performance with its observed case mix, and the "denominator" is the number of complications expected on the basis of the nation's performance with that hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital's performance given its case-mix to an average hospital's performance with the same case-mix. Thus a lower ratio indicates lower-than-expected complication or better quality and a higher ratio indicates higher-than-expected complication or worse quality.

The predicted hospital outcome (the numerator) is calculated by regressing the risk factors and the hospital-specific intercept on the risk of complications, multiplying the estimated regression coefficients by the patient characteristics in the hospital, transforming, and then summing over all patients attributed to the hospital to get a value. The expected number of complications (the denominator) is obtained by regressing the risk factors and a common intercept on the complications outcome using all hospitals in our sample, multiplying the subsequent estimated regression coefficients by the patient characteristics observed in the hospital, transforming, and then summing over all patients in the hospital to get a value.

Please see attachment for more details on the calculation algorithm.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
The method for discriminating hospital performance has not been determined. For the 6 publicly reported measures of hospital outcomes developed with similar methodology and reported on the CMS website www.hospitalcompare.hhs.gov, CMS currently estimates an interval estimate for each risk-standardized rate to characterize the amount of uncertainty associated with the rate, compares the interval estimate to the national crude rate for the outcome, and categorizes hospitals as "better than the US national rate," "worse than the US national rate," or "no different than the US national rate." However, the decision to publicly report this measure and the approach has not been determined.

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)*:
This measure is not based on a survery or sample.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Electronic administrative data/claims

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.***):**
The datasets used to create the measures are described below.

1. 2008 Part A (inpatient) data
   Part A inpatient data includes claims paid for Medicare inpatient hospital care, skilled nursing facility care, some home health agency services, and hospice care. For purposes of this project, Part A is used to refer to inpatient services only and includes data from 2 time periods:
   a. Index admission: Index admission data are based on the inclusion/exclusion criteria for THA/TKA, and comorbidities (if any) are identified from the secondary diagnoses associated with the index admission.

b. Pre-index: 12 months prior to the index admission ("pre-index").

2. 2008 Part A (outpatient) data – 12 months pre-index
Hospital outpatient refers to Medicare claims paid for the facility component of surgical or diagnostic procedures, emergency room care, and other non-inpatient services performed in a hospital outpatient department or ambulatory surgical/diagnostic center.

3. Part B data – 12 months pre-index
Part B data refers to Medicare claims for the services of physicians (regardless of setting) and other outpatient care, services, and supplies. For purposes of this project, Part B services included only face-to-face encounters between a care provider and patient. We thus do not include services such as laboratory tests, medical supplies, or other ambulatory services.

4. 2008 Medicare Enrollment Database
This database contains Medicare beneficiary demographic, benefit/coverage, enrollment status on admission, and vital status information. These data have previously been shown to accurately reflect patient vital status (Fleming Fisher et al., 1992).

Fleming C., Fisher ES, Chang CH, Bubolz D, Malenda J. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs Hospitals. Medical Care. 1992; 30(5): 377-91.

2a.26-28 **Data source/data collection instrument reference web page URL or attachment:**

2a.29-31 **Data dictionary/code table web page URL or attachment:** URL N/A
http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1182785083979

2a.32-35 **Level of Measurement/Analysis** (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency

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

2a.38-41 **Clinical Services** (*Healthcare services being measured, check all that apply*)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 **Data/sample** *(description of data/sample and size)*: Medicare Part A inpatient claims data for calendar year 2007 and 2008 were used to test reliability. The 2008 cohort included 290,329 admissions and the 2007 cohort included 294,697 admissions.

2b.2 **Analytic Method** *(type of reliability & rationale, method for testing)*:
The reliability of the model was tested using identical cohort inclusion/exclusion criteria for patients who underwent THA and/or TKA. We randomly selected 50% of the THA and/or TKA admissions that met all inclusion and exclusion criteria in 2008 and created a development sample, which we used to build the model. We used the remaining 50% of THA/TKA admissions in 2008 as the validation sample. We also used all qualifying THA and/or TKA admissions in 2007 data as an additional sample to validate the model. Model performance was assessed in the development dataset and both validation datasets. In addition, we will run the model in addtional datasets and compare the risk-standardized complication rates for each hospital.

2b.3 **Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Preliminary results indicate similar model performance in both cohorts (e.g., ROC=0.69 in the development cohort, 0.70 in the 2008 validation cohort, and 0.69 in the 2007 validation cohort).

2b
C ☐
P ☐
M ☐
N ☐

2c. Validity testing

2c.1 **Data/sample** *(description of data/sample and size)*: Face validity: model performance

2c.2 **Analytic Method** *(type of validity & rationale, method for testing)*:
During measure development, we consulted with representatives from potential users of this measure including clinicians, professional societies, payers, and consumers. We use this field to describe the role that these representatives played on the working group and Technical Expert Panel (TEP). We used a structured measure evaluation tool to assess face validity and other measure properties.

We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcomes measurement set forth in National Quality Forum (NQF) guidance for outcomes measures (National Quality Forum, 2010), CMS Measure Management System guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006). We obtained expert and stakeholder input on the measure through three mechanisms: first, through regular discussions with a working group; second, through a series of three conference calls with a national Technical Expert Panel (TEP); and third, through a public comment period.

Early in the development phase, we assembled a working group that included individuals with clinical and methodological expertise relevant to orthopedic quality measurement. We held regular conference calls throughout the development process, and the Yale team solicited detailed feedback and guidance on key clinical and methodological decisions pertaining to measure development. The working group provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader TEP.

In alignment with CMS' Measure Management System, YNHHSC/CORE also released a public call for nominations and convened a TEP. Potential members were also solicited via e-mail in consultation with the working group and CMS. The role of the TEP was to provide feedback on key methodological decisions made in consultation with the working group. The TEP was comprised of individuals with diverse perspectives and backgrounds including clinicians, consumers, hospitals, purchasers, and experts in quality improvement. Finally, we solicited public comment on the proposed measure through CMS' Measure Management System Public Comment site (https://www.cms.gov/MMS/17_CallforPublicComment.asp#TopOfPage). Public comments were summarized and publicly posted for 30 days. The resulting content was taken into consideration during the final stages of measure development.

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.nysna.org/images/pdfs/practice/nqf_ana_outcomes_draft10.pdf. Accessed August 19, 2010.

Krumholz HM,Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation. January 24, 2006 2006;113(3):456-462.

2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
The experts agree the measure accurately reflects the quality of care and distinguishes levels of quality for patients undergoing THA and/or TKA.

2c
C☐
P☐
M☐
N☐

2d. Exclusions Justified

2d.1 **Summary of Evidence supporting exclusion(s):**
Rationale for exclusions is described in "Denominator Exclusions

2d.2 **Citations for Evidence:**
See "Denominator Exclusions"

2d
C☐
P☐
M☐
N☐
NA
☐

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

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

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

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: 2008 Medicare Part A (inpatient) data, hospital outpatient data, and Part B data were used to identify candidate variables for risk adjustment. Specifically, Medicare Part A inpatient data was used to identify variables for risk adjustment in the index admission, while Part A outpatient and Part B data were used to identify variables for risk adjustment in the 12-month period preceding the index date of admission. As described in section 2b, we developed and validated the model in three separate cohorts to assess and compare model performance: (1) development sample of 145,206 admissions in 2008 data; (2) validation sample of 145,123 admissions in 2008 data; and (3) validation sample of 294,697 admissions in 2007 data.

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
This measure was fully risk-adjusted using a hierarchical logistic regression model to calculate hospital risk-standardized complication rates (RSCR). (see "risk adjustment methodology/variables" for additional details).
Approach to assessing model performance:
For the development and validation cohort, we computed five summary statistics for assessing model performance (Harrell, 2001):
(1) over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)
(2) predictive ability

(3) area under the receiver operating characteristic (ROC) curve

(4) distribution of residuals

(5) model chi-square (A test of statistical significance usually employed for categorical data to determine whether there is a good fit between the observed data and expected values; i.e., whether the differences between observed and expected values are attributable to true differences in characteristics or instead the result of chance variation.

F.E. Harrell and Y.C.T. Shih, Using full probability models to compute probabilities of actual interest to decision makers, Int. J. Technol. Assess. Health Care 17 (2001), pp. 17–26.

2e.3 Testing Results *(risk model performance metrics)*:
Performance Metrics in Development Cohort: Development cohort consisted of 145,206 patient stays at 3,221 hospitals (half of 2008 cohort), with a risk-adjusted median complication rate of 4.2%. The development model has strong discrimination and fit. The risk-standardized complication rate ranges from 2.5% to 8.6%, a range of 6.1%. Results are summarized below:

Over-fitting indices: (0,1)
Residuals lack of fit: <-2 = 0.0%; [-2, 0) = 95.8%; [0, 2) = 0.4%; [2+ = 3.8%
Model Chi-square [# of covariates]: 4401 [33]
Predictive ability (lowest decile %, highest decile %): (2, 15)
Area under the ROC curve = 0.69 (GLM)

The discrimination and the explained variation of the model are consistent with those of models currently used to publicly report condition specific rates of both mortality and readmission.

2e
C☐
P☐
M☐
N☐
NA ☐

Model Validation using 2008 Validation Cohort: 2008 validation cohort consisted of 145,123 admissions (other half of the 2008 cohort) randomly selected from 3,223 hospitals, with a risk-standardized median complication rate of 4.1%. The model performance was not substantively different in this validation sample, as compared to the development sample. Results are summarized below:

Over-fitting indices: (0.04, 1.02)
Residuals lack of fit: <-2 = 0.0%; [-2, 0) = 95.8%; [0, 2) = 0.4%; [2+ = 3.7%
Model Chi-square[# of covariates]: 4698 [33]
Predictive ability (lowest decile %, highest decile %):(2, 15)
Area under the ROC curve = 0.70

Model Validation using 2007 Validation Cohort: 2007 validation cohort consisted of 294,697 admissions from 3,300 hospitals. The model performance was not substantively different in this validation sample, as compared to the development sample. Results are summarized below:

Over-fitting indices: (0.002, 1.00)
Residuals lack of fit: <-2 = 0.0%; [-2, 0) = 95.7%; [0, 2) = 0.4%; [2+ = 3.9%
Model Chi-square[# of covariates]: 9236 [33]
Predictive ability (lowest decile %, highest decile %):(2, 15)
Area under the ROC curve = 0.69

We also examined the temporal variation of the standardized estimates and frequencies of the variables in the models. The frequencies and regression coefficients are fairly consistent over the two cohorts.

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

2f. Identification of Meaningful Differences in Performance

2f.1 **Data/sample from Testing or Current Use** *(description of data/sample and size)*: 2008 Medicare Part A inpatient claims data.

2f.2 **Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*:
Unadjusted median hospital-level complication rates following THA and/or TKA were assessed across hospitals.

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)*:
Median hospital-level risk-standardized complication rate was 4.2% with a range from 2.2-8.9%. This is likely a signal of differences in the quality of care received for patients undergoing THA and/or TKA. Total hip replacement and TKA are elective procedures typically performed on healthy patients. Therefore, complication rates are expected to be lower than that for an emergent procedure. The variation observed for complications is likely a signal that though rates may be relatively low there are differences in the quality of care delivered across hospitals that result in variation in outcomes.

2f C☐ P☐ M☐ N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 **Data/sample** *(description of data/sample and size)*: No comparable data source is available at this time. We will perform validity testing of the development model in data from a different time frame.

2g.2 **Analytic Method** *(type of analysis & rationale)*:
N/A

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

2g C☐ P☐ M☐ N☐ NA☐

2h. Disparities in Care

2h C☐

2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)*: This measure is not stratified. **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** There were no disparities detected during measure development. Please see "Summary of Data on Disparities by Population Group" for additional details.	**P☐ M☐ N☐ NA ☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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:** Not in use but testing completed **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** CMS plans to use the measures for public reporting and will propose the measures through rulemaking process. **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***):** This measure is not currently in use. **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)*: N/A **3a.5 Methods** *(e.g., focus group, survey, QI project)*: No consumer or other field testing has been completed at this time. However, this measure was systematically evaluated by an expert group of orthopedists and Technical Expert Panel (TEP) over a period of 8 months. Regular meetings were held throughout the development of this measure, during which we received input and feedback on key methodological and other measure decisions (see section 2c - Validity Testing for more details on process of TEP input). **3a.6 Results** *(qualitative and/or quantitative results and conclusions)*: The TEP agreed that the measure would be useful in informing consumers and hospitals.	**3a** **C☐ P☐ M☐ N☐**
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:** **5.1 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:** N/A	**3c** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**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*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	**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.** Using administrative claims variables for risk adjustment This measure uses variables from claims data submitted by hospitals to CMS for payment as "clinical" risk adjusters. Prior research has demonstrated that administrative claims data can be used to develop risk-adjusted outcomes measures for mortality following admission for myocardial infarction, heart failure, and death and that the models produce estimates of risk-standardized mortality rates (RSMRs) that are very similar to rates estimated by models based on chart data. This high level of agreement between the clinical and claims-based models supports the use of the claims-based models for public reporting. The models have also demonstrated consistent performance across years of claims data. The approach to identifying risk factors for patients also mitigates the potential limitations of claims data. Because not every diagnosis is	**4d** C☐ P☐ M☐ N☐

coded at every visit, we identified comorbid conditions for risk adjustment in inpatient, outpatient, and physician claims data coded in the year prior to admission, as well as those coded in the secondary diagnosis fields during the index admission. This strategy allows for comprehensive review of patients' medical histories. If a diagnosis appeared only once, in some visits and not others, it was included, minimizing the effect of incomplete coding.

We were careful, however, to include information about each patient's status at admission and not to adjust for possible complications of the admission. Although some codes, by definition, represent conditions that are present before admission (e.g. cancer), other codes and conditions cannot be distinguished from complications occurring during the index hospitalization (e.g. infection or shock). If these are secondary diagnoses from the index admission, then they are not adjusted for in the analysis.

Using administrative claims codes to define complications
This measure identifies complications in claims data. This approach is similar to that used in an ICD complications measure recently approved by NQF. In consultation with a technical expert panel, it was agreed that the codes and restrictions applied to certain complication definitions (i.e., requiring an intervention/procedure code in addition to the diagnosis code for the complication itself) were adequate for identifying clinically significant adverse events (outcomes). To further assess the accuracy of the administrative claims codes, we plan to conduct a validation study to determine whether the specific codes used to identify complications in Medicare claims reliably identify hip/knee complications documented in charts.

Potentially creating access barrier
Because THAs and TKAs are elective procedures, publicly reporting the measure could potentially reduce access to care for certain patients who may be healthy enough to undergo the procedure but who carry a higher risk for complications.

4e. Data Collection Strategy/Implementation

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

4e.2 Costs to implement the measure (*costs of data collection, fees associated with proprietary measures*)**:**
This measure uses claims data submitted by hospitals to CMS for payment, There are no costs associated with data collection, as hospitals are mandated by CMS to submit claims for reimbursement purposes. There is no additional cost/burden on hospitals.

4e.3 Evidence for costs:
N/A

4e.4 Business case documentation: Key points as noted in various sections of this document are as follows:

1. The median 30-day all-cause risk-standardized complication rate is high (4.2%)
2. There is substantial variation in risk-standardized complication rates across hospitals, ranging from 2.2-8.9%, respectively) (based on preliminary analysis of 2008 Part A inpatient claims data).
3. Quality of care should be addressed as THA and TKA procedures are associated with high volume and cost (relative to other elective procedures performed in the Medicare population).

4e
C☐ P☐ M☐ N☐

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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
Centers for Medicare & Medicaid Services (CMS), 500 Security Blvd, Mail Stop S3-02-01, Baltimore, Maryland, 21244-9045

Co.2 Point of Contact
Lein, Han, PhD, lein.han@cms.hhs.gov, 410-786-0205-

Measure Developer If different from Measure Steward
Co.3 Organization
Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/ CORE), 1 Church Street, Suite 200, New Haven, Connecticut, 06510

Co.4 Point of Contact
Laura, Grosso, PhD, laura.grosso@yale.edu, 203-764-5742-

Co.5 Submitter If different from Measure Steward POC
Laura, Grosso, PhD, laura.grosso@yale.edu, 203-764-5742-, Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/ CORE)

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.
Working Group
Role: To provide technical expertise on key methodological decisions during measure development.
Members:
Daniel J. Berry, MD
Professor of Orthopedics, Mayo Clinic College of Medicine
Chair, Department of Orthopædic Surgery, Mayo Clinic
Kevin J. Bozic, MD, MBA
Associate Professor and vice chair, Department of Orthopaedic Surgery at the University of California, San Francisco
Chair, Health Systems Committee, AAOS

Robert Bucholz, MD
Professor, Orthopædic Surgery, University of Texas Southwestern Medical Center
Past President, American Academy of Orthopædic Surgeons (AAOS)

Lisa Gale Suter, MD
Assistant Professor, Yale University School of Medicine, Rheumatology (West Haven Veterans Association Hospital)

Charles M. Turkelson, PhD
Director of Research and Scientific Affairs, AAOS

Lawrence Weis, MD
Assistant Professor, Yale Orthopædics and Rehabilitation, Yale University School of Medicine, Orthopædics (West Haven Veterans Association Hospital)

Technical Expert Panel
Role: To provide feedback on recommendations for measure development.
Members:
Mark L. Francis, MD
Professor of Medicine and Biomedical Sciences, Chief, Division of Rheumatology, Department of Internal Medicine, Texas Tech University Health Sciences Center

Cynthia Jacelon, PhD, RN, CRRN
Associate Professor, School of Nursing, University of Massachusetts
Association of Rehabilitation Nurses

Norman Johanson, MD
Chairman, Orthopedic Surgery, Drexel University College of Medicine

C. Kent Kwoh, MD
Professor of Medicine, University of Pittsburgh
Associate Chief and Director of Clinical Research,
Division of Rheumatology and Clinical Immunology

Courtland G. Lewis, MD
American Association of Orthopaedic Surgeons

Jay Lieberman, MD
Professor and Chairman, Department of Orthopedic Surgery, University of Connecticut Health Center
Director, New England Musculoskeletal Institute

Peter Lindenauer, MD, M.Sc.
Hospitalist and Health Services Researcher, Baystate Medical Center
Professor of Medicine, Tufts University

Russell Robbins, MD, MBA
Principal, Mercer´s Total Health Management
Barbara Schaffer
THA Patient

Nelson SooHoo, MD, MPH
Professor, University of California at Los Angeles

Steven H. Stern, MD
Vice President, Cardiology & Orthopedics/ Neuroscience, UnitedHealthcare

Richard E. White, Jr., MD
American Association of Hip and Knee Surgeons

Ad.2 If adapted, provide name of original measure:
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 Complication calculation algorithm.pdf
Date of Submission (*MM/DD/YY*)**:** 03/28/2011

THA/TKA Complication Calculation Algorithm

We estimate a generalized linear model and a hierarchical generalized linear model which accounts for the clustering of observations within hospitals. The generalized linear model (GLM) links the outcome to the patient-level risk factors,[20] Let Y_{ij} denote the outcome (equal to 1 if patient dies or has a complication, zero otherwise) for the j^{th} patient who had a THA/TKA procedure at the i^{th} hospital; $\mathbf{Z}_{ij}$ denotes a set of risk factors based on the data. Let I denote the total number of hospitals and n_i the number of index patient stays in hospital i. We assume the outcome is related linearly to the covariates via a known linked function, h, where

GLM $h(Y_{ij}) = \alpha + \beta\mathbf{Z}_{ij}$ (1)

and $\mathbf{Z}_{ij} = (Z_{1ij}, Z_{2ij}, \ldots, Z_{pij})$ is a set of p patient-specific covariates. In our case, h = the logit link.

To account for the natural clustering of observations within hospitals, we then estimate an HGLM that links the risk factors to the same outcome and a hospital-specific random effect,

HGLM $h(Y_{ij}) = \alpha_i + \beta\mathbf{Z}_{ij}$ (2)

$\alpha_i = \mu + \omega_i;\ \omega_i \sim N(0, \tau^2)$ (3)

where α_i represents the hospital-specific intercept, $\mathbf{Z}_{ij}$ is defined as above, μ the adjusted average outcome over all hospitals in the sample, and τ^2 the between-hospital variance component.[21] This model separates within-hospital variation from between-hospital variation. Both HGLMs and GLMs are estimated using the SAS software system (GLIMMIX and LOGISTIC procedures, respectively).

We first fit the GLM described in Equation (1) using the logit link.

Having identified the covariates that were selected, we next fit the HGLM described in Equations (2) and (3), again using the logit link function; e.g.,

Logit $(P(Y_{ij} = 1)) = \alpha_i + \beta\ \mathbf{Z}_{ij}$

$\alpha_i = \mu + \omega_i,\ \omega_i \sim N(0, \tau^2)$

where $\mathbf{Z}_{ij}$ consisted of the covariates retained in the GLM model. As before, $Y_{ij} = 1$ if patient j treated at hospital i had the event; 0 otherwise.

Hospital performance reporting

Using the set of risk factors in the GLM, we fit the HGLM defined by Equations (2) - (3) and estimate the parameters, $\hat{\mu}$, $\{\hat{\alpha}_i, \hat{\alpha}_2, \ldots, \hat{\alpha}_I\}$, $\hat{\beta}$, and $\hat{\tau}^2$. We calculate a standardized outcome, s_i, for each hospital by computing the ratio of the number of predicted complications to the number of expected complications, multiplied by the unadjusted overall complication rate, $\bar{y}$. Specifically, we calculate

Predicted $\hat{y}_{ij}(Z) = h^{-1}(\hat{\alpha}_i + \hat{\beta}\,\mathbf{Z}_{ij})$ (4)

Expected $\hat{e}_{ij}(Z) = h^{-1}(\hat{\mu} + \hat{\beta}\,\mathbf{Z}_{ij})$ (5)

$$\hat{s}_i(Z) = \frac{\sum_{j=1}^{n_i} \hat{y}_{ij}(Z)}{\sum_{j=1}^{n_i} \hat{e}_{ij}(Z)} \times \bar{y} \qquad (6)$$

If more (fewer) "predicted" cases than "expected" cases have the outcome in a hospital, then $\hat{s}_i$ will be higher (lower) than the unadjusted average. For each hospital, we compute an interval estimate of s_i to characterize the level of uncertainty around the point estimate using bootstrapping simulations. The point estimate and interval estimate can be used to characterize and compare hospital performance (e.g., higher than expected, as expected, or lower than expected).

Creating Interval Estimates

Because the statistic described in Equation (6) is a complex function of parameter estimates, we use re-sampling and simulation techniques to derive an interval estimate. The bootstrapping simulation has the advantage of avoiding unnecessary distributional assumptions.

Calculation Algorithm

Let *I* denote the total number of hospitals in the sample. We repeat steps 1 – 4 below for b = 1,2,…B times:

1. Sample *I* hospitals with replacement.

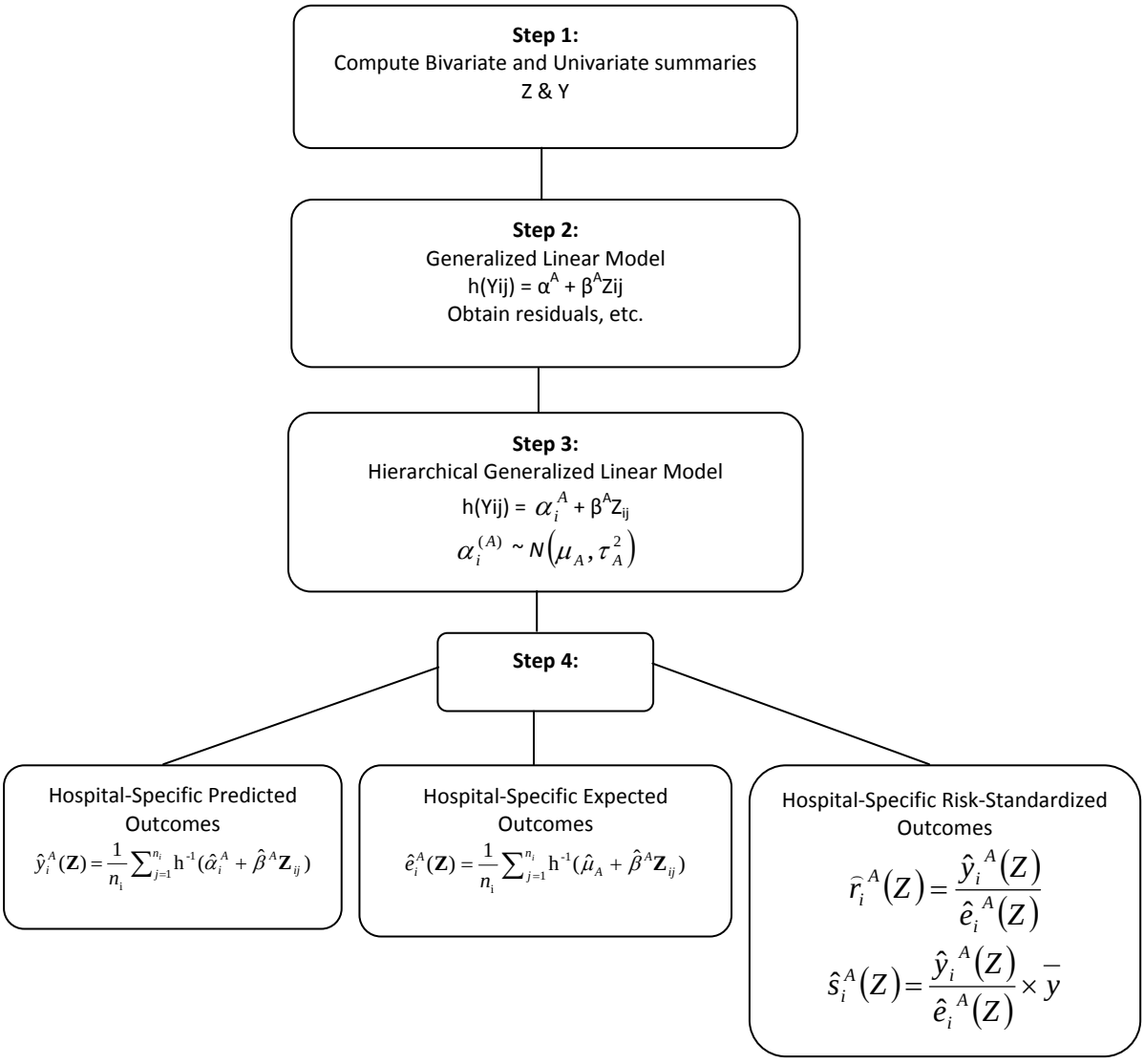
2. Fit the HGLM using all patients within each sampled hospital. We use as starting values the parameter estimates obtained by fitting the model to all hospitals. If some hospitals are selected more than once in a bootstrapped sample, we treat them as distinct so that we have *I* random effects to estimate the variance components. At the conclusion of Step 2, we have:
    a. $\hat{\beta}^{(b)}$ (the estimated regression coefficients of the risk factors).
    b. The parameters governing the random effects, hospital adjusted outcomes, distribution, $\hat{\mu}^{(b)}$ and $\hat{\tau}^{2(b)}$.
    c. The set of hospital-specific intercepts and corresponding variances, $\{\hat{\alpha}_i^{(b)}, \hat{\mathrm{var}}(\alpha_i^{(b)}); i = 1,2,\ldots,I\}$.

3. We generate a hospital random effect by sampling from the distribution of the hospital-specific distribution obtained in Step 2c. We approximate the distribution for each random effect by a normal distribution. Thus, we draw $\alpha_i^{(b*)} \sim N\left(\hat{\alpha}_i^{(b)}, \hat{\mathrm{var}}\left(\hat{\alpha}_i^{(b)}\right)\right)$ for the unique set of hospitals sampled in Step 1.

4. Within each unique hospital *i* sampled in Step 1, and for each case *j* in that hospital, we calculate $\hat{y}_{ij}^{(b)}$, $\hat{e}_{ij}^{(b)}$, and $\hat{s}_i(Z)^{(b)}$ where $\hat{\beta}^{(b)}$ and $\hat{\mu}^{(b)}$ are obtained from Step 2 and $\hat{\alpha}_i^{(b*)}$ is obtained from Step 3.

Ninety-five percent interval estimates (or alternative interval estimates) for the hospital-standardized outcome can be computed by identifying the 2.5th and 97.5th percentiles of the B estimates (or the percentiles corresponding to the alternative desired intervals).

Figure 1. Analysis Steps



Outcomes Measure: Hospital-level Risk-Standardized Complication Rates following Elective Total Hip Arthroplasty (THA) and Total Knee Arthroplasty (TKA)

Measure Methodology Report

Submitted By Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE):

Laura M. Grosso, PhD, MPH
Jeptha P. Curtis, MD
Zhenqui Lin, PhD
Lori L. Geary, MPH
Smitha Vellanky, MSc
Carol Oladele, MPH
Yongfei Wang, MS
Elizabeth E. Drye, MD, SM
Harlan M. Krumholz, MD, SM

Contract # HHSM-500-2008-0025I-MIDS Task Order T0001

Prepared For:

Centers for Medicare & Medicaid Services (CMS)

Submitted September 29, 2010

ACKNOWLEDGEMENTS

We would like to acknowledge the ongoing support from our working group. These individuals gave generously of their time, providing guidance on key clinical and methodological decisions. Without their expertise, this project would not have been possible. Working group members include:

Daniel J. Berry, MD
Professor of Orthopedics, Mayo Clinic College of Medicine
Chair, Department of Orthopædic Surgery, Mayo Clinic

Kevin J. Bozic, MD, MBA
Associate Professor and vice chair, Department of Orthopaedic Surgery at the University of California, San Francisco
Chair, Health Systems Committee, AAOS

Robert Bucholz, MD
Professor, Orthopædic Surgery, University of Texas Southwestern Medical Center
Past President, American Academy of Orthopædic Surgeons (AAOS)

Lisa Gale Suter, MD
Assistant Professor, Yale University School of Medicine, Rheumatology

Charles M. Turkelson, PhD
Director of Research and Scientific Affairs, AAOS

Lawrence Weis, MD
Assistant Professor, Yale Orthopædics and Rehabilitation, Yale University School of Medicine, Orthopædics (West Haven Veterans Association Hospital)

We would also like to thank the Technical Expert Panel members, who provided great support and insight on key measure decisions. These members include:

Mark L. Francis, MD
Professor of Medicine and Biomedical Sciences, Chief, Division of Rheumatology, Department of Internal Medicine, Texas Tech University Health Sciences Center, Texas Tech University, Health Sciences Center

Cynthia Jacelon, PhD, RN, CRRN
Associate Professor, School of Nursing, University of Massachusetts
Association of Rehabilitation Nurses

Norman Johanson, MD
Chairman, Orthopedic Surgery, Drexel University College of Medicine

C. Kent Kwoh, MD

Professor of Medicine, Associate Chief and Director of Clinical Research, Division of Rheumatology and Clinical Immunology University of Pittsburgh

Courtland G. Lewis, MD
American Association of Orthopaedic Surgeons

Jay Lieberman, MD
Professor and Chairman, Department of Orthopedic Surgery, University of Connecticut Health Center; Director, New England Musculoskeletal Institute

Peter Lindenauer, MD, M.Sc.
Hospitalist and Health Services Researcher, Baystate Medical Center; Professor of Medicine, Tufts University

Russell Robbins, MD, MBA
Principal, Mercer's Total Health Management

Barbara Schaffer
THA Patient

Nelson SooHoo, MD, MPH
Professor, University of California at Los Angeles

Steven H. Stern, MD
Vice President, Cardiology & Orthopedics/ Neuroscience, UnitedHealthcare

Richard E. White, Jr., MD
American Association of Hip and Knee Surgeons

In addition, we would like to acknowledge and thank Angela Merrill, Sandi Nelson, Marian Wrobel, and Eric Schone from Mathematica Policy Research, Inc. and Lein Han and Michael Rapp at the Centers for Medicare & Medicaid Services for their contributions to this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Overview of Measure

Total hip and knee arthroplasties (THA and TKA, respectively) are commonly performed procedures that improve quality of life. In 2003 there were 202,500 THAs and 402,100 TKAs performed[1] and the number of procedures performed has increased steadily over the past decade.[2-3]

Although these procedures dramatically improve quality of life, they are costly. In 2005 annual hospital charges totaled $3.95 billion and $7.42 billion for primary THA and TKA, respectively.[2] These costs are projected to increase by 340% to 17.4 billion for THA and by 450% to 40.8 billion for TKA by 2015.[2] Medicare is the single largest payer for these procedures, covering approximately two-thirds of all THAs and TKAs performed in the US.[3] Combined, THA and TKA procedures account for the largest procedural cost in the Medicare budget.[4]

Given the high volume and cost associated with these procedures (relative to other elective procedures performed in the Medicare population), it is imperative to address quality of care. Complications increase costs associated with THA and TKA and affect the quality, and potentially quantity, of life for patients. A quality measure to address complications following THA and TKA provides an opportunity to provide targets for efforts to improve the quality of care and reduce costs for patients undergoing these elective procedures.

CMS contracted with Yale-New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE) to develop hospital outcomes measures that reflect the quality of care for patients undergoing elective THA and TKA procedures and are suitable for public reporting. YNHHSC/CORE, in consultation with CMS and a working group of leading national orthopedic surgeons active in quality improvement, developed a hospital-level, risk-standardized measure of complication rates following elective THA and TKA procedures. The goal of the measure is to improve the quality of care delivered to patients undergoing THA and TKA procedures.

This report provides the background and detailed technical information on the measure. In brief, we developed a model that estimates hospital-specific, risk-standardized, complication rates following THA/TKA. We used Medicare claims data and linked it to CMS claims and enrollment data to identify complications after THA/TKA. To account for the clustering of observations within hospitals and differences in the number of admissions across hospitals, we used hierarchical logistic regression to estimate the risk-standardized complication rates (RSCRs).

This measure was developed concurrently with a second CMS outcomes measure – 30-day all-cause readmission following THA and/or TKA. These are complementary measures that assess separate domains of quality. The complications measure will inform quality improvement efforts targeted toward minimizing medical and surgical complications during surgery and in the recovery phase. The readmission measure captures an additional domain of care provided in the transition to outpatient settings. The readmission measure is presented in a separate technical report.

These two measures expand a set of hospital outcomes measures CMS has developed to improve hospital quality and meet its mandate under the Deficit Reduction Act (DRA) of 2005 to publicly report outcomes and efficiency measures on the consumer site, Hospital Compare (http://www.hospitalcompare.hhs.gov). CMS began publicly reporting acute myocardial infarction (AMI) and heart failure (HF) 30-day mortality measures as outcomes measures in June 2007, and added a pneumonia 30-day mortality measure in August 2008. In addition, CMS began publicly reporting 30-day readmission measures for AMI, HF, and pneumonia in July 2009.

1.2 Approach to Measure Development

We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside expert and public input. The measure is consistent with the technical approach to outcomes measurement set forth in National Quality Forum (NQF) guidance for outcomes measures [5], CMS Measure Management System guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes". [6] We obtained expert and stakeholder input on the measure through three mechanisms: first, through regular discussions with a working group; second, through a series of three conference calls with a national Technical Expert Panel (TEP); and third, through a public comment period.

Early in the development phase, we assembled a working group that included individuals with clinical and methodological expertise relevant to orthopedic quality measurement. We held regular conference calls throughout the development process and the Yale team solicited detailed feedback and guidance on key clinical and methodological decisions pertaining to measure development. The working group provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader TEP.

In alignment with CMS' Measure Management System, YNHHSC/CORE also released a public call for nominations and convened a TEP. Potential members were also solicited via e-mail in consultation with the working group and CMS. The role of the TEP was to provide feedback on key methodological decisions made in

consultation with the working group. The TEP was comprised of individuals with diverse perspectives and backgrounds and included clinicians, consumers, hospitals, purchasers, and experts in quality improvement. Finally, we solicited public comment on the proposed measure through CMS' Measure Management System Public Comment site (https://www.cms.gov/MMS/17_CallforPublicComment.asp#TopOfPage). Public comments were summarized and publicly posted for 30 days. The resulting content was taken into consideration during the final stages of measure development.

1.3 Importance of a Complications Measure

Measuring and reporting complication rates will inform health care providers about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care for Medicare patients. The measure will also provide patients with information that could guide their choices. In addition, it has the potential to lower health care costs associated with complications. The measure will increase transparency for consumers.

Complication rates have been shown to vary across hospitals, suggesting care can be improved. Prospective studies show risk adjusted rates for periprosthetic joint infection, a rare but devastating complication, vary between 2.3 to 1.6 percent after 1 and 2 years of follow-up respectively.[7-8] Ninety-day death rates following THA also range from 0.7 to 2.7 percent and are high for an elective procedure.[9-10] Rates for pulmonary embolism following TKA range from 0.5 and 0.9 percent.[10-13] Rates for wound infection in Medicare population based studies vary between 0.3 and 1.0 percent.[10, 12-13] Rates for septicemia range from 0.1%, during the index admission[14] to 0.3%, 90 days following discharge for primary TKA.[10] Rates for bleeding and hematoma following TKA range from 0.94[14] to 1.7%.[15]

The variation in complication rates across hospitals indicates there is room for quality improvement and targeted efforts to reduce these complications could result in significant cost savings.

2. METHODS

2.1 Overview

We developed a hospital-level complications quality measure for patients undergoing THA and TKA procedures. The model estimates hospital-level RSCRs using hierarchical generalized linear modeling (HGLM), to account for the clustering of patients within hospitals. To adjust for differences in hospital case mix, the model adjusts for patient risk factors, including age and comorbidities present at the time of admission.

We identified index admissions for inclusion in the measure via ICD-9 procedure codes for THA and TKA in 2008 Medicare Part A inpatient claims. Because there are no dates associated with procedure codes in Part A data, we use the date of the index admission as the starting point for all follow-up. We used Medicare Part A data for years 2008 and 2009 to identify complications associated with these claims. We identified Information on comorbid conditions for risk adjustment using ICD-9 codes in inpatient, outpatient, and part B Medicare claims data in the 12 months prior to the date of the index admission.

The measure calculates the hospital risk-standardized complication rate (RSCR) by producing a ratio of the number of "predicted" to the number of "expected" complications for each hospital and then multiplying the ratio by the national raw complication rate. For each hospital, the "numerator" of the ratio is the number of complications predicted on the basis of the hospital's performance with its observed case mix (using a hospital-specific estimate intercept term), and the "denominator" is the number of expected complications, based on the nation's performance using the hospital's observed case mix and the national intercept term. In other words, we estimate the complication rate based on each hospital's particular experience and divide it by the estimated complication rate had the hospital performed at the average level for all the hospitals.

The model estimates the hospital-specific intercept term used in the numerator based on how well each hospital performs relative to other hospitals with a similar case mix. Among hospitals with similar case mixes, hospitals that have a lower rate of complications will have a lower intercept term; hospitals that have a higher rate of complications will have a higher intercept term.

2.2 Data Sources

We obtained index admission and in-hospital comorbidity data from Medicare's Standard Analytic File (SAF). Comorbidities were also assessed using Part A inpatient and outpatient, and Part B physician office and hospital Medicare claims in the 12 months prior to admission. Enrollment and post-discharge mortality status were obtained from Medicare's enrollment database which contains beneficiary demographic, benefit/coverage, and vital status information.

2.3 Outcome Definition

After conducting a comprehensive literature review and in consultation with the working group, YNHHSC/CORE identified complications for potential inclusion in a complications measure. To be considered as candidates for inclusion in the outcome, the complications had to:

- Represent meaningful complications attributable to the THA/TKA procedures
- Be identifiable in administrative claims data
- Be fair to hospitals and physicians

Based on these criteria and in consultation with the working group, we identified several candidate complications for inclusion in a composite complications measure:

- Death
- Mechanical complications
- Periprosthetic joint infection
- Surgical site bleeding
- Wound infection
- Pulmonary embolism
- Acute myocardial infarction (AMI)
- Pneumonia
- Sepsis/septicemia
- Deep vein thrombosis (DVT)
- Urinary tract infection (UTI)

DVT and UTI were excluded based on working group feedback documented in the literature. We excluded DVT because experts advised that there is wide variability across hospitals in screening [16-17] and readmission practices for this complication. We excluded UTI because there is wide variability in diagnosing UTI, and the rates are likely inflated due to overdiagnosis in patients post THA/TKA [18] Working group members also noted that there is wide variability in readmission for UTI in US hospitals and wide variability in treatment for it.

Based on these considerations, we included the following complications in the measure:

- Death
- Mechanical complications
- Periprosthetic joint infection
- Surgical site bleeding
- Wound infection
- Pulmonary embolism
- AMI
- Pneumonia
- Sepsis/septicemia

Some of these complications have varying degrees of severity not conveyed in the ICD-9 codes, specifically, wound infection, periprosthetic joint infection, and surgical site bleeding. For example, the claims codes used to identify wound infection may reflect redness and swelling around the incision site, or a true wound infection, requiring incision and drainage. Thus, to capture clinically important complications and to reduce the likelihood of capturing miscoded complications, working group and TEP members recommended only counting these complications in the outcome if they are associated with accompanying ICD-9 procedure codes indicating that they were severe enough to require specific interventions. We therefore imposed additional coding requirements for these complications to set an appropriate threshold for severity.

We include the following complications in the outcome only if they are accompanied by the following procedure codes listed during the admission in which the complication occurred:

Periprosthetic joint infection

Presence of a periprosthetic joint infection code AND the presence of <u>at least one</u> of the following procedure codes

- Incision and drainage
- Revision
- Removal

Wound infection

Presence of a wound infection code AND the presence of <u>at least one</u> of the following procedure codes:

- Incision and drainage
- Revision
- Removal

Surgical site bleeding

Presence of a surgical site bleeding code AND the presence of the following procedure code:

- Incision and drainage

Please refer to Appendix A for complication-specific measure specifications.

2.4 Measure Timeframe

To determine the appropriate follow-up period, we obtained clinical input and examined 90-day trends in complication rates (Figures 1 and 2). Figure 1 conveys the week-by-week rates for mortality and surgical complications occurring from the date of index admission to 90 days post date of index admission. Figure 2

conveys the week-by-week rates for medical complications. These analyses indicate that most complications occur 7 days following the procedure, but level off at 30 days. Although a standardized period of follow-up is ideal, defining a single optimal period of assessment appropriate for a wide range of complications was challenging. For example, the working group and TEP agreed that mechanical complications and periprosthetic joint infection are still attributable to the procedure for up to 90 days following the procedure, while medical complications, such as AMI, are far less likely to be attributable to the procedure after 7 days. Both the working group and TEP advised that we establish complication-specific follow-up periods. Accordingly, we reviewed each complication with the working group and TEP and chose either a 7, 30, or 90 day follow-up period by consensus.

We observe two complications for 90 days: mechanical complications and periprosthetic joint infection as these complications are still attributable to the index THA/TKA for up to 90 days afterwards. Preliminary analyses indicate rates for mechanical complications are elevated until 90 days post the date of index admission. We observe four complications for 30 days: death, surgical site bleeding, wound infection, and pulmonary embolism as rates of these complications are elevated until approximately 30 days post the date of index admission. This finding was consistent with input from clinical experts. AMI, pneumonia, and sepsis/septicemia are followed to 7 days post date of index admission (Figure 2). These conditions are more likely to be attributable to procedure if they occur within the first week after the procedure.

Analyses indicate that the rate for these complications decreases sharply 7 days from the date of index admission and a 7 day follow-up period limits overlap with the 30-day all-cause readmission measure. The list of complications and their associated follow-up periods are listed in Table 1.

Figure 1. Trend in Mortality and Surgical Complication Rates (Medicare FFS Part A Inpatient Data, 2008)

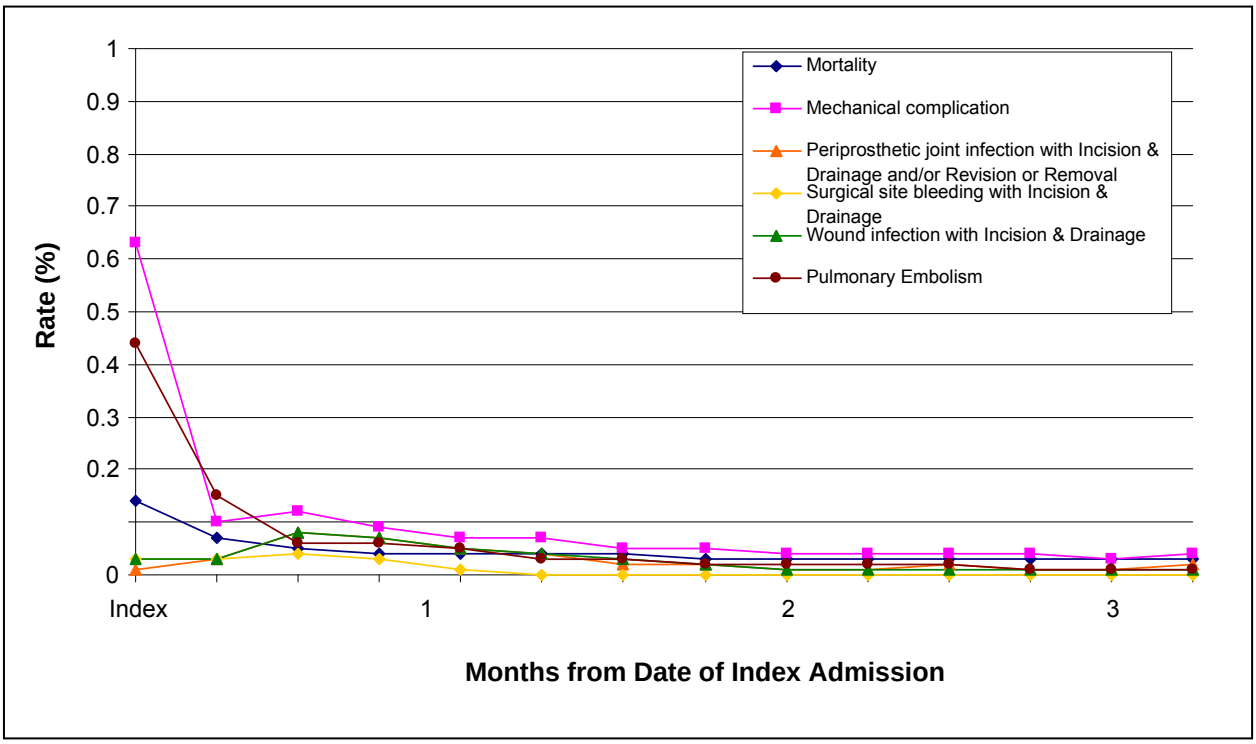


Figure 2. Trend in Medical Complication Rates (Medicare FFS Part A Inpatient Data, 2008)

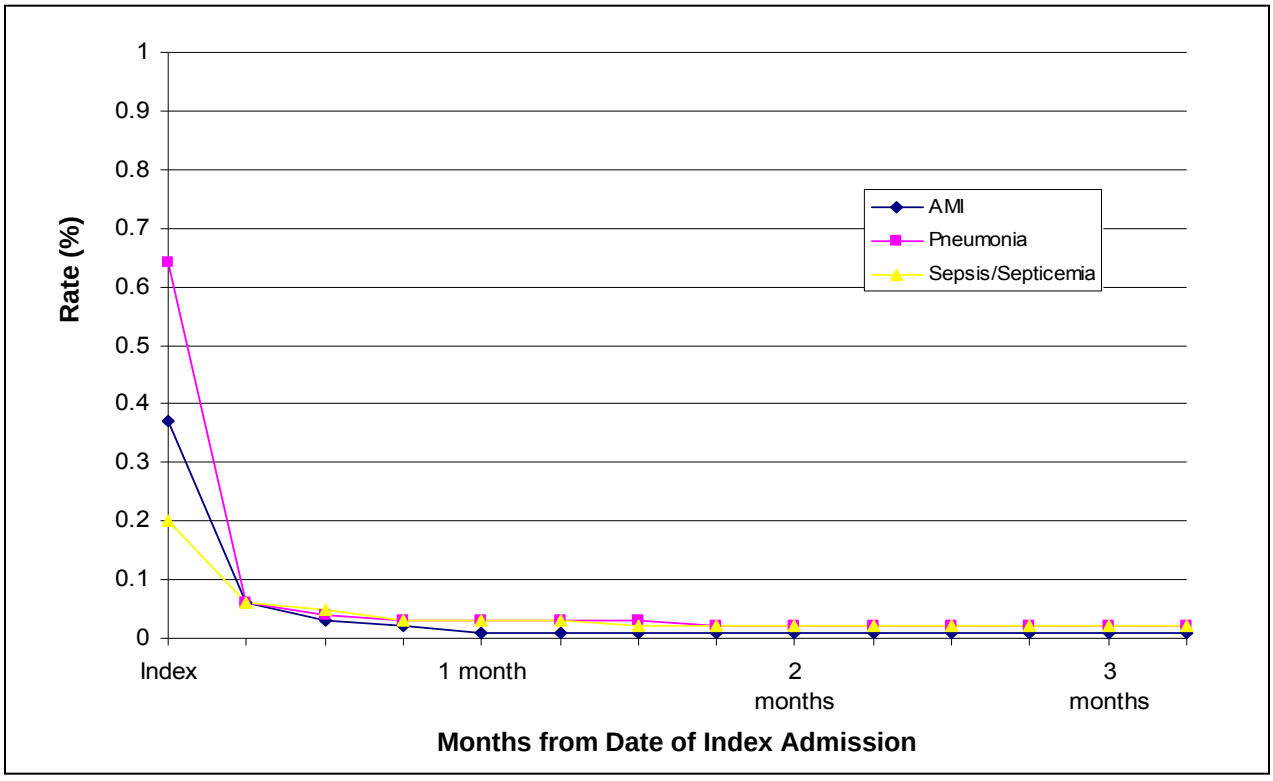


Table 1. Complication Follow-up Periods

Complication	Follow-up Period	Rationale
Death	30 days	• Still attributable to procedure
Mechanical complications	90 days	• Mechanical complications occurring 90 days post procedure can still be attributable to the index procedure • Data indicate that the rate is elevated until 90 days post procedure
Periprosthetic joint infection (PJI)	90 days	• Periprosthetic joint infections occurring 90 days post procedure can still be attributable to the index procedure • Although the rate tapers off after approximately 6 weeks, it remains slightly elevated until 90 days post procedure
Surgical site bleeding	30 days	• Consistent with clinical course • Data indicate that rate decreases after 30 days
Wound infection	30 days	• Consistent with clinical course • Data indicate that rate decreases after 30 days
Pulmonary embolism	30 days	• Consistent with clinical course • Data indicate that rate decreases after 30 days
AMI	7 days	• More likely to be attributable to procedure if it occurs within 7 days of procedure • Rate decreases sharply 7 days from admission and returns to baseline within 30 days • Limits overlap with 30-day all-cause readmission measure
Pneumonia	7 days	• More likely to be attributable to procedure if it occurs within 7 days of procedure • Rate decreases sharply 7 days from admission and returns to baseline within 30 days • Limits overlap with 30-day all-cause readmission measure
Sepsis/septicemia	7 days	• More likely to be attributable to procedure if it occurs within 7 days of procedure • Rate decreases 7 days from admission and returns to baseline within 30 days • Limits overlap with 30-day all-cause readmission measure

The working group and TEP recognized that a model using complication-specific timeframes may make measure interpretation more complex, but there was agreement that this potential disadvantage was offset by improvements to face validity and acceptability of the measure.

2.5 Cohort Definition

In consultation with the working group, we considered whether to develop separate measures for patients undergoing THA and TKA procedures or to combine patients undergoing either procedure into a single hospital quality measure. We combined these patient cohorts for the complications measure for several reasons, including:

- A large proportion of THA and TKA procedures are elective and performed in similar patient cohorts for similar indications (e.g., osteoarthritis)
- The same surgeons frequently perform both procedures
- Both procedures have similar lengths of stay
- The rates and types of complications are similar (Table 2)
- The mortality and readmission rates are similar (Table 2)
- Hospitals develop protocols/programs for lower extremity total joint arthroplasty, rather than for THA and TKA separately
- Combining admissions for both procedures will provide greater power to detect hospital-level variation to enable quality improvement

Table 2. Procedure Characteristics and Unadjusted Mortality, Readmission, and Complication Rates for THA and TKA (Medicare Inpatient Part A, 2008).

		Total Hip Replacement* (excludes partial hip replacement and hip fractures)	**Total Knee Replacement****
Procedure-related characteristics			
Number of Patients Receiving Procedure		97,130	240,517
Mean Length of Stay (SD)		3.8 (2.3)	3.6 (1.7)
Mean Patient Age (SD)		75.2 (6.6)	74.2 (6.1)
Number of Hospitals Performing Procedure		3083	3307
Median Number of Procedures Performed at Each Hospital (Q1-Q3)		16 (6 - 41)	40 (13 - 257)
Mortality		**% (5th-95th)**	**% (5th-95th)**
In-hospital Mortality	Patient level	0.2	0.1
	Hospital level: median	0 (0 - 0.9)	0 (0 - 0.6)
30-day Mortality	Patient level	0.5	0.3
	Hospital level: median	0 (0 - 2.9)	0 (0 - 1.7)
90-day Mortality	Patient level	0.9	0.5
	Hospital level: median	0 (0 - 5.6)	0 (0 - 3.0)
Readmission		**% (5th-95th)**	**% (5th-95th)**
30-day All-cause Readmission	Patient level	6.9	5.9
	Hospital level: median	5 (0 - 25)	5 (0 - 18)
90-day All-cause Readmission	Patient level	12.2	10.7
	Hospital level: median	11 (0 - 38)	10 (0 - 27)
Complications		**% (30-day / 90-day)**	**% (30-day / 90-**
Dislocation		0.8 / 1.1	0.1 / 0.1
DVT		0.1 /0.2	0.2 / 0.2
Hematoma		1.9 / 2.0	1.2 / 1.3
Periprosthetic Joint Infection		0.5 / 0.7	0.4 / 0.6
Postoperative infection		0.8 / 1.0	0.7 / 0.8
Pulmonary Embolism		0.5 / 0.7	0.8 / 1.0
Mechanical complication of internal orthopedic device, implant and graft		2.7 / 3.3	0.3 / 0.4
Venous thrombosis		0.1 / 0.2	0.1 / 0.1
Wound Infection		0.7 / 0.9	0.7 / 0.8
All complications combined		5.8 / 7.0	3.4 / 4.1
* Includes ICD-9 code 81.51 ** Incudes ICD-9 code 81.54			

Patients undergoing non-elective THA or TKA have greater risk of complications and receive a wider variety of surgical procedures than individuals undergoing elective THA or TKA. In consultation with the working group and with the goal of defining a comprehensive yet reasonably homogeneous cohort for quality assessment, we selected inclusion and exclusion criteria in order to identify patients undergoing elective THA and TKA for degenerative (either primary or secondary) arthritis.

Patients eligible for inclusion in the measure are those aged 65 and older admitted to non-federal acute care hospitals with an ICD-9 code for THA and/or TKA. Patients must have had continuous enrollment in Medicare fee-for-service (FFS) for one year prior to the date of index admission to ensure full data availability for risk adjustment. The flow chart depicting cohort selection is presented in Figure 3.

Eligible index admissions are identified using the following ICD-9-CM procedure codes in Medicare Part A Inpatient claims data:

- 81.51 Total Hip Arthroplasty
- 81.54 Total Knee Arthroplasty

2.6 Exclusion Criteria

We excluded the following patient stays from the cohort:

1. Patients with hip fractures
   Rationale: Patients with hip fractures have higher mortality, complication and readmission rates and the procedures are not elective

2. Patients undergoing revision procedures (with or without a concurrent THA/TKA)
   Rationale: Revision procedures may be performed at a disproportionately small number of hospitals and are associated with higher mortality, complication and readmission rates

3. Patients undergoing partial hip arthroplasty (PHA) procedures (with or without a concurrent THA/TKA)
   Rationale: Partial arthroplasties are primarily done for hip fractures and are typically performed on patients who are older, more frail, and with more comorbid conditions

4. Patients undergoing resurfacing procedures (with or without a concurrent THA/TKA)
   Rationale: Resurfacing procedures are a different type of procedure where only the joint's articular surface is replaced. A THA involves surgical removal of the neck of the femur (thighbone) and insertion of a stem deep

inside the bone to connect with the pelvic socket and liner. These procedures are typically performed on younger, healthier patients

5. Patients who were transferred in to the index hospital
   Rationale: If the patient is transferred from another acute care facility to the hospital where the index procedure occurs, it is likely that the procedure is not elective or that the admission is associated with an acute condition

6. Patients who leave the hospital against medical advice (AMA)
   Rationale: Hospitals and physicians do not have the opportunity to provide the highest quality care

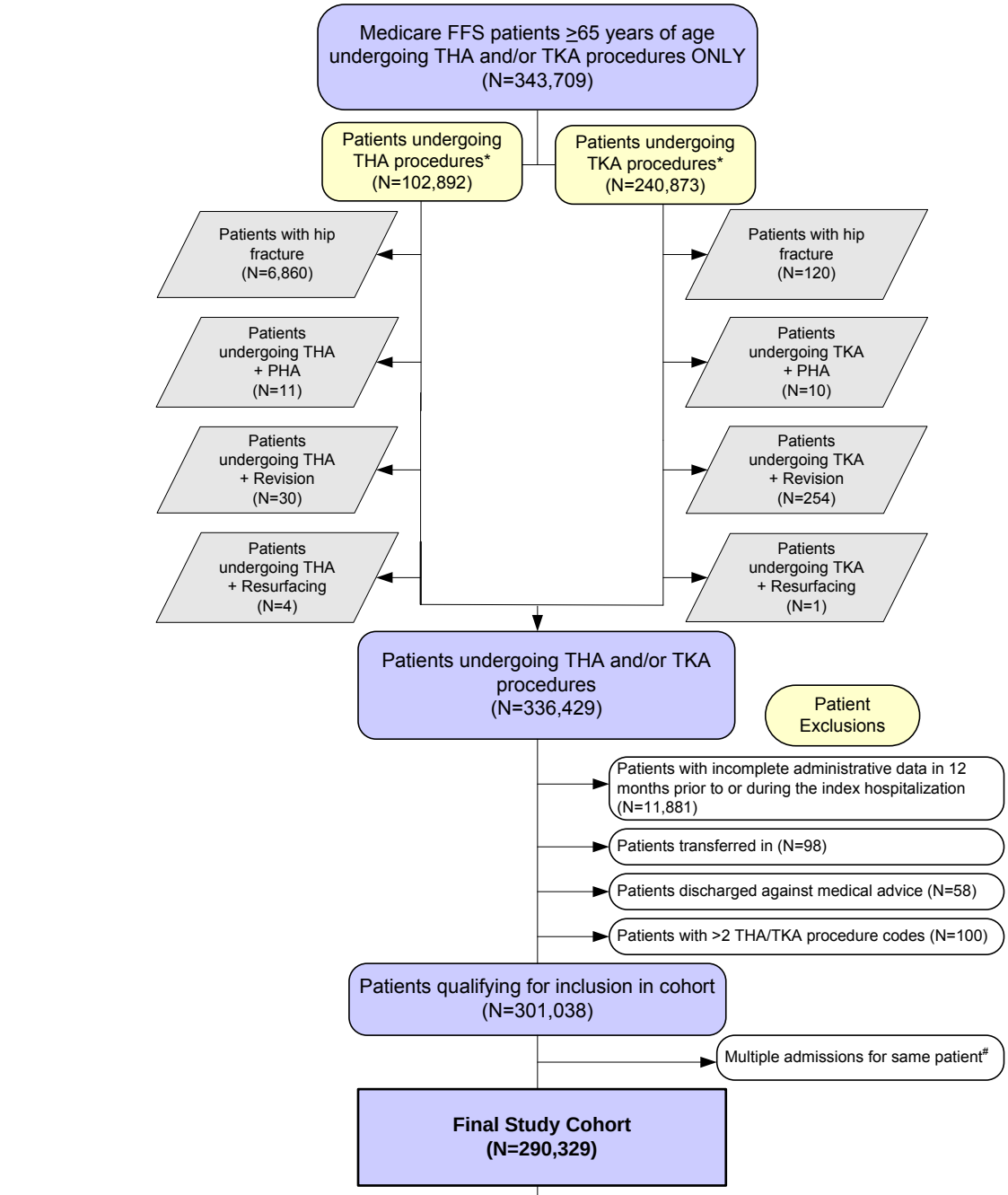
7. Patients with more than two THA/TKA procedures codes during the index hospitalization
   Rationale: It is unlikely that patients would receive more than two THA/TKA procedures in one hospitalization, and this may reflect a coding error

8. Patients with inconsistent or unknown mortality status or other unreliable data (e.g. date of death precedes admission date)
   Rationale: Outcome status is unreliable, although this is rare

9. Multiple admissions for these procedures for a single patient in the 12 months studied; one hospitalization per patient was randomly selected for inclusion after applying the other exclusion criteria
   Rationale: Observations are not independent; a patient is not eligible for the death outcome during the first admission, admitted later in the year for another procedure

Appendix B lists the ICD-9 codes for hip fracture, revision procedures, partial hip arthroplasty procedure, and resurfacing procedures.

Figure 3. Cohort for Model Development



Procedure	N (%)
1 THA procedure	83,066
1 TKA Procedure	197,646
THA + THA	305
TKA + TKA	9263
THA + TKA	49

*THA and TKA are presented separately for illustrative purposes and are not mutually exclusive

#Randomly selected and kept in the cohort one of multiple admissions for the same patient

2.7 Approach to Risk Adjustment

The goal of risk adjustment is to account for patient demographic and clinical characteristics while illuminating important quality differences. The model adjusts for case mix differences based on the clinical status of the patient at the time of admission. Conditions that may represent adverse outcomes due to care received during the index admission are not considered for inclusion in the risk adjusted model. Although they may increase the risk of mortality and complications, including them as covariates in a risk-adjusted model could attenuate the measure's ability to characterize the quality of care delivered by hospitals. Appendix C lists the conditions not adjusted for if they only appear in the index admission and not in the 12 months prior to admission. This methodology is consistent with NQF guidelines.

Consistent with NQF guidelines, the model does not adjust for socioeconomic status (SES), race, or ethnicity because risk-adjusting for SES would hold hospitals with a large proportion of low SES patients to a different standard of care than hospitals treating a larger proportion of high SES patients. Model does not adjust for patients' admission source and their discharge disposition either (e.g. skilled nursing facility) because these factors are associated with structure of the health care system.

2.8 Candidate and Final Risk-Adjustment Variables

Our goal was to develop a parsimonious model that included clinically relevant variables that are strongly associated with risk of complications. The candidate variables for the model were derived from: the index admission, with comorbidities identified from the index admission secondary diagnoses (excluding potential complications), 12-month pre-index inpatient Part A data, outpatient hospital data, and Part B physician data.

For administrative model development, we started with the 189 Condition Categories (CCs). CCs are clinically relevant diagnostic groups of the more than 15,000 ICD-9 codes.[19] We used the April 2010 version of the ICD-9 to CC assignment map, which is maintained by CMS and posted at www.qualitynet.org.

To select candidate variables, a team of clinicians reviewed all 189 CCs and excluded those that were not relevant to the Medicare population (Appendix D) or that were not clinically relevant to the complications outcome (e.g., attention deficit disorder, female infertility, cataract). Clinically relevant CCs were selected as candidate variables. CCs with high clinical relevance to the outcome were broken out and certain conditions within that CC were examined separately when clinically indicated. For example, obesity and morbid obesity are known risk factors for complications following THA/TKA. We reviewed these comorbidities and based on these analyses and expert feedback, morbid obesity was separated from CC 24 (obesity and other endocrine/metabolic/nutritional

disorders) and included in the risk adjusted model independently. Other CCs were combined into clinically coherent groups. Other candidate variables included age, sex, type of procedure (THA, TKA or both), and number of procedures (1 versus 2) and are listed in Table 3.

Table 3. THA/TKA Complications Model Candidate Variables

Category	Variable	ICD-9 Code(s) or CC(s)
Demographic	Age-65 (years above 65, continuous)	
	Sex	
Procedure	Type of procedure	ICD-9-CM 81.51 (THA) ICD-9-CM 81.54 (TKA)
	Number of procedures (1 versus 2)	
Comorbidities	Skeletal deformities	ICD-9-CM 755.63
	Post traumatic osteoarthritis	ICD-9-CM 716.15, 716.16
	Morbid obesity	ICD-9-CM 278.01
	History of Infection	CC 1, 3-6
	Septicemia/shock	CC 2
	Metastatic cancer and acute leukemia	CC 7
	Cancer	CC 8-10
	Respiratory/Heart/Digestive/Urinary/Other Neoplasms	CC 11-13
	Benign neoplasms of skin, breast, eye	CC 14
	Diabetes and DM complications	CC 15-20, 119, 120
	Protein-calorie malnutrition	CC 21
	Disorders of Fluid/Electrolyte/Acid-Base	CC 22, 23
	Obesity/disorders of thyroid, cholesterol, lipids	CC 24
	Liver and biliary disease	CC 25-30
	Intestinal Obstruction/Perforation	CC 31
	Pancreatic Disease	CC 32
	Inflammatory Bowel Disease	CC 33
	Peptic Ulcer, Hemorrhage, Other Specified Gastrointestinal Disorders	CC 34
	Appendicitis	CC 35
	Other Gastrointestinal Disorders	CC 36
	Bone/Joint/Muscle Infections/Necrosis	CC 37
	Rheumatoid Arthritis and Inflammatory Connective Tissue Disease	CC 38
	Disorders of the Vertebrae and Spinal Discs	CC 39
	Osteoarthritis of Hip and Knee	CC 40
	Osteoporosis and Other Bone/Cartilage Disorders	CC 41
	Congenital/Developmental Skeletal and Connective Tissue Disorders	CC 42
	Other Musculoskeletal and Connective Tissue Disorders	CC 43
	Severe Hematological Disorders	CC 44
	Disorders of Immunity	CC 45
	Coagulation Defects and Other Specified Hematological Disorders	CC 46
	Iron Deficiency and Other/Unspecified Anemias and Blood	CC 47

Category	Variable	ICD-9 Code(s) or CC(s)
	Disease	
	Delirium and Encephalopathy	CC 48
	Dementia and senility	CC 49, 50
	Drug/alcohol abuse/dependence/psychosis	CC 51-53
	Major psychiatric Disorders	CC 54-56
	Personality Disorders	CC 57
	Depression	CC 58
	Anxiety Disorders	CC 59
	Other psychiatric disorders	CC 60
	Mental retardation or developmental disability	CC 61-65
	Hemiplegia, paraplegia, paralysis, functional disability	CC 67-69, 100-102, 177-178
	Muscular Dystrophy	CC 70
	Polyneuropathy	CC 71
	Multiple Sclerosis	CC 72
	Parkinson's and Huntington's Diseases	CC 73
	Seizure Disorders and Convulsions	CC 74
	Coma, Brain Compression/Anoxic Damage	CC 75
	Mononeuropathy, Other Neurological Conditions/Injuries	CC 76
	Respirator Dependence/Tracheostomy Status	CC 77
	Respiratory Arrest	CC 78
	Cardio-Respiratory Failure and Shock	CC 79
	Congestive Heart Failure	CC 80
	Acute Coronary Syndrome	CC 81-82
	Chronic Atherosclerosis	CC 83-84
	Heart Infection/Inflammation, Except Rheumatic	CC 85
	Valvular and Rheumatic Heart Disease	CC 86
	Congenital cardiac/circulatory defect	CC 87-88
	Hypertension	CC 89, 91
	Hypertensive heart disease	CC 90
	Arrhythmias	CC 92, 93
	Other and Unspecified Heart Disease	CC 94
	Stroke	CC 95, 96
	Cerebrovascular disease	CC 97-99, 103
	Vascular or circulatory disease	CC 104-106
	Cystic fibrosis	CC 107
	COPD	CC 108
	Fibrosis of lung or other chronic lung disorder	CC 109
	Asthma	CC 110
	Pneumonia	CC 111-113
	Pleural effusion/pneumothorax	CC 114
	Other lung disorder	CC 115
	Legally Blind	CC 116
	Major eye infections/inflammations	CC 117
	Retinal detachments	CC 118
	Retinal Disorders, Except Detachment and Vascular Retinopathies	CC 121
	Glaucoma	CC 122
	Other Eye Disorders	CC 124

Category	Variable	ICD-9 Code(s) or CC(s)
	Significant Ear, Nose, and Throat Disorders	CC 125
	Hearing Loss	CC 126
	Other Ear, Nose, Throat, and Mouth Disorders	CC 127
	Kidney Transplant Status	CC 128
	End-stage renal disease or dialysis	CC 129, 130
	Renal Failure	CC 131
	Nephritis	CC 132
	Urinary Obstruction and Retention	CC 133
	Incontinence	CC 134
	Urinary Tract Infection	CC 135
	Other urinary tract disorders	CC 136
	Pelvic Inflammatory disease	CC 138
	Other female genital disorders	CC 139
	Male genital disorders	CC 140
	Decubitus ulcer or chronic skin ulcer	CC 148, 149
	Extensive burns	CC 150, 151
	Cellulitis, Local Skin Infection	CC 152
	Other Dermatological Disorders	CC 153
	Trauma	CC 154-156, 158-161
	Vertebral Fractures	CC 157
	Other Injuries	CC 162
	Poisonings and Allergic Reactions	CC 163
	Major Complications of Medical Care and Trauma	CC 164
	Other Complications of Medical Care	CC 165
	Major Symptoms, Abnormalities	CC 166
	Minor Symptoms, Signs, Findings	CC 167
	Major Organ Transplant Status	CC 174
	Other organ transplant/replacement	CC 175

To inform final variable selection, a modified approach to stepwise logistic regression was performed. A subsample of the data was used to create 500 “bootstrap” samples. For each sample, we ran a logistic stepwise regression that included the candidate variables. The results were summarized to show the percentage of times that each of the candidate variables was significantly associated with complications ($p<0.001$) in each of the 500 repeated samples (e.g., 70 percent would mean that the candidate variable was selected as significant at $p<0.001$ in 70 percent of the estimations). We also assessed the direction and magnitude of the regression coefficients.

The clinical team reviewed these results and decided to retain all risk adjustment variables above a 70% cutoff, because they demonstrated a relatively strong association with risk for complications and were clinically relevant. Additionally, specific variables with particular clinical relevance to the risk of complications were forced into the model (regardless of % selection) to ensure appropriate risk-adjustment for THA and TKA. These included:

Markers for end of life/frailty:

- decubitus ulcer (CC 148)
- dementia and senility (CC 49 and CC 50, respectively)
- metastatic cancer and acute leukemia (CC 7)
- protein-calorie malnutrition (CC 21)
- hemiplegia/paraplegia/paralysis/functional disability (CC 67-69, 100-102, 177-178)
- stroke (CC 95-96)

Diagnoses with potential asymmetry among hospitals that would impact the validity of the model:

- cancer (CC 8-12)

Final model variables are listed in Table 4.

Table 4. THA/TKA Complications Model Final Model Variables

Category	Variable	ICD-9 Code(s) or CC(s)
Demographic	Age-65 (years above 65, continuous)	
	Sex	
Procedure	Type of procedure	ICD-9-CM 81.51 (THA)
	Number of procedures (1 vs. 2)	
Comorbidities	Skeletal deformities	ICD-9-CM 755.63
	Post traumatic osteoarthritis	ICD-9-CM 716.15, 716.16
	Morbid obesity	ICD-9-CM 278.01
	Metastatic cancer and acute leukemia	CC 7
	Cancer	CC 8-10
	Respiratory/Heart/Digestive/Urinary/Other Neoplasms	CC 11-13
	Diabetes and DM complications	CC 15-20, 119, 120
	Protein-calorie malnutrition	CC 21
	Bone/Joint/Muscle Infections/Necrosis	CC 37
	Rheumatoid Arthritis and Inflammatory Connective Tissue Disease	CC 38
	Osteoarthritis of Hip or Knee	CC 40
	Osteoporosis and Other Bone/Cartilage Disorders	CC 41
	Dementia and senility	CC 49, 50
	Major psychiatric disorders	CC 54-56
	Hemiplegia, paraplegia, paralysis, functional disability	CC 67-69, 100-102, 177-178
	Cardio-Respiratory Failure and Shock	CC 79
	Chronic Atherosclerosis	CC 83-84
	Stroke	CC 95, 96
	Vascular or circulatory disease	CC 104-106
	COPD	CC 108
	Pneumonia	CC 111-113
	Pleural effusion/pneumothorax	CC 114
	End-stage renal disease or dialysis	CC 129, 130
	Renal Failure	CC 131
	Decubitus ulcer or chronic skin ulcer	CC 148, 149
	Trauma	CC 154-156, 158-161
	Vertebral Fractures	CC 157
	Other injuries	CC 162
	Major Complications of Medical Care and Trauma	CC 164

2.9 Statistical Approach to Model Development

We randomly selected 50% of the THA and/or TKA admissions that met all inclusion and exclusion criteria and created a development sample, which we used to build the model. We used the remaining 50% of THA/TKA admissions as the validation sample. We also used all qualifying THA and/or TKA admissions in

2007 data as an additional sample to validate the model. Model performance was assessed in the development dataset and both validation datasets.

Due to the natural clustering of hospitalizations within hospitals, we used hierarchical generalized linear models (HGLMs) to model the log-odds of death and complications. Death and complications were modeled as a function of patient-level demographic and clinical characteristics and a random hospital-specific intercept. This strategy accounts for within-hospital correlation of the observed outcomes and models the assumption that underlying differences in quality among the health care facilities being evaluated lead to systematic differences in outcomes.

We then calculated hospital risk-standardized complication rates (RSCRs) using a hierarchical logistic regression model (given the hierarchical nature of the data). These rates are calculated as the ratio of the predicted number of complications to the expected number of complications, multiplied by the national unadjusted complication rate. The expected number of complications for each hospital was estimated using that hospital's patient mix and the national intercept. Specifically, for each patient in the data set, the estimated regression coefficients are multiplied by the observed characteristics and the average of the hospital-specific intercepts is added to this quantity. Then, the quantity is transformed to the probability scale. For each patient within a hospital, these probabilities are summed. The predicted number of complications in each hospital employs a similar calculation. The predicted number of complications for each hospital is calculated by summing the predicted complication rates for all patients in the hospital. The predicted complication rate for each patient is calculated through the hierarchical model by applying the estimated regression coefficients to the patient characteristics observed and adding the hospital-specific intercept. In order to assess hospital performance in any specific year (e.g. the validation cohort), we re-estimate the model coefficients using that year's data.

More specifically, we estimate a generalized linear model and a hierarchical generalized linear model which accounts for the clustering of observations within hospitals. The generalized linear model (GLM) links the outcome to the patient-level risk factors,[20] Let Y_{ij} denote the outcome (equal to 1 if patient dies or has a complication, zero otherwise) for the j^{th} patient who had a THA/TKA procedure at the i^{th} hospital; $\mathbf{Z}_{ij}$ denotes a set of risk factors based on the data. Let I denote the total number of hospitals and n_i the number of index patient stays in hospital i. We assume the outcome is related linearly to the covariates via a known linked function, h, where

$$\text{GLM} \qquad h(Y_{ij}) = \alpha + \beta \mathbf{Z}_{ij} \qquad (1)$$

and $\mathbf{Z}_{ij} = (Z_{1ij}, Z_{2ij}, \ldots, Z_{pij})$ is a set of p patient-specific covariates. In our case, h = the logit link.

To account for the natural clustering of observations within hospitals, we then estimate an HGLM that links the risk factors to the same outcome and a hospital-specific random effect,

HGLM $h(Y_{ij}) = \alpha_i + \beta \mathbf{Z}_{ij}$ (2)

$\alpha_i = \mu + \omega_i; \quad \omega_i \sim N(0, \tau^2)$ (3)

where α_i represents the hospital-specific intercept, $\mathbf{Z}_{ij}$ is defined as above, μ the adjusted average outcome over all hospitals in the sample, and τ^2 the between-hospital variance component.[21] This model separates within-hospital variation from between-hospital variation. Both HGLMs and GLMs are estimated using the SAS software system (GLIMMIX and LOGISTIC procedures, respectively).

We first fit the GLM described in Equation (1) using the logit link.
Having identified the covariates that remained, we next fit the HGLM described in Equations (2) and (3), again using the logit link function; e.g.,

Logit $(P(Y_{ij} = 1)) = \alpha_i + \beta \mathbf{Z}_{ij}$

$\alpha_i = \mu + \omega_i, \quad \omega_i \sim N(0, \tau^2)$

where $\mathbf{Z}_{ij}$ consisted of the covariates retained in the GLM model. As before, Y_{ij} = 1 if patient *j* treated at hospital *i* had the event; 0 otherwise.

2.10 Hospital Performance Reporting

Using the set of risk factors in the GLM, we fit the HGLM defined by Equations (2) - (3) and estimate the parameters, $\hat{\mu}$, $\{\hat{\alpha}_i, \hat{\alpha}_2, ..., \hat{\alpha}_I\}$, $\hat{\beta}$, and $\hat{\tau}^2$. We calculate a standardized outcome, s_i, for each hospital by computing the ratio of the number of predicted complications to the number of expected complications, multiplied by the unadjusted overall complication rate, $\bar{y}$. Specifically, we calculate

Predicted $\hat{y}_{ij}(Z) = h^{-1}(\hat{\alpha}_i + \hat{\beta} \mathbf{Z}_{ij})$ (4)

Expected $\hat{e}_{ij}(Z) = h^{-1}(\hat{\mu} + \hat{\beta} \mathbf{Z}_{ij})$ (5)

$$\hat{s}_i(Z) = \frac{\sum_{j=1}^{n_i} \hat{y}_{ij}(Z)}{\sum_{j=1}^{n_i} \hat{e}_{ij}(Z)} \times \bar{y} \quad (6)$$

If more (fewer) "predicted" cases than "expected" cases have the outcome in a hospital, then $\hat{s}_i$ will be higher (lower) than the unadjusted average. For each

hospital, we compute an interval estimate of s_i to characterize the level of uncertainty around the point estimate using bootstrapping simulations. The point estimate and interval estimate can be used to characterize and compare hospital performance (e.g., higher than expected, as expected, or lower than expected).

2.10.1 Creating Interval Estimates

Because the statistic described in Equation (6) is a complex function of parameter estimates, we use re-sampling and simulation techniques to derive an interval estimate. The bootstrapping simulation has the advantage of avoiding unnecessary distributional assumptions.

2.10.2 Algorithm

Let *I* denote the total number of hospitals in the sample. We repeat steps 1 – 4 below for b = 1,2,…B times:

1. Sample *I* hospitals with replacement.

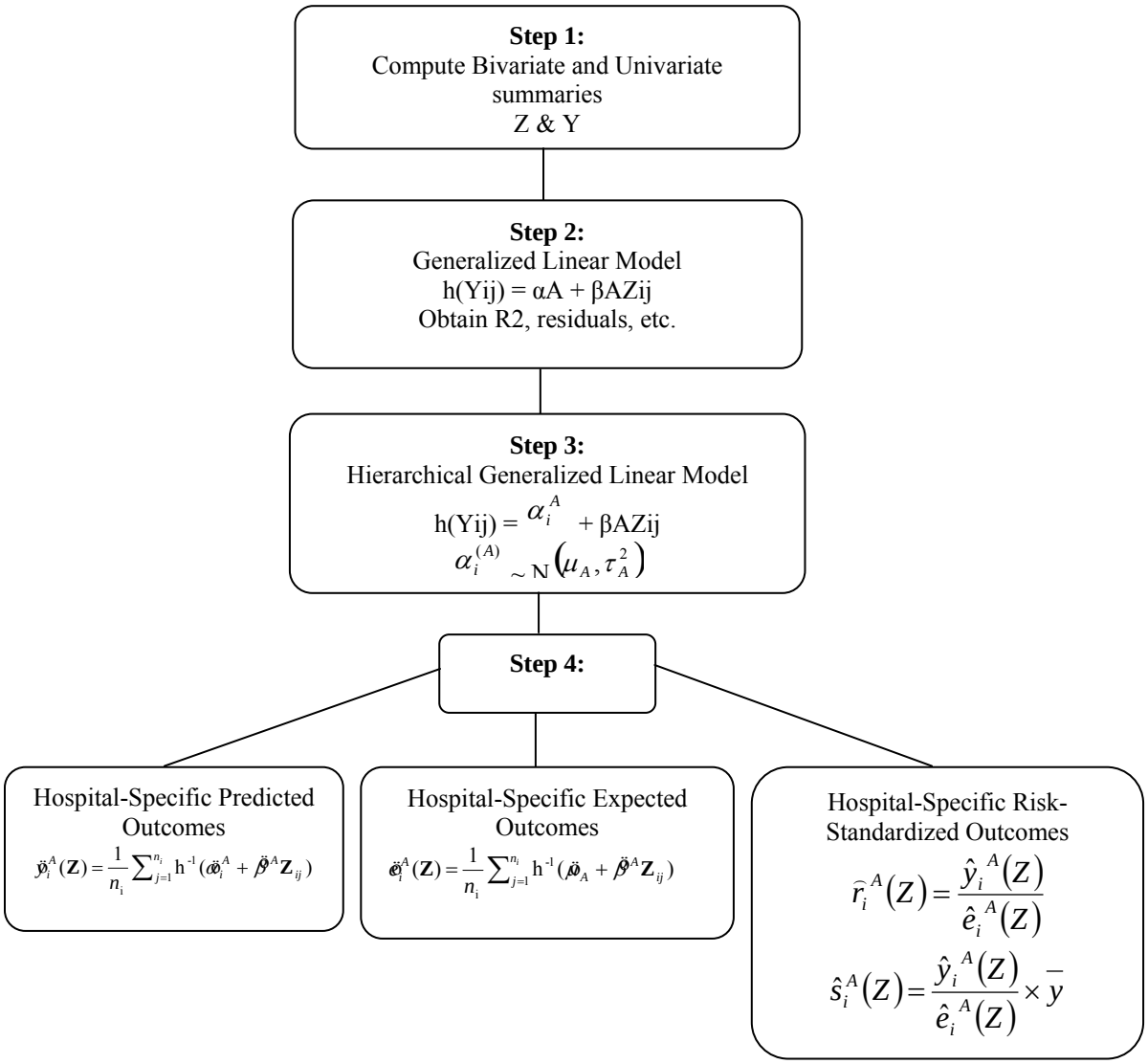
2. Fit the HGLM using all patients within each sampled hospital. We use as starting values the parameter estimates obtained by fitting the model to all hospitals. If some hospitals are selected more than once in a bootstrapped sample, we treat them as distinct so that we have *I* random effects to estimate the variance components. At the conclusion of Step 2, we have:
   a. $\hat{\beta}^{(b)}$ (the estimated regression coefficients of the risk factors).
   b. The parameters governing the random effects, hospital adjusted outcomes, distribution, $\hat{\mu}^{(b)}$ and $\hat{\tau}^{2(b)}$.
   c. The set of hospital-specific intercepts and corresponding variances, $\{\hat{\alpha}_i^{(b)}, \hat{\mathrm{var}}(\alpha_i^{(b)}); i = 1,2,\ldots,I\}$.

3. We generate a hospital random effect by sampling from the distribution of the hospital-specific distribution obtained in Step 2c. We approximate the distribution for each random effect by a normal distribution. Thus, we draw $\alpha_i^{(b*)} \sim N(\hat{\alpha}_i^{(b)}, \hat{\mathrm{var}}(\hat{\alpha}_i^{(b)}))$ for the unique set of hospitals sampled in Step 1.

4. Within each unique hospital *i* sampled in Step 1, and for each case *j* in that hospital, we calculate $\hat{y}_{ij}^{(b)}$, $\hat{e}_{ij}^{(b)}$, and $\hat{s}_i(Z)^{(b)}$ where $\hat{\beta}^{(b)}$ and $\hat{\mu}^{(b)}$ are obtained from Step 2 and $\hat{\alpha}_i^{(b*)}$ is obtained from Step 3.

Ninety-five percent interval estimates (or alternative interval estimates) for the hospital-standardized outcome can be computed by identifying the 2.5^{th} and 97.5^{th} percentiles of randomly half of the B estimates (or the percentiles corresponding to the alternative desired intervals). [22]

Figure 4. Analysis Steps



3. RESULTS

3.1 Model Results

3.1.1 Development and Validation Models

Tables 5 and 6 convey the developmental sample model results for GLM and HGLM analyses respectively. The standardized estimates are regression coefficients expressed in units of standard deviations and can range between -1 and 1, with ±1 indicating a perfect linear relationship and 0 indicating no linear relationship.[1] The estimated between-hospital variance in the adjusted log-odds of a complication(s) is 0.105, based on the 2008 full dataset. This result implies that the odds of a complication(s) for a high-complication hospital (+1 SD) are 1.91 times that in a low-complication hospital (-1 SD). If there were no differences between hospitals, the between-hospital variance would be 0 and the odds ratio would be 1.0.Table 7 conveys the GLM results for the validation sample.

3.1.2 Model Performance

We computed 6 summary statistics for assessing model performance [23] over-fitting indices[2], predictive ability, area under the receiver operating characteristic (ROC) curve, distribution of residuals, and model chi-square[3]. Table 8 conveys GLM model performance results for both the developmental and validation samples.

The models for both the development and validation samples have strong discrimination and fit. Model predictive ability ranges from 2% in the lowest predictive decile to 15% in the highest decile in both samples, indicating

---

[1] Standardized estimates are like correlation coefficients. We compute them in order to compare the size of the coefficients by standardizing the coefficients to be unitless.

[2] Over-fitting refers to the phenomenon in which a model well describes the relationship between predictive variables and outcome in the development dataset, but fails to provide valid predictions in new patients.

[3] Chi-Square – A test of statistical significance usually employed for categorical data to determine whether there is a good fit between the observed data and expected values; i.e., whether the differences between observed and expected values are attributable to true differences in characteristics or instead the result of chance variation. The formula for computing the chi-square is as follows:

$$\sum \frac{(O-E)^2}{E}$$

where O = observed value
E = expected value, and
degrees of freedom (df) = (rows-1)(columns-1)

the model can reasonably classify patients on the outcome, based on their risk. The area under the ROC curve (C statistic) is 0.69 for the development model and 0.70 for the validation model (Table 8).The discrimination ability is consistent with models currently used to publicly report condition specific rates of both mortality and readmission.

Table 9 conveys the standardized estimates by year of discharge in the full datasets for 2007 and 2008. There are no notable differences in the standardized estimates between the two years. Table 10 conveys the risk factor frequency for the development and validation samples by year of discharge. The prevalence of morbid obesity increased slightly to 3.36% in 2008, compared with 2.91% in 2007. There were no other notable changes in risk factor frequency over the two-year period.

Table 5. GLM Model Results for 2008 Development Sample (ROC=0.69)

Description	Estimate	Standard Error	Standardized Estimate	Odds Ratio	95% Confidence Interval for OR
Intercept	-3.58	0.06			
Demographics					
Age-65‡ (mean)	0.03	0.00	0.10	1.03	(1.03 – 1.04)
Male	0.09	0.03	0.02	1.10	(1.04 – 1.16)
THA/TKA Procedure					
THA procedure	0.53	0.03	0.13	1.70	(1.61 – 1.80)
Number of procedures (one vs. two)	0.51	0.07	0.05	1.67	(1.46 – 1.91)
Comorbid Conditions					
Skeletal deformities (ICD-9 code 755.63)	0.31	0.30	0.01	1.37	(0.77 – 2.45)
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16)	0.24	0.15	0.01	1.27	(0.94 – 1.73)
Morbid obesity (ICD-9 code 278.01)	0.17	0.07	0.02	1.19	(1.03 – 1.37)
Metastatic cancer and acute leukemia (CC 7)	0.38	0.13	0.02	1.46	(1.12 – 1.89)
Cancer (CC 8-10)	-0.06	0.04	-0.01	0.94	(0.87 – 1.02)
Respiratory/Heart/Digestive/Urinary/Other Neoplasms (CC 11-13)	-0.15	0.04	-0.03	0.86	(0.80 – 0.93)
Diabetes and DM complications (CC 15-20, 119, 120)	0.15	0.03	0.04	1.16	(1.09 – 1.22)
Protein-calorie malnutrition (CC 21)	0.84	0.10	0.04	2.32	(1.91 – 2.83)
Bone/Joint/Muscle Infections/Necrosis (CC 37)	0.00	0.06	0.00	1.00	(0.88 – 1.13)
Rheumatoid Arthritis and Inflammatory Connective Tissue Disease (CC 38)	0.03	0.05	0.00	1.03	(0.94 – 1.12)
Osteoarthritis of Hip or Knee (CC 40)	-0.61	0.05	-0.07	0.54	(0.49 – 0.60)
Osteoporosis and Other Bone/Cartilage Disorders (CC 41)	0.01	0.03	0.00	1.01	(0.95 – 1.08)
Dementia and senility (CC 49, 50)	0.17	0.05	0.02	1.19	(1.07 – 1.32)
Major psychiatric disorders (CC 54-56)	0.19	0.06	0.02	1.21	(1.07 – 1.36)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	0.18	0.09	0.01	1.20	(1.00 – 1.43)
Cardio-Respiratory Failure and Shock (CC 79)	-0.30	0.08	-0.02	0.74	(0.64 – 0.86)
Chronic Atherosclerosis (CC 83-84)	0.21	0.03	0.05	1.24	(1.17 – 1.31)
Stroke (CC 95, 96)	-0.10	0.08	-0.01	0.91	(0.78 – 1.06)
Vascular or circulatory disease (CC 104-106)	0.11	0.03	0.03	1.12	(1.05 – 1.19)
COPD (CC 108)	0.15	0.03	0.03	1.17	(1.09 – 1.25)
Pneumonia (CC 111-113)	1.53	0.04	0.19	4.61	(4.29 – 4.96)
Pleural effusion/pneumothorax (CC 114)	-0.37	0.09	-0.02	0.69	(0.59 – 0.82)
End-stage renal disease or dialysis (CC 129, 130)	0.74	0.20	0.02	2.09	(1.41 – 3.10)
Renal Failure (CC 131)	0.01	0.05	0.00	1.01	(0.91 – 1.11)
Decubitus ulcer or chronic skin ulcer (CC 148, 149)	0.24	0.13	0.01	1.27	(0.99 – 1.64)
Trauma (CC 154-156, 158-161)	0.70	0.05	0.08	2.02	(1.84 – 2.20)
Vertebral Fractures (CC 157)	0.12	0.09	0.01	1.13	(0.94 – 1.36)
Other injuries (CC162)	0.09	0.03	0.02	1.09	(1.03 – 1.16)
Major Complications of Medical Care and Trauma (CC 164)	0.45	0.05	0.05	1.57	(1.42 – 1.74)

Table 6. HGLM Model Results for 2008 Development Sample

Description	Estimate	Standard Error	T-Value	Pr > T-Value	Odds Ratio	95% Confidence Interval for OR
Intercept	-3.57	0.06	-61.36	<.0001		
Demographics						
Age-65‡ (mean)	0.03	0.002	14.54	<.0001	1.03	(1.03 – 1.04)
Male	0.09	0.03	3.31	0.001	1.10	(1.04 – 1.16)
THA/TKA Procedure						
THA procedure	0.54	0.03	19.58	<.0001	1.71	(1.62 – 1.81)
Number of procedures (one vs. two)	0.53	0.07	7.75	<.0001	1.69	(1.48 – 1.93)
Comorbid Conditions						
Skeletal deformities (ICD-9 code 755.63)	0.34	0.29	1.17	0.242	1.40	(0.80 – 2.47)
Post traumatic osteoarthritis (ICD-9	0.26	0.15	1.72	0.086	1.30	(0.96 – 1.74)
Morbid obesity (ICD-9 code 278.01)	0.18	0.07	2.49	0.013	1.19	(1.04 – 1.37)
Metastatic cancer and acute leukemia	0.38	0.13	2.91	0.004	1.46	(1.13 – 1.88)
Cancer (CC 8-10)	-0.06	0.04	-1.54	0.123	0.94	(0.87 – 1.02)
Respiratory/Heart/Digestive/Urinary/Other	-0.14	0.04	-4.02	<.0001	0.87	(0.81 – 0.93)
Diabetes and DM complications (CC 15-	0.14	0.03	4.82	<.0001	1.15	(1.09 – 1.22)
Protein-calorie malnutrition (CC 21)	0.84	0.10	8.54	<.0001	2.31	(1.90 – 2.79)
Bone/Joint/Muscle Infections/Necrosis	-0.01	0.06	-0.11	0.910	0.99	(0.88 – 1.12)
Rheumatoid Arthritis and Inflammatory	0.03	0.04	0.72	0.471	1.03	(0.95 – 1.13)
Osteoarthritis of Hip or Knee (CC 40)	-0.61	0.05	-12.76	<.0001	0.54	(0.49 – 0.59)
Osteoporosis and Other Bone/Cartilage	0.01	0.03	0.41	0.679	1.01	(0.95 – 1.08)
Dementia and senility (CC 49, 50)	0.17	0.05	3.19	0.001	1.18	(1.07 – 1.31)
Major psychiatric disorders (CC 54-56)	0.19	0.06	3.14	0.001	1.21	(1.07 – 1.35)
Hemiplegia, paraplegia, paralysis,	0.18	0.09	2.12	0.034	1.20	(1.01 – 1.43)
Cardio-Respiratory Failure and Shock	-0.30	0.07	-4.05	<.0001	0.74	(0.64 – 0.86)
Chronic Atherosclerosis (CC 83-84)	0.21	0.03	7.63	<.0001	1.24	(1.17 – 1.31)
Stroke (CC 95, 96)	-0.10	0.07	-1.28	0.199	0.91	(0.79 – 1.05)
Vascular or circulatory disease (CC 104-	0.11	0.03	3.84	0.0001	1.12	(1.06 – 1.19)
COPD (CC 108)	0.15	0.03	4.41	<.0001	1.16	(1.09 – 1.24)
Pneumonia (CC 111-113)	1.53	0.04	42.39	<.0001	4.62	(4.31 – 4.96)
Pleural effusion/pneumothorax (CC 114)	-0.37	0.08	-4.36	<.0001	0.69	(0.59 – 0.82)
End-stage renal disease or dialysis (CC	0.73	0.20	3.72	0.0002	2.07	(1.41 – 3.03)
Renal Failure (CC 131)	-0.001	0.05	-0.02	0.988	1.00	(0.91 – 1.10)
Decubitus ulcer or chronic skin ulcer (CC	0.24	0.13	1.90	0.058	1.27	(0.99 – 1.63)
Trauma (CC 154-156, 158-161)	0.70	0.04	15.99	<.0001	2.02	(1.86 – 2.20)
Vertebral Fractures (CC 157)	0.12	0.09	1.39	0.166	1.13	(0.95 – 1.35)
Other injuries (CC162)	0.08	0.03	2.84	0.005	1.09	(1.03 – 1.15)
Major Complications of Medical Care and Trauma (CC 164)	0.45	0.05	8.80	<.0001	1.56	(1.41 – 1.72)

Table 7. GLM Model Results for 2008 Validation Sample (ROC=0.70)

Label	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimates	Odds Ratio	95 % Confidence Interval for OR
Intercept	-3.62	0.06	3744.33	<.0001			
Demographics							
Age-65‡ (mean)	0.03	0.002	224.72	<.0001	0.11	1.03	(1.03 - 1.04)
Male	0.11	0.03	13.08	0.0003	0.03	1.11	(1.05 - 1.18)
THA/TKA Procedure							
THA procedure	0.56	0.03	384.24	<.0001	0.14	1.75	(1.65 - 1.85)
Number of procedures (one vs. two)	0.37	0.07	25.24	<.0001	0.04	1.45	(1.26 - 1.68)
Comorbid Conditions							
Skeletal deformities (ICD-9 code 755.63)	0.31	0.27	1.28	0.259	0.01	1.36	(0.80 - 2.31)
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16)	0.35	0.14	6.26	0.01	0.01	1.42	(1.08 - 1.87)
Morbid obesity (ICD-9 code 278.01)	0.40	0.07	35.90	<.0001	0.04	1.50	(1.31 - 1.71)
Metastatic cancer and acute leukemia (CC 7)	0.03	0.15	0.03	0.85	0.001	1.03	(0.76 - 1.39)
Cancer (CC 8-10)	-0.07	0.04	2.81	0.094	-0.01	0.93	(0.86 - 1.01)
Respiratory/Heart/Digestive/Urinary/Other Neoplasms (CC 11-13)	-0.09	0.04	6.33	0.012	-0.02	0.91	(0.85 - 0.98)
Diabetes and DM complications (CC 15-20, 119, 120)	0.12	0.03	14.93	0.000	0.03	1.12	(1.06 - 1.19)
Protein-calorie malnutrition (CC 21)	0.70	0.10	50.97	<.0001	0.03	2.02	(1.67 - 2.46)
Bone/Joint/Muscle Infections/Necrosis (CC 37)	0.02	0.07	0.12	0.734	0.00	1.02	(0.90 - 1.16)
Rheumatoid Arthritis and Inflammatory Connective Tissue Disease (CC 38)	-0.04	0.05	0.62	0.429	-0.01	0.96	(0.88 - 1.06)
Osteoarthritis of Hip or Knee (CC 40)	-0.66	0.05	180.53	<.0001	-0.08	0.52	(0.47 - 0.57)
Osteoporosis and Other Bone/Cartilage Disorders (CC 41)	-0.01	0.03	0.11	0.743	0.00	0.99	(0.93 - 1.05)
Dementia and senility (CC 49, 50)	0.17	0.05	9.99	0.002	0.02	1.19	(1.07 - 1.32)
Major psychiatric disorders (CC 54-56)	0.09	0.06	2.15	0.143	0.01	1.10	(.097 - 1.24)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	0.11	0.09	1.39	0.238	0.01	1.11	(0.93 - 1.32)
Cardio-Respiratory Failure and Shock (CC 79)	-0.25	0.07	11.51	0.001	-0.02	0.78	(0.67 - 0.90)
Chronic Atherosclerosis (CC 83-84)	0.19	0.03	44.25	<.0001	0.05	1.21	(1.15 - 1.29)
Stroke (CC 95, 96)	-0.01	0.08	0.01	0.917	0.00	0.99	(0.86 - 1.15)
Vascular or circulatory disease (CC 104-106)	0.11	0.03	11.57	0.001	0.02	1.11	(1.05 - 1.18)
COPD (CC 108)	0.15	0.03	18.16	<.0001	0.03	1.16	(1.08 - 1.24)
Pneumonia (CC 111-113)	1.55	0.04	1754.99	<.0001	0.20	4.72	(4.39 - 5.08)
Pleural effusion/pneumothorax (CC 114)	-0.26	0.08	9.85	0.002	-0.02	0.77	(0.65 - 0.91)
End-stage renal disease or dialysis (CC 129, 130)	0.42	0.20	4.35	0.037	0.01	1.53	(1.03 - 2.27)
Renal Failure (CC 131)	0.12	0.05	6.69	0.010	0.02	1.13	(1.03 - 1.24)
Decubitus ulcer or chronic skin ulcer (CC 148, 149)	0.15	0.13	1.38	0.240	0.01	1.17	(0.90- 1.50)
Trauma (CC 154-156, 158-161)	0.69	0.05	234.27	<.0001	0.08	2.00	(1.83 - 2.18)
Vertebral Fractures (CC 157)	0.10	0.09	1.22	0.269	0.01	1.11	(0.93 - 1.32)
Other injuries (CC162)	0.14	0.03	21.49	<.0001	0.03	1.15	(1.08 - 1.22)
Major Complications of Medical Care and Trauma (CC 164)	0.56	0.05	119.55	<.0001	0.06	1.74	(1.58- 1.93)

Table 8. Model Performance for GLM Model

Indices	Development Sample	Validation Sample	Validation Sample
Year	2008 (50%)	2008 (50%)	2007 (100%)
Number of Admissions	145,206	145,123	294,697
Number of Hospitals	3,221	3,223	3,300
Number of Complications	6148	6043	12,707
Calibration (γ0, γ1)[1]	(0, 1)	(0.04, 1.02)	(0.002, 1.00)
Discrimination -Predictive Ability (lowest decile %, highest decile %)	(2%, 15%)	(2%, 15%)	(2%, 15%)
Discrimination – Area Under Receiver Operator Curve	0.69	0.70	0.69
Residuals Lack of Fit (Pearson Residual Fall %)			
<-2	0	0	0
[-2, 0)	95.8	95.8	95.7
[0, 2)	0.4	0.4	0.4
[2+	3.8	3.7	3.9
Model Wald χ^2 [Number of Covariates][4]	4401 [33]	4698 [33]	9236 (33)

[1] Over-Fitting Indices (γ_0, γ_1) provide evidence of over-fitting and require several steps to calculate. Let *b* denote the *estimated vector* of regression coefficients. *Predicted Probabilities* ($\hat{p}$) = 1/(1+exp{-Xb}), and *Z* = *Xb* (e.g., the linear predictor that is a scalar value for everyone). A new logistic regression model that includes only an intercept and a slope by regressing the logits on Z is fitted in the validation sample; e.g., Logit(P(Y=1|Z)) = γ_0 + γ_1Z. Estimated values of γ_0 far from 0 and estimated values of γ_1 far from 1 provide evidence of over-fitting.

Table 9. Standardized Estimates for GLM Model by Year of Discharge (GLM)

Description	2008 (100%)			2007 (100%)		
	Standardized Estimates	Odds Ratio	95% Confidence Interval for OR	Standardized Estimates	Odds Ratio	95% Confidence Interval for OR
Demographics						
Age-65‡ (mean)	0.11	1.03	(1.03 - 1.04)	0.10	1.03	(1.03 – 1.04)
Male	0.03	1.11	(1.06 - 1.15)	0.02	1.10	(1.04 – 1.16)
THA/TKA Procedure						
THA procedure	0.14	1.73	(1.66 - 1.80)	0.13	1.70	(1.61 – 1.80)
Number of procedures (one vs. two)	0.04	1.56	(1.42 - 1.73)	0.05	1.67	(1.46 – 1.91)
Comorbid Conditions						
Skeletal deformities (ICD-9 code 755.63)	0.01	1.36	(0.92 - 2.02)	0.01	1.37	(0.77 – 2.45)
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16)	0.01	1.35	(1.10 - 1.66)	0.01	1.27	(0.94 – 1.73)
Morbid obesity (ICD-9 code 278.01)	0.03	1.34	(1.21 - 1.47)	0.02	1.19	(1.03 – 1.37)
Metastatic cancer and acute leukemia (CC 7)	0.01	1.24	(1.02 - 1.51)	0.02	1.46	(1.12 – 1.89)
Cancer (CC 8-10)	-0.01	0.94	(0.89 - 0.99)	-0.01	0.94	(0.87 – 1.02)
Respiratory/Heart/Digestive/Urinary/Other Neoplasms (CC 11-13)	-0.03	0.89	(0.85 - 0.93)	-0.03	0.86	(0.80 – 0.93)
Diabetes and DM complications (CC 15-20, 119, 120)	0.03	1.14	(1.09 - 1.19)	0.04	1.16	(1.09 – 1.23)
Protein-calorie malnutrition (CC 21)	0.03	2.16	(1.88 - 2.48)	0.04	2.32	(1.91 – 2.83)
Bone/Joint/Muscle Infections/Necrosis (CC 37)	0.00	1.01	(0.92 - 1.11)	0.00	1.00	(0.88 – 1.13)
Rheumatoid Arthritis and Inflammatory Connective Tissue Disease (CC 38)	0.00	1.00	(0.93 - 1.06)	0.00	1.03	(0.94 – 1.12)
Osteoarthritis of Hip or Knee (CC 40)	-0.07	0.53	(0.49 - 0.57)	-0.07	0.54	(0.49 – 0.60)
Osteoporosis and Other Bone/Cartilage Disorders (CC 41)	0.00	1.00	(0.96 - 1.05)	0.00	1.01	(0.95 – 1.08)
Dementia and senility (CC 49, 50)	0.02	1.19	(1.10 - 1.28)	0.02	1.19	(1.07 – 1.32)
Major psychiatric disorders (CC 54-56)	0.01	1.15	(1.06 - 1.25)	0.02	1.21	(1.07 – 1.36)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	0.01	1.15	(1.02 - 1.30)	0.01	1.20	(1.01 – 1.43)
Cardio-Respiratory Failure and Shock (CC 79)	-0.02	0.76	(0.69 - 0.85)	-0.02	0.74	(0.64 – 0.86)
Chronic Atherosclerosis (CC 83-84)	0.05	1.23	(1.18 - 1.28)	0.05	1.24	(1.17 – 1.31)
Stroke (CC 95, 96)	0.00	0.95	(0.85 - 1.06)	-0.01	0.91	(0.78 – 1.06)
Vascular or circulatory disease (CC 104-106)	0.03	1.12	(1.07 - 1.17)	0.03	1.12	(1.05 – 1.19)
COPD (CC 108)	0.03	1.16	(1.11 - 1.22)	0.03	1.17	(1.09 – 1.25)
Pneumonia (CC 111-113)	0.19	4.67	(4.43 - 4.91)	0.19	4.61	(4.29 – 4.96)
Pleural effusion/pneumothorax (CC 114)	-0.02	0.73	(0.65 - 0.82)	-0.02	0.69	(0.59 – 0.82)
End-stage renal disease or dialysis (CC 129, 130)	0.01	1.79	(1.35 - 2.36)	0.02	2.09	(1.41 – 3.10)
Renal Failure (CC 131)	0.01	1.07	(1.00 - 1.14)	0.00	1.01	(0.91 – 1.11)
Decubitus ulcer or chronic skin ulcer (CC 148, 149)	0.01	1.21	(1.01 - 1.45)	0.01	1.27	(0.99 – 1.64)
Trauma (CC 154-156, 158-161)	0.08	2.01	(1.88 - 2.14)	0.08	2.02	(1.84 – 2.20)
Vertebral Fractures (CC 157)	0.01	1.12	(0.98 - 1.27)	0.01	1.13	(0.94 – 1.36)
Other injuries (CC162)	0.03	1.12	(1.07 - 1.17)	0.02	1.09	(1.03 – 1.16)
Major Complications of Medical Care and Trauma (CC 164)	0.05	1.65	(1.54 - 1.78)	0.05	1.57	(1.42 – 1.74)

Table 10. Risk Factor Frequency by Year of Discharge (GLM)

Description	2008 Development Sample	2008 Validation Sample	2007 Validation Sample
Male	35.76	35.62	35.49
THA procedure	28.76	28.67	28.63
Number of procedures (one vs. two)	3.32	3.30	3.61
Skeletal deformities	0.13	0.14	0.14
Post traumatic osteoarthritis	0.49	0.56	0.49
Morbid obesity	3.36	3.40	2.91
Metastatic cancer and acute leukemia	0.64	0.58	0.65
Cancer	12.84	12.76	12.77
Respiratory/Heart/Digestive/Urinary/Other Neoplasms	17.87	18.02	17.75
Diabetes and DM complications	27.31	27.38	26.75
Protein-calorie malnutrition	0.58	0.67	0.54
Bone/Joint/Muscle Infections/Necrosis	2.97	2.84	3.12
Rheumatoid Arthritis and Inflammatory Connective Tissue Disease	8.52	8.56	8.31
Osteoarthritis of Hip or Knee	95.26	95.35	95.31
Osteoporosis and Other Bone/Cartilage Disorders	24.81	25.11	24.19
Dementia and senility	4.39	4.36	4.22
Major psychiatric disorders	3.66	3.79	3.57
Hemiplegia, paraplegia, paralysis, functional disability	1.54	1.63	1.51
Cardio-Respiratory Failure and Shock	2.05	2.07	1.98
Chronic Atherosclerosis	30.74	30.72	31.05
Stroke	2.45	2.44	2.50
Vascular or circulatory disease	22.47	22.61	22.09
COPD	14.65	14.65	15.16
Pneumonia	5.38	5.49	5.46
Pleural effusion/pneumothorax	1.47	1.49	1.47
End-stage renal disease or dialysis	0.14	0.15	0.15
Renal Failure	6.02	6.18	5.51
Decubitus ulcer or chronic skin ulcer	0.44	0.47	0.43
Trauma	5.08	5.13	5.00
Vertebral Fractures	1.30	1.37	1.30
Other injuries	27.57	27.71	27.66
Major Complications of Medical Care and Trauma	3.88	3.93	3.88

3.1.3 Unadjusted and Adjusted Complication Rate Distributions

Figures 5 and 6 display the frequency distributions of the hospital-specific complication rates, with and without risk-adjustment in the 2008 cohort. The unadjusted mean complication rate is 4.98 and ranges from 0 to 100% (Figure 5). The median unadjusted complication rate is 3.70%.

After adjusting for patient and clinical characteristics, the risk-standardized rates are more normally distributed (Figure 6) with a mean of 4.23, ranging from 2.20 to 8.88%. The median adjusted complication rate is 4.16%.

Figure 5. Unadjusted Hospital Complication Rates (2008 Sample; N=3,311 Hospitals)

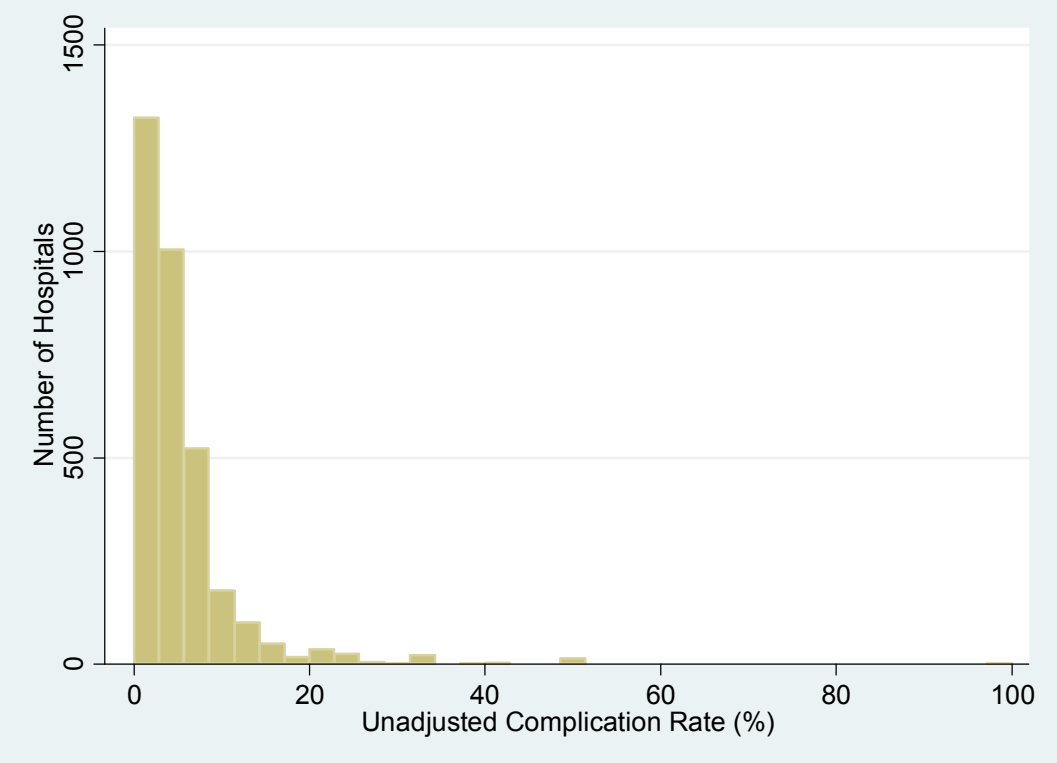
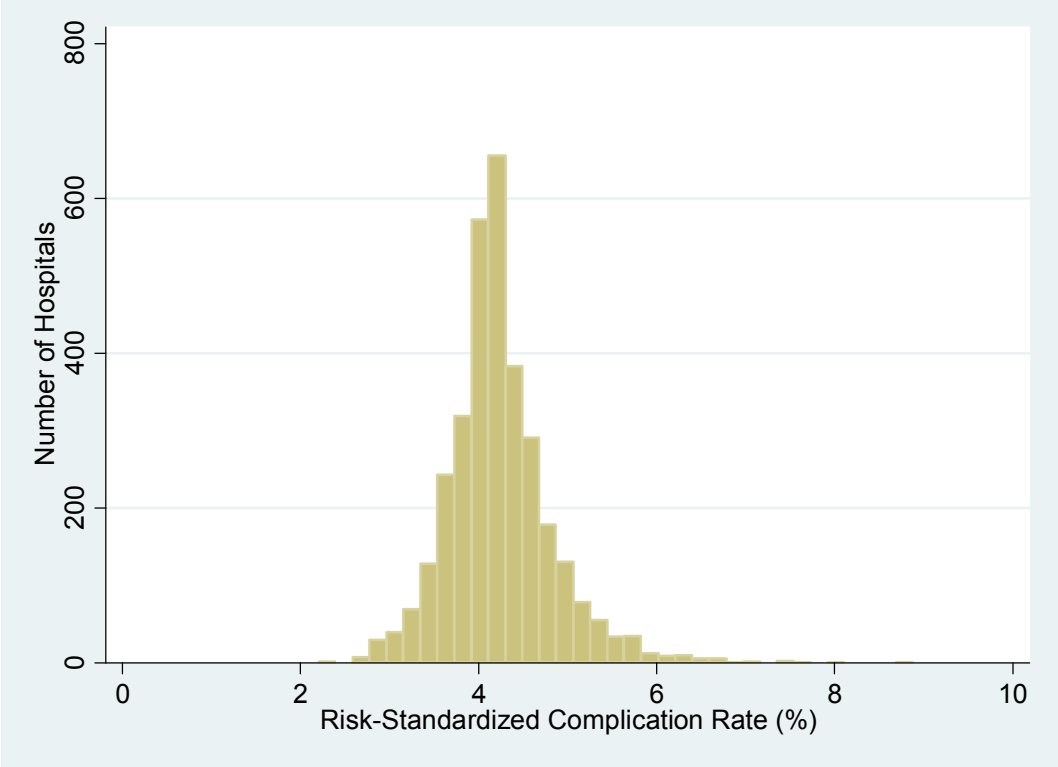


Figure 6. Distribution of Hospital Risk-Standardized Complication Rates (2008 Sample; N=3,311 Hospitals) – HGLM



4. MAIN FINDINGS / SUMMARY

The proposed measure of death and complications has the potential to significantly improve the quality of care delivered to patients undergoing elective primary THA and TKA procedures. Risk-standardized complication rates can be used for targeted quality improvement efforts by hospitals to decrease rates for death and complications post THA and TKA. The risk standardized model meets recognized standards for outcomes measurement and was developed with extensive input from clinicians and experts in measure development. The cohort for inclusion in the measure is appropriately defined, consisting of patients undergoing elective primary THA and/or TKA. The definitions for the complications, the complication-specific follow-up periods, and the risk-adjustment methodology all have strong face validity, which may facilitate physician acceptance. We excluded covariates that are not appropriate for inclusion in a quality measure, including race, socioeconomic status, and physician and hospital-level variables (e.g., procedural volume). The hierarchical modeling accounts for the clustering of patients within hospitals and differences in sample size across hospitals, thereby allowing for valid comparisons across hospitals. In summary, we present a claims-based model of death and complications post THA/TKA that is suitable for public reporting.

5. REFERENCES

1. Kurtz S,Ong K,Lau E,Mowat F,Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am.* Apr 2007;89(4):780-785.
2. Kurtz SM,Ong KL,Schmier J, et al. Future clinical and economic impact of revision total hip and knee arthroplasty. *J Bone Joint Surg Am.* Oct 2007;89 Suppl 3:144-151.
3. Ong KL,Mowat FS,Chan N, et al. Economic burden of revision hip and knee arthroplasty in Medicare enrollees. *Clin Orthop Relat Res.* May 2006;446:22-28.
4. Bozic KJ,Rubash HE,Sculco TP,Berry DJ. An analysis of medicare payment policy for total joint arthroplasty. *J Arthroplasty.* Sep 2008;23(6 Suppl 1):133-138.
5. National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.nysna.org/images/pdfs/practice/nqf_ana_outcomes_draft10.pdf. Accessed August 19, 2010.
6. Krumholz HM,Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* January 24, 2006 2006;113(3):456-462.
7. Kurtz S,Ong K,Lau E, et al. Prosthetic joint infection risk after TKA in the Medicare population. *Clin Orthop Relat Res.* 2010;468:5.
8. Bongartz T,Halligan CS,Osmon DR, et al. Incidence and risk factors of prosthetic joint infection after total hip or knee replacement in patients with rheumatoid arthritis. *Arthritis Rheum.* Dec 15 2008;59(12):1713-1720.
9. Soohoo NF,Farng E,Lieberman JR,Chambers L,Zingmond DS. Factors That Predict Short-term Complication Rates After Total Hip Arthroplasty. *Clin Orthop Relat Res.* Sep 2010;468(9):2363-2371.
10. Cram P,Vaughan-Sarrazin MS,Wolf B,Katz JN,Rosenthal GE. A comparison of total hip and knee replacement in specialty and general hospitals. *J Bone Joint Surg Am.* Aug 2007;89(8):1675-1684.
11. Khatod M,Inacio M,Paxton EW, et al. Knee replacement: epidemiology, outcomes, and trends in Southern California: 17,080 replacements from 1995 through 2004. *Acta Orthop.* Dec 2008;79(6):812-819.
12. Mahomed NN,Barrett JA,Katz JN, et al. Rates and outcomes of primary and revision total hip replacement in the United States medicare population. *J Bone Joint Surg Am.* Jan 2003;85-A(1):27-32.

13. Solomon DH,Chibnik LB,Losina E, et al. Development of a preliminary index that predicts adverse events after total knee replacement. *Arthritis & Rheumatism.* 2006;54(5):1536-1542.
14. Browne J,Cook C,Hofmann A,Bolognesi M. Postoperative morbidity and mortality following total knee arthroplasty with computer navigation. *Knee.* Mar 2010;17(2):152-156.
15. Huddleston JI,Maloney WJ,Wang Y, et al. Adverse Events After Total Knee Arthroplasty: A National Medicare Study. *The Journal of Arthroplasty.* 2009;24(6, Supplement 1):95-100.
16. Geerts WH,Pineo GF,Heit JA, et al. Prevention of Venous Thromboembolism. *Chest.* September 1, 2004 2004;126(3 suppl):338S-400S.
17. Pierce C,Haut E,Kardooni S, et al. Surveillance bias and deep vein thrombosis in the national trauma data bank: the more we look, the more we find. *J Trauma.* 2008;64:6.
18. Woodford HJ,George J. Diagnosis and Management of Urinary Tract Infection in Hospitalized Older People. *Journal of the American Geriatrics Society.* 2009;57(1):107-114.
19. Pope G,Ellis R,Ash A, et al. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. *Health Care Financing Review.* 2000;21(3):26.
20. McCallagh PNJ. *Generalized Linear Models*. London: Chapman and Hall; 1989.
21. Daniels M,Gatsonis C. Hierarchical Generalized Linear Models in the Analysis of Variations in Health Care Utilization. *Journal of the American Statistical Association.* 1999;94(445):14.
22. Normand S,Wang Y,Krumholz H. Assessing surrogacy of data sources for institutional comparisons. *Health Services and Outcomes Research Methodology.* 2007;7:79-96.
23. Harrell F,Shih Y. Using full probablility models to compute probabiliites of actual interest to decision makers. *Int J Technol Assess Health Care.* Winter 2001;17(1):10.

6. APPENDIX

6.1 Appendix A: Complication-Specific Measure Specifications

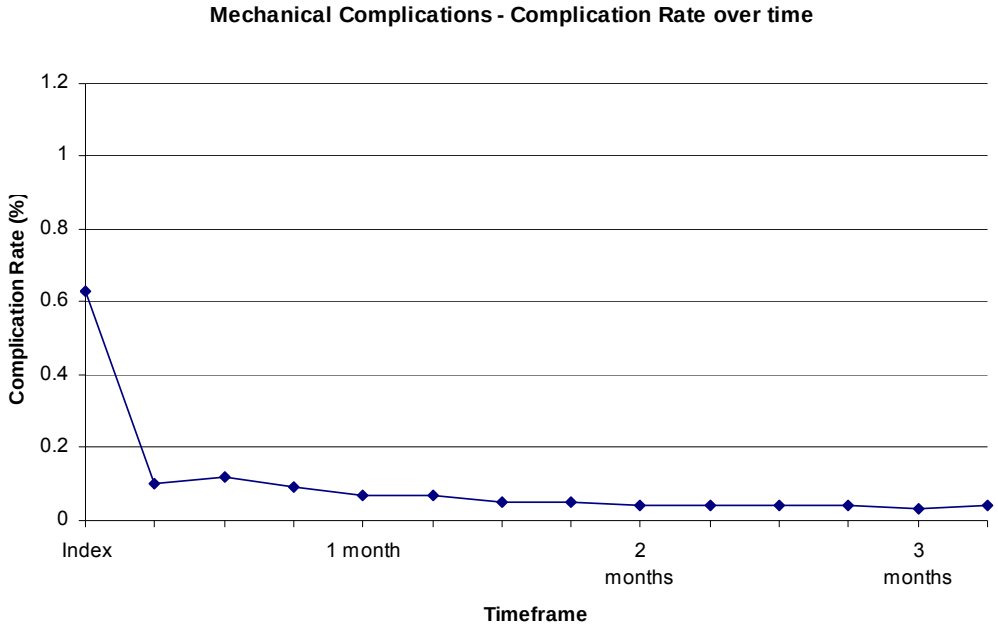
MECHANICAL COMPLICATIONS

Complication ICD-9 Code*	Description
996.4[1]	Mechanical complication of internal orthopedic device implant and graft
996.40[2]	Unspecified mechanical complication of internal orthopedic device, implant, and graft
996.41[2]	Mechanical loosening of prosthetic joint
996.42[2]	Dislocation of prosthetic joint
996.44[2]	Peri-prosthetic fracture around prosthetic joint
996.47[2]	Other mechanical complication of prosthetic joint implant
996.49[2]	Other mechanical complication of other internal orthopedic device, implant, and graft

When to Count as Complication	
Index Admission	**Rationale**
• Presence of any mechanical complication code listed above in a primary or secondary diagnosis field	• These codes identify mechanical complications related to the index procedure
Readmission	
• Presence of any mechanical complication code listed above in a primary or secondary diagnosis field	• These codes identify all mechanical complications, including those identified at the time of a readmission (even though mechanical complication may not be the primary reason for that readmission), since all are likely to be procedure-related
Follow-up Period for Complications Measure	
• During index admission or within 90 days from admission date	• Data indicate that the rate is elevated until 90 days post procedure • Mechanical complications occurring 90 days post procedure can still be attributable to the index procedure

[1] Weaver F, Hynes D, Hopkinson W, Wixson R, Khuri S, Daley J, Henderson W. (2003). Preoperative risks and outcomes of hip and knee arthroplasty in the Veterans Health Administration. *J Arthroplasty*, 18(6): 693-708.
[2] Memtsoudis S, Gonzalez Ella Valle A, Besculides M, Gaber L, Sculco T. (2008). In-hospital complications and mortality of unilateral, bilateral, and revision TKA. *Clin Orthop Relat Res*, 466:2617-2627.

*NOTE: Mechanical complication codes not used: 996.43, 996.45, 996.46



Data Source: Medicare Part A Inpatient Data, 2008

PERIPROSTHETIC JOINT INFECTION

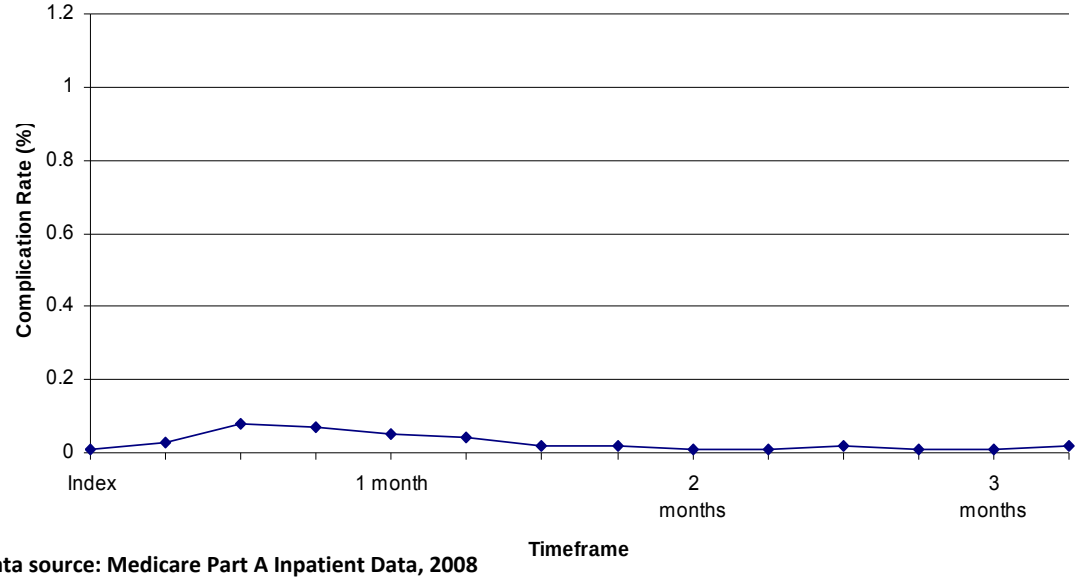
Complication ICD-9 Code	Description
996.66[3]	Infection and inflammatory reaction due to internal joint prosthesis

Intervention ICD-9 Code	Description
86.22	Excisional debridement of wound, infection, or burn
86.28	Nonexcisional debridement of wound, infection, or burn
86.04	Other incision with drainage of skin and subcutaneous tissue
81.53	Revise Hip Replacement, NOS
81.55	Revision of Knee replacement, NOS
81.59	Revision of joint replacement of lower extremity, not elsewhere classified
00.70	REV Hip Repl-acetab/fem
00.71	REV Hip Repl-acetab comp
00.72	REV Hip Repl-fem comp
00.73	REV Hip Repl-liner/head
00.80	Replacement of femoral, tibial, and patellar components (all components)
00.81	Replacement of tibial baseplate and tibial insert (liner)
00.82	Revision of knee replacement, femoral component
00.83	Revision of knee replacement, patellar component
00.84	Revision of total knee replacement, tibial insert (liner)
80.05	Arthrotomy for removal of prosthesis, hip
80.06	Arthrotomy for removal of prosthesis, knee
80.09	Arthrotomy for removal of prosthesis, other unspecified sites

[3] Thomas C, Cadwallader HL, Riley TV. (2004). Surgical-site infections after orthopaedic surgery: statewide surveillance using linked administrative databases. *J Hosp Infect*, (57(1): 25-30.

When to Count as Complication	
Index Admission	**Rationale**
• Presence of periprosthetic joint infection code listed above in a primary or secondary diagnosis field AND the presence of <u>at least one</u> of the following procedure codes: o Incision and drainage o Revision o Removal	• These codes identify periprosthetic joint infection related to the index procedure • Requiring an intervention sets an appropriate threshold for severity and will therefore more likely capture true joint infections and reduce false positives
Readmission	
• Presence of periprosthetic joint infection code listed above in a primary or secondary diagnosis field AND the presence of <u>at least one</u> of the following procedure codes: o Incision and drainage o Revision o Removal	• These codes identify all periprosthetic joint infections, including those identified at the time of a readmission (even though PJI may not be the primary reason for that readmission), since all are likely to be procedure-related
Follow-up Period for Complications Measure	
• During index admission or within 90 days from admission date	• Although the rate tapers off after approximately 6 weeks, it remains slightly elevated until 90 days post procedure • Periprosthetic joint infections occurring 90 days post procedure can still be attributable to the index procedure

Periprosthetic joint infection with Incision & Drainage and/or Revision/Removal - Complication Rate over Time



Data source: Medicare Part A Inpatient Data, 2008

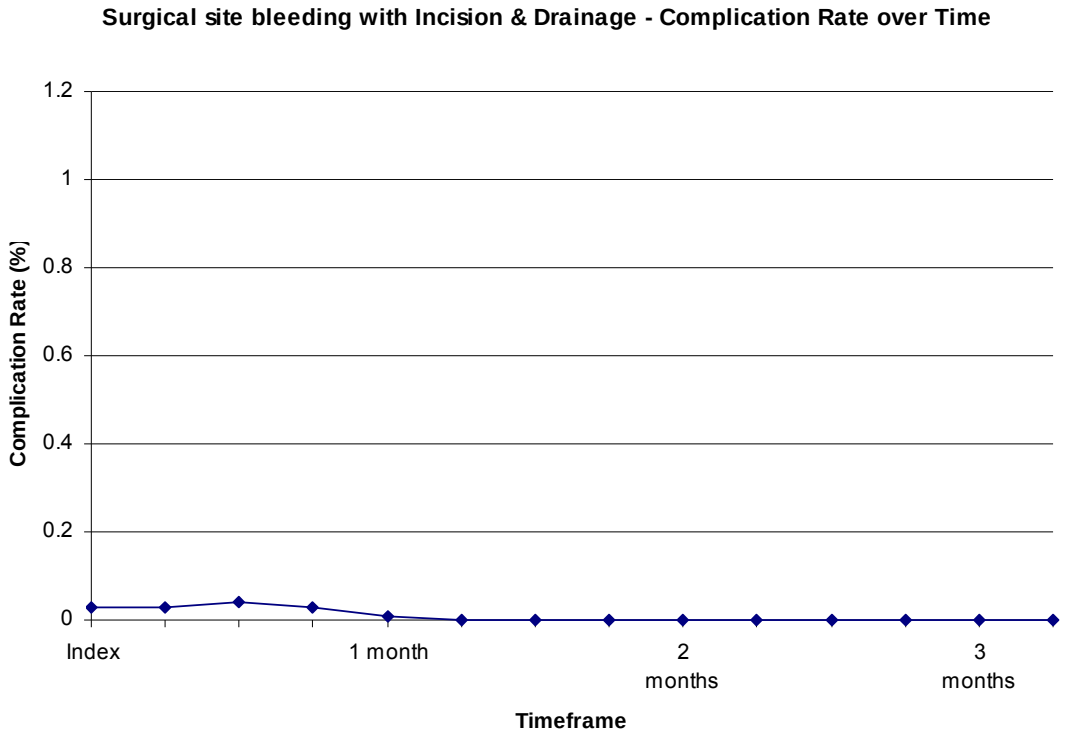
SURGICAL SITE BLEEDING

Complication ICD-9 Code	Description
998.1[4,5,6]	Hemorrhage or hematoma complicating a procedure not elsewhere classified
998.11[1,3,7,8]	Hemorrhage complicating a procedure
998.12[1,3,4,5]	Hematoma complicating a procedure
998.13[3]	Seroma complicating a procedure
286.5[5]	Bleeding from anticoagulation
719.10[1]	Hemarthrosis site unspecified
719.16[1]	Hemarthrosis involving lower leg
719.17[1]	Hemarthrosis involving ankle and foot

Intervention ICD-9 Code	Description
86.04	Other incision with drainage of skin and subcutaneous tissue

When to Count as Complication	Rationale
Index Admission	
• Presence of any bleeding code listed above in a primary or secondary diagnosis field AND: o procedure code for incision and drainage	• These codes identify surgical site bleeding related to the index procedure • Requiring an intervention sets an appropriate threshold for severity and will therefore more likely capture true surgical site bleeding and reduce false positives
Readmission	
• Presence of any bleeding code listed above in the primary or secondary diagnosis fields AND: o procedure code for incision and drainage	• These codes identify all surgical site bleeds, including those identified at the time of a readmission (even though bleeding may not be the primary reason for that readmission), since all are likely to be procedure-related
Follow-up Period for Complications Measure	
• During index admission or within 30 days from admission date	• Data indicate that rate decreases after 30 days • Consistent with clinical course

[4] Bozic K, Vail T, Pekow P, Maselli J, Lindenauer P, Auerbach A. (2009). Does aspirin have a role in venous thromboembolism prophylaxis in total knee arthroplasty patients? *J Arthroplasty*, 00(0): 1-8.
[5] Memtsoudis S, Gonzalez Ella Valle A, Besculides M, Gaber L, Sculco T. (2008). In-hospital complications and mortality of unilateral, bilateral, and revision TKA. *Clin Orthop Relat Res*, 466:2617-2627.
[6] Deyo R, Martin B, Kreuter W, Jarvik J, Mirza S. (2010). Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA*, 303(13): 1259-65.
[7] *Version 4.1 technical documentation* AHRQ Quality Indicators. December, 2009. Agency for Healthcare Research and Quality, Rockville, MD. http://www.qualityindicators.ahrq.gov/TechnicalSpecs41.htm
[8] Weaver F, Hynes D, Hopkinson W, Wixson R, Khuri S, Daley J, Henderson W. (2003). Preoperative risks and outcomes of hip and knee arthroplasty in the Veterans Health Administration. *J Arthroplasty*, 18(6): 693-708.



Data Source: Medicare Inpatient Part A Data, 2008

WOUND INFECTION

Complication ICD-9 Code*	Description
998.6[2,9]	Persistent postoperative fistula not elsewhere classified
998.83[2,3,10]	Non-healing surgical wound
998.3[4]	Disruption of wound
998.30[2,3,4]	Disruption of wound, unspecified
998.31[2,3,4]	Disruption of internal operation (surgical) wound
998.32[2,3,4]	Disruption of external operation (surgical) wound
998.33	Disruption of traumatic wound repair
998.5[2,3,4,11]	Postoperative infection not elsewhere classified
998.51[4]	Infected postoperative seroma
998.59[4,12]	Other postoperative infection
996.67[7]	Infection and inflammatory reaction due to other internal orthopedic device implant and graft

Intervention ICD-9 Code	Description
86.22	Excisional debridement of wound, infection, or burn
86.28	Nonexcisional debridement of wound, infection, or burn
86.04	Other incision with drainage of skin and subcutaneous tissue
81.53	Revise Hip Replacement, NOS
81.55	Revision of Knee replacement, NOS
81.59	Revision of joint replacement of lower extremity, not elsewhere classified
00.70	REV Hip Repl-acetab/fem
00.71	REV Hip Repl-acetab comp
00.72	REV Hip Repl-fem comp
00.73	REV Hip Repl-liner/head
00.80	Replacement of femoral, tibial, and patellar components (all components)
00.81	Replacement of tibial baseplate and tibial insert (liner)
00.82	Revision of knee replacement, femoral component
00.83	Revision of knee replacement, patellar component
00.84	Revision of total knee replacement, tibial insert (liner)
80.05	Arthrotomy for removal of prosthesis, hip
80.06	Arthrotomy for removal of prosthesis, knee
80.09	Arthrotomy for removal of prosthesis, other unspecified sites

[9] Memtsoudis S, Gonzalez Ella Valle A, Besculides M, Gaber L, Sculco T. (2008). In-hospital complications and mortality of unilateral, bilateral, and revision TKA. *Clin Orthop Relat Res*, 466:2617-2627.

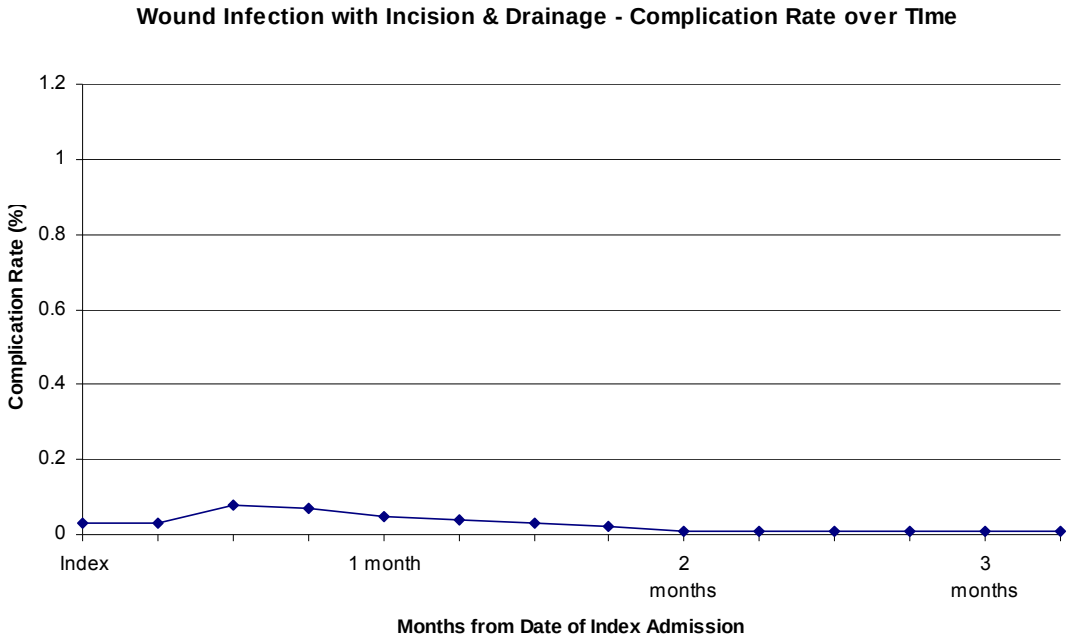
[10] Deyo R, Martin B, Kreuter W, Jarvik J, Mirza S. (2010). Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA*, 303(13): 1259-65.

[11] Thomas C, Cadwallader HL, Riley TV. (2004). Surgical-site infections after orthopaedic surgery: statewide surveillance using linked administrative databases. *J Hosp Infect*, (57(1): 25-30.

[12] Centers for Medicare and Medicaid Services No-Pay List

*NOTE: Wound infection codes not used: 890.0, 890.1, 890.2, 891.0, 891.1, 891.2, 894.1, 894.2, 998.89, 999.3, 999.31, 999.39, 686.9, 682.5, 682.6

When to Count as Complication	
Index Admission	**Rationale**
• Presence of any wound infection code listed above in a primary or secondary diagnosis field AND the presence of <u>at least one</u> of the following procedure codes: o Incision and drainage o Revision o Removal	• These codes identify wound infection related to the index procedure • Requiring an intervention sets an appropriate threshold for severity and will therefore capture true wound infections and reduce false positives
Readmission	
• Presence of any wound infection code listed above in a primary or secondary diagnosis field AND the presence of <u>at least one</u> of the following procedure codes: o Incision and drainage o Revision o Removal	• These codes identify all wound infections, including those identified at the time of a readmission (even though wound infection may not be the primary reason for that readmission), since all are likely to be procedure-related
Follow-up Period for Complications Measure	
• During index admission or within 30 days from admission date	• Data indicate that rate decreases after 30 days • Consistent with clinical course



Data Source: Medicare Inpatient Data, 2008

PULMONARY EMBOLISM (PE)

Complication	ICD-9 Code	Description
	415.1[13,14,15,16,17,18]	Pulmonary embolism and infarction
	415.11[1,2,3,6]	Iatrogenic pulmonary embolism and infarction
	415.19[1,2,3,6]	Other pulmonary embolism and infarction

When to Count as Complication	
Index Admission	**Rationale**
• Presence of any pulmonary embolism code listed in the primary or secondary diagnosis fields	• These codes identify PE related to the index procedure
Readmission	
• Presence of any pulmonary embolism code listed above in the primary or secondary diagnosis fields	• These codes identify all PEs, including those identified at the time of a readmission (even though PE may not be the primary reason for that readmission), since all are likely to be procedure-related
Follow-up Period for Complications Measure	
• During index admission or within 30 days from admission date	• Data indicate that rate decreases after 30 days • Consistent with clinical course

[13] *Version 4.1 technical documentation* AHRQ Quality Indicators. December, 2009. Agency for Healthcare Research and Quality, Rockville, MD. http://www.qualityindicators.ahrq.gov/TechnicalSpecs41.htm

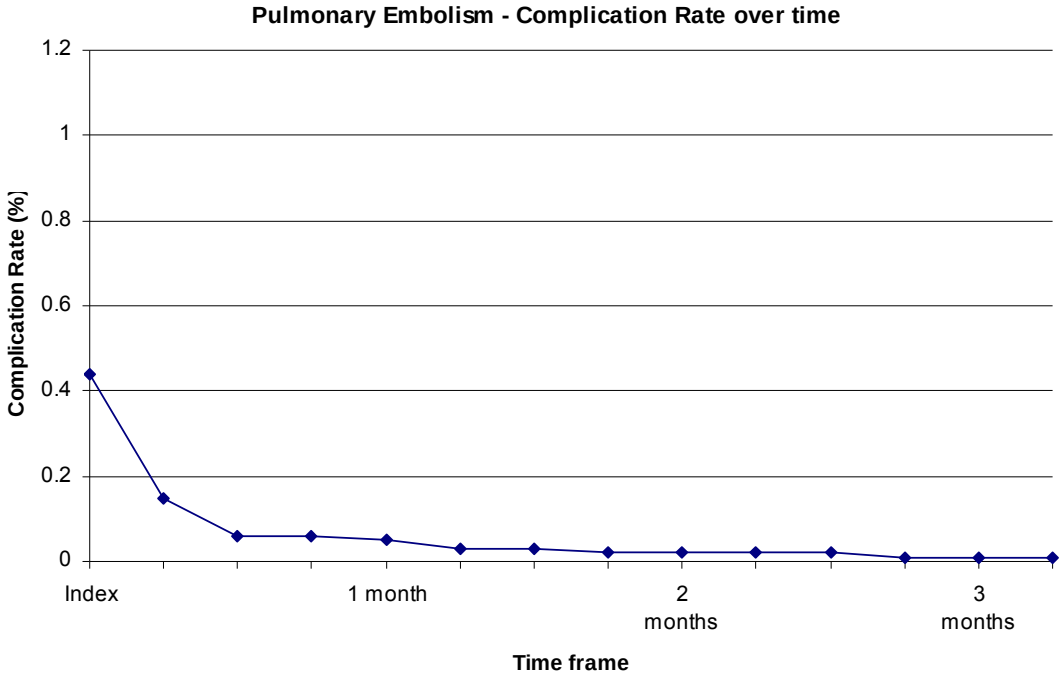
[14] Solomon D, Chibnik L, Losina E, Huang J, Fossel A, Husni E, Katz J. (2006). Development of a preliminary index that predicts adverse events after total knee replacement. *Arthritis Rheum*, 54(5): 1536-1542.

[15] Huddleston J, Maloney W, Wang Y, Verzier N, Hunt D, Herndon J. (2009). Adverse events after total knee arthroplasty. J Arthroplasty, 24(6): 95-100.

[16] Memtsoudis S, Gonzalez Ella Valle A, Besculides M, Gaber L, Sculco T. (2008). In-hospital complications and mortality of unilateral, bilateral, and revision TKA. *Clin Orthop Relat Res*, 466:2617-2627.

[17] Weaver F, Hynes D, Hopkinson W, Wixson R, Khuri S, Daley J, Henderson W. (2003). Preoperative risks and outcomes of hip and knee arthroplasty in the Veterans Health Administration. *J Arthroplasty*, 18(6): 693-708.

[18] Deyo R, Martin B, Kreuter W, Jarvik J, Mirza S. (2010). Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA*, 303(13): 1259-65.



Data Source: Medicare Inpatient Part A Data, 2008

ACUTE MYOCARDIAL INFARCTION (AMI)

Complication ICD-9 Code	Description
*410[19,20]	Acute myocardial infarction
410.0[1,21]	Acute myocardial infarction of anterolateral wall
410.00[1]	Acute myocardial infarction of anterolateral wall episode of care unspecified
410.01[1]	Acute myocardial infarction of anterolateral wall initial episode of care
410.1[1,3]	Acute myocardial infarction of other anterior wall
410.10[1]	Acute myocardial infarction of other anterior wall episode of care unspecified
410.11[1]	Acute myocardial infarction of other anterior wall initial episode of care
410.2[1,3]	Acute myocardial infarction of inferolateral wall
410.20[1]	Acute myocardial infarction of inferolateral wall episode of care unspecified
410.21[1]	Acute myocardial infarction of inferolateral wall initial episode of care
410.3[1,3]	Acute myocardial infarction of inferoposterior wall
410.30[1]	Acute myocardial infarction of inferoposterior wall episode of care unspecified
410.31[1]	Acute myocardial infarction of inferoposterior wall initial episode of care
410.4[1,3]	Acute myocardial infarction of other inferior wall
410.40[1]	Acute myocardial infarction of other inferior wall episode of care unspecified
410.41[1]	Acute myocardial infarction of other inferior wall initial episode of care
410.5[1,3]	Acute myocardial infarction of other lateral wall
410.50[1]	Acute myocardial infarction of other lateral wall episode of care unspecified
410.51[1]	Acute myocardial infarction of other lateral wall initial episode of care
410.6[1,3]	True posterior wall infarction
410.60[1]	True posterior wall infarction episode of care unspecified
410.61[1]	True posterior wall infarction initial episode of care
410.7[1,3]	Subendocardial infarction
410.70[1]	Subendocardial infarction episode of care unspecified
410.71[1]	Subendocardial infarction initial episode of care
410.8[1,3]	Acute myocardial infarction of other specified sites
410.80[1]	Acute myocardial infarction of other specified sites episode of care unspecified
410.81[1]	Acute myocardial infarction of other specified sites initial episode of care
410.9[1,3]	Acute myocardial infarction of unspecified site
410.90[1]	Acute myocardial infarction of unspecified site episode of care unspecified
410.91[1]	Acute myocardial infarction of unspecified site initial episode of care

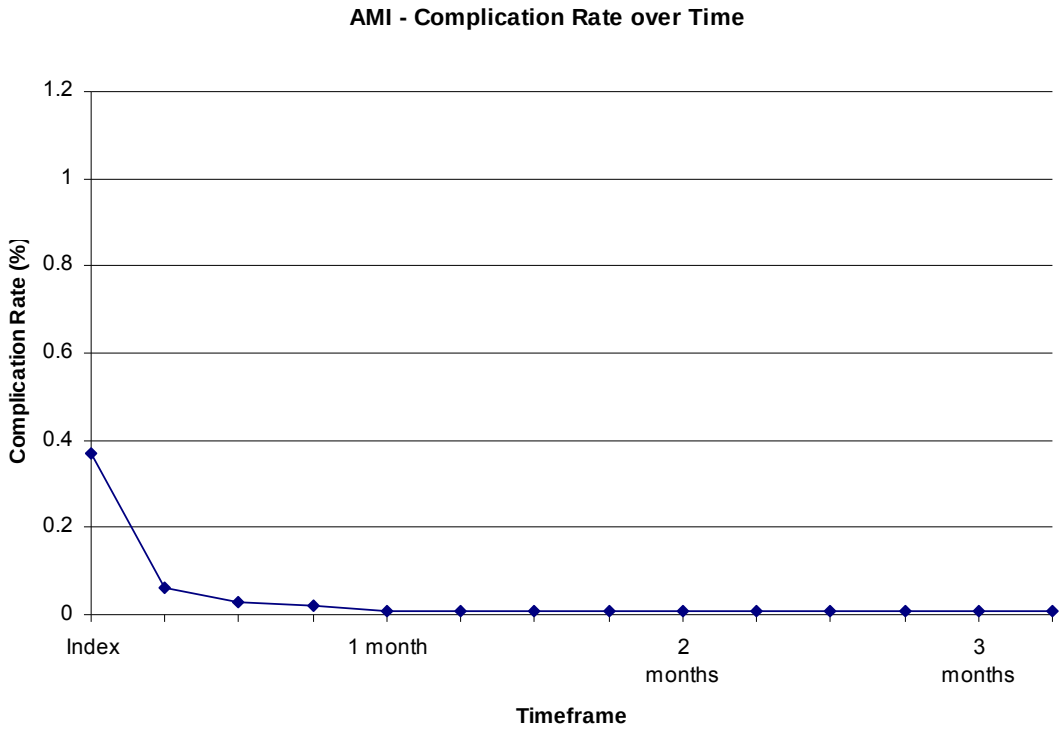
[19] Yale/CORE cohort definition for pneumonia

[20] Weaver F, Hynes D, Hopkinson W, Wixson R, Khuri S, Daley J, Henderson W. (2003). Preoperative risks and outcomes of hip and knee arthroplasty in the Veterans Health Administration. *J Arthroplasty*, 18(6): 693-708.

[21] Deyo R, Martin B, Kreuter W, Jarvik J, Mirza S. (2010). Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA*, 303(13): 1259-65.

* **NOTE: Excludes the following code: 0**410.x2

When to Count as Complication	
Index Admission	**Rationale**
• Presence of any AMI code listed above in a primary or secondary diagnosis field	• These codes identify AMI related to the index procedure
Readmission	
• Presence of any AMI code listed above in a primary field only	• These codes identify AMI's that were the primary reason for a readmission • AMIs that are secondary diagnoses in readmissions may represent a history of AMI or a complication of the second admission
Follow-up Period for Complications Measure	
• During index admission or within 7 days from index admission date	• More likely to be attributable to procedure if it occurs within 7 days of procedure • Rate decreases sharply 7 days from admission and returns to baseline within 30 days • Limits overlap with 30-day all-cause readmission measure



Data source: Medicare Part A Inpatient Data, 2008

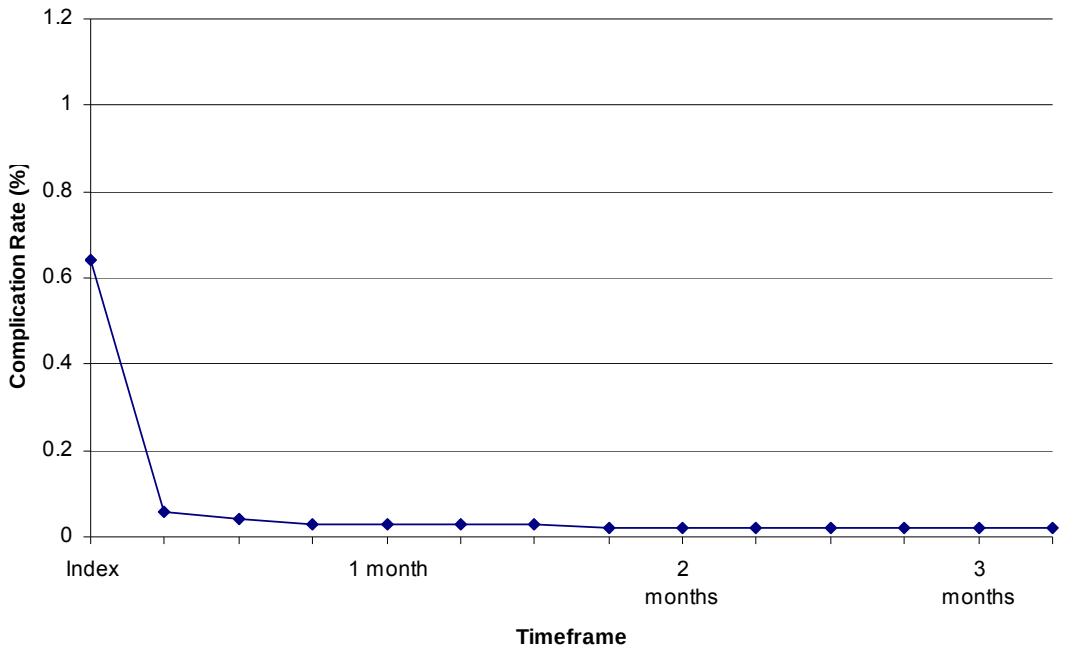
PNEUMONIA

Complication ICD-9 Code	Description
480[22]	Viral pneumonia
480.0[1]	Pneumonia due to adenovirus
480.1[1]	Pneumonia due to respiratory syncytial virus
480.2[1]	Pneumonia due to parainfluenza virus
480.3[1]	Pneumonia due to sars-associated coronavirus
480.8[1]	Pneumonia due to other virus not elsewhere classified
480.9[1]	Viral pneumonia unspecified
481[1,23,24,25,26]	Pneumococcal pneumonia
482[4,5]	Other Bacterial Pneumonia
482.0[1,5]	Pneumonia due to klebsiella pneumoniae
482.1[1,5]	Pneumonia due to pseudomonas
482.2[1,2,3,5]	Pneumonia due to hemophilus influenzae (h. influenzae)
482.3	Pneumonia due to streptococcus
482.30[1,2,3,5]	Pneumonia due to streptococcus unspecified
482.31[1,2,3,5]	Pneumonia due to streptococcus group a
482.32[1,2,3,5]	Pneumonia due to streptococcus group b
482.39[1,2,3,5]	Pneumonia due to other streptococcus
482.4	Pneumonia due to staphylococcus
482.40[1,5]	Pneumonia due to staphylococcus unspecified
482.41[1,2,3,5]	Methicillin susceptible pneumonia due to staphylococcus aureus
482.42[5]	Methicillin resistant pneumonia due to staphylococcus aureus
482.49[1,5]	Other staphylococcus pneumonia
482.81[1,5]	Pneumonia due to anaerobes
482.82[1,5]	Pneumonia due to escherichia coli [e.coli]
482.83[1,5]	Pneumonia due to other gram-negative bacteria
482.84[1,5]	Pneumonia due to legionnaires' disease
482.89[1,5]	Pneumonia due to other specified bacteria
482.9[1,2,3,5]	Bacterial pneumonia unspecified
483[1,2,3]	Pneumonia due to other specified organism
483.0[1]	Pneumonia due to mycoplasma pneumoniae
483.1[1]	Pneumonia due to chlamydia
483.8[1]	Pneumonia due to other specified organism
485[1-5]	Bronchopneumonia organism unspecified
486[1-5]	Pneumonia organism unspecified
487.0[1]	Influenza with pneumonia
507.0[4]	Pneumonitis due to inhalation of food or vomitus

[22] Yale/CORE cohort definition for pneumonia
[23] *Version 4.1 technical documentation* AHRQ Quality Indicators. December, 2009. Agency for Healthcare Research and Quality, Rockville, MD. http://www.qualityindicators.ahrq.gov/TechnicalSpecs41.htm
[24] National Quality Forum Endorsed Standard-Bacterial Pneumonia.
[25] Weaver F, Hynes D, Hopkinson W, Wixson R, Khuri S, Daley J, Henderson W. (2003). Preoperative risks and outcomes of hip and knee arthroplasty in the Veterans Health Administration. *J Arthroplasty*, 18(6): 693-708.
[26] Deyo R, Martin B, Kreuter W, Jarvik J, Mirza S. (2010). Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA*, 303(13): 1259-65.

When to Count as Complication	
Index Admission	**Rationale**
• Presence of any pneumonia code listed above in a primary or secondary diagnosis field	• These codes identify pneumonia related to the index procedure
Readmission	
• Presence of any pneumonia code listed above in a primary diagnosis field only	• These codes identify pneumonias that were the primary reason for a readmission • Pneumonias that are secondary diagnoses in readmissions may represent a history of pneumonia or a complication of the second admission
Follow-up Period for Complications Measure	
• During index admission or within 7 days from index admission date	• More likely to be attributable to procedure if it occurs within 7 days of procedure • Rate decreases sharply 7 days from admission and returns to baseline within 30 days • Limits overlap with 30-day all-cause readmission measure

Pneumonia - Complication Rate over Time



Data source: Medicare Part A Inpatient Data, 2008

SEPSIS/SEPTICEMIA

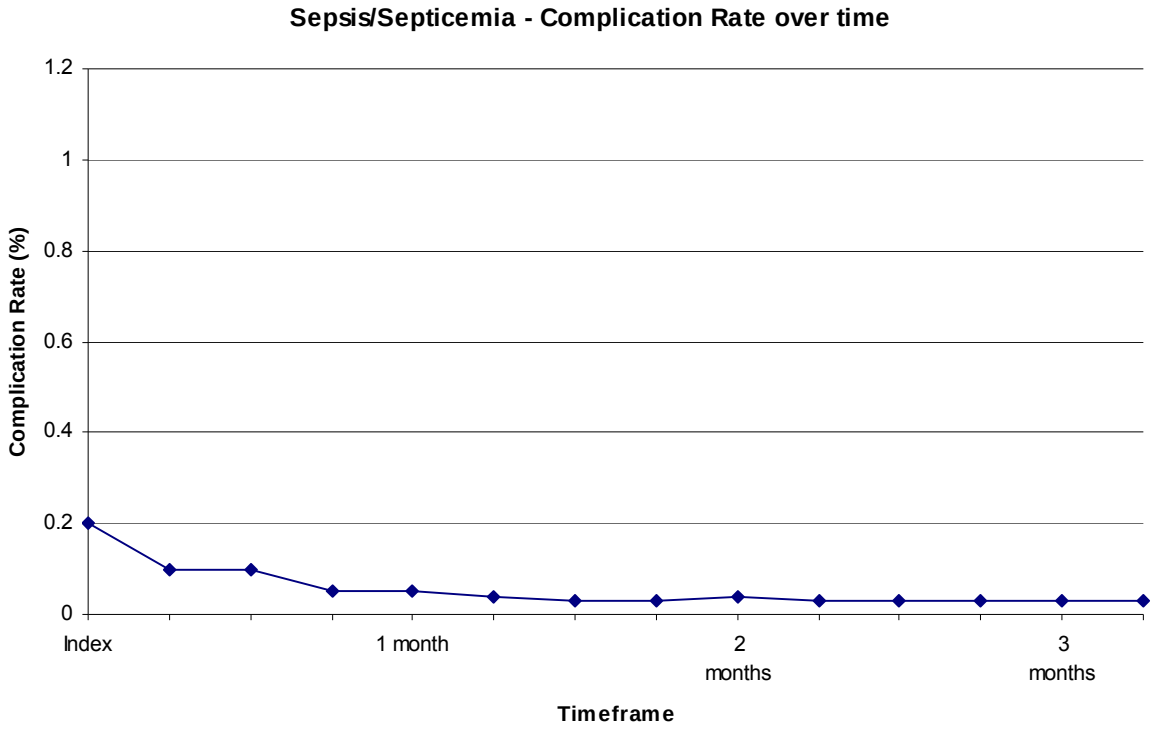
Complications	ICD-9 Code	Description
038[27]		Septicemia
	038.0[28,29]	Streptococcal septicemia
	038.1[2,3]	Staphylococcal septicemia
	038.10[2,3]	Staphylococcal septicemia unspecified
	038.11[2,3]	Methicillin susceptible staphylococcus aureus septicemia
	038.12[2,3]	Methicillin resistant staphylococcus aureus septicemia
	038.19[2,3]	Other staphylococcal septicemia
	038.2[2,3]	Pneumococcal septicemia
	038.3[2,3]	Septicemia due to anerobes
	038.4[2,3]	Septicemia due to other gram-negative organisms
	038.40[2,3]	Septicemia due to gram negative organisms unspecified
	038.41[2,3]	Septicemia due to h. influenzae
	038.42[2,3]	Septicemia due to e. coli
	038.43[2,3]	Septicemia due to pseudomonas
	038.44[2,3]	Septicemia due to serratia
	038.49[2,3]	Other septicemia due to gram-negative organisms
	038.8[2,3]	Other specified septicemias
	038.9[2,3]	Unspecified septicemia
785.52[2,3]		Septic shock
785.59[2,3]		Other shock without trauma
790.7		Bacteremia
995.91[2,3]		Systemic inflammatory response syndrome due to infectious process w/out organ dysfunction
995.92[2,3]		Systemic inflammatory response syndrome due to infectious process with organ dysfunction
998.0[2,3]		Postoperative shock not elsewhere classified
998.59		Post procedural sepsis

[27] Weaver F, Hynes D, Hopkinson W, Wixson R, Khuri S, Daley J, Henderson W. (2003). Preoperative risks and outcomes of hip and knee arthroplasty in the Veterans Health Administration. *J Arthroplasty*, 18(6): 693-708.

[28] *Version 4.1 technical documentation* AHRQ Quality Indicators. December, 2009. Agency for Healthcare Research and Quality, Rockville, MD. http://www.qualityindicators.ahrq.gov/TechnicalSpecs41.htm

[29] Solomon D, Chibnik L, Losina E, Huang J, Fossel A, Husni E, Katz J. (2006). Development of a preliminary index that predicts adverse events after total knee replacement. *Arthritis Rheum*, 54(5): 1536-1542.

When to Count as Complication	
Index Admission	**Rationale**
• Presence of any sepsis/septicemia code listed above in a primary or secondary diagnosis field	• These codes identify sepsis/septicemia related to the index procedure
Readmission	
• Presence of any sepsis/septicemia code listed above in a primary diagnosis or secondary diagnosis field	• Sepsis/septicemia rates will be underestimated if identified using primary diagnosis field only, as these codes are found more frequently in the secondary diagnosis fields • Primary field may indicate the source of sepsis/septicemia
Follow-up Period for Complications Measure	
• During index admission or within 7 days from index admission date	• More likely to be attributable to procedure if it occurs within 7 days of procedure • Rate decreases 7 days from admission and returns to baseline within 30 days • Limits overlap with 30-day all-cause readmission measure



Data source: Medicare Part A Inpatient Data, 2008

6.2 Appendix B: ICD-9-CM Codes for Hip Fracture, Revision Procedures, Partial Hip Arthroplasty, and Resurfacing Procedure

ICD-9-CM Codes for Hip Fracture	
733.1	Pathologic fracture
733.10	Pathological fracture unspecified site
733.14	Pathological fracture of neck of femur
733.15	Pathological fracture of other specified part of femur
733.19	Pathological fracture of other specified site
733.8	Malunion and nonunion of fracture
733.81	Malunion of fracture
733.82	Nonunion of fracture
733.95	Stress fracture of other bone
733.96	Stress fracture of femoral neck
733.97	Stress fracture of shaft of femur
808.0	Closed fracture of acetabulum
808.1	Open fracture of acetabulum
820.00	Fracture of unspecified intracapsular section of neck of femur closed
820.01	Fracture of epiphysis (separation) (upper) of neck of femur closed
820.02	Fracture of midcervical section of femur closed
820.03	Fracture of base of neck of femur closed
820.09	Other transcervical fracture of femur closed
820.10	Fracture of unspecified intracapsular section of neck of femur open
820.11	Fracture of epiphysis (separation) (upper) of neck of femur open
820.12	Fracture of midcervical section of femur open
820.13	Fracture of base of neck of femur open
820.19	Other transcervical fracture of femur open
820.20	Fracture of unspecified trochanteric section of femur closed
820.21	Fracture of intertrochanteric section of femur closed
820.22	Fracture of subtrochanteric section of femur closed
820.30	Fracture of unspecified trochanteric section of femur open
820.31	Fracture of intertrochanteric section of femur open
820.32	Fracture of subtrochanteric section of femur open
820.8	Fracture of unspecified part of neck of femur closed
820.9	Fracture of unspecified part of neck of femur open
821	Fracture of other and unspecified parts of femur
821.0	Fracture of shaft or unspecified part of femur closed
821.00	Fracture of unspecified part of femur closed
821.01	Fracture of shaft of femur closed
821.1	Fracture of shaft or unspecified part of femur open
821.10	Fracture of unspecified part of femur open
821.11	Fracture of shaft of femur open

ICD-9-CM Codes for THA and TKA Revision Procedures	
00.70	REV Hip Repl-acetab/fem OCT05
00.71	REV Hip Repl-acetab comp OCT05
00.72	REV Hip Repl-fem comp OCT05
00.73	REV Hip Repl-liner/head OCT05

00.80	Replacement of femoral, tibial, and patellar components (all components)
00.81	Replacement of tibial baseplate and tibial insert (liner)
00.82	Revision of knee replacement, femoral component
00.83	Revision of knee replacement, patellar component
00.84	Revision of total knee replacement, tibial insert (liner)
81.53	Revise Hip Replacement, NOS
81.55	Revision of Knee replacement, NOS
81.59	Revision of joint replacement of lower extremity, not elsewhere classified

ICD-9-CM Code for Partial Hip Arthroplasty Procedure	
81.52	Partial Hip Replacement

ICD-9-CM Codes for THA Resurfacing Procedure	
00.85	Resurfacing hip, total, acetabulum and femoral head, hip resurfacing arthroplasty, total
00.86	Resurfacing hip, partial, femoral head, hip resurfacing arthroplasty, NOS, hip resurfacing arthroplasty, partial, femoral head
00.87	Resurfacing hip, partial, acetabulum, hip resurfacing arthroplasty, partial, acetabulum

6.3 Appendix C Conditions That May Represent Adverse Outcomes of Care Received During Index Admission.

CC	Description
2	Septicemia/Shock
6	Other Infectious Diseases
17	Diabetes with Acute Complications
23	Disorders of Fluid/Electrolyte/Acid-Base
24	Other Endocrine/Metabolic/Nutritional Disorders
31	Intestinal Obstruction/Perforation
34	Peptic Ulcer, Hemorrhage, Other Specified Gastrointestinal Disorders
36	Other Gastrointestinal Disorders
37	Bone/Joint/Muscle Infections/Necrosis
43	Other Musculoskeletal and Connective Tissue Disorders
46	Coagulation Defects and Other Specified Hematological Disorders
47	Iron Deficiency and Other/Unspecified Anemias and Blood Disease
48	Delirium and Encephalopathy
51	Drug/Alcohol Psychosis
75	Coma, Brain Compression/Anoxic Damage
76	Mononeuropathy, Other Neurological Conditions/Injuries
77	Respirator Dependence/Tracheostomy Status
78	Respiratory Arrest
79	Cardio-respiratory failure and shock
80	Congestive heart failure
81	Acute myocardial infarction
82	Unstable angina
85	Heart Infection/Inflammation, Except Rheumatic
95	Cerebral Hemorrhage
96	Ischemic or Unspecified Stroke
97	Precerebral Arterial Occlusion and Transient Cerebral Ischemia
100	Hemiplegia/Hemiparesis
101	Cerebral Palsy and Other Paralytic Syndromes
102	Speech, Language, Cognitive, Perceptual
104	Vascular Disease with Complications
105	Vascular Disease
106	Other Circulatory Disease
111	Aspiration and Specified Bacterial Pneumonias
112	Pneumococcal Pneumonia, Emphysema, Lung Abscess
114	Pleural Effusion/Pneumothorax
130	Dialysis Status
131	Renal failure
132	Nephritis
133	Urinary Obstruction and Retention
135	Urinary Tract Infection
148	Decubitus Ulcer of Skin
152	Cellulitis, Local Skin Infection
154	Severe Head Injury
155	Major Head Injury
156	Concussion or Unspecified Head Injury

CC	Description
157	Vertebral Fractures
158	Hip Fracture/Dislocation
159	Major Fracture, Except of Skull, Vertebrae, or Hip
160	Internal Injuries
161	Traumatic Amputation
162	Other Injuries
163	Poisonings and Allergic Reactions
164	Major Complications of Medical Care and Trauma
165	Other Complications of Medical Care
175	Other Organ Transplant/Replacement
177	Amputation Status, Lower Limb/Amputation
178	Amputation Status, Upper Limb

6.4 Appendix D CCs Not Considered for Risk Adjustment

CC	Description	Rationale
66	Attention Deficit Disorder	Pediatric ; Low frequency
123	Cataracts	Marker of clinical practice, not clinical relevant
129	End Stage Renal Disease	Not included in CMS-HCC Model
137	Female Infertility	Irrelevant to Medicare FFS Population
141	Ectopic Pregnancy	Irrelevant to Medicare FFS Population
142	Miscarriage/Abortion	Irrelevant to Medicare FFS Population
143	Completed Pregnancy with Major Complications	Irrelevant to Medicare FFS Population
144	Completed Pregnancy with Complications	Irrelevant to Medicare FFS Population
145	Completed Pregnancy without Complication	Irrelevant to Medicare FFS Population
146	Uncompleted Pregnancy with Complications	Irrelevant to Medicare FFS Population
147	Uncompleted Pregnancy with No or Minor Complications	Irrelevant to Medicare FFS Population
168	Extremely Low Birthweight Neonates	Fetal Effects; Irrelevant to Medicare FFS Population
169	Very Low Birthweight Neonates	Fetal Effects; Irrelevant to Medicare FFS Population
170	Serious Perinatal Problems Affecting Newborn	Fetal Effects; Irrelevant to Medicare FFS Population
171	Other Perinatal Problems Affecting Newborn	Fetal Effects; Irrelevant to Medicare FFS Population
172	Normal, Single Birth	Fetal Effects; Irrelevant to Medicare FFS Population
173	Major Organ Transplant	Not included in CMS-HCC Model
176	Artificial Openings for Feeding or Elimination	CC too heterogeneous; Mix of disparate codes
179	Post-Surgical States/Aftercare/Elective	CC too heterogeneous; Mix of disparate codes
180	Radiation Therapy	CC too heterogeneous; Mix of disparate codes
181	Chemotherapy	CC too heterogeneous; Mix of disparate codes
182	Rehabilitation	CC too heterogeneous; Mix of disparate codes
183	Screening/Observation/Special Exams	CC too heterogeneous; Mix of disparate codes
184	History of Disease	CC too heterogeneous; Mix of disparate codes
185	Oxygen	Not included in CMS-HCC Model; DME
186	CPAP/IPPB/Nebulizers	Not included in CMS-HCC Model; DME
187	Patient Lifts, Power Operated Vehicles, Beds	Not included in CMS-HCC Model; DME
188	Wheelchairs, Commodes	Not included in CMS-HCC Model; DME
189	Walkers	Not included in CMS-HCC Model; DME

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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 subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. 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 subcriterion 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 subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, 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 subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1551 **NQF Project:** Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Hospital-level 30-day all-cause risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA)
De.2 **Brief description of measure***:* This measure estimates hospital 30-day RSRRs following elective primary THA and TKA in patients 65 years and older. The measure uses Medicare claims data to develop a hospital-level RSRR for THA and TKA and will include patients readmitted for any reason within 30 days of discharge date of the index admission. Some patients are admitted within 30 days of the index hospitalization to undergo another elective THA/TKA procedure. These are considered planned readmissions and are NOT counted in the measure as readmissions.
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** This measure is paired with a complications measure for THA and TKA.
De.4 **National Priority Partners Priority Area:** Care coordination, Safety De.5 **IOM Quality Domain**: Effectiveness, Patient-centered, Efficiency, Safety De.6 **Consumer Care Need:** Getting better, Living with illness

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:**	**A** Y☐ N☐

A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary A.4 **Measure Steward Agreement attached:**	
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, Internal quality improvement	**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, Frequently performed procedure, High resource use, Other 1a.2 High cost 1a.3 **Summary of Evidence of High Impact:** Primary elective THA and TKA are beneficial procedures that greatly improve the quality of life for patients who choose to undergo these procedures (Hawker et al., 1998). However, these high volume procedures are expensive and are associated with significant readmission rates. High Readmission Rate We conducted analyses using 2008 Medicare Part A inpatient claims data and found a median 30-day risk-standardized hospital readmission rate of 6.1%. This rate is high considering these are elective procedures typically performed on younger, healthier patients, as compared to other Medicare patients. High Volume THA and TKA are priority areas for outcomes measure development, as they are commonly performed procedures in the US. In 2003 there were 202,500 primary hip arthroplasties and 402,100 primary total knee	**1a** C☐ P☐ M☐ N☐

arthroplasties performed (Kurtz et al., 2007). The number of procedures performed has increased steadily over the past decade (Kurtz et al., 2007; Ong et al., 2006).

High Cost
Although these procedures can dramatically improve patient health-related quality-of-life, they are costly. In 2005 annual hospital charges totaled $3.95 billion and $7.42 billion for primary THA and TKA, respectively (Kurtz et al., 2007). These costs are projected to increase by 340% to 17.4 billion for THA and by 450% to 40.8 billion for TKA by 2015 (Kurtz et al., 2007). Medicare is the single largest payer for these procedures, covering approximately two-thirds of all THAs and TKAs performed in the US (Ong et al., 2006). THA and TKA procedures combined account for the largest procedural cost in the Medicare budget (Bozic et al., 2008).

1a.4 **Citations for Evidence of High Impact:** Bozic KJ, Rubash HE, Sculco TP, Berry DJ. An analysis of medicare payment policy for total joint arthroplasty. Journal of Arthroplasty. 2008;23(6 Suppl 1):133-138.

Hawker GJ, Wright J, Coyte P, Paul J, Dittus R, Croxford B, et al. Health-related quality of life after knee replacement. J Bone Joint Surg Am. 1998; 80:163-73.

Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am. Apr 2007;89(4):780-785.

Kurtz SM, Ong KL, Schmier J, et al. Future clinical and economic impact of revision total hip and knee arthroplasty. J Bone Joint Surg Am. Oct 2007;89 Suppl 3:144-151.

Ong KL, Mowat FS, Chan N, Lau E, Halpern MT, Kurtz SM. Economic burden of revision hip and knee arthroplasty in Medicare enrollees. Clin Orthop Relat Res. May 2006;446:22-28.

1b. Opportunity for Improvement

1b.1 **Benefits (improvements in quality) envisioned by use of this measure:** THA and TKA are priority areas for outcomes measure development, as they are costly and commonly performed procedures. Hospital readmission is an outcome that is likely attributable to care processes and is an important outcome for patients. Measuring and reporting readmission rates will inform health care providers about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices. Furthermore, the measure will increase transparency for consumers and has the potential to lower health care costs associated with readmissions.

1b.2 **Summary of data demonstrating performance gap (variation or overall poor performance) across providers:**
Readmission rates are high, given these are elective procedures and there is marked variation in rates across hospitals. The unadjusted mean readmission rate was 6.78% and ranged from 0% to 100% across 3,310 hospitals in 2008. Even after adjustment for patient and clinical characteristics, the mean readmission rate was 6.30%, ranging from 3.06% to 50.94%. Because these are elective procedures that are performed on relatively healthy patients, readmission rates are expected to be lower in these patients as compared to patients admitted for an emergent procedure.

The literature indicates there is considerable variation in practice patterns, patient outcomes, and adherence to payer-defined practice guidelines for both THA and TKA (Bozic et al 2008; Ong et al 2008). Our analyses are consistent with this evidence. In 2008, 30-day adjusted readmission rates ranged from 3.06% to 50.94%. This variation likely indicates differences in the quality of care received across hospitals. These findings suggest that many readmissions could potentially be prevented.

1b.3 **Citations for data on performance gap:**
Bozic KJ, Chiu V. Quality Measurement and Public Reporting in Total Joint Replacement. The Journal of Replacement. 2008; 23:146-149.

Ong K, Lau E, Manley M, Kurtz S. Effect of procedure duration on total hip arthroplasty and total knee arthroplasty survivorship in the United States Medicare population. J Arthroplasty. 2008; 23(6): 127-132.

1b
C☐
P☐
M☐
N☐

1b.4 Summary of Data on disparities by population group:
We conducted analyses to explore disparities by SES. We used Medicaid eligibility status identified in the Medicare claims enrollment database (EDB) as a proxy for SES. This approach is consistent with prior research as well as NQF recommendations (http://www.nysna.org/images/pdfs/practice/nqf_ana_outcomes_draft10.pdf). Patients were categorized into two groups, based on their eligibility status for Medicaid (yes/no). The Medicaid eligible population represents lower SES status. Analyses demonstrated that although SES is a significant predictor of readmission at the patient level, it does not affect overall hospital performance in the risk-adjusted readmission model. Consistent with NQF guidelines, this measure does not risk-adjust for SES factors.

1b.5 Citations for data on Disparities:
N/A

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 measure will calculate 30-day all-cause hospital-level readmission rates after elective primary THA and/or TKA. The goal is to reduce readmission rates post hospitalization for elective THA/TKA. It addresses an outcome for a commonly performed, high cost procedure performed for a priority condition (osteoarthritis) and may lead to reduced morbidity and mortality.

1c.2-3. Type of Evidence: Expert opinion, Systematic synthesis of research

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Readmission is an outcome that reflects the quality of healthcare for patients undergoing a primary elective THA and/or TKA procedure. However, evidence regarding the relationship between healthcare processes (including inpatient and post-discharge care) and readmissions for this population is sparse. A systematic review of the literature did not identify any existing statistical models to compare hospital-level readmission for patients admitted for an elective THA or TKA. However, a working group and technical expert panel (TEP) of orthopedists, rheumatologists, consumer and purchaser perspective, disparities experts, and quality improvement experts were consulted in confirming that readmission is an outcome likely attributable to care processes see section 2c for details) and that hospital-level readmission rates could be improved.

Research has shown that readmission rates are influenced by the quality of inpatient and outpatient care, as well as hospital system characteristics, such as the bed capacity of the local health care system (Fisher et al. 1994). In addition, specific hospital processes such as discharge planning, medication reconciliation, and coordination of outpatient care have been shown to affect readmission rates (Nelson et al. 2000).

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
N/A

1c.6 Method for rating evidence: N/A

1c.7 Summary of Controversy/Contradictory Evidence: All-cause readmission

This measure calculates 30-day all cause readmission rate. An alternative approach would be to calculate readmissions for procedure-specific complications (e.g. mechanical complications, revision, wound infection, surgical site bleeding). In consultation with an expert panel, we decided on all-cause readmission (except for planned readmissions), rather than procedure-specific readmission for several reasons. First, from the patient perspective, readmission for any reason is likely to be an undesirable outcome of care. Second, readmissions not associated with a procedure-specific diagnosis may still be related to inpatient care and patients' transitions to non-acute setting. Examples include errors in medication reconciliation, inadequate follow-up, and failure to ensure that patients discharged home have adequate support. Third, a readmission measure will complement the complications measure for patients undergoing TKA/THA that is submitted to NQF. Using all-cause readmission will, however, undoubtedly include a mix of unavoidable and avoidable readmissions. However, the goal of the measure is not to reduce readmissions to zero, but to decrease the readmission rates across hospitals. Readmissions within 30 days after discharge from an elective procedure

1c
C☐
P☐
M☐
N☐

are likely attributable to the care received during the index admission.

Planned Readmissions

Some patients are admitted within 30 days of the index hospitalization to undergo another THA/TKA procedure. Some of these are considered planned readmissions and we do NOT count them as readmissions in the measure. If a patient undergoes a second primary THA/TKA and is admitted to the hospital within 30 days of the discharge date for the index admission, and the admission is associated with a primary discharge diagnosis of osteoarthritis, rheumatoid arthritis, osteonecrosis, and arthropathy (excluding septic arthropathy), the readmission is likely planned and is not counted as a readmission in the measure.

Use of Hierarchical Generalized Linear Modeling
Hierarchical modeling for hospital outcomes measurement is the appropriate statistical approach for hospital outcomes measures given the structure of the data and the underlying assumption of such measures, which is that hospital quality of care influences 30-day readmission rates. However, CMS frequently receives comments and questions about this approach, so we are concisely reiterating the rationale for and merits of using hierarchical logistic regression. Patients are clustered within hospitals and, as such, have a shared exposure to the hospital quality and processes. The use of hierarchical modeling accounts for the clustering of patients within hospitals. Second, hierarchical models distinguish within-hospital variation and between-hospital variation to estimate the hospital's contribution to the risk of readmission. This allows for an estimation of the hospital's influence on patient outcomes. Finally, within hierarchical models we can account for both differences in case mix and sample size to fairly profile hospital performance. If we did not use hierarchical modeling we could overestimate variation and potentially misclassify hospitals' performance. Accurately estimating variation is an important objective for models used in public reporting and potentially used in value-based purchasing programs.

1c.8 Citations for Evidence (*other than guidelines*): Fisher ES, Wennberg JE, Stukel TA, Sharp SM. Hospital readmission rates for cohorts of Medicare beneficiaries in Boston and New Haven. N Engl J Med. 1994;331(15):989-995.

Nelson EA, Maruish ME, Axler JL. Effects of discharge planning and compliance with outpatient appointments on readmission rates. Psychiatr Serv. 2000;51(7):885-889.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number*):
Not applicable-we didn't cite any clinical practice guidelines because this is an outcomes measure, not a process of care measure.

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 subcriteria 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***):** This outcome measure does not have a traditional numerator and denominator like a core process measure (e.g., percentage of adult patients with diabetes aged 18-75 years receiving one or more hemoglobin A1c tests per year); thus, we are using this field to define readmissions. The outcome for this measure is a readmission to any acute care hospital, for any reason occurring within 30 days of the discharge date of the index hospitalization. We do not count planned readmissions in the outcome (see numerator details). **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** 30 days from discharge date of index hospitalization **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** A readmission to any acute care hospital for any reason within 30 days of the discharge date of index hospitalization. Planned (elective) readmissions: We do not count readmissions in the measure that are associated with a subsequent "planned" THA/TKA procedure within 30-days of discharge from index hospitalization. Some patients may elect to stage their orthopedic replacement procedures across hospitalizations (for example, a patient may have the left and right knees replaced within one or two weeks of each other, potentially across multiple hospitalizations). In consultation with an expert panel we define planned readmissions as a second admission with an ICD-9 procedure code for THA or TKA AND a primary discharge diagnosis of osteoarthritis, rheumatoid arthritis, osteonecrosis, or arthropathy (excluding septic arthropathy). The criteria for identifying a subsequent planned THA and/or TKA is as follows: 1. Admission with at least one of the following ICD-9 procedure codes within 30 days of discharge date of index hospitalization: • 81.51 – Primary total hip replacement • 81.54 – Primary total knee replacement, AND 2. A principal diagnosis code of one the following ICD-9 codes for osteoarthritis, rheumatoid arthritis, osteonecrosis, or arthropathy: • 714, 714.0, 714.1, 714.2, 714.3, 714.30, 714.31, 714.32, 714.33, 714.4, 714.8, 714.89, 714.9, 715, 715.0, 715.00, 715.09, 715.1, 715.10, 715.15, 715.16, 715.18, 715.2, 715.20, 715.25, 715.26, 715.28, 715.3, 715.30, 715.35, 715.36, 715.38, 715.8, 715.80, 715.89, 715.9, 715.90, 715.95, 715.96, 715.98, 716.5, 716.50, 716.55, 716.56, 716.58, 716.59, 716.8, 716.80, 716.85, 716.86, 716.88, 716.89, 716.9, 716.90, 716.95, 716.96, 716.98, 716.99, 733.42, 733.43	
2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):** The target population for this measure includes admissions for patients at least 65 years of age undergoing primary THA and/or TKA procedures. **2a.5 Target population gender:** Female, Male **2a.6 Target population age range:** 65 years of age and older **2a.7 Denominator Time Window (***The time period in which cases are eligible for inclusion in the*	**2a-specs** C☐ P☐ M☐ N☐

denominator):
This measure was developed using claims data from calendar year 2007 and 2008. The time period for public reporting has not been determined.

2a.8 **Denominator Details (***All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
The denominator includes patients aged 65 and older admitted to non-federal acute care hospitals for an elective, primary THA and/or TKA in 2007 and 2008. Patients are eligible for inclusion in the denominator if they had a THA and/or a TKA AND had continuous enrollment in Medicare FFS one year prior to the date of index admission.

This cohort is defined using the following ICD-9-CM procedure codes identified in Medicare Part A Inpatient claims data:
81.51 Total Hip Arthroplasty
81.54 Total Knee Arthroplasty

2a.9 **Denominator Exclusions (***Brief text description of exclusions from the target population***):** Patients will be excluded from the cohort if they meet any of the followed criteria:

1. Patients with hip fractures
Presence of one of the following diagnosis codes: 733.1, 733.10, 733.14, 733.15, 733.19, 733.8, 733.81, 733.82, 733.95, 733.96, 733.97, 808.0, 808.1, 820.00, 820.01, 820.02, 820.03, 820.09, 820.10, 820.11, 820.12, 820.13, 820.19, 820.20, 820.21, 820.22, 820.30, 820.31, 820.32, 820.8, 820.9, 821, 821.0, 821.00, 821.01, 821.1, 821.10, 821.11
Rationale: Patients with hip fractures have higher mortality, complication and readmission rates and the procedure (THA) is generally not elective.

2. Patients undergoing revision procedures (with or without a concurrent THA/TKA)
Presence of one of the following procedure codes: 81.53, 81.55, 81.59, 00.70, 00.71, 00.72, 00.73, 00.80, 00.81, 00.82, 00.83, 00.84
Rationale: Revision procedures may be performed at a disproportionately small number of hospitals and are associated with higher mortality, complication, and readmission rates.

3. Patients undergoing partial hip arthroplasty procedures (with or without a concurrent THA/TKA)
Presence of the following procedure code: 81.52
Rationale: Partial arthroplasties are primarily done for hip fractures and are typically performed on patients who are older, more frail, and with more comorbid conditions.

4. Patients undergoing resurfacing procedures (with or without a concurrent THA/TKA)
Presence of one of the following procedure codes: 00.85, 00.86, 00.87
Rationale: Resurfacing procedures are a different type of procedure which are typically performed on younger, healthier patients.

5. Patients without at least 30-days post-discharge enrolment in Medicare
Rationale: The 30-day readmission outcome cannot be assessed for the standardized time period.

6. Patients who are transferred in to the index hospital
Rationale: If the patient is transferred from another acute care facility to the hospital where the index procedure occurs, it is likely that the procedure is not elective.

7. Patients who were admitted for the index procedure and subsequently transferred to another acute care facility
Rationale: Attribution of readmission to the index hospital would not be possible in these cases, since the index hospital performed the procedure but another hospital discharged the patient to the non-acute care setting.

8. Patients who leave against medical advice (AMA)
Rationale: Hospitals and physicians do not have the opportunity to provide the highest quality care for these patients.

9. Patients with more than two THA/TKA procedures codes during the index hospitalization
Rationale: Patients with more than two procedure codes for THA/TKA are excluded because it is rare that a patient would have 3 arthroplasty procedures done at one time. This is likely to be a coding error.

10. Patients who die during the index admission
Rationale: Patients who die during the initial hospitalization are not eligible for readmission.

Additional otherwise qualifying THA and/or TKA admissions that occurred within 30 days of discharge date of an earlier index admission are not considered as index admission. They are considered as potential readmissions. Any THA and/or TKA admission is either an index admission or a potential readmission, but not both.

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
See "Denominator Exclusion" section

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
This measure is not stratified.

2a.12-13 Risk Adjustment Type: Risk-adjustment devised specifically for this measure/condition

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
The measure estimates hospital-level 30-day all-cause RSRRs using hierarchical logistic regression models. In brief, the approach simultaneously models outcomes at two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand et al., 2007). To model the log-odds of 30-day all-cause readmission at the patient level, the model adjusts for age, sex, and selected clinical covariates. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of readmission at the hospital, after accounting for case mix. If there were no differences among hospitals, then after adjusting for case mix, the hospital intercepts should be identical across all hospitals.

The measure adjusts for key variables that are clinically relevant and have strong relationships with the outcome (e.g. demographic factors, disease severity indicators, and indicators of frailty). For each patient, covariates are obtained from Medicare claims extending 12 months prior to and including the index admission. The model adjusts for case mix differences based on the clinical status of the patient at the time of admission. We use condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis and procedure codes. We do not risk-adjust for CCs that are possible adverse events of care and that are only recorded in the index admission. In addition, only comorbidities that convey information about the patient at that time or in the 12-months prior, and not complications that arise during the course of the hospitalization are included in the risk-adjustment. The risk adjustment model included 33 variables which are listed below:

Demographics
1. Age-65 (years above 65, continuous)
2. Sex

TKA/THA Procedure
3. THA procedure
4. Number of procedures (2 vs.1)

Clinical Risk Factors
5. History of Infection (CC 1, 3-6)
6. Metastatic cancer and acute leukemia (CC 7)
7. Cancer (CC 8-12)
8. Diabetes and DM complications (CC 15-20, 119, 120)
9. Protein-calorie malnutrition (CC 21)

10. Disorders of Fluid/Electrolyte/Acid-Base (CC 22, 23)
11. Rheumatoid Arthritis and Inflammatory Connective Tissue Disease (CC 38)
12. Severe Hematological Disorders (CC 44)
13. Dementia and senility (CC 49, 50)
14. Major psychiatric disorders (CC 54-56)
15. Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
16. Polyneuropathy (CC 71)
17. Congestive Heart Failure (CC 80)
18. Chronic Atherosclerosis (CC 83-84)
19. Hypertension (CC 89, 91)
20. Arrhythmias (CC 92, 93)
21. Stroke (CC 95, 96)
22. Vascular or circulatory disease (CC 104-106)
23. COPD (CC 108)
24. Pneumonia (CC 111-113)
25. End-stage renal disease or dialysis (CC 129, 130)
26. Renal Failure (CC 131)
27. Decubitus ulcer or chronic skin ulcer (CC 148, 149)
28. Cellulitis, Local Skin Infection (CC 152)
29. Other Injuries (CC162)
30. Major Symptoms, Abnormalities (CC 166)
31. Skeletal Deformities (ICD-9 code 755.63)
32. Post Traumatic Osteoarthritis (ICD-9 codes 716.15, 716.16)
33. Morbid Obesity (ICD-9 code 278.01)

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22 (2): 206-226.

2a.15-17 Detailed risk model available Web page URL or attachment: Attachment THA-TKA Readmission Technical Report.pdf

2a.18-19 Type of Score: Rate/proportion

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***):**
The RSRR is calculated as the ratio of the number of "predicted" to the number of "expected" readmissions, multiplied by the national unadjusted readmission rate. For each hospital, the "numerator" of the ratio is the number of readmissions within 30 days predicted on the basis of the hospital's performance with its observed case mix, and the "denominator" is the number of readmissions expected on the basis of the nation's performance with that hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital's performance given its case-mix to an average hospital's performance with the same case-mix. Thus a lower ratio indicates lower-than-expected readmission or better quality and a higher ratio indicates higher-than-expected readmission or worse quality.

The predicted hospital outcome (the numerator) is calculated by regressing the risk factors and the hospital-specific intercept on the risk of readmission, multiplying the estimated regression coefficients by the patient characteristics in the hospital, transforming, and then summing over all patients attributed to the hospital to get a value. The expected number of readmissions (the denominator) is obtained by regressing the risk factors and a common intercept on the readmission outcome using all hospitals in our sample, multiplying the subsequent estimated regression coefficients by the patient characteristics observed in the hospital, transforming, and then summing over all patients in the hospital to get a value.

Please see attachment for more details on the calculation algorithm.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
The method for discriminating hospital performance has not been determined. For the six publicly reported measures of hospital outcomes developed with similar methodology and reported on the CMS website www.hospitalcompare.hhs.gov, CMS currently estimates an interval estimate for each risk-standardized rate to characterize the amount of uncertainty associated with the rate, compares the interval estimate to the

national crude rate for the outcome, and categorizes hospitals as "better than the US national rate," "worse than the US national rate," or "no different than the US national rate." However, the decision to publicly report this measure and the approach has not been determined.

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)*:
This measure is not based on a survey or sample.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Electronic administrative data/claims

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.***):**
We obtained index admission, readmission, and in-hospital comorbidity data from Medicare's Standard Analytic File (SAF). Comorbidities were also assessed using Part A inpatient, outpatient, and Part B office visit Medicare claims in the 12 months prior to index admission. Enrollment and post-discharge mortality status were obtained from Medicare's enrollment database which contains beneficiary demographic, benefit/coverage, and vital status information.

1. 2008 Part A (inpatient) data
Part A inpatient data includes claims for Medicare inpatient hospital care, skilled nursing facility care, some home health agency services, and hospice care. For purposes of this project, Part A is used to refer to inpatient services only and includes data from 2 time periods:
a. Index admission: Index admission data are based on the inclusion/exclusion criteria for THA/TKA, and comorbidities (if any) are identified from the secondary diagnoses associated with the index admission.
b. Pre-index: 12 months prior to the index admission ("pre-index").

2. 2008 Part A (outpatient) data – 12 months pre-index
Hospital outpatient refers to Medicare claims paid for the facility component of surgical or diagnostic procedures, emergency room care, and other non-inpatient services performed in a hospital outpatient department or ambulatory surgical/diagnostic center.

3. Part B data – 12 months pre-index
Part B data refers to Medicare claims for the services of physicians (regardless of setting) and other outpatient care, services, and supplies. For purposes of this project, Part B services included only face-to-face encounters between a care provider and patient. We thus do not include services such as laboratory tests, medical supplies, or other ambulatory services.

2a.26-28 Data source/data collection instrument reference web page URL or attachment:

2a.29-31 Data dictionary/code table web page URL or attachment: URL N/A
http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1182785083979

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency

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

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)***:** Medicare Part A inpatient claims data for calendar year 2007 and 2008 were used to test reliability. The 2008 cohort included 296,224 admissions and the 2007 cohort included 300,338 admissions.

2b
C☐
P☐
M☐
N☐

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
The reliability of the model was tested using identical cohort inclusion/exclusion criteria for patients who underwent THA and/or TKA. We randomly selected 50% of the THA and/or TKA admissions that met all inclusion and exclusion criteria in 2008 and created a development sample, which we used to build the model. We used the remaining 50% of THA/TKA admissions in 2008 as the validation sample. We also used all qualifying THA and/or TKA admissions in 2007 data as an additional sample to validate the model. Model performance was assessed in the development dataset and both validation datasets. In addition we will run the model in additional datasets and compare the risk-standardized readmission rates for each hospital.

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Preliminary results indicate similar model performance in the three cohorts (e.g., ROC=0.64 in all models). See additional results for these cohorts in the "testing results" section below.

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: Face validity: model performance.

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
During measure development, we consulted with representatives from potential users of this measure including clinicians, professional societies, payers, and consumers. We use this field to describe the role that these representatives played on the working group and Technical Expert Panel (TEP). We used a structured measure evaluation tool to assess face validity and other measure properties.

We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcomes measurement set forth in National Quality Forum (NQF) guidance for outcomes measures (National Quality Forum, 2010), CMS Measure Management System guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006). We obtained expert and stakeholder input on the measure through three mechanisms: first, through regular discussions with a working group; second, through a series of three conference calls with a national Technical Expert Panel (TEP); and third, through a public comment period.

Early in the development phase, we assembled a working group that included individuals with clinical and methodological expertise relevant to orthopedic quality measurement. We held regular conference calls throughout the development process, and the Yale team solicited detailed feedback and guidance on key clinical and methodological decisions pertaining to measure development. The working group provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader TEP.

In alignment with CMS' Measure Management System, YNHHSC/CORE also released a public call for nominations and convened a TEP. Potential members were also solicited via e-mail in consultation with the working group and CMS. The role of the TEP was to provide feedback on key methodological decisions made in consultation with the working group. The TEP was comprised of individuals with diverse perspectives and backgrounds including clinicians, consumers, hospitals, purchasers, and experts in quality improvement. Finally, we solicited public comment on the proposed measure through CMS' Measure Management System Public Comment site (https://www.cms.gov/MMS/17_CallforPublicComment.asp#TopOfPage). Public comments were summarized and publicly posted for 30 days. The resulting content was taken into consideration during the final stages of measure development.

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.nysna.org/images/pdfs/practice/nqf_ana_outcomes_draft10.pdf. Accessed August 19, 2010.

Krumholz HM,Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and

2c
C☐
P☐
M☐
N☐

the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation. January 24, 2006 2006;113(3):456-462. **2c.3 Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*: The experts agree the measure accurately reflects the quality of care and distinguishes levels of quality for patients undergoing THA and/or TKA.	
2d. Exclusions Justified **2d.1 Summary of Evidence supporting exclusion(s):** Rationale for exclusion is described in "Denominator Exclusions." **2d.2 Citations for Evidence:** See "Denominator Exclusions" **2d.3 Data/sample** *(description of data/sample and size)*: N/A **2d.4 Analytic Method** *(type analysis & rationale)*: N/A **2d.5 Testing Results** *(e.g., frequency, variability, sensitivity analyses)*: N/A	**2d** C☐ P☐ M☐ N☐ NA ☐
2e. Risk Adjustment for Outcomes/ Resource Use Measures **2e.1 Data/sample** *(description of data/sample and size)*: 2008 Medicare Part A inpatient and outpatient data and Part B outpatient data are used to identify candidate variables for risk adjustment. Specifically, Medicare Part A inpatient data is used to identify variables for risk adjustment in the index admission. Part A outpatient and Part B data are used to identify comorbid conditions to include in the risk adjustment in the 12-month period preceding the index date of admission. As described in section 2b, we developed and validated the model in three separate cohorts to assess and compare model performance: (1) development sample of 148,132 admissions in 2008 data; (2) validation sample of 148,092 in 2008 data; and (3) validation sample of 300,338 admissions in 2007 data. **2e.2 Analytic Method** *(type of risk adjustment, analysis, & rationale)*: This measure is fully risk-adjusted using a hierarchical logistic regression model to calculate hospital RSRRs. (see "risk adjustment methodology" for additional details). Approach to assessing model performance: For the development and validation cohorts, we computed five summary statistics for assessing model performance (Harrell, 2001): (1) over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients) (2) predictive ability (3) area under the receiver operating characteristic (ROC) curve (4) distribution of residuals (5) model chi-square (A test of statistical significance usually employed for categorical data to determine whether there is a good fit between the observed data and expected values; i.e., whether the differences between observed and expected values are attributable to true differences in characteristics or instead the result of chance variation. F.E. Harrell and Y.C.T. Shih, Using full probability models to compute probabilities of actual interest to	**2e** C☐ P☐ M☐ N☐ NA ☐

decision makers, Int. J. Technol. Assess. Health Care 17 (2001), pp. 17–26.

2e.3 Testing Results *(risk model performance metrics)*:
Performance Metrics in Development Cohort: Development cohort consisted of 148,132 patient stays at 3,223 hospitals (half of 2008 cohort), with a risk-adjusted median readmission rate of 6.04%. The development model has strong discrimination and fit. The risk-standardized readmission rate ranges from 3.2% to 46.8%, a range of 43.6%. Results are summarized below:

Over-fitting indices: (0,1)
Residuals lack of fit: <-2 = 0.0%; [-2, 0) = 93.8%; [0, 2) = 0.1%; [2+ = 6.0%
Model Chi-square [# of covariates]: 2492 [33]
Predictive ability (lowest decile %, highest decile %): (2.4, 13.4)
Area under the ROC curve = 0.65 (GLM)

The discrimination and the explained variation of the model are consistent with those of models currently used to publicly report condition specific rates of both mortality and readmission.

Model Validation using 2008 Validation Cohort: 2008 Validation cohort consisted of 148,092 admissions (other half of the 2008 cohort) randomly selected from 3,213 hospitals, with a risk-standardized median readmission rate of 6.02%. The model performance was not substantively different in this validation sample, as compared to the development sample. Results are summarized below:

Over-fitting indices: (-0.06, 0.98)
Residuals lack of fit: <-2 = 0.0%; [-2, 0) = 93.8%; [0, 2) = 0.1%; [2+ = 6.0%
Model Chi-square[# of covariates]: 2406 [33]
Predictive ability (lowest decile %, highest decile %):(2.6, 13.2)
Area under the ROC curve = 0.64

Model Validation using 2007 Validation Cohort: 2007 validation cohort consisted of 300,338 admissions from 3,295 hospitals. The model performance was not substantively different in this validation sample, as compared to the development sample. Results are summarized below:

Over-fitting indices: (-0.11, 0.94)
Residuals lack of fit: <-2 = 0.0%; [-2, 0) = 93.6%; [0, 2) = 0.1%; [2+ = 6.2%
Model Chi-square[# of covariates]: 4596 [33]
Predictive ability (lowest decile %, highest decile %):(2.8, 13.4)
Area under the ROC curve = 0.64

We also examined the temporal variation of the standardized estimates and frequencies of the variables in the models. The frequencies and regression coefficients are fairly consistent over the three cohorts.

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

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: 2008 Medicare Part A inpatient claims data

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
Unadjusted median hospital-level readmission rates following THA and/or TKA were assessed across hospitals.

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)*:
Median hospital-level risk-standardized readmission rate of 2008 was 6.06% and ranged from 3.06% to 50.94%. This is likely a signal of differences in the quality of care received for patients undergoing THA and/or TKA. Total hip replacement and TKA are elective procedures typically performed on healthy patients.

2f
C☐
P☐
M☐
N☐

Therefore, readmission rates are expected to be lower than that for an emergent procedures and conditions. The variation observed for readmissions is likely a signal that though rates may be relatively low there are differences in the quality of care delivered across hospitals that result in variation in outcomes.	
2g. Comparability of Multiple Data Sources/Methods 2g.1 **Data/sample** *(description of data/sample and size)*: No comparable data source is available at this time. We will perform validity testing of the development model in data from a different time frame. 2g.2 **Analytic Method** *(type of analysis & rationale)*: N/A 2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)*: N/A	**2g** C☐ P☐ M☐ N☐ NA ☐
2h. Disparities in Care 2h.1 **If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: This measure is not stratified. 2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** There were no hospital-level disparities detected during measure development. Please see "Summary of Data on Disparities by Population Group" for additional information.	**2h** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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:** Not in use but testing completed 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*): CMS plans to use the measures for public reporting and will propose the measures through rulemaking process. 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*): The measure is not currently in use. **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)*: N/A 3a.5 **Methods** *(e.g., focus group, survey, QI project)*: No consumer or other field testing has been completed at this time. However, this measure was systematically evaluated by an expert group of orthopedists and a TEP over a period of eight months. Regular meetings were held throughout the development of this measure, during which we received input and	**3a** C☐ P☐ M☐ N☐

feedback on key methodological and other measure decisions (see section 2c-Validity Testing for more details on process of TEP input). 3a.6 **Results** *(qualitative and/or quantitative results and conclusions)*: The TEP agreed that the measure would be useful in informing consumers and hospitals.	
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:** 5.1 **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:** N/A	**3c** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**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*) Yes 4b.2 **If not, specify the near-term path to achieve electronic capture by most providers.**	**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** C☐ P☐ M☐ N☐

4c.2 If yes, provide justification.	**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.** Using administrative claims variables for risk adjustment This measure uses variables from claims data submitted by hospitals to CMS for payment as "clinical" risk adjusters. Prior research has demonstrated that administrative claims data can be used to develop risk-adjusted outcomes measures for mortality following admission for myocardial infarction, heart failure, and death and that the models produce estimates of risk-standardized mortality rates (RSMRs) that are very similar to rates estimated by models based on chart data. This high level of agreement between the clinical and claims-based models supports the use of the claims-based models for public reporting. The models have also demonstrated consistent performance across years of claims data. The approach to identifying risk factors for patients also mitigates the potential limitations of claims data. Because not every diagnosis is coded at every visit, we identified comorbid conditions for risk adjustment in inpatient, outpatient, and physician claims data coded in the year prior to admission, as well as those coded in the secondary diagnosis fields during the index admission. This strategy allows for comprehensive review of patients' medical histories. If a diagnosis appeared only once, in some visits and not others, it was included, minimizing the effect of incomplete coding. Conditions that may represent adverse outcomes due to care received during the index admission are not considered for inclusion in the risk-adjusted model. Although they may increase the risk of readmission, including them as covariate in a risk-adjusted model could attenuate the measure's ability to characterize the quality of care delivered by hospitals Potentially creating access barrier These are elective procedures, and therefore publicly reporting this measure could reduce access to care for certain populations, particularly for patients who may be healthy enough to undergo the procedure but who carry a higher risk for readmission. We do not anticipate this; however, we recommend close monitoring of any unexpected consequences, once the measure is implemented.	**4d** **C**☐ **P**☐ **M**☐ **N**☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** N/A **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** This measure uses claims data submitted by hospitals to CMS for payment, There are no costs associated with data collection, as hospitals are mandated by CMS to submit claims for reimbursement purposes. There is no additional cost/burden on hospitals. **4e.3 Evidence for costs:** N/A **4e.4 Business case documentation:** Key points as noted in various sections of this document are as follows: 1. The median 30-day all-cause risk-standardized readmission rate is high (6.06%) 2. There is substantial variation in risk-standardized readmission rates across hospitals, ranging from 3.06% to 50.94%. 3. Reducing readmission is a key focus of the health care reform bill. 4. Quality of care should be addressed as THA and TKA procedures are associated with high volume and cost (relative to other elective procedures performed in the Medicare population).	**4e** **C**☐ **P**☐ **M**☐ **N**☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria 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
Centers for Medicare & Medicaid Services (CMS), 7500 Security Blvd, Mail Stop S3-02-01, Baltimore, Maryland, 21244-9045

Co.2 Point of Contact
Lein, Han, PhD, lein.han@cms.hhs.gov, 410-786-0205-

Measure Developer If different from Measure Steward
Co.3 Organization
New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE), 1 Church Street Suite 200, New Haven, Connecticut, 06510

Co.4 Point of Contact
Laura, Grosso, PhD, laura.grosso@yale.edu, 203-764-5742-

Co.5 Submitter If different from Measure Steward POC
Laura, Grosso, PhD, laura.grosso@yale.edu, 203-764-5742-, New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE)

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.
Working Group

Role: To provide technical expertise on key methodological decisions during measure development.

Members:

Daniel J. Berry, MD
Professor of Orthopedics, Mayo Clinic College of Medicine
Chair, Department of Orthopædic Surgery, Mayo Clinic

Kevin J. Bozic, MD, MBA
Associate Professor and vice chair, Department of Orthopaedic Surgery at the University of California, San Francisco
Chair, Health Systems Committee, AAOS

Robert Bucholz, MD

Professor, Orthopædic Surgery, University of Texas Southwestern Medical Center
Past President, American Academy of Orthopædic Surgeons (AAOS)

Lisa Gale Suter, MD
Assistant Professor, Yale University School of Medicine, Rheumatology (West Haven Veterans Association Hospital)

Charles M. Turkelson, PhD
Director of Research and Scientific Affairs, AAOS

Lawrence Weis, MD
Assistant Professor, Yale Orthopædics and Rehabilitation, Yale University School of Medicine, Orthopædics (West Haven Veterans Association Hospital)

Technical Expert Panel

Role: To provide feedback on recommendations for measure development.

Members:

Mark L. Francis, MD
Professor of Medicine and Biomedical Sciences, Chief, Division of Rheumatology, Department of Internal Medicine, Texas Tech University Health Sciences Center

Cynthia Jacelon, PhD, RN, CRRN
Associate Professor, School of Nursing, University of Massachusetts
Association of Rehabilitation Nurses

Norman Johanson, MD
Chairman, Orthopedic Surgery, Drexel University College of Medicine

C. Kent Kwoh, MD
Professor of Medicine, University of Pittsburgh
Associate Chief and Director of Clinical Research,
Division of Rheumatology and Clinical Immunology

Courtland G. Lewis, MD
American Association of Orthopaedic Surgeons

Jay Lieberman, MD
Professor and Chairman, Department of Orthopedic Surgery, University of Connecticut Health Center,
Director, New England Musculoskeletal Institute

Peter Lindenauer, MD, M.Sc.
Hospitalist and Health Services Researcher, Baystate Medical Center
Professor of Medicine, Tufts University

Russell Robbins, MD, MBA
Principal, Mercer´s Total Health Management

Barbara Schaffer
THA Patient

Nelson SooHoo, MD, MPH
Professor, University of California at Los Angeles

Steven H. Stern, MD
Vice President, Cardiology & Orthopedics/ Neuroscience, UnitedHealthcare

Richard E. White, Jr., MD American Association of Hip and Knee Surgeons
Ad.2 If adapted, provide name of original measure: **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 Readmission calculation algorithm.pdf
Date of Submission (*MM/DD/YY*)**:** 03/28/2011

THA/TKA Readmission Calculation Algorithm

We estimate a generalized linear model and a hierarchical generalized linear model which accounts for the clustering of observations within hospitals. The generalized linear model (GLM) links the outcome to the patient-level risk factors,[20] Let Y_{ij} denote the outcome (equal to 1 if patient has a readmission, zero otherwise) for the j^{th} patient who had a THA/TKA procedure at the i^{th} hospital; $\mathbf{Z}_{ij}$ denotes a set of risk factors based on the data. Let I denote the total number of hospitals and n_i the number of index patient stays in hospital i. We assume the outcome is related linearly to the covariates via a known linked function, h, where

GLM $h(Y_{ij}) = \alpha + \beta\mathbf{Z}_{ij}$ (1)

and $\mathbf{Z}_{ij} = (Z_{1ij}, Z_{2ij}, \ldots, Z_{pij})$ is a set of p patient-specific covariates. In our case, h = the logit link.

To account for the natural clustering of observations within hospitals, we then estimate an HGLM that links the risk factors to the same outcome and a hospital-specific random effect,

HGLM $h(Y_{ij}) = \alpha_i + \beta\mathbf{Z}_{ij}$ (2)

$\alpha_i = \mu + \omega_i;\ \omega_i \sim N(0, \tau^2)$ (3)

where α_i represents the hospital-specific intercept, $\mathbf{Z}_{ij}$ is defined as above, μ the adjusted average outcome over all hospitals in the sample, and τ^2 the between-hospital variance component.[21] This model separates within-hospital variation from between-hospital variation. Both HGLMs and GLMs are estimated using the SAS software system (GLIMMIX and LOGISTIC procedures, respectively).

We first fit the GLM described in Equation (1) using the logit link.

Having identified the covariates that were selected, we next fit the HGLM described in Equations (2) and (3), again using the logit link function; e.g.,

Logit $(P(Y_{ij} = 1)) = \alpha_i + \beta\ \mathbf{Z}_{ij}$

$\alpha_i = \mu + \omega_i,\ \omega_i \sim N(0, \tau^2)$

where $\mathbf{Z}_{ij}$ consisted of the covariates retained in the GLM model. As before, $Y_{ij} = 1$ if patient j treated at hospital i had the event; 0 otherwise.

Hospital performance reporting

Using the set of risk factors in the GLM, we fit the HGLM defined by Equations (2) - (3) and estimate the parameters, $\hat{\mu}$, $\{\hat{\alpha}_i, \hat{\alpha}_2, \ldots, \hat{\alpha}_I\}$, $\hat{\beta}$, and $\hat{\tau}^2$. We calculate a standardized outcome, s_i, for each hospital by computing the ratio of the number of predicted readmissions to the number of expected readmissions, multiplied by the unadjusted overall readmission rate, $\bar{y}$. Specifically, we calculate

Predicted $\hat{y}_{ij}(Z) = h^{-1}(\hat{\alpha}_i + \hat{\beta}\,\mathbf{Z}_{ij})$ (4)

Expected $\hat{e}_{ij}(Z) = h^{-1}(\hat{\mu} + \hat{\beta}\,\mathbf{Z}_{ij})$ (5)

$$\hat{s}_i(Z) = \frac{\sum_{j=1}^{n_i} \hat{y}_{ij}(Z)}{\sum_{j=1}^{n_i} \hat{e}_{ij}(Z)} \times \bar{y} \qquad (6)$$

If more (fewer) "predicted" cases than "expected" cases have the outcome in a hospital, then $\hat{s}_i$ will be higher (lower) than the unadjusted average. For each hospital, we compute an interval estimate of s_i to characterize the level of uncertainty around the point estimate using bootstrapping simulations. The point estimate and interval estimate can be used to characterize and compare hospital performance (e.g., higher than expected, as expected, or lower than expected).

Creating Interval Estimates

Because the statistic described in Equation (6) is a complex function of parameter estimates, we use re-sampling and simulation techniques to derive an interval estimate. The bootstrapping simulation has the advantage of avoiding unnecessary distributional assumptions.

Calculation Algorithm

Let *I* denote the total number of hospitals in the sample. We repeat steps 1 – 4 below for b = 1,2,…B times:

1. Sample *I* hospitals with replacement.

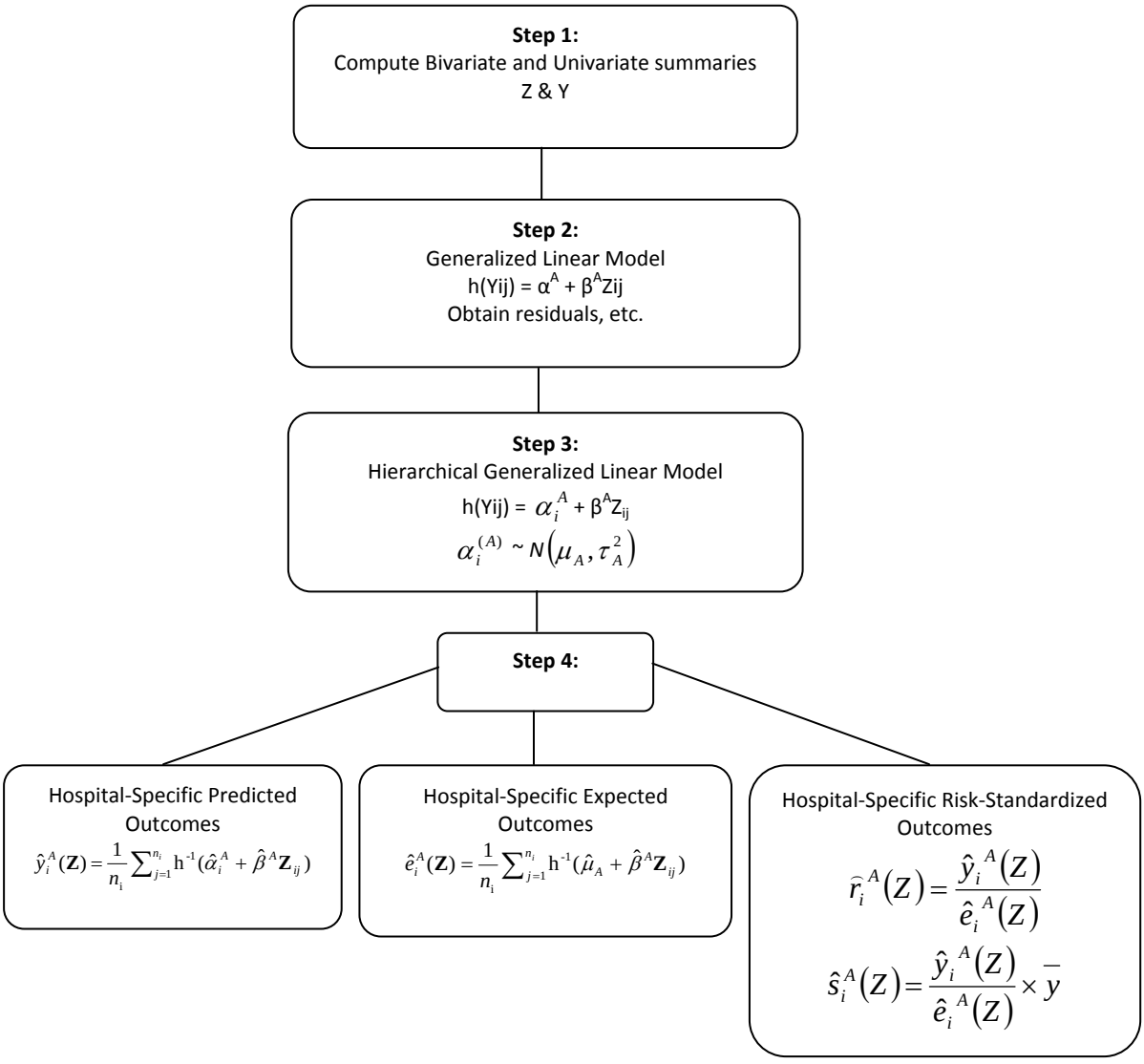
2. Fit the HGLM using all patients within each sampled hospital. We use as starting values the parameter estimates obtained by fitting the model to all hospitals. If some hospitals are selected more than once in a bootstrapped sample, we treat them as distinct so that we have *I* random effects to estimate the variance components. At the conclusion of Step 2, we have:
    a. $\hat{\beta}^{(b)}$ (the estimated regression coefficients of the risk factors).
    b. The parameters governing the random effects, hospital adjusted outcomes, distribution, $\hat{\mu}^{(b)}$ and $\hat{\tau}^{2(b)}$.
    c. The set of hospital-specific intercepts and corresponding variances, $\{\hat{\alpha}_i^{(b)}, \hat{\mathrm{var}}(\alpha_i^{(b)}); i = 1,2,\ldots,I\}$.

3. We generate a hospital random effect by sampling from the distribution of the hospital-specific distribution obtained in Step 2c. We approximate the distribution for each random effect by a normal distribution. Thus, we draw $\alpha_i^{(b*)} \sim N\left(\hat{\alpha}_i^{(b)}, \hat{\mathrm{var}}\left(\hat{\alpha}_i^{(b)}\right)\right)$ for the unique set of hospitals sampled in Step 1.

4. Within each unique hospital *i* sampled in Step 1, and for each case *j* in that hospital, we calculate $\hat{y}_{ij}^{(b)}$, $\hat{e}_{ij}^{(b)}$, and $\hat{s}_i(Z)^{(b)}$ where $\hat{\beta}^{(b)}$ and $\hat{\mu}^{(b)}$ are obtained from Step 2 and $\hat{\alpha}_i^{(b*)}$ is obtained from Step 3.

Ninety-five percent interval estimates (or alternative interval estimates) for the hospital-standardized outcome can be computed by identifying the 2.5th and 97.5th percentiles of the B estimates (or the percentiles corresponding to the alternative desired intervals).

Figure 1. Analysis Steps



Hospital-level 30-Day All-Cause Risk-Standardized Readmission Following Elective Total Hip Arthroplasty (THA) and Total Knee Arthroplasty (TKA)

Measure Methodology Report

Submitted By Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE):

Laura M. Grosso, PhD, MPH
Jeptha P. Curtis, MD
Zhenqiu Lin, PhD
Lori L. Geary, MPH
Smitha Vellanky, MSc
Carol Oladele, MPH
Yongfei Wang, MS
Elizabeth E. Drye, MD, SM
Harlan M. Krumholz, MD, SM

Contract # HHSM-500-2008-0025I-MIDS Task Order T0001

Prepared For:

Centers for Medicare & Medicaid Services (CMS)

Submitted September 29, 2010

ACKNOWLEDGEMENTS

We would like to acknowledge the ongoing support from our working group. These individuals gave generously of their time, providing guidance on key clinical and methodological decisions. Without their expertise, this project would not have been possible. Working group members include:

Daniel J. Berry, MD
Professor of Orthopedics, Mayo Clinic College of Medicine
Chair, Department of Orthopædic Surgery, Mayo Clinic

Kevin J. Bozic, MD, MBA
Associate Professor and vice chair, Department of Orthopaedic Surgery at the University of California, San Francisco
Chair, Health Systems Committee, AAOS

Robert Bucholz, MD
Professor, Orthopædic Surgery, University of Texas Southwestern Medical Center
Past President, American Academy of Orthopædic Surgeons (AAOS)

Lisa Gale Suter, MD
Assistant Professor, Yale University School of Medicine, Rheumatology

Charles M. Turkelson, PhD
Director of Research and Scientific Affairs, AAOS

Lawrence Weis, MD
Assistant Professor, Yale Orthopædics and Rehabilitation, Yale University School of Medicine, Orthopædics (West Haven Veterans Association Hospital)

We would also like to thank the Technical Expert Panel members, who provided great support and insight on key measure decisions. These members include:

Mark L. Francis, MD
Professor of Medicine and Biomedical Sciences, Chief, Division of Rheumatology, Department of Internal Medicine, Texas Tech University Health Sciences CenterTexas Tech University, Health Sciences Center

Cynthia Jacelon, PhD, RN, CRRN
Associate Professor, School of Nursing, University of Massachusetts
Association of Rehabilitation Nurses

Norman Johanson, MD
Chairman, Orthopedic Surgery, Drexel University College of Medicine

C. Kent Kwoh, MD

Professor of Medicine, Associate Chief and Director of Clinical Research, Division of Rheumatology and Clinical Immunology University of Pittsburgh

Courtland G. Lewis, MD
American Association of Orthopaedic Surgeons

Jay Lieberman, MD
Professor and Chairman, Department of Orthopedic Surgery, University of Connecticut Health Center; Director, New England Musculoskeletal Institute

Peter Lindenauer, MD, M.Sc.
Hospitalist and Health Services Researcher, Baystate Medical Center; Professor of Medicine, Tufts University

Russell Robbins, MD, MBA
Principal, Mercer's Total Health Management

Barbara Schaffer
THA Patient

Nelson SooHoo, MD, MPH
Professor, University of California at Los Angeles

Steven H. Stern, MD
Vice President, Cardiology & Orthopedics/ Neuroscience, UnitedHealthcare

Richard E. White, Jr., MD
American Association of Hip and Knee Surgeons

In addition, we would like to acknowledge and thank Angela Merrill, Sandi Nelson, Marian Wrobel, and Eric Schone from Mathematica Policy Research, Inc. and Lein Han and Michael Rapp at the Centers for Medicare & Medicaid Services for their contributions to this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Overview of Measure

Total hip and knee arthroplasties (THA and TKA, respectively) are priority areas for outcomes measure development, as they are commonly performed procedures that improve quality of life. In 2003 there were 202,500 THAs and 402,100 TKAs performed[1] and the number of procedures performed has increased steadily over the past decade.[2-3]

Although these procedures dramatically improve quality of life, they are costly. In 2005 annual hospital charges totaled $3.95 billion and $7.42 billion for primary THA and TKA, respectively.[2] These costs are projected to increase by 340% to 17.4 billion for THA and by 450% to 40.8 billion for TKA by 2015.[2] Medicare is the single largest payer for these procedures, covering approximately two-thirds of all THAs and TKAs performed in the US.[3] Combined, THA and TKA procedures account for the largest procedural cost in the Medicare budget.[4]

Given the high volume and cost associated with these procedures (relative to other elective procedures performed in the Medicare population), it is imperative to address quality of care. Readmissions increase costs associated with THA and TKA and affect the quality, and potentially quantity, of life for patients. A quality measure to address readmission following THA and TKA provides an opportunity to provide targets for efforts to improve the quality of care and reduce costs for patients undergoing these elective procedures.

CMS contracted with Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE) to develop hospital outcomes measures that reflect the quality of care for patients undergoing elective THA and TKA procedures and are suitable for public reporting. YNHHS/CORE, in consultation with CMS, developed a model that estimates hospital-specific, risk-standardized, 30-day all-cause readmission rates following elective THA and TKA procedures. The goal of this readmission measure is to improve the quality of care delivered to patients undergoing THA and TKA procedures.

This report provides the background and detailed technical information on the measure. In brief, we developed a model that estimates hospital-specific, risk-standardized, 30-day all-cause readmission rates following THA/TKA. We used Medicare claims data and linked them to CMS claims and enrollment data to identify readmissions within 30 days from the discharge date of the index admission. To account for the clustering of observations within hospitals and

differences in the number of admissions across hospitals, we used hierarchical logistic regression to estimate the risk-standardized readmission rates (RSRRs).

This measure was developed concurrently with a second CMS outcomes measure-hospital risk-standardized complications for THA/TKA procedures. These are complementary measures that assess separate domains of quality. The complications measure will inform quality improvement efforts targeted toward minimizing medical and surgical complications during surgery and in the recovery phase. The readmission measure captures an additional domain of care provided in the transitions to outpatient settings. The complications measure is presented in a separate technical report.

These two measures expand a set of hospital outcomes measures CMS has developed to improve hospital quality and meet its mandate under the Deficit Reduction Act (DRA) of 2005 to publicly report outcomes and efficiency measures on the consumer Web site, Hospital Compare (http://www.hospitalcompare.hhs.gov). CMS began publicly reporting acute myocardial infarction (AMI) and heart failure (HF) 30-day mortality measures as outcomes measures in June 2007, and added a pneumonia 30-day mortality measure in August 2008. In addition, CMS began publicly reporting 30-day readmission measures for AMI, HF, and pneumonia in July 2009.

1.2 Approach to Measure Development

We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcomes measurement set forth in National Quality Forum (NQF) guidance for outcomes measures,[5] CMS Measure Management System guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes."[6] We obtained expert and stakeholder input on the measure through three mechanisms: first, through regular discussions with a working group; second, through a series of three conference calls with a national Technical Expert Panel (TEP); and third, through a public comment period.

Early in the development phase, we assembled a working group that included individuals with clinical and methodological expertise relevant to orthopedic quality measurement. We held regular conference calls throughout the development process and the Yale team solicited detailed feedback and guidance on key clinical and methodological decisions pertaining to measure development. The working group provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader TEP.

In alignment with CMS' Measure Management System, YNHHSC/CORE also released a public call for nominations and convened a TEP. Potential members were also solicited via e-mail in consultation with the working group and CMS. The role of the TEP was to provide feedback on key methodological decisions made in consultation with the working group. The TEP was comprised of individuals with diverse perspectives and backgrounds including clinicians, consumers, hospitals, purchasers, and experts in quality improvement. Finally, we solicited public comment on the proposed measure through CMS' Measure Management System Public Comment site (https://www.cms.gov/MMS/17_CallforPublicComment.asp#TopOfPage). Public comments were summarized and publicly posted for 30 days. The resulting content was taken into consideration during the final stages of measure development.

1.3 Importance of a Readmission Measure

THA and TKA are priority areas for outcomes measure development, as they are costly and commonly performed procedures. Hospital readmission is an outcome that is likely attributable to care processes and is an important outcome for patients. Measuring and reporting readmission rates will inform health care providers about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices. Furthermore, the measure will increase transparency for consumers and has the potential to lower health care costs associated with readmissions.

Research has shown that readmission rates are influenced by the quality of inpatient and outpatient care, as well as hospital system characteristics, such as the bed capacity of the local health care system.[7] In addition, specific hospital processes such as discharge planning, medication reconciliation, and coordination of outpatient care have been shown to affect readmission rates.[8]

Preliminary analyses using 2008 Medicare Part A inpatient claims indicate that readmission rates post THA and TKA are high for elective procedures and vary across hospitals. Preliminary analyses indicated the median 30-day risk-standardized hospital readmission rate was 6.1%. This rate is high given these are elective procedures typically performed on healthier Medicare patients. Additionally, results demonstrated that the rates varied across hospitals (5^{th} percentile, 4.6%; 95^{th} percentile, 8.4%) indicating there is room for quality improvement.

2. METHODS

2.1 Overview

We developed a hospital-level 30-day, all-cause risk-standardized measure of readmission to any acute care hospital following THA/TKA. We developed this model for all inpatient admissions with a primary elective THA and/or TKA using hierarchical generalized linear modeling (HGLM), to account for the clustering of patients within hospitals. To adjust for differences in hospital case mix, the model adjusted for patient risk factors, including age and comorbidities present at the time of admission. A detailed description of the risk-adjustment variables and the measure methodology is in Sections 2.6 and 2.7.

We identified index admissions for inclusion in the measure via ICD-9 procedure codes for THA and TKA in 2008 Medicare Part A inpatient claims. Because there are no dates associated with procedure codes in Part A data, we use the date of the index admission as the starting point for all follow-up. We used Medicare Part A data for years 2008 and 2009 to identify readmissions. We identified Information on comorbid conditions for risk adjustment using ICD-9 codes in inpatient, outpatient, and part B Medicare claims data in the 12 months prior to the date of the index admission.

The measure calculates the hospital risk-standardized readmission rate (RSRR) by producing a ratio of the number of "predicted" to the number of "expected" readmissions for each hospital and then multiplying the ratio by the national raw readmission rate. For each hospital, the "numerator" of the ratio is the number of readmissions predicted on the basis of the hospital's performance with its observed case mix (using an estimated hospital-specific intercept term), and the "denominator" is the number of expected readmissions, based on the nation's performance using the hospital's observed case mix and the national intercept term.

The model estimates the hospital-specific intercept term used in the numerator based on how well each hospital performs relative to other hospitals with a similar case mix. Among hospitals with similar case mixes, hospitals that have a lower rate of readmission will have a lower intercept term; hospitals that have a higher rate of readmission will have a higher intercept term.

2.2 Data Sources

We obtained index admission, readmission, and in-hospital comorbidity data from Medicare's Standard Analytic File (SAF). Comorbidities were also assessed using Part A inpatient, outpatient, and Part B office visit Medicare claims in the 12 months prior to index admission. Enrollment and post-discharge mortality

status were obtained from Medicare's enrollment database which contains beneficiary demographic, benefit/coverage, and vital status information.

2.3 Outcome Definition

The outcome for this measure is 30-day all-cause readmission. We define a readmission as a subsequent acute care hospital inpatient admission within 30 days of the discharge date for the index admission.

2.3.1 Planned Readmissions

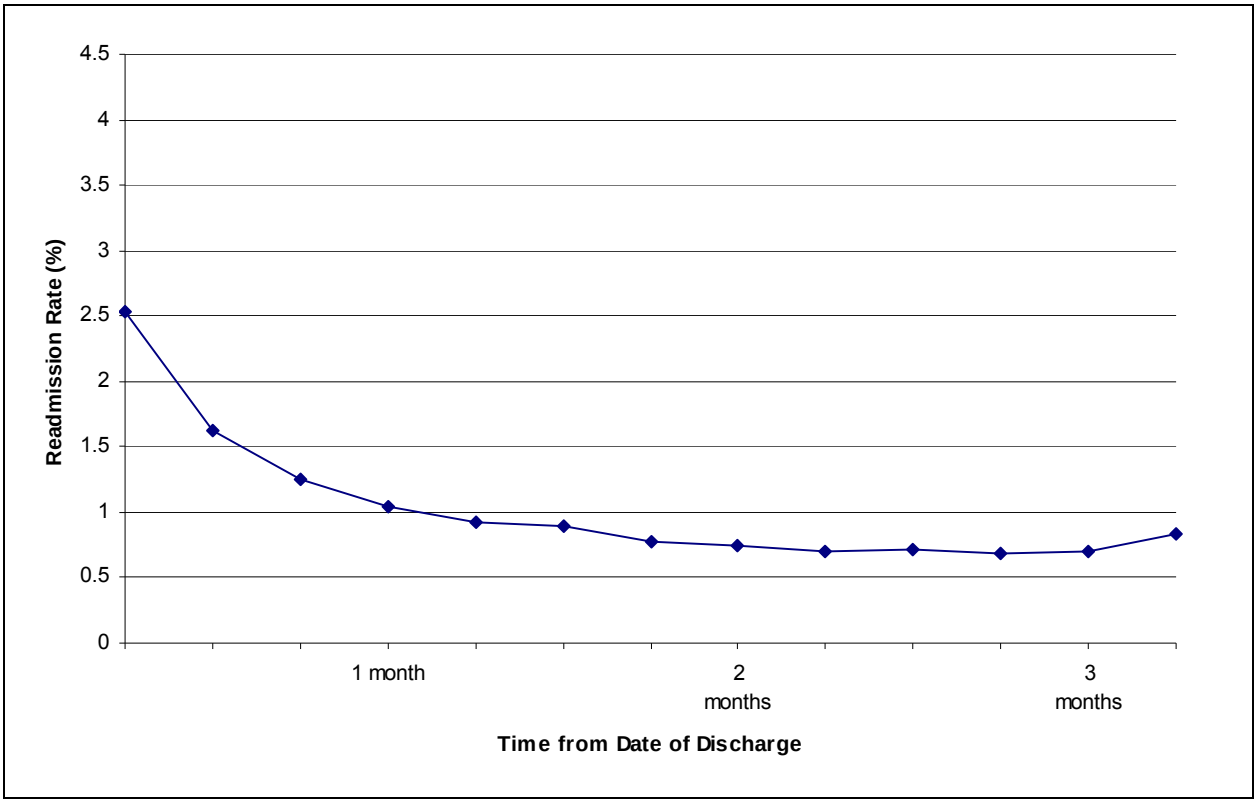
Some patients are admitted within 30 days of the index hospitalization to undergo another THA/TKA procedure. Some of these are considered planned readmissions and we do NOT count them as readmissions in the measure. If a patient undergoes a second primary THA/TKA and is admitted to the hospital within 30 days of the discharge date for the index admission, and the admission is associated with a primary discharge diagnosis of osteoarthritis, rheumatoid arthritis, osteonecrosis, and arthropathy (excluding septic arthropathy), the readmission is likely planned and is not counted as a readmission in the measure. Appendix A lists the ICD-9 codes used to identify these discharge diagnoses.

2.3.2 30-Day Timeframe

A 30-day timeframe is clinically sensible and is a meaningful timeframe for hospitals because readmissions are more likely attributable to care received within the index hospitalization *and* during the transition to the outpatient setting. For example, hospitals, in collaboration with their medical communities, take actions to reduce readmission, such as: ensure patients are clinically ready at discharge; reduce risk of infection; reconcile medications; improve communications among providers involved in transition of care; encourage strategies that promote disease management principles; and educate patients about symptoms to monitor, whom to contact with questions, and where and when to seek follow-up care. Finally, this timeframe is consistent with the other readmission measures approved by the National Quality Forum (NQF).

Based on preliminary analyses of the hazard of readmission over a 90-day period, risk of readmission is highest within the first two weeks after the discharge date from the index admission (Figure 1). The rate plateaus between 30 and 45 days post discharge, suggesting that a 30-day timeframe would capture the period of highest risk of readmission.

Figure 1. Hazard of Readmission Following THA/TKA (Medicare Part A Inpatient, 2008)



2.3.3 All-Cause Readmission

We used all-cause readmission (excluding planned readmissions), rather than readmission for procedural complications for several reasons. First, from the patient perspective, readmission for any reason is likely to be an undesirable outcome of care after elective surgery. Second, readmissions not directly related to the procedure may still be a result of the care received during the index hospitalization. For example, a patient who underwent a THA/TKA who develops a hospital-acquired infection may ultimately be readmitted for sepsis. It would be inappropriate to treat this readmission as unrelated to the care the patient received for the procedure. Another patient might experience a procedure-related complication following his THA or TKA, which may go untreated and result in renal failure. The resulting readmission for renal failure could have been prevented with higher quality of care during the admission for the THA/TKA that could have reduced the risk for the complication. Furthermore, the range of potentially avoidable readmissions also includes those not directly related to the procedures such as those resulting from poor communication or inadequate follow-up. As such, creating a

comprehensive list of potential complications related to THA/TKA would be arbitrary and, ultimately, impossible to implement. Using all-cause readmission, on the other hand, will undoubtedly include a mix of unavoidable and avoidable readmissions. Thus, the goal of this measure is not to reduce readmissions to zero, but to instead assess hospital performance relative to what is expected given the performance of other hospitals with similar case mixes.

2.4 Cohort Definition

In consultation with the working group, we considered whether to develop separate measures for patients undergoing THA and TKA procedures or to combine patients undergoing either procedure into a single hospital quality measure. We combined these patient cohorts for several reasons, including:

- A large proportion of THA and TKA procedures are elective and performed in similar patient cohorts for similar indications (e.g., osteoarthritis)
- The same surgeons frequently perform both procedures
- Both procedures have similar lengths of stay
- The rates and types of complications are similar (Table 1)
- The mortality and readmission rates are similar (Table 1)
- Hospitals develop protocols/programs for lower extremity total joint arthroplasty, rather than for THA and TKA separately
- Combining admissions for both procedures will provide greater power to detect hospital-level variation to enable quality improvement

Table 1. Procedure Characteristics and Unadjusted Mortality, Readmission, and Complication Rates for THA and TKA (Medicare Inpatient Part A, 2008).

		Total Hip Replacement* (excludes partial hip replacement and hip fractures)	Total Knee Replacement**
Procedure-related characteristics			
Number of Patients Receiving Procedure		97,130	240,517
Mean Length of Stay (SD)		3.8 (2.3)	3.6 (1.7)
Mean Patient Age (SD)		75.2 (6.6)	74.2 (6.1)
Number of Hospitals Performing Procedure		3083	3307
Median Number of Procedures Performed at Each Hospital (Q1-Q3)		16 (6 - 41)	40 (13 - 257)
Mortality		**% (5th-95th)**	**% (5th-95th)**
In-hospital Mortality	Patient level	0.2	0.1
	Hospital level: median	0 (0 - 0.9)	0 (0 - 0.6)
30-day Mortality	Patient level	0.5	0.3
	Hospital level: median	0 (0 - 2.9)	0 (0 - 1.7)
90-day Mortality	Patient level	0.9	0.5
	Hospital level: median	0 (0 - 5.6)	0 (0 - 3.0)
Readmission		**% (5th-95th)**	**% (5th-95th)**
30-day All-cause Readmission		6.9	5.9
	Hospital level: median	5 (0 - 25)	5 (0 - 18)
90-day All-cause Readmission		12.2	10.7
	Hospital level: median	11 (0 - 38)	10 (0 - 27)
Complications		**% (30-day / 90-day)**	**% (30-day / 90-day)**
Dislocation		0.8 / 1.1	0.1 / 0.1
DVT		0.1 /0.2	0.2 / 0.2
Hematoma		1.9 / 2.0	1.2 / 1.3
Periprosthetic Joint Infection		0.5 / 0.7	0.4 / 0.6
Postoperative infection		0.8 / 1.0	0.7 / 0.8
Pulmonary Embolism		0.5 / 0.7	0.8 / 1.0
Mechanical complication of internal orthopedic device, implant		2.7 / 3.3	0.3 / 0.4
Venous thrombosis		0.1 / 0.2	0.1 / 0.1
Wound Infection		0.7 / 0.9	0.7 / 0.8
All complications combined		5.8 / 7.0	3.4 / 4.1
* Includes ICD-9 code 81.51 ** Includes ICD-9 code 81.54			

Patients undergoing non-elective THA or TKA have greater risk of complications and receive a wider variety of surgical procedures than individuals undergoing elective THA or TKA. In consultation with the working group and with the goal of defining a comprehensive yet reasonably homogeneous cohort for quality

assessment, we selected inclusion and exclusion criteria in order to identify patients undergoing elective THA and TKA for degenerative (either primary or secondary) arthritis.

Patients eligible for inclusion in the measure are those aged 65 and older admitted to non-federal acute care hospitals with an ICD-9 code for THA and/or TKA. Patients must have had continuous enrollment in Medicare fee-for-service (FFS) for one year prior to the date of index admission to ensure full data availability for risk adjustment. The flow chart depicting cohort selection is presented in Figure 2.

Eligible index admissions are identified using the following ICD-9-CM procedure codes in Medicare Part A Inpatient claims data:

- 81.51 Total Hip Arthroplasty
- 81.54 Total Knee Arthroplasty

2.5 Exclusion Criteria

1. Patients with hip fractures
   Rationale: Patients with hip fractures have higher mortality, complication, and readmission rates and the procedures are not elective

2. Patients undergoing revision procedures (with or without a concurrent THA/TKA)
   Rationale: Revision procedures may be performed at a disproportionately small number of hospitals and are associated with higher mortality, complication and readmission rates

3. Patients undergoing partial hip arthroplasty (PHA) procedures (with or without a concurrent THA/TKA)
   Rationale: Partial arthroplasties are primarily done for hip fractures and are typically performed on patients who are older, more frail, and with more comorbid conditions

4. Patients undergoing resurfacing procedures (with or without a concurrent THA/TKA)
   Rationale: Resurfacing procedures are a different type of procedure where only the joint's articular surface is replaced. A THA involves surgical removal of the neck of the femur (thighbone) and insertion of a stem deep inside the bone to connect with the pelvic socket and liner. Furthermore, resurfacing procedures are typically performed on younger, healthier patients

5. Patients who were transferred in to the index hospital

Rationale: If the patient is transferred from another acute care facility to the hospital where the index procedure occurs, it is likely that the procedure is not elective.

6. Patients who were admitted for the index procedure and subsequently transferred to another acute care facility
   Rationale: Attribution of readmission to the index hospital would not be possible in these cases, since the index hospital performed the procedure but another hospital discharged the patient to the non-acute care setting.

7. Patients who leave the hospital against medical advice (AMA)
   Rationale: Hospitals and physicians do not have the opportunity to provide the highest quality care for these patients.

8. Patients with more than two THA/TKA procedures codes during the index hospitalization
   Rationale: It is unlikely that patients would receive more than two THA/TKA procedures in one hospitalization, and this may reflect a coding error.

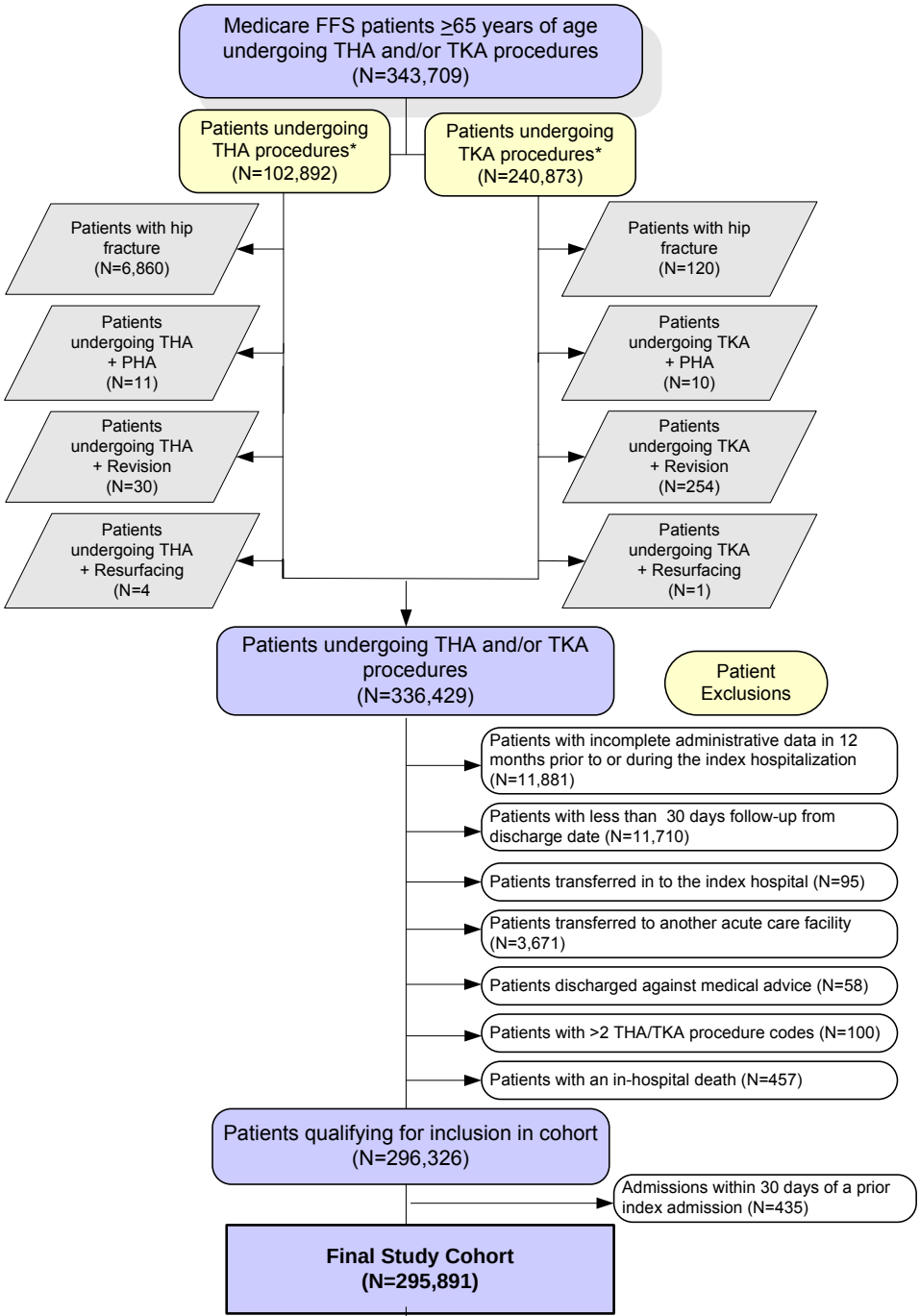
9. Patients without at least 30-days post-discharge enrollment in Medicare FFS.
   Rationale: The 30-day readmission outcome cannot be assessed for the standardized time period.

10. Patients with inconsistent or unknown mortality status or other unreliable data (e.g. date of death precedes admission date)
    Rationale: Outcome status is unreliable, although this is rare.

11. Patients who die during the index admission
    Rationale: Patients who die during the initial hospitalization are not eligible for readmission.

Appendix B lists the ICD-9-CM codes for hip fracture, revision procedures, partial hip arthroplasty procedure, and resurfacing procedures.

Figure 2. Cohort for Model Development



Procedure	N (%)
1 THA procedure	84,094
1 TKA Procedure	202,286
THA + THA	304
TKA + TKA	9,158
THA + TKA	49

*THA and TKA are presented separately for illustrative purposes and are not mutually exclusive

2.6 Approach to Risk Adjustment

The goal of risk adjustment is to account for patient demographic and clinical characteristics while illuminating important quality differences. The model adjusts for case mix differences based on the clinical status of the patient at the time of admission. Conditions that may represent adverse outcomes due to care received during the index admission are not considered for inclusion in the risk adjusted model. Although they may increase the risk of readmission, including them as covariates in a risk-adjusted model could attenuate the measure's ability to characterize the quality of care delivered by hospitals. Appendix C lists the conditions not adjusted for if they only appear in the index admission and not in the 12 months prior to admission. This methodology is consistent with NQF guidelines.

Consistent with NQF guidelines, the model does not adjust for socioeconomic status (SES), race, or ethnicity because risk-adjusting for SES would hold hospitals with a large proportion of low SES patients to a different standard of care than hospitals treating a larger proportion of high SES patients. The model does not adjust for patients' admission source and their discharge disposition either (e.g. skilled nursing facility) because these factors are associated with structure of the health care system, and may reflect the quality of care delivered by the system.

2.7 Candidate and Final Risk-adjustment Variables

Our goal was to develop a parsimonious model that included clinically relevant variables that are strongly associated with risk of readmission. The candidate variables for the model were derived from: the index admission, with comorbidities identified from the index admission secondary diagnoses (excluding potential complications), 12-month pre-index inpatient Part A data, outpatient hospital data, and Part B physician data.

For administrative model development, we started with the 189 Condition Categories (CCs). CCs are clinically relevant diagnostic groups of the more than 15,000 ICD-9 codes.[9] We used the April 2010 version of the ICD-9 to CC assignment map, which is maintained by CMS and posted at www.qualitynet.org.

To select candidate variables, a team of clinicians reviewed all 189 CCs and excluded those that were not relevant to the Medicare population (Appendix D) or that were not clinically relevant to the readmission outcome (e.g., attention deficit disorder, female infertility, cataract). Clinically relevant CCs were selected as candidate variables. CCs with high clinical relevance to the outcome were broken out and certain conditions within that CC were examined separately when clinically indicated. For example, obesity and morbid obesity are known risk factors for complications and readmission following THA/TKA. We reviewed these comorbidities and based on these analyses and expert feedback, morbid

obesity was separated from CC 24 (obesity and other endocrine/metabolic/nutritional disorders) and included in the risk adjusted model independently. Other CCs were combined into clinically coherent groups. Other candidate variables included age, sex, type of procedure (THA, TKA, both), and number of procedures (one versus two) and are listed in Table 2.

Table 2. THA/TKA Readmission Model Candidate Variables

Category	Variable	ICD-9 Code(s) or CC(s)
Demographic	Age-65 (years above 65, continuous)	
	Sex	
Procedure	Type of procedure	ICD-9-CM 81.51 (THA) ICD-9-CM 81.54 (TKA)
	Number of procedures (one versus two)	
Comorbidities	Skeletal deformities	ICD-9-CM 755.63
	Post traumatic osteoarthritis	ICD-9-CM 716.15, 716.16
	Morbid obesity	ICD-9-CM 278.01
	History of Infection	CC 1, 3-6
	Septicemia/shock	CC 2
	Metastatic cancer and acute leukemia	CC 7
	Cancer	CC 8-12
	Other neoplasms	CC 13
	Benign neoplasms of skin, breast, eye	CC 14
	Diabetes and DM complications	CC 15-20, 119, 120
	Protein-calorie malnutrition	CC 21
	Disorders of Fluid/Electrolyte/Acid-Base	CC 22, 23
	Obesity/disorders of thyroid, cholesterol, lipids	CC 24
	Liver and biliary disease	CC 25-30
	Intestinal Obstruction/Perforation	CC 31
	Pancreatic Disease	CC 32
	Inflammatory Bowel Disease	CC 33
	Peptic Ulcer, Hemorrhage, Other Specified Gastrointestinal Disorders	CC 34
	Appendicitis	CC 35
	Other Gastrointestinal Disorders	CC 36
	Bone/Joint/Muscle Infections/Necrosis	CC 37
	Rheumatoid Arthritis and Inflammatory Connective Tissue Disease	CC 38
	Disorders of the Vertebrae and Spinal Discs	CC 39
	Osteoarthritis of Hip and Knee	CC 40
	Osteoporosis and Other Bone/Cartilage Disorders	CC 41
	Congenital/Developmental Skeletal and Connective Tissue Disorders	CC 42
	Other Musculoskeletal and Connective Tissue Disorders	CC 43
	Severe Hematological Disorders	CC 44
	Disorders of Immunity	CC 45

Category	Variable	ICD-9 Code(s) or CC(s)
	Coagulation Defects and Other Specified Hematological Disorders	CC 46
	Iron Deficiency and Other/Unspecified Anemias and Blood Disease	CC 47
	Delirium and Encephalopathy	CC 48
	Dementia and senility	CC 49, 50
	Drug/alcohol abuse/dependence/psychosis	CC 51-53
	Major psychiatric Disorders	CC 54-56
	Personality Disorders	CC 57
	Depression	CC 58
	Anxiety Disorders	CC 59
	Other psychiatric disorders	CC 60
	Mental retardation or developmental disability	CC 61-65
	Hemiplegia, paraplegia, paralysis, functional disability	CC 67-69, 100-102, 177-178
	Muscular Dystrophy	CC 70
	Polyneuropathy	CC 71
	Multiple Sclerosis	CC 72
	Parkinson's and Huntington's Diseases	CC 73
	Seizure Disorders and Convulsions	CC 74
	Coma, Brain Compression/Anoxic Damage	CC 75
	Mononeuropathy, Other Neurological Conditions/Injuries	CC 76
	Respirator Dependence/Tracheostomy Status	CC 77
	Respiratory Arrest	CC 78
	Cardio-Respiratory Failure and Shock	CC 79
	Congestive Heart Failure	CC 80
	Acute Coronary Syndrome	CC 81-82
	Chronic Atherosclerosis	CC 83-84
	Heart Infection/Inflammation, Except Rheumatic	CC 85
	Valvular and Rheumatic Heart Disease	CC 86
	Congenital cardiac/circulatory defect	CC 87-88
	Hypertension	CC 89, 91
	Hypertensive heart disease	CC 90
	Arrhythmias	CC 92, 93
	Other and Unspecified Heart Disease	CC 94
	Stroke	CC 95, 96
	Cerebrovascular disease	CC 97-99, 103
	Vascular or circulatory disease	CC 104-106
	Cystic fibrosis	CC 107
	COPD	CC 108
	Fibrosis of lung or other chronic lung disorder	CC 109
	Asthma	CC 110
	Pneumonia	CC 111-113
	Pleural effusion/pneumothorax	CC 114
	Other lung disorder	CC 115
	Legally Blind	CC 116
	Major eye infections/inflammations	CC 117
	Retinal detachments	CC 118

Category	Variable	ICD-9 Code(s) or CC(s)
	Retinal Disorders, Except Detachment and Vascular Retinopathies	CC 121
	Glaucoma	CC 122
	Other Eye Disorders	CC 124
	Significant Ear, Nose, and Throat Disorders	CC 125
	Hearing Loss	CC 126
	Other Ear, Nose, Throat, and Mouth Disorders	CC 127
	Kidney Transplant Status	CC 128
	End-stage renal disease or dialysis	CC 129, 130
	Renal Failure	CC 131
	Nephritis	CC 132
	Urinary Obstruction and Retention	CC 133
	Incontinence	CC 134
	Urinary Tract Infection	CC 135
	Other urinary tract disorders	CC 136
	Pelvic Inflammatory disease	CC 138
	Other female genital disorders	CC 139
	Male genital disorders	CC 140
	Decubitus ulcer or chronic skin ulcer	CC 148, 149
	Extensive burns	CC 150, 151
	Cellulitis, Local Skin Infection	CC 152
	Other Dermatological Disorders	CC 153
	Trauma	CC 154-156, 158-161
	Vertebral Fractures	CC 157
	Other Injuries	CC 162
	Poisonings and Allergic Reactions	CC 163
	Major Complications of Medical Care and Trauma	CC 164
	Other Complications of Medical Care	CC 165
	Major Symptoms, Abnormalities	CC 166
	Minor Symptoms, Signs, Findings	CC 167
	Major Organ Transplant Status	CC 174
	Other organ transplant/replacement	CC 175

To inform final variable selection, a modified approach to stepwise logistic regression was performed. A subsample of the data was used to create 500 "bootstrap" samples. For each sample, we ran a logistic stepwise regression that included the candidate variables. The results were summarized to show the percentage of times that each of the candidate variables was significantly associated with readmission ($p<0.001$) in each of the 500 repeated samples (e.g., 70 percent would mean that the candidate variable was selected as significant at $p<0.001$ in 70 percent of the estimations). We also assessed the direction and magnitude of the regression coefficients.

The clinical team reviewed these results and decided to retain all risk adjustment variables above a 70% cutoff, because they demonstrated a relatively strong association with risk for readmission and were clinically relevant. Additionally,

specific variables with particular clinical relevance to the risk of readmission were forced into the model (regardless of % selection) to ensure appropriate risk-adjustment for THA and TKA. These included:

Markers for end of life/frailty:

- decubitus ulcer (CC 148)
- dementia and senility (CC 49 and CC 50, respectively)
- metastatic cancer and acute leukemia (CC 7)
- protein-calorie malnutrition (CC 21)
- hemiplegia/paraplegia/paralysis/functional disability (CC 67-69, 100-102, 177-178)
- stroke (CC 95-96)

Diagnoses with potential asymmetry among hospitals that would impact the validity of the model:

- cancer (CC 8-12)

Final model variables are listed in Table 3.

Table 3. THA/TKA Readmission Final Model Variables

Category	Variable	ICD-9 Code(s) or CC(s)
Demographic	Age-65 (years above 65, continuous)	
	Sex	
Procedure	Type of procedure	ICD-9-CM 81.51 (THA)
	Number of procedures (1 vs. 2)	
Comorbidities	Skeletal deformities	ICD-9-CM 755.63
	Post traumatic osteoarthritis	ICD-9-CM 716.15, 716.16
	Morbid obesity	ICD-9-CM 278.01
	History of Infection	CC 1, 3-6
	Metastatic cancer and acute leukemia	CC 7
	Cancer	CC 8-12
	Diabetes and DM complications	CC 15-20, 119, 120
	Protein-calorie malnutrition	CC 21
	Disorders of Fluid/Electrolyte/Acid-Base	CC 22, 23
	Joint/Muscle Infections/Necrosis	CC 37
	Rheumatoid Arthritis and Inflammatory Connective Tissue Disease	CC 38
	Severe Hematological Disorders	CC 44
	Dementia and senility	CC 49. 50
	Major psychiatric disorders	CC 54-56
	Hemiplegia, paraplegia, paralysis, functional disability	CC 67-69, 100-102, 177-178
	Polyneuropathy	CC 71
	Congestive Heart Failure	CC 80
	Chronic Atherosclerosis	CC 83-84
	Hypertension	CC 89, 91
	Arrhythmias	CC 92, 93
	Stroke	CC 95, 96
	Vascular or circulatory disease	CC 104-106
	COPD	CC 108
	Pneumonia	CC 111-113
	End-stage renal disease or dialysis	CC 129, 130
	Renal Failure	CC 131
	Decubitus ulcer or chronic skin ulcer	CC 148, 149
	Cellulitis, Local Skin Infection	CC 152
	Other injuries	CC 162
	Major Symptoms, Abnormalities	CC 166

2.8 Statistical Approach to Model Development

We randomly selected 50% of the THA and/or TKA admissions that met all inclusion and exclusion criteria and created a development sample, which we used to build the model. We used the remaining 50% of THA/TKA admissions as the validation sample. We also used all qualifying THA and/or TKA admissions in 2007 data as an additional sample to validate the model. Model performance was assessed in the development dataset and both validation datasets.

Due to the natural clustering of observations within hospitals, we used hierarchical generalized linear models (HGLMs) to model the log-odds of readmission. Readmission was modeled as a function of patient-level demographic and clinical characteristics and a random hospital-specific intercept. This strategy accounts for within-hospital correlation of the observed outcomes and models the assumption that underlying differences in quality among the health care facilities being evaluated lead to systematic differences in outcomes.

We then calculated hospital risk-standardized readmission rates (RSRRs) using a hierarchical logistic regression model (given the hierarchical nature of the data). These rates are calculated as the ratio of the predicted number of readmissions to the expected number of readmissions, multiplied by the national unadjusted readmission rate. The expected number of readmissions for each hospital was estimated using that hospital's patient mix and the national intercept. Specifically, for each patient in the data set, the estimated regression coefficients are multiplied by the observed characteristics and the average of the hospital-specific intercepts is added to this quantity. Then, the quantity is transformed to the probability scale. For each patient within a hospital, these probabilities are summed. The predicted number of readmissions in each hospital employs a similar calculation. The predicted number of readmissions for each hospital is calculated by summing the predicted readmission rates for all patients in the hospital. The predicted readmission rate for each patient is calculated through the hierarchical model by applying the estimated regression coefficients to the patient characteristics observed and adding the hospital-specific intercept. In order to assess hospital performance in any specific year (e.g. the validation cohort), we re-estimate the model coefficients using that year's data.

More specifically, we estimate a generalized linear model and a hierarchical generalized linear model which accounts for the clustering of observations within hospitals. The generalized linear model (GLM) links the outcome to the patient-level risk factors.[10] Let Y_{ij} denote the outcome (equal to 1 if patient dies or has a complication, zero otherwise) for the j^{th} patient who had a THA/TKA procedure at the i^{th} hospital; $\mathbf{Z}_{ij}$ denotes a set of risk factors based on the data. Let I denote the total number of hospitals and n_i the number of index patient stays in hospital i. We assume the outcome is related linearly to the covariates via a known linked function, h, where

GLM $h(Y_{ij}) = \alpha + \beta \mathbf{Z}_{ij}$ (1)

and $\mathbf{Z}_{ij} = (Z_{1ij}, Z_{2ij}, \ldots, Z_{pij})$ is a set of p patient-specific covariates. In our case, h = the logit link.

To account for the natural clustering of observations within hospitals, we then estimate an HGLM that links the risk factors to the same outcome and a hospital-specific random effect,

HGLM $h(Y_{ij}) = \alpha_i + \beta \mathbf{Z}_{ij}$ (2)

$\alpha_i = \mu + \omega_i; \quad \omega_i \sim N(0, \tau^2)$ (3)

where α_i represents the hospital-specific intercept, $\mathbf{Z}_{ij}$ is defined as above, μ the adjusted average outcome over all hospitals in the sample, and τ^2 the between-hospital variance component.[11] This model separates within-hospital variation from between-hospital variation. Both HGLMs and GLMs are estimated using the SAS software system (GLIMMIX and LOGISTIC procedures, respectively).

We first fit the GLM described in Equation (1) using the logit link.
Having identified the covariates that remained, we next fit the HGLM described in Equations (2) and (3), again using the logit link function; e.g.,

Logit $(P(Y_{ij} = 1)) = \alpha_i + \beta \mathbf{Z}_{ij}$

$\alpha_i = \mu + \omega_i, \quad \omega_i \sim N(0, \tau^2)$

where $\mathbf{Z}_{ij}$ consisted of the covariates retained in the GLM model. As before, Y_{ij} = 1 if patient j treated at hospital i had the event; 0 otherwise.

2.9 Hospital Performance Reporting

Using the set of risk factors in the GLM, we fit the HGLM defined by Equations (2) - (3) and estimate the parameters, $\hat{\mu}$, $\{\hat{\alpha}_i, \hat{\alpha}_2, ..., \hat{\alpha}_I\}$, $\hat{\beta}$, and $\hat{\tau}^2$. We calculate a standardized outcome, s_i, for each hospital by computing the ratio of the number of predicted readmissions to the number of expected readmissions, multiplied by the unadjusted overall readmission rate, $\bar{y}$. Specifically, we calculate

Predicted $\hat{y}_{ij}(Z) = h^{-1}(\hat{\alpha}_i + \hat{\beta} \mathbf{Z}_{ij})$ (4)

Expected $\hat{e}_{ij}(Z) = h^{-1}(\hat{\mu} + \hat{\beta} \mathbf{Z}_{ij})$ (5)

$$\hat{s}_i(Z) = \frac{\sum_{j=1}^{n_i} \hat{y}_{ij}(Z)}{\sum_{j=1}^{n_i} \hat{e}_{ij}(Z)} \times \bar{y} \qquad (6)$$

If more (fewer) "predicted" cases than "expected" cases have the outcome in a hospital, then $\hat{s}_i$ will be higher (lower) than the unadjusted average. For each hospital, we compute an interval estimate of s_i to characterize the level of uncertainty around the point estimate using bootstrapping simulations. The point estimate and interval estimate can be used to characterize and compare hospital performance (e.g., higher than expected, as expected, or lower than expected).

2.9.1 Creating Interval Estimates

Because the statistic described in Equation (6) is a complex function of parameter estimates, we use re-sampling and simulation techniques to derive an interval estimate. The bootstrapping simulation has the advantage of avoiding unnecessary distributional assumptions.

2.9.2 Algorithm

Let *I* denote the total number of hospitals in the sample. We repeat steps 1 – 4 below for b = 1,2,…B times:

1. Sample *I* hospitals with replacement.

2. Fit the HGLM using all patients within each sampled hospital. We use as starting values the parameter estimates obtained by fitting the model to all hospitals. If some hospitals are selected more than once in a bootstrapped sample, we treat them as distinct so that we have *I* random effects to estimate the variance components. At the conclusion of Step 2, we have:
   a. $\hat{\beta}^{(b)}$ (the estimated regression coefficients of the risk factors).
   b. The parameters governing the random effects, hospital adjusted outcomes, distribution, $\hat{\mu}^{(b)}$ and $\hat{\tau}^{2(b)}$.
   c. The set of hospital-specific intercepts and corresponding variances, $\{\hat{\alpha}_i^{(b)}, \hat{\text{var}}(\alpha_i^{(b)}); i = 1,2,\ldots,I\}$.

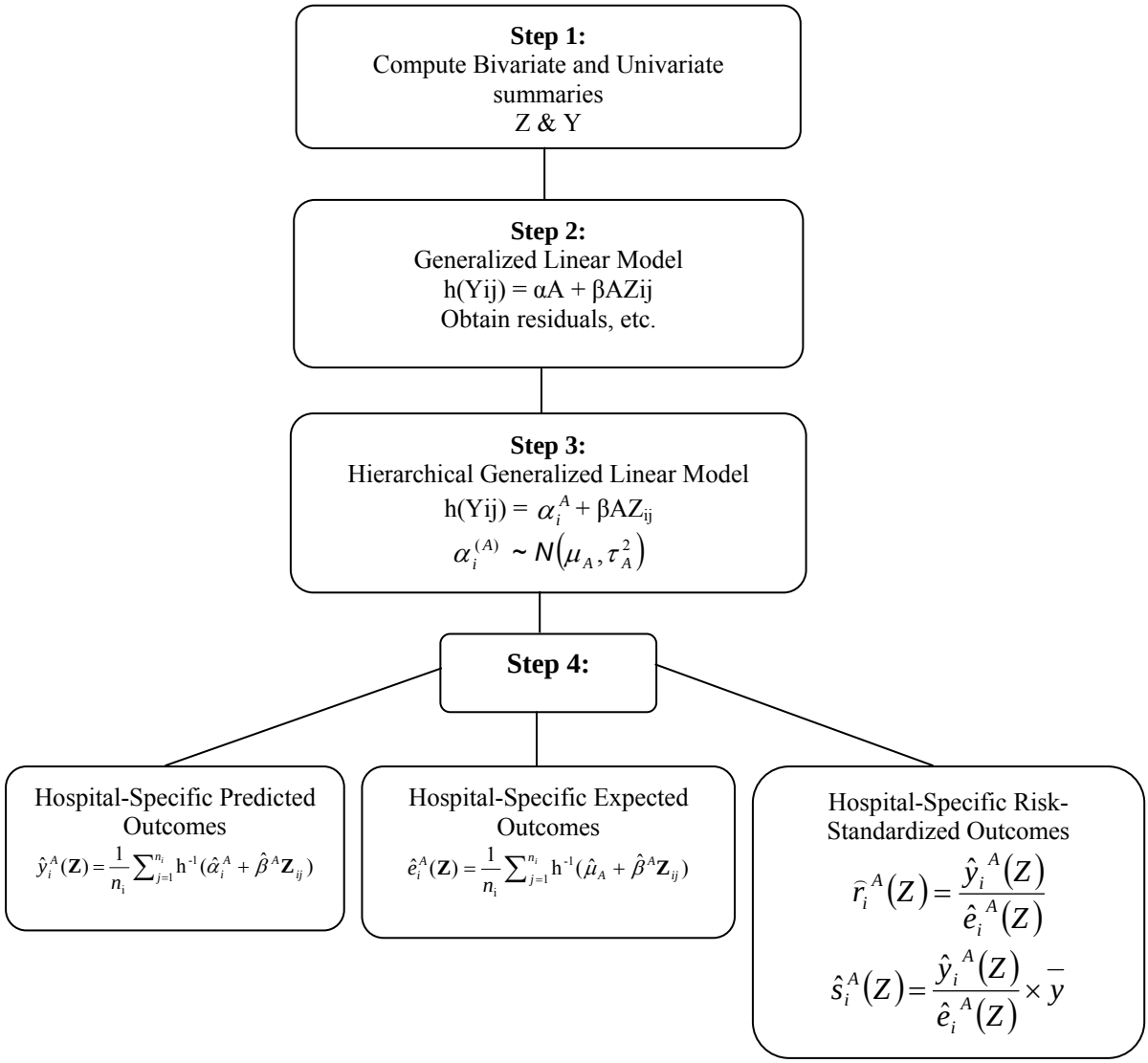
3. We generate a hospital random effect by sampling from the distribution of the hospital-specific distribution obtained in Step 2c. We approximate the distribution for each random effect by a normal

distribution. Thus, we draw $\alpha_i^{(b*)} \sim N\left(\hat{\alpha}_i^{(b)}, \hat{\text{var}}\left(\hat{\alpha}_i^{(b)}\right)\right)$ for the unique set of hospitals sampled in Step 1.

4. Within each unique hospital *i* sampled in Step 1, and for each case *j* in that hospital, we calculate $\hat{y}_{ij}^{(b)}$, $\hat{e}_{ij}^{(b)}$, and $\hat{s}_i\left(Z\right)^{(b)}$ where $\hat{\beta}^{(b)}$ and $\hat{\mu}^{(b)}$ are obtained from Step 2 and $\hat{\alpha}_i^{(b*)}$ is obtained from Step 3.

Ninety-five percent interval estimates (or alternative interval estimates) for the hospital-standardized outcome can be computed by identifying the 2.5th and 97.5th percentiles of randomly half of the B estimates (or the percentiles corresponding to the alternative desired intervals).[12]

Figure 3. Analysis Steps



3. RESULTS

3.1 Model Results

3.1.1 Development and Validation Models

Table 4 conveys the GLM model results for the 2008 development sample. The standardized estimates are regression coefficients expressed in units of standard deviations and can range between -1 and 1, with ±1 indicating a perfect linear relationship and 0 indicating no linear relationship.[1] The area under the receiver operating characteristic (ROC) curve is 0.64 indicating good discrimination.

Table 5 conveys the HGLM model results for the 2008 development sample. The T value is the parameter estimate divided by its standard error, and its associated probability indicates whether the variable is significantly associated with the outcome. The estimated between-hospital variance in the adjusted log-odds of readmission is 0.156 based on the 2008 full dataset. This result implies that the odds of readmission for patients at a high-readmission hospital (+1 SD) are 2.20 times that in a low-readmission hospital (-1 SD). If there were no differences between hospitals, the between-hospital variance would be 0 and the odds ratio would be 1.0.Table 6 conveys the GLM results for the validation sample.

3.1.2 Model Performance

Table 7 conveys model performance results for both the developmental and validation samples. We computed the following summary statistics for assessing model performance[13]: over-fitting indices[2], predictive ability, area under the (ROC) curve, distribution of residuals, and model chi-square[3]. The models for both the development and validation samples

---

[1] Standardized estimates are like correlation coefficients. We compute them in order to compare the size of the coefficients by standardizing the coefficients to be unitless.

[2] Over-fitting refers to the phenomenon in which a model well describes the relationship between predictive variables and outcome in the development dataset, but fails to provide valid predictions in new patients.

[3] Chi-Square – A test of statistical significance usually employed for categorical data to determine whether there is a good fit between the observed data and expected values; i.e., whether the differences between observed and expected values are attributable to true differences in characteristics or instead the result of chance variation. The formula for computing the chi-square is as follows:

$$\sum \frac{(O-E)^2}{E}$$

where O = observed value
E = expected value, and

have strong discrimination and fit. Model predictive ability ranges from 2.5% in the lowest predictive decile to 13.2% in the highest decile in the development sample and the validation sample from 2008. Predictive ability ranges from 2.8% in the lowest predictive decile to 13.3% in the highest decile in the 2007 validation sample, indicating the model can reasonably classify patients on the outcome, based on their risk. The area under the ROC curve (C statistic) is 0.64 for the development model and for both validation models (Table 7). The discrimination ability is consistent with models currently used to publicly report condition specific rates of mortality and readmission.

Table 8 conveys the standardized estimates by year of discharge in the full datasets for 2007 and 2008. There are no notable differences in the standardized estimates between the two years. Table 9 conveys the risk factor frequency for the development and validation samples by year of discharge. The prevalence of morbid obesity increased slightly to 3.5% in 2008, compared with 2.97% in 2007. There were no other notable changes in risk factor frequency over the two-year period.

degrees of freedom (df) = (rows-1)(columns-1)

Table 4. GLM Model Results for 2008 Development Sample (ROC = 0.64)

Description	Estimate	Standard Error	Standardized Estimates	Odds Ratio	95% Confidence Interval for OR
Intercept	-3.86	0.04			
Demographics					
Age-65‡ (mean)	0.03	0.00	0.11	1.03	(1.03 – 1.04)
Male	0.09	0.02	0.02	1.10	(1.05 – 1.15)
THA/TKA Procedure					
THA procedure	0.14	0.02	0.03	1.15	(1.10 – 1.21)
Number of procedures (one vs. two)	0.19	0.06	0.02	1.21	(1.08 – 1.37)
Comorbid Conditions					
Skeletal deformities (ICD-9 code 755.63)	0.12	0.28	0.00	1.13	(0.65 – 1.96)
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16)	-0.10	0.15	0.00	0.90	(0.67 – 1.22)
Morbid obesity (ICD-9 code 278.01)	0.26	0.06	0.03	1.30	(1.16 – 1.44)
History of infection (CC 1, 3-6)	0.11	0.03	0.02	1.11	(1.05 – 1.17)
Metastatic cancer and acute leukemia (CC 7)	0.06	0.13	0.00	1.06	(0.82 – 1.37)
Cancer (CC 8-12)	-0.05	0.03	-0.01	0.95	(0.90 – 1.00)
Diabetes and DM complications (CC 15-20, 119, 120)	0.14	0.02	0.03	1.15	(1.10 – 1.21)
Protein-calorie malnutrition (CC 21)	0.29	0.10	0.01	1.33	(1.08 – 1.63)
Disorders of fluid/electrolyte/acid-base (CC 22, 23)	0.15	0.03	0.03	1.16	(1.09 – 1.23)
Rheumatoid arthritis and inflammatory connective tissue Disease (CC 38)	0.08	0.04	0.01	1.08	(1.00 – 1.16)
Severe hematological disorders (CC 44)	0.43	0.10	0.02	1.54	(1.28 – 1.86)
Dementia and senility (CC 49, 50)	0.14	0.05	0.02	1.15	(1.05 – 1.26)
Major psychiatric disorders (CC 54-56)	0.23	0.05	0.02	1.26	(1.15 – 1.40)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	0.13	0.08	0.01	1.14	(0.99 – 1.32)
Polyneuropathy (CC 71)	0.17	0.04	0.02	1.19	(1.10 – 1.29)
Congestive heart failure (CC 80)	0.16	0.03	0.03	1.17	(1.10 – 1.25)
Chronic atherosclerosis (CC 83-84)	0.23	0.02	0.06	1.26	(1.20 – 1.32)
Hypertension (CC 89, 91)	0.18	0.03	0.04	1.19	(1.12 – 1.28)
Arrhythmias (CC 92, 93)	0.18	0.03	0.04	1.20	(1.14 – 1.26)
Stroke (CC 95, 96)	0.04	0.06	0.00	1.04	(0.92 – 1.18)
Vascular or circulatory disease (CC 104-106)	0.12	0.03	0.03	1.13	(1.07 – 1.18)
COPD (CC 108)	0.29	0.03	0.06	1.34	(1.27 – 1.42)
Pneumonia (CC 111-113)	0.17	0.04	0.02	1.19	(1.10 – 1.29)
End-stage renal disease or dialysis (CC 129, 130)	0.62	0.18	0.01	1.86	(1.30 – 2.65)
Renal failure (CC 131)	0.18	0.04	0.02	1.19	(1.10 – 1.29)
Decubitus ulcer or chronic skin ulcer (CC 148, 149)	0.15	0.06	0.01	1.16	(1.04 – 1.29)
Cellulitis, local skin infection (CC 152)	0.16	0.04	0.02	1.17	(1.09 – 1.26)
Other injuries (CC162)	0.16	0.02	0.04	1.18	(1.12 – 1.24)
Major symptoms, abnormalities (CC 166)	0.13	0.02	0.04	1.14	(1.09 – 1.19)

Table 5. HGLM Model Results for 2008 Development Sample

Description	Estimate	Standard Error	T-Value	Pr > T-Value	Odds Ratio	95% Confidence Interval for OR
Intercept	-3.87	0.04	-93.90	<.0001		
Demographics						
Age-65‡ (mean)	0.03	0.00	18.99	<.0001	1.04	(1.03 – 1.04)
Male	0.10	0.02	4.36	<.0001	1.11	(1.06 – 1.16)
THA/TKA Procedure						
THA procedure	0.14	0.02	5.93	<.0001	1.15	(1.10 – 1.20)
Number of procedures (one vs. two)	0.20	0.06	3.34	0.001	1.23	(1.09 – 1.38)
Comorbid Conditions						
Skeletal deformities (ICD-9 code 755.63)	0.14	0.28	0.51	0.610	1.15	(0.67 – 1.97)
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16)	-0.10	0.15	-0.65	0.517	0.91	(0.68 – 1.22)
Morbid obesity (ICD-9 code 278.01)	0.24	0.05	4.41	<.0001	1.27	(1.14 – 1.42)
History of infection (CC 1, 3-6)	0.10	0.03	3.82	0.000	1.11	(1.05 – 1.17)
Metastatic cancer and acute leukemia (CC 7)	0.08	0.13	0.61	0.543	1.08	(0.84 – 1.39)
Cancer (CC 8-12)	-0.06	0.03	-2.20	0.028	0.94	(0.89 – 0.99)
Diabetes and DM complications (CC 15-20, 119, 120)	0.13	0.02	5.63	<.0001	1.14	(1.09 – 1.20)
Protein-calorie malnutrition (CC 21)	0.30	0.10	2.90	0.004	1.35	(1.10 – 1.65)
Disorders of fluid/electrolyte/acid-base (CC 22, 23)	0.14	0.03	4.39	<.0001	1.15	(1.08 – 1.22)
Rheumatoid arthritis and inflammatory connective Tissue disease (CC 38)	0.08	0.04	2.09	0.037	1.08	(1.00 – 1.16)
Severe hematological disorders (CC 44)	0.44	0.09	4.66	<.0001	1.55	(1.29 – 1.86)
Dementia and senility (CC 49, 50)	0.15	0.05	3.28	0.001	1.16	(1.06 – 1.27)
Major psychiatric disorders (CC 54-56)	0.24	0.05	4.86	<.0001	1.27	(1.15 – 1.40)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	0.13	0.07	1.81	0.070	1.14	(0.99 – 1.32)
Polyneuropathy (CC 71)	0.18	0.04	4.38	<.0001	1.20	(1.10 – 1.30)
Congestive heart failure (CC 80)	0.17	0.03	5.23	<.0001	1.19	(1.11 – 1.27)
Chronic atherosclerosis (CC 83-84)	0.23	0.02	9.62	<.0001	1.26	(1.20 – 1.32)
Hypertension (CC 89, 91)	0.17	0.03	5.29	<.0001	1.19	(1.11 – 1.27)
Arrhythmias (CC 92, 93)	0.19	0.03	7.61	<.0001	1.21	(1.15 – 1.27)
Stroke (CC 95, 96)	0.04	0.06	0.72	0.474	1.04	(0.93 – 1.18)
Vascular or circulatory disease (CC 104-106)	0.12	0.03	4.61	<.0001	1.12	(1.07 – 1.18)
COPD (CC 108)	0.29	0.03	10.55	<.0001	1.34	(1.27 – 1.41)
Pneumonia (CC 111-113)	0.18	0.04	4.39	<.0001	1.19	(1.10 – 1.29)
End-stage renal disease or dialysis (CC 129, 130)	0.60	0.18	3.32	0.001	1.82	(1.28 – 2.58)
Renal failure (CC 131)	0.18	0.04	4.56	<.0001	1.19	(1.11 – 1.29)
Decubitus ulcer or chronic skin ulcer (CC 148, 149)	0.15	0.06	2.78	0.005	1.17	(1.05 – 1.30)
Cellulitis, local skin infection (CC 152)	0.16	0.04	4.53	<.0001	1.18	(1.10 – 1.26)
Other injuries (CC162)	0.16	0.02	6.92	<.0001	1.18	(1.12 – 1.23)
Major symptoms, abnormalities (CC 166)	0.12	0.02	4.97	<.0001	1.13	(1.07 – 1.18)

Table 6. GLM Model Results for 2008 Validation Sample (ROC = 0.64)

Label	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimates	Odds Ratio	95% Confidence Interval for OR
Intercept	-3.85	0.04	8920.02	<.0001			
Demographics							
Age-65‡ (mean)	0.03	0.00	314.50	<.0001	0.11	1.03	(1.03 – 1.04)
Male	0.11	0.02	22.57	<.0001	0.03	1.12	(1.07 – 1.17)
THA/TKA Procedure							
THA procedure	0.13	0.02	29.15	<.0001	0.03	1.14	(1.09 – 1.19)
Number of procedures (one vs. two)	0.24	0.06	16.01	<.0001	0.02	1.27	(1.13 – 1.43)
Comorbid Conditions							
Skeletal deformities (ICD-9 code 755.63)	0.01	0.29	0.00	0.965	0.00	1.01	(0.57 – 1.79)
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16)	-0.03	0.15	0.04	0.833	0.00	0.97	(0.73 – 1.30)
Morbid obesity (ICD-9 code 278.01)	0.24	0.06	17.52	<.0001	0.02	1.27	(1.13 – 1.41)
History of Infection (CC 1, 3-6)	0.11	0.03	15.98	<.0001	0.02	1.12	(1.06 – 1.18)
Metastatic cancer and acute leukemia (CC 7)	0.29	0.11	6.19	0.013	0.01	1.33	(1.03 – 1.67)
Cancer (CC 8-12)	-0.02	0.03	0.48	0.489	0.00	0.98	(0.93 – 1.04)
Diabetes and DM complications (CC 15-20, 119, 120)	0.12	0.02	22.91	<.0001	0.03	1.12	(1.07- 1.18)
Protein-calorie malnutrition (CC 21)	0.03	0.11	0.08	0.779	0.00	1.03	(0.83 – 1.28)
Disorders of fluid/electrolyte/acid-base (CC 22, 23)	0.14	0.03	20.62	<.0001	0.03	1.15	(1.08 – 1.23)
Rheumatoid arthritis and inflammatory connective tissue disease (CC 38)	0.11	0.04	8.92	0.003	0.02	1.12	(1.04 – 1.20)
Severe hematological disorders (CC 44)	0.38	0.10	16.20	<.0001	0.02	1.47	(1.22 – 1.77)
Dementia and senility (CC 49, 50)	0.25	0.04	32.73	<.0001	0.03	1.29	(1.18 – 1.41)
Major psychiatric disorders (CC 54-56)	0.34	0.05	48.31	<.0001	0.04	1.40	(1.27 – 1.54)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	0.11	0.07	2.15	0.143	0.01	1.12	(0.96 – 1.29)
Polyneuropathy (CC 71)	0.12	0.04	8.61	0.003	0.02	1.13	(1.04 – 1.23)
Congestive heart failure (CC 80)	0.26	0.03	64.55	<.0001	0.04	1.30	(1.22 – 1.38)
Chronic atherosclerosis (CC 83-84)	0.22	0.02	85.83	<.0001	0.06	1.25	(1.19 – 1.31)
Hypertension (CC 89, 91)	0.19	0.03	31.46	<.0001	0.04	1.21	(1.13 – 1.29)
Arrhythmias (CC 92, 93)	0.13	0.03	25.99	<.0001	0.03	1.14	(1.08 – 1.20)
Stroke (CC 95, 96)	0.06	0.06	1.08	0.299	0.01	1.07	(0.95 – 1.20)
Vascular or circulatory disease (CC 104-106)	0.09	0.03	13.76	0.000	0.02	1.10	(1.05 – 1.16)
COPD (CC 108)	0.21	0.03	53.50	<.0001	0.04	1.23	(1.16 – 1.30)
Pneumonia (CC 111-113)	0.23	0.04	33.33	<.0001	0.03	1.26	(1.17 – 1.37)
End-stage renal disease or dialysis (CC 129, 130)	0.87	0.17	26.58	<.0001	0.02	2.40	(1.72 – 3.34)
Renal failure (CC 131)	0.21	0.04	28.03	<.0001	0.03	1.23	(1.14 – 1.33)
Decubitus ulcer or chronic skin ulcer (CC 148, 149)	0.14	0.06	6.07	0.014	0.01	1.15	(1.03 – 1.28)
Cellulitis, local skin infection (CC 152)	0.13	0.04	12.74	0.000	0.02	1.14	(1.06 – 1.23)
Other injuries (CC162)	0.14	0.02	32.21	<.0001	0.03	1.15	(1.09 – 1.20)
Major symptoms, abnormalities (CC 166)	0.15	0.02	40.49	<.0001	0.04	1.17	(1.11 – 1.22)

Table 7. Model Performance for GLM Model

Indices	Development Sample	Validation Sample	Validation Sample
Year	2008 (50%)	2008 (50%)	2007 (100%)
Number of Admissions	147, 959	147,932	300,012
Number of Hospitals	3,227	3,225	3,295
Number of Readmissions	8,978	9,116	19,007
Calibration (γ0, γ1)[1]	(0, 1)	(0.01, 1.00)	(-0.08, 0.95)
Discrimination -Predictive Ability (lowest decile %, highest decile %)	(2.5%, 13.2%)	(2.5%, 13.2%)	(2.8%, 13.3%)
Discrimination – Area Under Receiver Operator Curve	0.64	0.64	0.64
Residuals Lack of Fit (Pearson Residual Fall %)			
<-2	0	0	0
[-2, 0)	93.9	93.8	93.7
[0, 2)	0.1	0.1	0.1
[2+	5.9	6.0	6.2
Model Wald χ^2 [Number of Covariates]	2346 [33]	2462 [33]	4546 [33]

[1] Over-Fitting Indices (γ_0, γ_1) provide evidence of over-fitting and require several steps to calculate. Let *b* denote the *estimated vector* of regression coefficients. *Predicted Probabilities* ($\hat{p}$) = 1/(1+exp{-Xb}), and *Z* = *Xb* (e.g., the linear predictor that is a scalar value for everyone). A new logistic regression model that includes only an intercept and a slope by regressing the logits on Z is fitted in the validation sample; e.g., Logit(P(Y=1|Z)) = $\gamma_0 + \gamma_1 Z$. Estimated values of γ_0 far from 0 and estimated values of γ_1 far from 1 provide evidence of over-fitting.

Table 8. Standardized Estimates by Year of Discharge (GLM)

Description	2008 (100%)			2007 (100%)		
	Standardized Estimates	OR	95% CI for Odds Ratio	Standardized Estimates	OR	95% CI for Odds Ratio
Demographics						
Age-65‡ (mean)	0.11	1.03	(1.03 - 1.04)	0.11	1.03	(1.03 – 1.04)
Male	0.03	1.11	(1.07 - 1.14)	0.02	1.10	(1.05 – 1.15)
THA/TKA Procedure						
THA procedure	0.03	1.14	(1.11 - 1.18)	0.03	1.15	(1.10 – 1.21)
Number of procedures (one vs. two)	0.02	1.24	(1.14 - 1.35)	0.02	1.21	(1.08 – 1.37)
Comorbid Conditions						
Skeletal deformities (ICD-9 code 755.63)	0.00	1.07	(0.72 - 1.59)	0.00	1.13	(0.65 – 1.96)
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16)	0.00	0.94	(0.76 - 1.15)	0.00	0.90	(0.67 – 1.22)
Morbid obesity (ICD-9 code 278.01)	0.02	1.28	(1.19 - 1.38)	0.03	1.30	(1.16 – 1.45)
History of Infection (CC 1, 3-6)	0.02	1.11	(1.07 - 1.16)	0.02	1.11	(1.05 – 1.17)
Metastatic cancer and acute leukemia (CC 7)	0.01	1.20	(1.01 - 1.42)	0.00	1.06	(0.82 – 1.37)
Cancer (CC 8-12)	-0.01	0.96	(0.93 - 1.00)	-0.01	0.95	(0.90 – 1.00)
Diabetes and DM complications (CC 15-20, 119, 120)	0.03	1.14	(1.10 - 1.18)	0.03	1.15	(1.10 – 1.21)
Protein-calorie malnutrition (CC 21)	0.01	1.17	(1.01 - 1.36)	0.01	1.33	(1.08 – 1.63)
Disorders of Fluid/Electrolyte/Acid-Base (CC 22, 23)	0.03	1.16	(1.11- 1.21)	0.03	1.16	(1.09 – 1.23)
Rheumatoid Arthritis and Inflammatory Connective Tissue Disease (CC 38)	0.01	1.10	(1.04 - 1.16)	0.01	1.08	(1.00 – 1.16)
Severe Hematological Disorders (CC 44)	0.02	1.51	(1.32 - 1.72)	0.02	1.54	(1.28 – 1.86)
Dementia and senility (CC 49, 50)	0.02	1.22	(1.15 - 1.30)	0.02	1.15	(1.05 – 1.26)
Major psychiatric disorders (CC 54-56)	0.03	1.33	(1.25 - 1.43)	0.02	1.26	(1.15 – 1.40)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	0.01	1.13	(1.02 - 1.25)	0.01	1.14	(0.99 – 1.32)
Polyneuropathy (CC 71)	0.02	1.16	(1.10 - 1.23)	0.02	1.19	(1.10 – 1.29)
Congestive Heart Failure (CC 80)	0.03	1.24	(1.18 - 1.29)	0.03	1.17	(1.10 – 1.25)
Chronic Atherosclerosis (CC 83-84)	0.06	1.26	(1.21 - 1.30)	0.06	1.26	(1.20 – 1.32)
Hypertension (CC 89, 91)	0.04	1.20	(1.15 - 1.26)	0.04	1.19	(1.12 – 1.28)
Arrhythmias (CC 92, 93)	0.04	1.17	(1.13 - 1.21)	0.04	1.20	(1.14 – 1.26)
Stroke (CC 95, 96)	0.00	1.05	(0.97 - 1.15)	0.00	1.04	(0.92 – 1.18)
Vascular or circulatory disease (CC 104-106)	0.02	1.11	(1.07 - 1.15)	0.03	1.13	(1.07 – 1.18)
COPD (CC 108)	0.05	1.28	(1.23 - 1.33)	0.06	1.34	(1.27 – 1.42)
Pneumonia (CC 111-113)	0.03	1.22	(1.16 - 1.29)	0.02	1.19	(1.10 – 1.29)
End-stage renal disease or dialysis (CC 129, 130)	0.02	2.10	(1.65 - 2.68)	0.01	1.86	(1.30 – 2.65)
Renal Failure (CC 131)	0.03	1.21	(1.15 - 1.28)	0.02	1.19	(1.10 – 1.29)
Decubitus ulcer or chronic skin ulcer (CC 148, 149)	0.01	1.15	(1.07 - 1.25)	0.01	1.16	(1.04 – 1.29)
Cellulitis, Local Skin Infection (CC 152)	0.02	1.16	(1.10 - 1.22)	0.02	1.17	(1.09 – 1.26)
Other injuries (CC162)	0.04	1.16	(1.12 - 1.20)	0.04	1.18	(1.12 – 1.24)
Major Symptoms, Abnormalities (CC 166)	0.04	1.15	(1.11 - 1.19)	0.04	1.14	(1.09 – 1.19)

Table 9. Risk Factor Frequency by Year of Discharge (GLM)

Description	2008 Development Sample	2008 Validation Sample	2007 Validation Sample
Male	35.93	35.56	35.57
THA procedure	28.40	28.68	28.49
Number of procedures (one vs. two)	3.21	3.22	3.52
Skeletal deformities (ICD-9 code 755.63)	0.13	0.14	0.15
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16)	0.51	0.53	0.49
Morbid obesity (ICD-9 code 278.01)	3.50	3.42	2.97
History of Infection (CC 1, 3-6)	17.88	17.89	17.63
Metastatic cancer and acute leukemia (CC 7)	0.58	0.62	0.64
Cancer (CC 8-12)	18.73	18.73	18.65
Diabetes and DM complications (CC 15-20, 119, 120)	27.31	27.36	26.70
Protein-calorie malnutrition (CC 21)	0.61	0.63	0.55
Disorders of Fluid/Electrolyte/Acid-Base (CC 22, 23)	11.98	11.97	11.85
Rheumatoid Arthritis and Inflammatory Connective Tissue Disease (CC 38)	8.61	8.55	8.34
Severe Hematological Disorders (CC 44)	0.74	0.76	0.73
Dementia and senility (CC 49, 50)	4.33	4.33	4.19
Major psychiatric disorders (CC 54-56)	3.69	3.72	3.56
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	1.58	1.59	1.50
Polyneuropathy (CC 71)	5.67	5.64	5.50
Congestive Heart Failure (CC 80)	9.55	9.74	9.81
Chronic Atherosclerosis (CC 83-84)	30.55	30.63	30.90
Hypertension (CC 89, 91)	82.59	82.52	82.07
Arrhythmias (CC 92, 93)	22.40	22.25	21.90
Stroke (CC 95, 96)	2.41	2.43	2.49
Vascular or circulatory disease (CC 104-106)	22.67	22.44	22.15
COPD (CC 108)	14.63	14.65	15.15
Pneumonia (CC 111-113)	5.40	5.44	5.43
End-stage renal disease or dialysis (CC 129, 130)	0.14	0.14	0.14
Renal Failure (CC 131)	6.10	6.12	5.53
Decubitus ulcer or chronic skin ulcer (CC 148, 149)	2.68	2.73	2.74
Cellulitis, Local Skin Infection (CC 152)	7.78	7.89	7.65
Other injuries (CC162)	27.32	27.69	27.50
Major Symptoms, Abnormalities (CC 166)	52.62	52.42	52.76

3.1.3 Unadjusted and Adjusted Readmission Rate Distributions

Figure 4 displays the unadjusted frequency distribution of the hospital-specific readmission rates in the 2008 cohort. The unadjusted mean readmission rate is 6.72% and ranged from 0% to 100% across the 3,310 hospitals. The median unadjusted readmission rate is 6.72%.

After adjusting for patient and clinical characteristics, the risk-standardized rates are more normally distributed (Figure 5) with a mean of 6.25%, ranging from 3.03% to 50.97%. The median adjusted readmission rate is 6.01%.

Figure 4. Unadjusted Hospital Readmission Rates (2008 Sample; N=3,310 Hospitals)

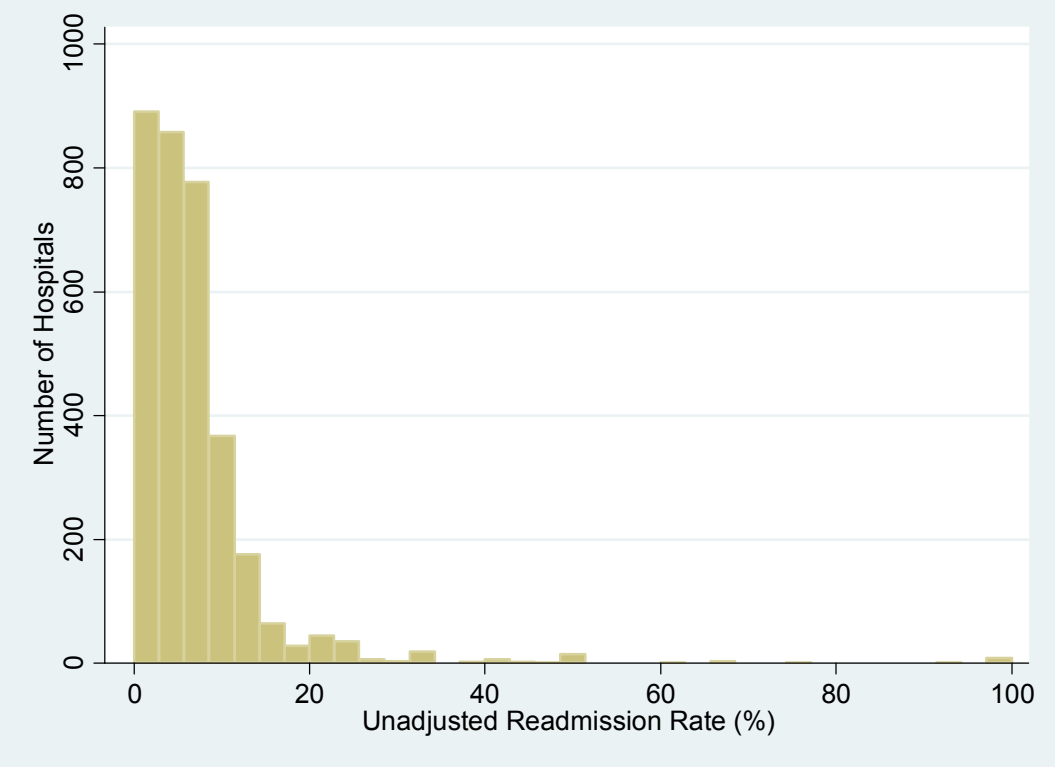
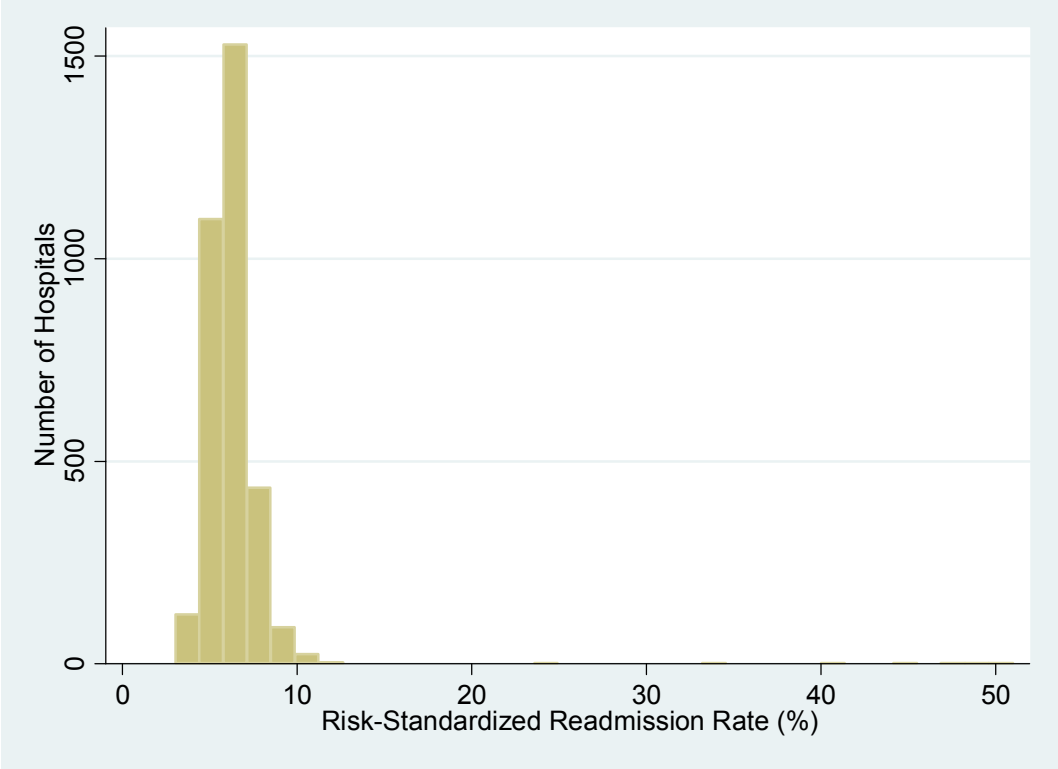


Figure 5. Distribution of Hospital Risk-Standardized Readmission Rates (2008 Sample; N=3,310 Hospitals) – HGLM



4. MAIN FINDINGS / SUMMARY

The proposed 30-day all-cause readmission measure has the potential to significantly improve the quality of care delivered to patients undergoing elective primary THA and TKA procedures. The risk standardized model is consistent with the consensus standards for publicly reported outcomes measures, and can be implemented using available data. This measure was developed with extensive input from experts with clinical and methodological expertise relevant to orthopedic quality measurement. The study sample is appropriately defined, consisting of patients undergoing elective primary THA and/or TKA and will allow for valid comparisons of hospital quality. We excluded covariates that are not appropriate for inclusion in a quality measure such as race, socioeconomic status, and hospital-level factors (e.g., hospital bed size and volume of THA/TKA procedures). The hierarchical modeling accounts for hospital case mix, the clustering of patients within hospitals and differences in sample size across hospitals, thereby making the measure suitable for public reporting.

5. REFERENCES

1. Kurtz S,Ong K,Lau E,Mowat F,Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am.* Apr 2007;89(4):780-785.
2. Kurtz SM,Ong KL,Schmier J, et al. Future clinical and economic impact of revision total hip and knee arthroplasty. *J Bone Joint Surg Am.* Oct 2007;89 Suppl 3:144-151.
3. Ong KL,Mowat FS,Chan N, et al. Economic burden of revision hip and knee arthroplasty in Medicare enrollees. *Clin Orthop Relat Res.* May 2006;446:22-28.
4. Bozic KJ,Rubash HE,Sculco TP,Berry DJ. An analysis of medicare payment policy for total joint arthroplasty. *J Arthroplasty.* Sep 2008;23(6 Suppl 1):133-138.
5. National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.nysna.org/images/pdfs/practice/nqf_ana_outcomes_draft10.pdf . Accessed August 19, 2010.
6. Krumholz HM,Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* January 24, 2006 2006;113(3):456-462.
7. Fisher ES,Wennberg JE,Stukel TA,Sharp SM. Hospital Readmission Rates for Cohorts of Medicare Beneficiaries in Boston and New Haven. *New England Journal of Medicine.* 1994;331(15):989-995.
8. Nelson EA,Maruish ME,Axler JL. Effects of Discharge Planning and Compliance With Outpatient Appointments on Readmission Rates. *Psychiatr Serv.* July 1 2000;51(7):885-889.
9. Pope G,Ellis R,Ash A, et al. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. *Health Care Financing Review.* 2000;21(3):26.
10. McCallagh PNJ. *Generalized Linear Models*. London: Chapman and Hall; 1989.
11. Daniels MJ,Gatsonis C. Hierarchical Generalized Linear Models in the Analysis of Variations in Health Care Utilization. *Journal of the American Statistical Association.* 1999;94(445):29-42.
12. Normand S,Wang Y,Krumholz H. Assessing surrogacy of data sources for institutional comparisons. *Health Services and Outcomes Research Methodology.* 2007;7:79-96.

13. Harrell F,Shih Y. Using full probablility models to compute probabiliites of actual interest to decision makers. *Int J Technol Assess Health Care.* Winter 2001;17(1):10.

6. APPENDIX

6.1 Appendix A: ICD-9-CM Codes for Osteoarthritis, Rheumatoid Arthritis, Osteonecrosis, and Arthropathy

Rheumatoid Arthritis	
714.0	Rheumatoid arthritis
714	Rheumatoid arthritis and other inflammatory polyarthropathies
714.1	Felty's syndrome
714.2	Other rheumatoid arthritis with visceral or systemic involvement
714.3	Juvenile chronic polyarthritis
714.30	Chronic or unspecified polyarticular juvenile rheumatoid arthritis
714.31	Acute polyarticular juvenile rheumatoid arthritis
714.32	Pauciarticular juvenile rheumatoid arthritis
714.33	Monoarticular juvenile rheumatoid arthritis
714.4	Chronic postrheumatic arthropathy
714.8	Other specified inflammatory polyarthropathies
714.89	Other specified inflammatory polyarthropathies
714.9	Unspecified inflammatory polyarthropathy

Osteoarthritis	
715	Osteoarthrosis and allied disorders
715.0	Osteoarthrosis generalized
715.00	Osteoarthrosis generalized involving unspecified site
715.09	Osteoarthrosis generalized involving multiple sites
715.1	Osteoarthrosis localized primary
715.10	Osteoarthrosis localized primary involving unspecified site
715.15	Osteoarthrosis localized primary involving pelvic region and thigh
715.16	Osteoarthrosis localized primary involving lower leg
715.18	Osteoarthrosis localized primary involving other specified sites
715.2	Osteoarthrosis localized secondary
715.20	Osteoarthrosis localized secondary involving unspecified site
715.25	Osteoarthrosis localized secondary involving pelvic region and thigh
715.26	Osteoarthrosis localized secondary involving lower leg
715.28	Osteoarthrosis localized secondary involving other specified sites
715.3	Osteoarthrosis localized not specified whether primary or secondary
715.30	Osteoarthrosis localized not specified whether primary or secondary involving unspecified site
715.35	Osteoarthrosis localized not specified whether primary or secondary involving pelvic region and thigh
715.36	Osteoarthrosis localized not specified whether primary or secondary involving lower leg
715.38	Osteoarthrosis localized not specified whether primary or secondary involving other specified sites
715.8	Osteoarthrosis involving or with mention of more than one site but not specified as generalized
715.80	Osteoarthrosis involving or with more than one site but not specified as generalized and involving unspecified site
715.89	Osteoarthrosis involving or with multiple sites but not specified as generalized

715.9	Osteoarthrosis unspecified whether generalized or localized
715.90	Osteoarthrosis unspecified whether generalized or localized involving unspecified site
715.95	Osteoarthrosis unspecified whether generalized or localized involving pelvic region and thigh
715.96	Osteoarthrosis unspecified whether generalized or localized involving lower leg
715.98	Osteoarthrosis unspecified whether generalized or localized involving other specified sites

Arthropathy	
716.5	Unspecified polyarthropathy or polyarthritis
716.50	Unspecified polyarthropathy or polyarthritis site unspecified
716.55	Unspecified polyarthropathy or polyarthritis involving pelvic region and thigh
716.56	Unspecified polyarthropathy or polyarthritis involving lower leg
716.58	Unspecified polyarthropathy or polyarthritis involving other specified sites
716.59	Unspecified polyarthropathy or polyarthritis involving multiple sites
716.8	Other specified arthropathy
716.80	Other specified arthropathy no site specified
716.85	Other specified arthropathy involving pelvic region and thigh
716.86	Other specified arthropathy involving lower leg
716.88	Other specified arthropathy involving other specified sites
716.89	Other specified arthropathy involving multiple sites
716.9	Unspecified arthropathy
716.90	Unspecified arthropathy site unspecified
716.95	Unspecified arthropathy involving pelvic region and thigh
716.96	Unspecified arthropathy involving lower leg
716.98	Unspecified arthropathy involving other specified sites
716.99	Unspecified arthropathy involving multiple sites

Osteonecrosis	
733.42	Aseptic necrosis of head and neck of femur
733.43	Aseptic necrosis of medial femoral condyle

6.2 Appendix B: ICD-9-CM Codes for Hip Fracture, Revision Procedures, Partial Hip Arthroplasty, and Resurfacing Procedure

ICD-9-CM Codes for Hip Fracture	
733.1	Pathologic fracture
733.10	Pathological fracture unspecified site
733.14	Pathological fracture of neck of femur
733.15	Pathological fracture of other specified part of femur
733.19	Pathological fracture of other specified site
733.8	Malunion and nonunion of fracture
733.81	Malunion of fracture
733.82	Nonunion of fracture
733.95	Stress fracture of other bone
733.96	Stress fracture of femoral neck
733.97	Stress fracture of shaft of femur
808.0	Closed fracture of acetabulum
808.1	Open fracture of acetabulum
820.00	Fracture of unspecified intracapsular section of neck of femur closed
820.01	Fracture of epiphysis (separation) (upper) of neck of femur closed
820.02	Fracture of midcervical section of femur closed
820.03	Fracture of base of neck of femur closed
820.09	Other transcervical fracture of femur closed
820.10	Fracture of unspecified intracapsular section of neck of femur open
820.11	Fracture of epiphysis (separation) (upper) of neck of femur open
820.12	Fracture of midcervical section of femur open
820.13	Fracture of base of neck of femur open
820.19	Other transcervical fracture of femur open
820.20	Fracture of unspecified trochanteric section of femur closed
820.21	Fracture of intertrochanteric section of femur closed
820.22	Fracture of subtrochanteric section of femur closed
820.30	Fracture of unspecified trochanteric section of femur open
820.31	Fracture of intertrochanteric section of femur open
820.32	Fracture of subtrochanteric section of femur open
820.8	Fracture of unspecified part of neck of femur closed
820.9	Fracture of unspecified part of neck of femur open
821	Fracture of other and unspecified parts of femur
821.0	Fracture of shaft or unspecified part of femur closed
821.00	Fracture of unspecified part of femur closed
821.01	Fracture of shaft of femur closed
821.1	Fracture of shaft or unspecified part of femur open
821.10	Fracture of unspecified part of femur open
821.11	Fracture of shaft of femur open

ICD-9-CM Codes for THA and TKA Revision Procedures	
00.70	REV Hip Repl-acetab/fem OCT05
00.71	REV Hip Repl-acetab comp OCT05
00.72	REV Hip Repl-fem comp OCT05
00.73	REV Hip Repl-liner/head OCT05

00.80	Replacement of femoral, tibial, and patellar components (all components)
00.81	Replacement of tibial baseplate and tibial insert (liner)
00.82	Revision of knee replacement, femoral component
00.83	Revision of knee replacement, patellar component
00.84	Revision of total knee replacement, tibial insert (liner)
81.53	Revise Hip Replacement, NOS
81.55	Revision of Knee replacement, NOS
81.59	Revision of joint replacement of lower extremity, not elsewhere classified

ICD-9-CM Code for Partial Hip Arthroplasty Procedure	
81.52	Partial Hip Replacement

ICD-9-CM Codes for THA Resurfacing Procedure	
00.85	Resurfacing hip, total, acetabulum and femoral head, hip resurfacing arthroplasty, total
00.86	Resurfacing hip, partial, femoral head, hip resurfacing arthroplasty, NOS, hip resurfacing arthroplasty, partial, femoral head
00.87	Resurfacing hip, partial, acetabulum, hip resurfacing arthroplasty, partial, acetabulum

6.3 Appendix C: Conditions That May Represent Adverse Outcomes of Care Received During Index Admission.

CC	Description
2	Septicemia/Shock
6	Other Infectious Diseases
17	Diabetes with Acute Complications
23	Disorders of Fluid/Electrolyte/Acid-Base
24	Other Endocrine/Metabolic/Nutritional Disorders
31	Intestinal Obstruction/Perforation
34	Peptic Ulcer, Hemorrhage, Other Specified Gastrointestinal Disorders
36	Other Gastrointestinal Disorders
37	Bone/Joint/Muscle Infections/Necrosis
43	Other Musculoskeletal and Connective Tissue Disorders
46	Coagulation Defects and Other Specified Hematological Disorders
47	Iron Deficiency and Other/Unspecified Anemias and Blood Disease
48	Delirium and Encephalopathy
51	Drug/Alcohol Psychosis
75	Coma, Brain Compression/Anoxic Damage
76	Mononeuropathy, Other Neurological Conditions/Injuries
77	Respirator Dependence/Tracheostomy Status
78	Respiratory Arrest
79	Cardio-respiratory failure and shock
80	Congestive heart failure
81	Acute myocardial infarction
82	Unstable angina
85	Heart Infection/Inflammation, Except Rheumatic
95	Cerebral Hemorrhage
96	Ischemic or Unspecified Stroke
97	Precerebral Arterial Occlusion and Transient Cerebral Ischemia
100	Hemiplegia/Hemiparesis
101	Cerebral Palsy and Other Paralytic Syndromes
102	Speech, Language, Cognitive, Perceptual
104	Vascular Disease with Complications
105	Vascular Disease
106	Other Circulatory Disease
111	Aspiration and Specified Bacterial Pneumonias
112	Pneumococcal Pneumonia, Emphysema, Lung Abscess
114	Pleural Effusion/Pneumothorax
130	Dialysis Status
131	Renal failure
132	Nephritis
133	Urinary Obstruction and Retention
135	Urinary Tract Infection
148	Decubitus Ulcer of Skin
152	Cellulitis, Local Skin Infection
154	Severe Head Injury
155	Major Head Injury
156	Concussion or Unspecified Head Injury

CC	Description
157	Vertebral Fractures
158	Hip Fracture/Dislocation
159	Major Fracture, Except of Skull, Vertebrae, or Hip
160	Internal Injuries
161	Traumatic Amputation
162	Other Injuries
163	Poisonings and Allergic Reactions
164	Major Complications of Medical Care and Trauma
165	Other Complications of Medical Care
175	Other Organ Transplant/Replacement
177	Amputation Status, Lower Limb/Amputation
178	Amputation Status, Upper Limb

6.4 Appendix D: CCs Not Considered for Risk Adjustment

CC	Description	Rationale
66	Attention Deficit Disorder	Pediatric ; Low frequency
123	Cataracts	Marker of clinical practice, not clinical relevant
129	End Stage Renal Disease	Not included in CMS-HCC Model
137	Female Infertility	Irrelevant to Medicare FFS Population
141	Ectopic Pregnancy	Irrelevant to Medicare FFS Population
142	Miscarriage/Abortion	Irrelevant to Medicare FFS Population
143	Completed Pregnancy with Major Complications	Irrelevant to Medicare FFS Population
144	Completed Pregnancy with Complications	Irrelevant to Medicare FFS Population
145	Completed Pregnancy without Complication	Irrelevant to Medicare FFS Population
146	Uncompleted Pregnancy with Complications	Irrelevant to Medicare FFS Population
147	Uncompleted Pregnancy with No or Minor Complications	Irrelevant to Medicare FFS Population
168	Extremely Low Birthweight Neonates	Fetal Effects; Irrelevant to Medicare FFS Population
169	Very Low Birthweight Neonates	Fetal Effects; Irrelevant to Medicare FFS Population
170	Serious Perinatal Problems Affecting Newborn	Fetal Effects; Irrelevant to Medicare FFS Population
171	Other Perinatal Problems Affecting Newborn	Fetal Effects; Irrelevant to Medicare FFS Population
172	Normal, Single Birth	Fetal Effects; Irrelevant to Medicare FFS Population
173	Major Organ Transplant	Not included in CMS-HCC Model
176	Artificial Openings for Feeding or Elimination	CC too heterogeneous; Mix of disparate codes
179	Post-Surgical States/Aftercare/Elective	CC too heterogeneous; Mix of disparate codes
180	Radiation Therapy	CC too heterogeneous; Mix of disparate codes
181	Chemotherapy	CC too heterogeneous; Mix of disparate codes
182	Rehabilitation	CC too heterogeneous; Mix of disparate codes
183	Screening/Observation/Special Exams	CC too heterogeneous; Mix of disparate codes
184	History of Disease	CC too heterogeneous; Mix of disparate codes
185	Oxygen	Not included in CMS-HCC Model; DME
186	CPAP/IPPB/Nebulizers	Not included in CMS-HCC Model; DME
187	Patient Lifts, Power Operated Vehicles, Beds	Not included in CMS-HCC Model; DME
188	Wheelchairs, Commodes	Not included in CMS-HCC Model; DME
189	Walkers	Not included in CMS-HCC Model; DME