NATIONAL QUALITY FORUM

Measure Submission and Evaluation Worksheet 5.0

This form contains the information submitted by measure developers/stewards, organized according to NQF's measure evaluation criteria and process. The evaluation criteria, evaluation guidance documents, and a blank online submission form are available on the <u>submitting standards web page</u>.

NQF #: 0362 NQF Project: Patient Safety Measures-Complications Project

(for Endorsement Maintenance Review)

Original Endorsement Date: May 15, 2008 Most Recent Endorsement Date: May 15, 2008

BRIEF MEASURE INFORMATION

De.1 Measure Title: Foreign Body left after procedure (PDI 3)

Co.1.1 Measure Steward: Agency for Healthcare Research and Quality

De.2 Brief Description of Measure: Count of discharges with foreign body left in during procedure in medical and surgical discharges among patients less than 18 years and not MDC 14 (pregnancy, childbirth, and puerperium)

2a1.1 Numerator Statement: Discharges under age 18 with ICD-9-CM codes for foreign body left in during procedure in any secondary diagnosis field of medical and surgical discharges defined by specific DRGs or MS-DRGs where several exclusions are applied to the numerator. (Details of the numerator, medical and surgical discharges DRGs and MS-DRGs, and exclusions appear in 2a1.3).

2a1.4 Denominator Statement: Not applicable

2a1.8 Denominator Exclusions: Not applicable

1.1 Measure Type: Outcome 2a1. 25-26 Data Source: Administrative claims 2a1.33 Level of Analysis: Facility

1.2-1.4 Is this measure paired with another measure? No

De.3 If included in a composite,	please identify the composite measure	(title and NQF number	if endorsed):
Not applicable			

STAFF NOTES (issues or questions regarding any criteria)

Comments on Conditions for Consideration:

Is the measure untested?	Yes No	If untested, explain how it meets	criteria for consideration for time-limited
endorsement:			

1a. Specific national health goal/priority identified by DHHS or NPP addressed by the measure (*check De.5*):
5. Similar/related <u>endorsed</u> or submitted measures (*check 5.1*):

Other Criteria:

Staff Reviewer Name(s):

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See <u>guidance on evidence</u>.

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: (The measure directly addresses a specific national health goal/priority identified by DHHS or NPP, or some other high impact aspect of healthcare.) **De.4 Subject/Topic Areas** (Check all the areas that apply): Surgery : General Surgery **De.5 Cross Cutting Areas** (Check all the areas that apply): Safety : Complications 1a.1 Demonstrated High Impact Aspect of Healthcare: Patient/societal consequences of poor quality 1a.2 If "Other," please describe: **1a.3 Summary of Evidence of High Impact** (*Provide epidemiologic or resource use data*): PDI 03 functions appropriately in pediatric populations to identify adverse events that are associated with excess length of stay and total charges. Using the Nationwide Inpatient Sample from 1998-2005 and the KIDS Inpatient Database from 1997, 2000, and 2003, Camp and colleagues identified 413 unique records flagged by PDI 03, and matched them with 1,227 control records with the same age, race, gender, and hospital identification code. Multiple regression analyses were performed for inpatient mortality, length of stay and total hospital charges (controlling for procedure category, admission type, and insurance status in the matched case-control data set). Adjusted for procedure category, admission type, and insurance status, patients with PDI 03 had Multivariable regression analysis revealed that patients with PDI 3 had an 8-day longer mean length of stay (95% CI, 5.6-10.3) and \$35,681 higher total hospital charges (95% CI, \$22,358-\$49,004), but were not more likely to die (odds ratio, 1.07), compared with patients without PDI 03. In a similar study using nearest-neighbor propensity score matching in the Pediatric Health Information System database from 2006 (an administrative database with data from 38 academic, nonprofit pediatric hospitals affiliated with the Child Health Corporation of America), Kronman and colleagues reported mean excess length of stay of 14.3 days and mean excess total charges of \$144,889 for each PDI 03 case, relative to matched controls. The excess charges came from all hospital cost centers, including pharmacy (\$14,777), supplies (\$5,681), laboratory (\$15,420), imaging (\$2,970), and other clinical activities (\$2,219). 1a.4 Citations for Evidence of High Impact cited in 1a.3: (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;121(6):e1653-e1659. (3) Camp M, Chang DC, Zhang Y, Chrouser K, Colombani PM, Abdullah F. Risk factors and outcomes for foreign body left during a procedure: analysis of 413 incidents after 1 946 831 operations in children. Arch Surg 2010;145(11):1085-1090. 1b. Opportunity for Improvement: H M L I (There is a demonstrated performance gap - variability or overall less than optimal performance) 1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure: This indicator is intended to flag cases of a foreign body accidentally left in a patient during a procedure. **1b.2 Summary of Data Demonstrating Performance Gap** (Variation or overall less than optimal performance across providers): [For Maintenance – Descriptive statistics for performance results for this measure - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc. This is a "sentinel event" indicator, which is described as a count rather than a rate. Accordingly, it is impossible to provide a distribution of scores across measured entities. 1b.3 Citations for Data on Performance Gap: [For Maintenance – Description of the data or sample for measure results reported in 1b.2 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included Source: Agency for Healthcare Research and Quality (AHRQ), Center for Delivery, Organization, and Markets, Healthcare Cost and Utilization Project, Nationwide Inpatient Sample, 2007, and AHRQ Quality Indicators, version 3.1. 1b.4 Summary of Data on Disparities by Population Group: [For Maintenance – Descriptive statistics for performance results for this measure by population group This is a "sentinel event" indicator, which is described as a count rather than a rate. Accordingly, it is impossible to provide descriptive statistics for performance results across population groups.

1b.5 Citations for Data on Disparities Cited in 1b.4: [For Maintenance – Description of the data or sample for measure results

1	reported in 1b.4 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities							
<i>included</i>] Source: Agency for Healthcare Research and Quality (AHRQ), Center for Delivery, Organization, and Markets, Healthcare Cost and Utilization Project, Nationwide Inpatient Sample, 2007, and AHRQ Quality Indicators, version 3.1.								
1c. Evidence (Measure focus is a health outcome OR meets the criteria for quantity, quality, consistency of the body of evidence.) Is the measure focus a health outcome? Yes No If not a health outcome, rate the body of evidence.								
Quantity: H M	Quantity: H M L I Quality: H M L I Consistency: H M L I							
Quantity Quality	Consistency	Does the measure pass subcriterion1c?						
M-H M-H	м-н	Yes						
L M-H	М	Yes IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise No						
M-H L	M-H	Yes IF potential benefits to patients clearly outweigh potential harms: otherwise No						
L-M-H L-M-H	L	No 🗌						
		s relationship to at least tervention, or service Does the measure pass subcriterion1c?						
Ine healthcare structure, process, intervention, or service Yes IF rationale supports relationship 1c.1 Structure-Process-Outcome Relationship (Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, intermediate clinical outcome, intermediate clinical outcome, intermediate clinical outcome-health outcome): Foreign body left during procedure is a health outcome measure. This measure captures illness or injury resulting from a foreign body that was left during a procedure, when the retention of that foreign body was not part of the originally planned procedure. These events are considered to be almost entirely preventable. For example, the 2011 Update of the NQF Serious Reportable Events in Healthcare includes this specification of "Unintended retention of a foreign object in a patient after surgery or other invasive procedure" "includes medical or surgical items intentionally placed by provider(s) that are unintentionally left in place. Excludes (a) objects present prior to surgery or other rinvasive procedure that are intentionally left in place. (b) objects intentionally left in mixer. Commission. It is also classified as a "hospital-acquired condition" by the Center for Medicare & Medicaid Services, based on a similar ICD-9-CM specification as PDI 03. Retained surgical items have become a major focus of safety improvement effort within health care organizations. Current thinking is summarized as follows by The Joint Commission: "The consequences extend beyond clinical complications and often include additional financial burdens such as extended lost time from with, additional expenses related to frequent follow up visits, and additional medicatons. Such complications also present a significant financial isk to he								

distractions (e.g., extraneous conversation, music, unnecessary interruptions). The circulator must be a registered nurse [R]. Radiographic imaging is not a substitute for performing accurate count procedures. Count procedures may be omitted or modified in an extreme patient emergency. This exception will be documented in the patient's medical record and when the patient's condition allows, radiographic imaging should be obtained to rule out the possibility of an unintentionally retained foreign object. "Perioperative protocol. Health care protocol. 2009 Sep (revised 2010 Oct). NGC:008119 Institute for Clinical Systems Improvement See also recent publications and guidelines from the "Nothing Left Behind" project (http://nothingleftbehind.org/) and the Association of Perioperative Registered Nurses, which offers "Recommended Practices for Prevention of Retained Surgical Items" (http://www.aorn.org/News/August2010News/RSI).

Our literature review identified four empirical studies supporting the relationship of this health outcome to processes of care. In the medical field, one research team used a computer-assisted material testing machine to study the mechanical properties of three different types of epidural catheters: polyurethane, radiopaque and clear nylon. In this blinded study, polyurethane was the least prone to breakage, whereas radiopaque catheters had the highest stiffness (Ates, Yucesoy et al. 2000).

At least two case-control studies have evaluated risk factors associated with retained objects after major surgery. The first study involved retrospective review of medical records flagged using the PDI 03 numerator code. In the multivariate analysis, factors associated with a significantly higher risk of retained foreign bodies were the total number of major procedures performed [OR 1.6; 95% CI 1.1-2.3, P=0.008] and an incorrect count [OR 16.2; 95% CI 1.3-197.8; P=0.02], highlighting the importance of ending surgery with correct instrument and sponge counts (Lincourt, Harrell et al. 2007). In the second study, cases were drawn from a large malpractice insurer representing one third of the physicians in Massachusetts from 1985 to 2001. Emergency surgery [RR 8.8; 95% CI 2.4-31.9], unplanned change in the operation [RR 4.1; 95% CI 1.4-12.4], and body-mass index [RR for each one unit interval 1.1; 95% CI 1.0-1.2] were among the significant risk factors for retained objects (Gawande, Studdert et al. 2003). Finally, in a review of lawsuits claiming retained objects during surgery, 10% involved procedures during which no sponge count was performed. In the case of non-vaginal surgery, a falsely correct sponge count was documented in 76% of the claims (Kaiser, Friedman et al. 1996). Falsely correct sponge counts were attributed to team fatigue, difficult operations, sponges "sticking together," or a poor counting system. Incorrect sponge counts that were accepted prior to closure resulted from either surgeons' dismissing the incorrect count without re-exploring the wound, or nursing staff allowing an incorrect count to be accepted. Ates, Y., C. A. Yucesoy, et al. (2000). "The mechanical properties of intact and traumatized epidural catheters." Anesth Analg 90(2): 393-9.

Gawande, A. A., D. M. Studdert, et al. (2003). "Risk factors for retained instruments and sponges after surgery." N Engl J Med 348(3): 229-35.

Kaiser, C. W., S. Friedman, et al. (1996). "The retained surgical sponge." Ann Surg 224(1): 79-84.

Lincourt, A. E., A. Harrell, et al. (2007). "Retained foreign bodies after surgery." J Surg Res 138(2): 170-4.

The exact proportion of PDI 03 events that is preventable, with optimal surgical technique, is unknown. However, in one series of 45 confirmed cases from 28 participating hospitals in the National Association of Children's Hospitals and Related Institutions (Scanlon MC, Harris JM II, Levy F, et al. Evaluation of the agency for healthcare research and quality pediatric quality indicators. Pediatrics 2008; 121:e1723–31), 20 (44%) were deemed preventable, 9 (20%) were deemed nonpreventable, and 16 (36%) were characterized as having uncertain preventability. This determination was made independently by clinicians at each site, who lacked formal training but were guided by teleconference discussions. The authors concluded that the average children's hospital in the US reports 0.2-0.3 preventable PDI 01 events each year. In a previous review of 45 cases from 14 children's hospitals (Scanlon MC, Miller M, Harris JM, Schulz K, Sedman A. Targeted chart review of pediatric patient safety events identified by the Agency for Healthcare Research and Quality's patient safety indicators methodology. J Patient Saf 2006; 2:191-7), using similar methods, 51% were deemed preventable, 27% were deemed unpreventable, and 22% were classified as "unable to determine."

1c.2-3 Type of Evidence (Check all that apply): Clinical Practice Guideline

1c.4 Directness of Evidence to the Specified Measure (State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population): Not applicable

1c.5 Quantity of Studies in the Body of Evidence (Total number of studies, not articles): Not applicable

1c.6 Quality of Body of Evidence (Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b)

directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events): Not applicable

1c.7 Consistency of Results across Studies (Summarize the consistency of the magnitude and direction of the effect): Not applicable

1c.8 Net Benefit (Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms): Not applicable

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? No

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: Not applicable

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: Not applicable

1c.13 Grade Assigned to the Body of Evidence: Not applicable

1c.14 Summary of Controversy/Contradictory Evidence: Not applicable

1c.15 Citations for Evidence other than Guidelines (Guidelines addressed below): Not applicable

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> (Including guideline # and/or page #): Not applicable

1c.17 Clinical Practice Guideline Citation: Not applicable

1c.18 National Guideline Clearinghouse or other URL: Not applicable

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? No

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: Not applicable

1c.23 Grade Assigned to the Recommendation: Not applicable

1c.24 Rationale for Using this Guideline Over Others: Not applicable

Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: Moderate 1c.26 Quality: Moderate1c.27 Consistency: Moderate

Was the threshold criterion, *Importance to Measure and Report*, met? (1a & 1b must be rated moderate or high and 1c yes) Yes No Provide rationale based on specific subcriteria:

For a new measure if the Committee votes NO, then STOP. For a measure undergoing endorsement maintenance, if the Committee votes NO because of 1b. (no opportunity for

improvement), it may be considered for continued endorsement and all criteria need to be evaluated.

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, <u>as specified</u>, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See <u>guidance on measure testing</u>.

S.1 Measure Web Page (In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained). Do you have a web page where current detailed specifications for <u>this</u> measure can be obtained? Yes

S.2 If yes, provide web page URL: http://qualityindicators.ahrq.gov/modules/pdi_resources.aspx

2a. RELIABILITY. Precise Specifications and Reliability Testing: H M L

2a1. Precise Measure Specifications. (The measure specifications precise and unambiguous.)

2a1.1 Numerator Statement (Brief, narrative description of the measure focus or what is being measured about the target population, e.g., cases from the target population with the target process, condition, event, or outcome): Discharges under age 18 with ICD-9-CM codes for foreign body left in during procedure in any secondary diagnosis field of medical and surgical discharges defined by specific DRGs or MS-DRGs where several exclusions are applied to the numerator. (Details of the numerator, medical and surgical discharges DRGs and MS-DRGs, and exclusions appear in 2a1.3). **2a1.2 Numerator Time Window** (The time period in which the target process, condition, event, or outcome is eligible for inclusion): User may specify the time window; generally one calendar year 2a1.3 Numerator Details (All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptors, and/or specific data collection items/responses: ICD-9-CM Foreign body left in during procedure diagnosis codes: 9984 FOREIGN BODY ACCIDENTALLY LEFT DURING A PROCEDURE 9987 ACUTE REACTIONS TO FOREIGN SUBSTANCE ACCIDENTALLY LEFT DURING A PROCEDURE Foreign body left in during: E8710 SURGICAL OPERATION

E8711 INFUSION OR TRANSFUSION E8712 KIDNEY DIALYSIS OR OTHER PERFUSION E8713 INJECTION OR VACCINATION E8714 ENDOSCOPIC EXAMINATION E8715 ASPIRATION OF FLUID OR TISSUE, PUNCTURE, AND CATHETERIZATION E8716 HEART CATHETERIZATION E8717 **REMOVAL OF CATHETER OR PACKING** E8718 OTHER SPECIFIED PROCEDURES E8719

UNSPECIFIED PROCEDURE See Pediatric Quality Indicators Appendices: - Appendix B – Surgical DRGs - Appendix C – Surgical MS-DRGs - Appendix D – Medical DRGs - Appendix E – Medical MS-DRGs Numerator exclusions: - with ICD-9-CM codes for foreign body left in during procedure in the principal diagnosis field or secondary diagnosis field if present on admission - normal newborn - newborns weighing less than 500 grams (Birth Weight Category 1) - MDC 14 (pregnancy, childbirth, and puerperium) - with missing discharge gender (SEX=missing), age (AGE=missing), guarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1=missing) See Pediatric Quality Indicators Appendices: - Appendix I - Definitions of Neonate, Newborn, Normal Newborn, and Outborn - Appendix L – Low Birth Weight Categories Link to PDI appendices: http://qualityindicators.ahrg.gov/Downloads/Software/SAS/V43/TechnicalSpecifications/PDI%20Appendices.pdf 2a1.4 Denominator Statement (Brief, narrative description of the target population being measured): Not applicable 2a1.5 Target Population Category (Check all the populations for which the measure is specified and tested if any): Children's Health **2a1.6 Denominator Time Window** (The time period in which cases are eligible for inclusion): Not applicable 2a1.7 Denominator Details (All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses): Not applicable **2a1.8 Denominator Exclusions** (Brief narrative description of exclusions from the target population): Not applicable 2a1.9 Denominator Exclusion Details (All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses): Not applicable 2a1.10 Stratification Details/Variables (All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses): Not applicable 2a1.11 Risk Adjustment Type (Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13): No risk adjustment or risk stratification 2a1.12 If "Other," please describe: 2a1.13 Statistical Risk Model and Variables (Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.):

Not applicable

2a1.14-16 Detailed Risk Model Available at Web page URL (or attachment). Include coefficients, equations, codes with descriptors, definitions, and/or specific data collection items/responses. Attach documents only if they are not available on a webpage and keep attached file to 5 MB or less. NQF strongly prefers you make documents available at a Web page URL. Please supply login/password if needed:

2a1.17-18. Type of Score: Count

2a1.19 Interpretation of Score (Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score): Better quality = Lower score

2a1.20 Calculation Algorithm/Measure Logic(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.):

Identify cases meeting the target outcome. Exclude cases meeting the exclusion criteria. Count the number of case at the hospital level.

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment:

URL

http://qualityindicators.ahrq.gov/Downloads/Resources/Publications/2011/QI%20Empirical%20Methods%2005-03-11.pdf Not applicable

2a1.24 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

2a1.25 Data Source (Check all the sources for which the measure is specified and tested). If other, please describe: Administrative claims

2a1.26 Data Source/Data Collection Instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.*): HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality, Rockville, MD.

2a1.27-29 Data Source/data Collection Instrument Reference Web Page URL or Attachment: URL http://www.hcup-us.ahrq.gov/sidoverview.jsp Not applicable

2a1.30-32 Data Dictionary/Code Table Web Page URL or Attachment:

URL

http://qualityindicators.ahrq.gov/Downloads/Software/WinQI/V43/AHRQ%20QI%20Software%20Instructions,%20WinQI.pdf Not applicable

2a1.33 Level of Analysis (Check the levels of analysis for which the measure is specified and tested): Facility

2a1.34-35 Care Setting (Check all the settings for which the measure is specified and tested): Hospital/Acute Care Facility

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample (Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included):

HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2008. Agency for Healthcare Research and Quality, Rockville, MD. Includes approximately 6 million pediatric discharges for 2,500 hospitals.

2a2.2 Analytic Method (Describe method of reliability testing & rationale): Not applicable

2a2.3 Testing Results (*Reliability statistics, assessment of adequacy in the context of norms for the test conducted*): Because this indicator is expressed as a count, no reliability testing was conducted.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H M L I

2b1.1 Describe how the measure specifications (measure focus, target population, and exclusions) are consistent with the evidence cited in support of the measure focus (criterion 1c) and identify any differences from the evidence: Similar to the NQF Serious Reportable Event for "Unintended retention of a foreign object in a patient after surgery or other invasive procedure". PDI 03 is intended to capture "occurrences of unintended retention of objects at any point after the surgery/procedure ends regardless of setting (post anesthesia recovery unit, surgical suite, emergency department, patient bedside) and regardless of whether the object is to be removed after discovery" and "unintentionally retained objects (including such things as wound packing material, sponges, catheter tips, trocars, guide wires) in all applicable settings." The Editorial Advisory Board for Coding Clinics for ICD-9-CM reinforced this alignment in 2009 (26(1):18) by instructing the coding community to follow current NQF guidance related to when surgery ends in assigning the ICD-9-CM code 998.4 (Foreign body accidentally left during a procedure). Specifically, according to the 2011 update, "Surgery ends after all incisions or procedural access routes have been closed in their entirety," device(s) such as probes or instruments have been removed, and, if relevant, final surgical counts confirming accuracy of counts and resolving any discrepancies have concluded and the patient has been taken from the operating/procedure room." The coding community may not yet be aware of NQF's 2011 change to the definition of when surgery ends, because the previous definition permitted a wider range of interpretation ("Surgery ends after counts have concluded, the surgical incision has been closed, and/or operative device(s) such as probes have been removed, regardless of setting"), but educational efforts are now underway. However, there is one substantive difference between the PDI 03 measure specification and the corresponding NQF SRE specification. The NQF SRE specification excludes "objects not present prior to surgery/procedure that are intentionally left in when the risk of removal exceeds the risk of retention (such as microneedles, broken screws)." ICD-9-CM does not recognize this exclusion, based on the principle that retention of the foreign body was not part of the originally planned procedure (and was not the original intent of the surgeon), and therefore it represents a complication of that procedure, even if the surgeon later decides that the risk of removal exceeds the risk of retention. See this guestion and answer in Coding Clinics for ICD-9-CM (1st guarter 2011): "During surgery a needle was placed along the right lateral aspect of the ring of the aortic valve and passed through tissue deep in the recess between the aortic root and the left atrium to try to cinch the valve down. The suture broke from the needle and the needle was lost within this tissue. Multiple attempts to find it were unsuccessful. An x-ray was obtained which did not reveal the needle. The chest was closed and a second x-ray showed that the needle was where the surgeon thought, to the right of the aortic valve. The chest was reopened but the needle still could not be located or palpated. The surgeon decided that continuing to look for the needle was likely to cause the patient more harm than good. Therefore, the chest was reapproximated... Assign code 998.4, Foreign body accidentally left during a procedure. Although the surgeon made the decision to leave the needle because continuing to search for it might cause harm to the patient, it was not the intent of the original procedure to leave a foreign body behind."

2b2. Validity Testing. (Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)

2b2.1 Data/Sample (Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included):

The most recent study of the criterion validity of PDI 03 was based on a consecutive sample of 72 flagged cases from 28 participating hospitals in the National Association of Children's Hospitals and Related Institutions (NACHRI) from 2003 through 2005 (Scanlon MC, Harris JM II, Levy F, et al. Evaluation of the agency for healthcare research and quality pediatric quality indicators. Pediatrics 2008; 121:e1723–31). Records were reviewed independently by clinicians at each site, who lacked formal training but were guided by teleconference discussions. A previous review of 45 flagged cases from 14 self-selected children's hospitals in the NACHRI Pediatric PSI Collaborative (Scanlon MC, Miller M, Harris JM, Schulz K, Sedman A. Targeted chart review of pediatric patient safety events identified by the Agency for Healthcare Research and Quality's patient safety indicators methodology. J Patient Saf 2006; 2:191-7) used similar methods.

2b2.2 Analytic Method (Describe method of validity testing and rationale; if face validity, describe systematic assessment): Forty-four distinct professional clinical organizations and hospital associations were invited to submit nominations. These organizations were selected based on the applicability of the specialty or subspecialty to the candidate quality indicators. Nineteen

organizations submitted nominations: Ambulatory Pediatric Association, American Academy of Allergy Asthma and Immunology, American Academy of Family Physicians, American Academy of Pediatrics, American College of Chest Physicians, American College of Nurse-Midwives, American Society of Pediatric Hematology/Oncology, American Society of Pediatric Nephrology, California Academy of Family Physicians, Child Health Corporation of America, National Association of Children's Hospitals and Related Institutions, National Association of Pediatric Nurse Practitioners, Pediatric Infectious Diseases Society, Society for Academic Emergency Medicine, Society for Adolescent Medicine, Society for Pediatric Anesthesia, Society of Critical Care Medicine, Society of Pediatric Nurses, and Society of Thoracic Surgeons.

These professional organizations nominated a total of 125 clinicians. All nominees were invited to participate, if eligible, in the evaluation of indicators available in Phase I and Phase II. In order to be eligible to participate, nominees were required to spend at least 30% of their work time on patient care, including hospitalized patients. From the 70 nominees accepting the invitation; five clinicians were ineligible to participate. Nominees were asked to provide information regarding their practice characteristics, including specialty, subspecialty, and setting (i.e., urban vs. rural location, region of country, and service to underserved populations), primary hospital of practice (i.e., funding source), and involvement in education (i.e., clinical training, academic affiliation).

To ensure appropriate clinical expertise on each panel, we identified the specialties that would be required to properly evaluate the indicators assigned to that panel. Panelists were selected so that each panel had diverse membership in terms of practice characteristics and setting. Thus, when a specific geographic area or type of clinician (e.g. academic) was over-represented by the pool of eligible nominees, randomly drawn members from that specific sub-group were contacted first to fill the panels. In addition, conference call scheduling logistics influenced assignments. From the 65 eligible nominees, 45 individuals accepted our invitation to participate on a specific panel.

Four panels were formed to evaluate indicators grouped as follows: Medical and surgical indicators, surgical only indicators, neonatal indicators and prevention indicators. All panels had diversity in the geographic location of panelists, and their type of practice.

2b2.3 Testing Results (Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment):

Face validity was systematically assessed using an expert panel process, as described in our original submission documents (McDonald K, Romano P, Davies S, Haberland C, Geppert J, Ku A, Choudry K. Measures of Pediatric Health Care Quality Based on Hospital Administrative Data: The Pediatric Quality Indicators. Rockville, MD: Agency for Healthcare Research and Quality, 2006). Specifically, this indicator was reviewed by a pediatric specialty panel with eleven pediatric clinicians, including one general pediatrician, one hospitalist, one critical care physician, one neonatologist, one infectious disease specialist, one hematologist, one cardiothoracic surgeon, one emergency medicine specialist, one interventional radiologist, and two surgeons. Median ratings were 8 (on a scale of 1-9) with agreement on usefulness for internal quality improvement, 8 with agreement for comparative reporting, and 8 with agreement for preventability.

The larger, more recent study of criterion validity, published in 2008, estimated a PPV of 63%, which is better than the PPV estimates for the adult version of this indicator (e.g., 45% [95% CI 35% to 56%] and 52%). This finding may not be surprising if children are less likely than adults to have retained foreign bodies from prior surgical interventions or trauma. Fewer details are reported from the earlier (2006) study, but Table 1 in that paper suggests a PPV of at least 51% to 73%.

Some false positives were due to complications that were actually present on admission (i.e., 11 of 27 false positives in the NACHRI study), which would automatically be excluded by users with "present on admission" (POA) data. Adjusting for the availability of POA data, the estimated PPV in the 2008 NACHRI study was 74%. The remaining false positives were described as drains or sutures that were purposely left in the body.

Due to the extreme rarity of this event, no studies have been able to assess the sensitivity of PDI 03. However, user feedback regarding internal review of flagged cases is continuously solicited, and has not revealed major concerns about false negative cases.

POTENTIAL THREATS TO VALIDITY. (All potential threats to validity were appropriately tested with adequate results.)

2b3. Measure Exclusions. (Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.)

2b3.1 Data/Sample for analysis of exclusions (Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included): Exclude cases:

with ICD-9-CM codes for foreign body left in during procedure in the principal diagnosis field or secondary diagnosis field if present

on admission

- normal newborn
- newborns weighing less than 500 grams (Birth Weight Category 1)
- MDC 14

- cases with select fields with missing data: missing gender (SEX=missig, age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1=missing)

2b3.2 Analytic Method (Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference):

Count of cases with principal diagnosis field or secondary diagnosis field if present on admission

If the user's data lacks present on admission information, then the likelihood that the outcome of interest and the covariates are present on admission is estimated using a Markov Chain Monte Carlo (MCMC) estimation procedure. That likelihood is then used to adjust the observed count.

2b3.3 Results (Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses): Of 51 cases identified with the outcome of interest, 24 were present on admission

2b4. Risk Adjustment Strategy. (For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.)

2b4.1 Data/Sample (Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included): Not applicable

2b4.2 Analytic Method (Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables):

Not applicable

2b4.3 Testing Results (Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot. and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata): Not applicable

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: Not applicable

2b5. Identification of Meaningful Differences in Performance. (The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.)

2b5.1 Data/Sample (Describe the data or sample including number of measured entities: number of patients: dates of data; if a sample, characteristics of the entities included):

HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2008. Agency for Healthcare Research and Quality, Rockville, MD. Includes approximately 6 million pediatric discharges for 2,500 hospitals.

2b5.2 Analytic Method (Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance):

Sum the number of cases identified with the outcome of interest by hospital

2b5.3 Results (Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance):

The measure is a serious reportable event so differences in peformance are not applicable

2b6. Comparability of Multiple Data Sources/Methods. (If specified for more than one data source, the various approaches result in comparable scores.)

2b6.1 Data/Sample (Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included): Not applicable

2b6.2 Analytic Method (Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure): Not applicable

2b6.3 Testing Results (Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted):

 Not applicable

 2c. Disparities in Care:
 H M L I NA (If applicable, the measure specifications allow identification of disparities.)

 2c.1 If measure is stratified for disparities, provide stratified results (Scores by stratified categories/cohorts): In regard to figures below:

 1st figure: estimate

 2nd figure: standard error

 3rd figure: p value relative to marked group (marked group = "c")

 4th figure: p value: current year relative to prior year

 Key:

 "c":

 Reference for p-value test statistics

 "*":

 Data do not meet criteria for statistical reliability, data quality, or confidentiality

 "DNC":

 Patient characteristic:

 Age groups for pediatric conditions

 0-4 c.0.027 0.007 DNC

0-4 c 0.027 0.007 DNC 5-9 * * DNC 10-14 * * * DNC 15-17 * * * DNC

Gender: Male c 0.037 0.010 DNC Female * * * DNC

Median income of patient's ZIP code: First quartile (lowest income) 0.052 0.015 * DNC Second quartile * * * DNC Third quartile * * * DNC Fourth quartile (highest income) c * * DNC

Location of patient residence (NCHS): Large central metropolitan 0.055 0.016 * DNC Large fringe metropolitan c ** DNC Medium metropolitan *** DNC Small metropolitan *** DNC Micropolitan *** DNC Not metropolitan or micropolitan *** DNC

Expected payment source: Private insurance c 0.038 0.011 DNC Medicare * * * DNC Medicaid 0.035 0.010 0.823 DNC Other insurance * * * DNC Uninsured / self-pay / no charge * * * DNC

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:

Not applicable

2.1-2.3 Supplemental Testing Methodology Information: URL

http://qualityindicators.ahrq.gov/Downloads/Modules_Non_Software/Modules%20Development%20Bullet/psi_development.zip Not applicable

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met? (*Reliability and Validity must be rated moderate or high*) Yes No Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

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)

C.1 Intended Purpose/ Use (Check all the purposes and/or uses for which the measure is intended): Public Reporting, Quality Improvement (Internal to the specific organization)

3.1 Current Use (Check all that apply; for any that are checked, provide the specific program information in the following *questions*): Public Reporting, Quality Improvement (Internal to the specific organization)

3a. Usefulness for Public Reporting: H M L I I (*The measure is meaningful, understandable and useful for public reporting.*)

3a.1. Use in Public Reporting - disclosure of performance results to the public at large (*If used in a public reporting program, provide name of program(s), locations, Web page URL(s)).* <u>If not publicly reported in a national or community program</u>, state the reason AND plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement: [*For <u>Maintenance</u> – If not publicly reported, describe progress made toward achieving disclosure of performance results to the public at large and expected date for public reporting; provide rationale why continued endorsement should be considered.*]

This measure is used for public reporting in 3 realms.

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

Iowa (Iowa Healthcare Collaborative) Iowa Healthcare Collaborative http://www.ihconline.org/aspx/publicreporting/iowareport.aspx

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

3a.2.Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. <u>If usefulness was demonstrated</u> (e.g., focus group, cognitive testing), describe the data, method, and results: 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
3.2 Use for other Accountability Functions (payment, certification, accreditation). If used in a public accountability program, provide name of program(s), locations, Web page URL(s): Not applicable
3b. Usefulness for Quality Improvement: H M L L I (<i>The measure is meaningful, understandable and useful for quality improvement.</i>)
 3b.1. Use in QI. If used in quality improvement program, provide name of program(s), locations, Web page URL(s): [For <u>Maintenance</u> – If not used for QI, indicate the reasons and describe progress toward using performance results for improvement]. The Pediatric Quality Indicators (PDIs) are a set of measures that can be used with hospital inpatient discharge data to provide a perspective on the quality of pediatric healthcare. Specifically, PDIs screen for problems that pediatric patients experience as a result of exposure to the healthcare system and that may be amenable to prevention by changes at the system or provider level.
Development of quality indicators for the pediatric population involves many of the same challenges associated with the development of quality indicators for the adult population. These challenges include the need to carefully define indicators using administrative data, establish validity and reliability, detect bias and design appropriate risk adjustment, and overcome challenges of implementation and use. However, the special population of children invokes additional, special challenges. Four factors— differential epidemiology of child healthcare relative to adult healthcare, dependency, demographics, and development—can pervade all aspects of children's healthcare; simply applying adult indicators to younger age ranges is insufficient.
This PDIs focus on potentially preventable complications and iatrogenic events for pediatric patients treated in hospitals, and on preventable hospitalizations among pediatric patients.
The PDIs apply to the special characteristics of the pediatric population; screen for problems that pediatric patients experience as a result of exposure to the healthcare system and that may be amenable to prevention by changes at the provider level or area level; and, help to evaluate preventive care for children in an outpatient setting, and most children are rarely hospitalized. http://qualityindicators.ahrq.gov/modules/pdi_overview.aspx
The following are several entities that use the measure in quality improvement:
1) Child Health Corporation of America (CHCA) CHCA reports performance in all PDIs to its 42 member hospitals for their tracking and use in quality improvement. CHCA members are large freestanding pediatric hospitals.
 National Association of Children's Hospitals and Related Institutions (NACHRI) As a benefit of membership, NACHRI reports all provider level PDIs to its approximately 85 member children's hospitals for their quality improvement applications.
3) University Healthcare Consortium (UHC) UHC is an alliance of 103 academic medical centers and 219 of their affiliated hospitals. UHC reports this and other AHRQ QIs to their member hospitals for their internal quality improvement purposes.

3b.2. Provide rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. If usefulness was demonstrated (e.g., QI initiative), describe the data, method and results: The AHRQ QI support line receives approximately 150 user gueries per month and almost 50 user per month download the AHRQ QLPDI software. Users have used the PDI since the release in 2006. Users can identify cases with the serious reportable event and conduct root cause analysis Overall, to what extent was the criterion, Usability, met? H M L I Provide rationale based on specific subcriteria: 4. FEASIBILITY Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria) 4a. Data Generated as a Byproduct of Care Processes: H M L I 4a.1-2 How are the data elements needed to compute measure scores generated? (Check all that apply). Data used in the measure are: Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims) 4b. Electronic Sources: H M L I 4b.1 Are the data elements needed for the measure as specified available electronically (Elements that are needed to compute measure scores are in defined, computer-readable fields): ALL data elements in electronic claims 4b.2 If ALL data elements are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources: 4c. Susceptibility to Inaccuracies, Errors, or Unintended Consequences: H M L I 4c.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results: Coding professionals follow detail guidelines, are subject to training and credentialing requirements, peer review and audit. 4d. Data Collection Strategy/Implementation: H M L I A.2 Please check if either of the following apply (regarding proprietary measures): 4d.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 and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues (e.g., fees for use of proprietary measures): The AHRQ QI software has been publicly available at no cost since 2001; Users have over ten years of experience using the AHRQ QI software in SAS and Windows Overall, to what extent was the criterion, Feasibility, met? H M L I Provide rationale based on specific subcriteria: **OVERALL SUITABILITY FOR ENDORSEMENT** Does the measure meet all the NQF criteria for endorsement? Yes No Rationale: If the Committee votes No. STOP.

If the Committee votes Yes, the final recommendation is contingent on comparison to related and competing measures.

5. COMPARISON TO RELATED AND COMPETING MEASURES

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure before a final recommendation is made.

5.1 If there are related measures (*either same measure focus or target population*) or competing measures (*both the same measure focus and same target population*), list the NQF # and title of all related and/or competing measures:

5a. Harmonization

5a.1 If this measure has EITHER the same measure focus OR the same target population as <u>NQF-endorsed measure(s)</u>: Are the measure specifications completely harmonized?

5a.2 If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden:

5b. Competing Measure(s)

5b.1 If this measure has both the same measure focus and the same target population as NQF-endorsed measure(s): Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (*Provide analyses when possible*):

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner): Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Co.2 Point of Contact: John, Bott, Contractor, AHRQ Quality Indicators Measure Expert Center for Delivery, Organization and Markets, John.Bott@ahrq.hhs.gov, 301-427-1317-

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

Co.4 Point of Contact: John, Bott, Contractor, AHRQ Quality Indicators Measure Expert Center for Delivery, Organization and Markets, John.Bott@ahrq.hhs.gov, 301-427-1317-

Co.5 Submitter: John, Bott, Contractor, AHRQ Quality Indicators Measure Expert Center for Delivery, Organization and Markets, John.Bott@ahrq.hhs.gov, 301-427-1317-, Agency for Healthcare Research and Quality

Co.6 Additional organizations that sponsored/participated in measure development: University of California-Davis Stanford University Battelle Memorial Institute

Co.7 Public Contact: John, Bott, Contractor, AHRQ Quality Indicators Measure Expert Center for Delivery, Organization and Markets, John.Bott@ahrq.hhs.gov, 301-427-1317-, Agency for Healthcare Research and Quality

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.

Multi-specialty Panel and Surgical Panel members are listed in the technical report: http://qualityindicators.ahrq.gov/Downloads/Modules_Non_Software/Modules%20Development%20Bullet/pdi_development.zip

Ad.2 If adapted, provide title of original measure, NQF # if endorsed, and measure steward. Briefly describe the reasons for adapting the original measure and any work with the original measure steward: This indicator was originally proposed by

lezzoni et al. as part of the Complications Screening Program (CSP "sentinel events")

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.3 Year the measure was first released: 2006

Ad.4 Month and Year of most recent revision: 08, 2011

Ad.5 What is your frequency for review/update of this measure? Annual

Ad.6 When is the next scheduled review/update for this measure? 12, 2011

Ad.7 Copyright statement: Not applicable

Ad.8 Disclaimers: Not applicable

Ad.9 Additional Information/Comments: Not applicable

Date of Submission (MM/DD/YY): 09/14/2011