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 submitting standards web page.

<table>
<thead>
<tr>
<th><strong>NQF #: 1769</strong></th>
<th><strong>NQF Project:</strong> Perinatal and Reproductive Health Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>(for Endorsement Maintenance Review)</td>
<td></td>
</tr>
<tr>
<td><strong>Original Endorsement Date:</strong></td>
<td><strong>Most Recent Endorsement Date:</strong></td>
</tr>
</tbody>
</table>

**BRIEF MEASURE INFORMATION**

**De.1 Measure Title:** Adverse Outcome Index

**Co.1 Measure Steward:** Beth Israel Deaconess Medical Center

**De.2 Brief Description of Measure:** The rate and severity of adverse events in the obstetric population during their delivery hospitalization

**2a1.1 Numerator Statement:** Any delivery with one or more of the adverse events

**2a1.4 Denominator Statement:** Total deliveries occurring during the time frame under review

**2a1.8 Denominator Exclusions:** None

**1.1 Measure Type:** Composite

**2a1.25-26 Data Source:** Administrative claims, Electronic Clinical Data, Electronic Clinical Data: Electronic Health Record, Electronic Clinical Data: Pharmacy, Paper Records

**2a1.33 Level of Analysis:** Clinician: Team, 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):** Adverse Outcome Index

**STAFF NOTES** *(issues or questions regarding any criteria)*

**Comments on Conditions for Consideration:**

**E.4 If component measures of the composite are aggregate-level measures, all must be either NQF-endorsed or submitted for consideration for NQF endorsement** Some or all component measures are not NQF-endorsed and have been submitted using the online measure submission tool

**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 *endorsed* or submitted measures *(check 5.1)*:

**Other Criteria:**

**Staff Reviewer Name(s):**

**1. IMPACT, OPPORTUNITY, EVIDENCE - 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.* (composite measure evaluation criteria)

(for NQF staff use) Specific NPP goal:

| 1d.1 Describe the purpose/objective of the composite measure: | The purpose of this measure is to allow for standardized measurement of adverse events in obstetrics that are related to care provided in labor and delivery so that the impact of quality of improvement efforts can be determined. |
| 1d.2 Describe the quality construct used in developing the composite: | Each of the ten measures represent an adverse outcome of clinical care during the delivery process. Individually the rates of these adverse events can be very low, such that getting a reasonable picture of a hospital’s obstetric quality would require many observations or need to be measured over an extended period of time. Such a requirement does not lend itself to tracking and monitoring quality so that corrective interventions can take place on a timely basis. Grouping the ten measures together provides a more complete picture of the continuum of the care to the mother and her infant and the weighting system adds a robust logic to the fact that some of the ten events are clearly more serious than others. |

1e.1 Describe how the component measures/items are consistent with and representative of the quality construct:  
Grouping the ten measures together provides a more complete picture of the continuum of the care to the mother and her infant and the weighting system adds a robust logic to the fact that some of the ten events are clearly more serious than others.

Each of the ten measures represent an adverse outcome of clinical care during the delivery process. Individually the rates of these adverse events can be very low such that getting a reasonable picture of a hospital’s obstetric quality would require many observations or need to be measured over an extended period of time.

If the component measures are combined at the patient level, complete 1a, 1b, and 1c.

If the component measures are combined at the aggregate level, skip to criterion 2, Scientific Acceptability of Measure Properties (individual measures are either NQF-endorsed or submitted individually).

1a. High Impact:  

- **H**: 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.)
- **M**: Moderate Impact
- **L**: Low Impact
- **I**: Insufficient

1a.1 Demonstrated High Impact Aspect of Healthcare:  
Affects large numbers, Frequently performed procedure, 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:  
There are more than 4,000,000 deliveries per year in the US and delivery is the most frequent reason for a hospital admission.

1a.4 Citations for Evidence of High Impact cited in 1a.3:  
NCHS:  
Toward Improving the Outcome of Pregnancy III, 2010 March of Dimes National Office

1b. Opportunity for Improvement:  

- **H**: High Impact (There is a demonstrated performance gap - variability or overall less than optimal performance)
- **M**: Moderate Impact
- **L**: Low Impact
- **I**: Insufficient

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:  
The ability to accurately and easily measure meaningful and reproducible outcomes is fundamental to quality improvement and patient safety efforts. These outcomes must be clinically relevant and occur with a high enough frequency to be responsive to improvement efforts. Without such measures, one can not adequately determine the impact of changes in care have on patients.
The adverse outcome index (~5-9%) has a rate that allows detection of changes in a relative short interval. The severity scoring also allows inferences to be made about the ability to prevent or respond to adverse events in order to lessen the degree of harm. Finally the use of administrative data bases has proven to be accurate and easy, which means that this quality indicator can be measured without significant chart review.

1b.2 Summary of Data Demonstrating Performance Gap (Variation or overall less than optimal performance across providers):

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Subcriterion</th>
<th>Rate the body of evidence.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>IF rationale supports relationship to at least one healthcare structure, process, intervention, or service</td>
<td>Does the measure pass subcriterion1c?</td>
</tr>
<tr>
<td>No</td>
<td>IF additional research unlikely to change conclusion that benefits to patients outweigh harms</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>IF potential benefits to patients clearly outweigh potential harms</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service</td>
<td></td>
</tr>
</tbody>
</table>

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]

6) Hamilton MD, E et al. Revisiting the Perinatal Adverse Outcome Index; Patient Safety and Quality Healthcare, May/June 2011, 837
1c.2-3 Type of Evidence (Check all that apply):
Selected individual studies (rather than entire body of evidence)

1c.4 Exclusions Justified  See individual adverse event submissions

1c.5 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):
The original randomized control study conducted by the Department of Defense (DoD) and Beth Israel Deaconess Medical Center (BIDMC) was an analysis of the impact of team training on adverse events occurring in L&D. The Adverse Outcome Index was developed by an expert panel including representatives from each participating hospital, ACOG, ASOAP, AF Institute of Pathology, US Navy BUMED, Office of Surgeon General and TRICARE. The expert panel identified the 10 adverse events and work with the ACOG Quality Committee to assign the weights. The 10 events and assigned weights are included in the current AOI.

1c.6 Quantity of Studies in the Body of Evidence (Total number of studies, not articles): Original DoD/BIDMC study; we are aware of a number of Collaboratives (Maryland, Premier, Greater NY Hospital Association, North Bronx Healthcare Network, North Shore-LIJ Health System) that have or are using the AOI for QI improvement initiatives and a number of hospitals that have used and reported on their use of the AOI.

1c.7 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): The original DoD/BIDMC study relied on abstracted data and included a Baseline period of two months before team training (intervention) and 5 months post team training. The study found a significant change in only one process measure and while there was a decline in the AOI from the Baseline to Follow-up period, the change was not significant. The study concluded that the time frame to assess the true impact of team training may have been to brief; current experience with the AOI has shown that there is generally a 6-9 month period required to realize the full impact of team training and movement in the AOI. Using administrative data to calculate the AOI, allows hospitals to look at a longer baseline period (1-2 years) and begin follow-up monitoring when the entire team is fully trained and continue monitoring indefinitely.

1c.8 Consistency of Results across Studies (Summarize the consistency of the magnitude and direction of the effect): Use of the AOI has shown remarkable consistency with most facilities improving between Baseline and Follow-up. Beginning performance often shows notable variation and often the hospitals with the highest Baseline AOI will improve the most. Hospitals without strong leadership or that experience a change in leadership often have difficulty moving their rates. A number of hospitals will usually improve their rates by improving the accuracy of their coding or through changes in practice patterns (NICU admits). The studies published using the AOI as a measure of quality have consistently demonstrated that the AOI is responsive to quality improvement efforts, and tracks with other quality measures.

1c.9 Net Benefit (Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms): Reducing the frequency of any of the adverse events is a net benefit and improvement in quality. Hospitals with high AOI rates due to documentation and/or coding issues rather than quality issues have the opportunity to easily correct their administrative data to more accurately reflect the quality of their care—an additional net benefit to the hospital under review.

1c.10 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded?  No

1c.11 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.12 System Used for Grading the Body of Evidence:  USPSTF

1c.13 If other, identify and describe the grading scale with definitions:
NQF #1769 Adverse Outcome Index

1c.14 Grade Assigned to the Body of Evidence:

1c.15 Summary of Controversy/Contradictory Evidence:

1c.16 Citations for Evidence other than Guidelines (Guidelines addressed below):

1c.17 Quote verbatim, the specific guideline recommendation (Including guideline # and/or page #): n/a

1c.18 Clinical Practice Guideline Citation:

1c.19 National Guidelines Clearinghouse or other URL:

1c.20 Grading of Strength of Guideline Recommendation. Has the recommendation been graded?  No

1c.21 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: n/a

1c.22 System Used for Grading the Strength of Guideline Recommendation: Other

1c.23 If other, identify and describe the grading scale with definitions: Expert panel from original study and current workgroup Consensus panels were developed for the identification of measures to be included in the composite measure (AOI). This included vetting each of the individual measures.

Peter E. Nielsen, MD; Marlene B. Goldman, ScD; Susan Mann, MD; David E. Shapiro, Ph.D.; Ronald G. Marcus, MB, BCh.; Stephen D. Pratt, MD; Penny Greenberg, RN; Munish Gupta, MD; Patricia McNamee, RN, MS; Mary Salisbury, RN, MSN; David J. Birnbach, MD; Paul A. Gluck, MD; Mark D. Pearlman, MD; Heidi King, MS; David N. Turnburg, MD, MPH; Benjamin P. Sachs, MB, BS.; Lauren Bales, MD; Naval Medical Center Camp Pendleton; Ronald Burkman, MD, Baystate Medical Center; Cynthia Brumfield, MD University of Alabama at Birmingham; Peter Cherouny, MD, University of Vermont- Fletcher Allen Health Care; Jack Cooley, MD, National Naval Medical Center; Harold Fox, MD, Johns Hopkins Medical Center; Elizabeth Golladay, MD, Tripler Army Medical Center; Lynn Leventis, MD, Naval Medical Center San Diego; Robert Lorenz, MD, William Beaumont Hospital; William Lucky, MD, Baptist Hospital of Miami; Patrick Nugent, MD, South Shore Hospital; Spike Lipschitz, MD, South Shore Hospital; Chris Stolle, MD, Naval Medical Center of Portsmouth; Cosmas van DeVEn, MD, University of Michigan Medical Center; Frank Witter, MD, Johns Hopkins Medical Center, Eileen Hemman, EdD. and Tom Bennettedti, MD, Suzrine Walker, RN, MPH and Thomas Strandjord, MD from the University of Washington. In addition, representatives from the American Congress of Obstetrics and Gynecology, the American Society for Obstetric Anesthesia and Perinatology, the American Society of Anesthesiologists, the Association of Women’s Health, Obstetric and Neonatal Nurses, the Armed Forces Institute of Pathology, the U.S. Navy Bureau of Medicine and Surgery, the Office of the Surgeon General, U.S. Army and TRICARE were present.

National Perinatal Information Center, Providence Rhode Island assisted with translating the AOI into administrative data specifications.

1c.24 Grade Assigned to the Recommendation:

1c.25 Rationale for Using this Guideline Over Others:

Based on the NQF descriptions for rating the evidence, what was the developer’s assessment of the quantity, quality, and consistency of the body of evidence?

1c.26 Quantity: High  1c.27 Quality: High  1c.28 Consistency: High

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

In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained.

| S.1 | Do you have a web page where current detailed specifications for this measure can be obtained? | Yes |
| S.2 | If yes, provide web page URL: | www.npic.org |

### 2a. Precisely Specified

#### 2a.0.1 Components of the Composite.
(List the components, i.e., domains/sub-composites, individual measures. If component measures are NQF-endorsed, include NQF measure number; if not NQF-endorsed, provide date of submission to NQF)

If the composite measure cannot be specified with a numerator and denominator, please consult with NQF staff. If the component measures are combined at the aggregate level, do not include the individual measure specifications below.

#### 2a.1 Composite 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):

Any delivery with one or more of the adverse events

#### 2a.2 Numerator Time Window
(The time period in which the target process, condition, event, or outcome is eligible for inclusion):

During the delivery admission

#### 2a.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: Any delivery or inborn admission with one or more of the following events: #1) in-hospital maternal death (MS Delivery DRGs 765-768 and 774-775 and discharge disp= died), #2) intrapartum neonatal death: inborns only with a discharge disposition of died within 7 days of birth, >= 2500 grams and >=37 weeks excluding congenital anomalies (Dx codes 740-759.9, fetal hydrops-778.0 and dwarfism-259.4); #3) uterine rupture: deliveries, MS DRGs 765-768,774-775, with 665.1 in first or second associated code position only) #4) unplanned maternal admission to the ICU: Delivery MS DRG 765-768,774-775 and dx code with 5th digit =2 and any DX code 640-677 with an ICU day of care or charge OR Delivery MS DRG 765-768,774-775 and dx codes with 5th digit =2 and any DX code 640-677 and discharged to another hospital (UB 04 disp=02)OR Delivery MS DRG 765-768,774-775 and dx codes with 5th digit =2 and any DX code 640-677 and one of the following codes: 96.04,96.05, 96.06,96.7,93.90,93.91,93.93; #5) birth trauma (inborns >=2000 grams with dx codes 767.0,767.11,767.3,767.4,767.5,767.6,767.7, and excluding 767.51; #6) unanticipated operative procedure (Delivery DRGs 765-768,774-775 and one of the following in the first or second associated procedure field: 75.92,69.02,54.61,38.86,39.98,69.52); #7) Admission to the NICU of an inborn within 1 day of birth for greater than a day and >= 2500 grams and >=37 weeks GA (excludes cases with congenital anomalies (Dx codes 740-759.9, fetal hydrops-778.0, dwarfism-259.4 and neonatal abstinence syndrome 779.5) OR Inborns inborn >= 2500 grams and >=37 weeks GA and transferred out (UB 04 disp= 02) within 1 day of birth excluding cases with congenital anomalies (Dx codes 740-759.9, fetal hydrops-778.0, dwarfism-259.4 and neonatal abstinence syndrome 779.5; #8) APGAR 5 < 7, inborn >= 2500 grams and >=37 weeks GA and APGAR 5< 7(excludes cases with congenital anomalies (Dx codes 740-759.9, fetal hydrops-778.0, or dwarfism-259.4; #9) Maternal blood transfusion (Delivery DRGs 765-768,774-775 with procedure code 99.03,99.04,99.05,99.07,99.08 or transfusion indicator =1 (yes); #10) 3rd or 4th degree laceration (Delivery DRGs 765-768,774-775 with dx codes 664.2 and 664.3 SEE ATTACHED ICD 10 CROSSWALK FOR 9 ADVERSE EVENTS WITH ICD 9 CODE SPECIFICATIONS

### 2a.4 Composite Denominator Statement
(Brief, narrative description of the target population being measured):

Total deliveries occurring during the time frame under review

### 2a.5 Target Population Category
(Check all the populations for which the measure is specified and tested if any): Maternal Care

### 2a.6 Denominator Time Window
(The time period in which cases are eligible for inclusion):

During the delivery/birth discharge

### 2a.7 Denominator Details
(All information required to identify and calculate the target population/denominator such as definitions,
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a1.8</td>
<td><strong>Denominator Exclusions</strong> <em>(Brief narrative description of exclusions from the target population):</em> None</td>
</tr>
<tr>
<td>2a1.9</td>
<td><strong>Denominator Exclusion Details</strong> <em>(All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses):</em> n/a</td>
</tr>
<tr>
<td>2a1.10</td>
<td><strong>Stratification Details/Variables</strong> <em>(All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses):</em> n/a</td>
</tr>
<tr>
<td>2a1.11</td>
<td><strong>Risk Adjustment Type</strong> <em>(Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13):</em> No risk adjustment or risk stratification 2a1.12 <em>If &quot;Other,&quot; please describe:</em></td>
</tr>
<tr>
<td>2a1.13</td>
<td><strong>Statistical Risk Model and Variables</strong> <em>(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.):</em> n/a</td>
</tr>
<tr>
<td>2a1.14-16</td>
<td><strong>Detailed Risk Model Available at Web page URL</strong> <em>(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:</em></td>
</tr>
<tr>
<td>2a1.17</td>
<td><strong>Type of Score:</strong> Rate/proportion</td>
</tr>
<tr>
<td>2a1.19</td>
<td><strong>Interpretation of Score</strong> <em>(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):</em> Better quality = Lower score</td>
</tr>
<tr>
<td>2a1.20</td>
<td><strong>Method of Scoring Other</strong></td>
</tr>
<tr>
<td>2a1.21</td>
<td>If &quot;other&quot; scoring method, describe: The adverse outcome index is the percent of deliveries complicated by one or more of the 10 adverse events.</td>
</tr>
<tr>
<td>2a1.22</td>
<td><strong>Missing Component Score</strong> <em>(Indicate how missing component scores are handled):</em> Some events counts (APGAR5&lt;7 or maternal blood transfusions) may not be available on the administrative data set but are always on some other electronic that can be submitted and merged in to the administrative data using a linking variable such as medical record number. Rarely do we have missing data for these variables.</td>
</tr>
<tr>
<td>2a1.23</td>
<td><strong>Weighting:</strong> Differential</td>
</tr>
<tr>
<td>2a1.24</td>
<td>If differential weighting, describe: Each of the ten adverse events has a different weight reflecting severity of the event: Maternal mortality = 750; in-hospital inborn mortality=400; uterine rupture= 100; unplanned maternal ICU admission= 65; birth trauma= 60; maternal unplanned procedure: 40; inborn ICU admission= 35; APGAR 5&lt;7 = 25; maternal blood transfusion= 20; 3rd or 4th degree laceration= 5</td>
</tr>
<tr>
<td>2a1.25</td>
<td><strong>Calculation Algorithm/Measure Logic</strong> <em>(Describe the calculation of the measure score as an ordered sequence of steps)</em></td>
</tr>
</tbody>
</table>
The Adverse Outcome Index (AOI) is a simple rate. It is the percent of deliveries complicated by one or more of the 10 adverse events described above. The Weighted Adverse Outcome Score (WAOS) is calculated by multiplying each event by its weight, summing all weights and dividing by the number of deliveries. The Severity Index (SI) is calculated by multiplying all the events by its weight, summing all the weights and dividing by the number of cases with an adverse event (numerator for the AOI).

2a1.26 Calculation Algorithm/Measure Logic Diagram URL or attachment:
Attachment
AOI_V2.2.docx

2a1.27 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):
n/a

2a1.28 Data Source (Check all the sources for which the measure is specified and tested). If other, please describe:
Administrative claims, Electronic Clinical Data, Electronic Clinical Data: Electronic Health Record, Electronic Clinical Data: Pharmacy, Paper Records

2a1.29 Data Source/Data Collection Instrument (Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.): Data are generally available in the UB 04 administrative data set. Some items such as APGAR5 score and maternal blood transfusion may need to be pulled from clinical/pharmacy data sets.

2a1.30-32 Data Source/data Collection Instrument Reference Web Page URL or Attachment:
URL
www.npic.org

2a1.33-35 Data Dictionary/Code Table Web Page URL or Attachment:
Attachment
AOI_ICD10_codes.pdf

2a1.36 Level of Analysis (Check the levels of analysis for which the measure is specified and tested): Clinician: Team, Facility

2a1.37 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):
There are two Collaboratives and approximately 30-40 individual hospitals that are using or have used the AOI in their QI programs, a total of ~ 100 (through NPIC/QAS). Total deliveries/inborns analyzed is greater than 300,000. Each hospital identified a Baseline period of 1-2 years, starting with discharges as early as 2006; for some hospitals the follow-up analysis is still ongoing. Each hospital submits their administrative data set (UB 04) for all mothers and neonates 0-28 days old admission. The hospitals included all levels of care (OB Level I-III), teaching and non-teaching, urban and rural.

2a2.2 Analytic Method (Describe method of reliability testing & rationale):
The AOI algorithm was applied to the discharge data set for the time period of interest, generally quarterly. Counts of case by the 10 adverse events by quarter were generated and the Adverse Outcome Index, Weighted Adverse Outcome Score, and Severity Index calculated. Each hospital reviewed their count of cases and in most cases pulled the charts for the numerator cases in order to validate the accuracy of the coded administrative data. Errors were infrequent but when they did occur, hospitals were able to submit corrected data. Chart review to validate the coded data is the best method to insure accuracy of documentation and coding.
of administrative data.

2a2.3 **Testing Results** *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted):*

We have found the consistency between the chart data and administrative data to be consistently high based on feedback from the participating hospitals and few requests to change the coded data.

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

2b1. **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:*

Hospitalization for a delivery is the most frequent reason for hospitalization. All mothers and infants go through the delivery process where an adverse event outcome can have potential long term impact. The 10 adverse events in the AOI range from the most serious (death to the mother and/or infant) to much less serious so the differential weighting accommodates the variation in the seriousness of the adverse event. The exclusions for each adverse event accommodate those conditions where the provider has little or no control over the outcome such as quality QI efforts would not improve the score.

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):*

There are two Collaboratives and approximately 30-40 individual hospitals that are using or have used the AOI in their QI programs, a total of ~ 100 (through NPIC/QAS). Total deliveries/inborns analyzed is greater than 300,000. Each hospital identified a Baseline period of 1-2 years, starting with discharges as early as 2006; for some hospitals the follow-up analysis is still on-going. Each hospital submits their administrative data set (UB 04) for all mothers and neonates 0-28 days old admission. The hospitals included all levels of care (OB Level I-III), teaching and non-teaching, urban and rural.

2b2.2 **Analytic Method** *(Describe method of validity testing and rationale; if face validity, describe systematic assessment):*

Each of the adverse events within the AOI may indicate poor team coordination and/or communication. Use of the AOI in the context of a collaborative or by an individual hospital identifies the volume and type of adverse events such that the team can identify whether there is a provider documentation problem, coding problem or quality problem. The exercise of performing chart review validation helps discriminate the source(s) of the problem and once improvements in documentation and coding are addressed, results in an excellent measure of quality of care in L&D.

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):*

Inaccuracies between the medical record data and the administrative data are generally rare for the most serious adverse events—mortality, uterine rupture, unplanned operative procedures. Data on lacerations and admissions to the NICU are usually the most reviewed and adjusted. Using the AOI count of adverse events and allowing for adjustments to inaccurate data allows providers and teams to have confidence in the use of administrative data in the calculation of the AOI.

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

If the component measures are combined at the patient level, complete 2b

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):*

The original DoD/BIDMC study identified categories of patients where adverse events were not preventable and therefore need to be excluded from the AOI calculations. When the AOI was being translated into an algorithm to be used with administrative data, BIDMC cases identified using the algorithm were matched against cases identified during the study period and the differences reconciled.

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

Reconciling each case with an adverse event resulted in a fairly strong overlap between the administrative cases and abstracted
study cases. There were some cases on the administrative list that were not on the abstracted case and vice versa. Each case was reviewed by analysts at BIDMC and NPIC/QAS resulting in refinement to the algorithm as well as improved identification of exclusions.

2b3.3 Results (Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):

If the component measures are combined at the patient level and include outcomes, complete 2e

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): n/a

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

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): n/a

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

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):

There are two Collaboratives and approximately 30-40 individual hospitals that are using or have used the AOI in their QI programs, a total of ~ 100 (through NPIC/QAS). Total deliveries/inborns analyzed is greater than 300,000. Each hospital identified a Baseline period of 1-2 years, starting with discharges as early as 2006; for some hospitals the follow-up analysis is still on-going. Each hospital submits their administrative data set (UB 04) for all mothers and neonates 0-28 days old admission. The hospitals included all levels of care (OB Level I-III), teaching and non-teaching, urban and rural.

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

Baseline and Follow-up calculations are made for the AOI, WAOS and SI for each hospital. The Baseline period is usually 4-8 quarters prior to an QI initiative- team training, simulation, NICHD common language, IHI bundle compliance training etc and a Follow-up period after the intervention. Percent change in each rate, test of statistically significant of trend and in comparison to a Baseline comparative rate and target benchmark are all calculated.

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):

Summary data for 61 hospitals participating in a collaborative or team training program show the Baseline ranges for the AOI, WAOS and SI of .031-.130, .51-7.86 and 12.11-41.24 respectively. Follow-up ranges for the AOI were: .026-.088; for WAOS of .55-2.42 and the SI of 13.27-42.22.

One collaborative of 16 hospitals, 5 improved on all 3 scores, 6 improved at least two scores. The overall rate of improvement for the AOI and WAOS was a decrease by 6.3% and .7 % respectively. The SI showed an increase of 2.7%.

The second collaborative of 20 hospitals had an average decrease for the AOI and WAOS of 2.08 % and 1.8% respectively; the SI increased by 2.9%. 6 of the 20 showed improvement on all three scores; 5 improvement on 2 scores, 5 on 1 score and 4 declined.
### 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):*

When the AOI was being translated into an algorithm to be used with administrative data, BIDMC cases identified using the algorithm were matched against cases identified during the study period and the differences reconciled. Hospitals currently using the AOI receive numerator case lists so they reconcile their adverse event counts with chart review.

**2b6.2 Analytic Method** *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure):*

Chart review validation or data from other sources (pharmacy, blood bank) shows a high degree of correlation with administrative data. We have used the administrative data algorithm exclusively since the original study allowing hospitals to submit supplemental data from other files when necessary. Other than in the first review with BIDMC we have not tested the comparability of scores using data exclusively from other sources.

**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):*

In the original review with BIDMC data, the overlap was very strong. Hospitals currently using the AOI, perform chart review comparisons regularly and find there is little discrepancy. As hospitals move toward a more integrated EHR, the discrepancy should be largely removed.

### 2c. Disparities in Care

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>M</th>
<th>L</th>
<th>I</th>
<th>NA</th>
</tr>
</thead>
</table>

*(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): n/a*

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

n/a

### 2.1-2.3 Supplemental Testing Methodology Information:

- 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), Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

**3.1 Current Use** *(Check all that apply; for any that are checked, provide the specific program information in the following questions):* Quality Improvement with Benchmarking (external benchmarking to multiple organizations), Quality Improvement (Internal to the specific organization)

**3a. Usefulness for Public Reporting**: H [ ] M [ ] L [ ] 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)). If not publicly reported in a national or community program, state the reason AND plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement: [For Maintenance – 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.]

Currently the AOI is reported back to and across collaboratives that are involved in QI however those data are not publicly reported. It is anticipated as the AOI receives wider endorsement, it will begin to reported more broadly and perhaps publicly.

3a.2. Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. If usefulness was demonstrated (e.g., focus group, cognitive testing), describe the data, method, and results: The measures, particularly the AOI, can be well understood as the total rate of adverse events for a hospital. With concise explanation, the healthcare audience should be able to understand that the WAOS is a relative measure of the seriousness of the adverse events from one hospital to another. Feedback reports to collaborative hospitals and their leadership were well understood with minimal explanation.

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): While not currently being used for activities other than QI, the AOI could be adopted by The Joint Commission for accreditation as part of the Perinatal Care Measure set and with the ever widening use of pay for performance measures, it is also a measure that could find its way to a P4P panel. As a composite measure it is more robust than a series of measures that have very low prevalences and wide variability.

3b. Usefulness for Quality Improvement: H [ ] M [ ] L [ ] I [ ]
(The measure is meaningful, understandable and useful for quality improvement.)

3b.1. Use in QI. If used in quality improvement program, provide name of program(s), locations, Web page URL(s): [For Maintenance – If not used for QI, indicate the reasons and describe progress toward using performance results for improvement].

1) Team Performance Plus (http://www.rmfstrategies.com/tpp/);
3) Premier Perinatal Patient Safety Initiative (http://www.premierinc.com/risk/tools-services/perinatal/index.jsp) and
4) NPIC/QAS (http://www.npic.org/contracts/QualityIndicators.php)
5) Greater NY Hospital Association, www.gnyha.org
6) North Bronx Healthcare Network

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 ability to accurately and easily measure meaningful and reproducible outcomes is fundamental to quality improvement and patient safety efforts. These outcomes must be clinically relevant and occur with a high enough frequency to be responsive to improvement efforts. Without such measures, on can not adequately determine the impact of changes in care have on patients. The adverse outcome index (~5-9%) has a rate that allows detection of changes in a relative short interval. The severity scoring also allows inferences to be made about the ability to prevent or respond to adverse events in order to lessen the degree of harm. Finally the use of administrative data bases has proven to be accurate and easy, which means that this quality indicator can be measured without significant chart review.


Overall, to what extent was the criterion, Usability, met? H [ ] M [ ] L [ ] I [ ]
Provide rationale based on specific subcriteria:
### Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. *(evaluation criteria)*

<table>
<thead>
<tr>
<th>Data Generated as a Byproduct of Care Processes:</th>
<th>H</th>
<th>M</th>
<th>L</th>
<th>I</th>
</tr>
</thead>
</table>

4a. How are the data elements needed to compute measure scores generated? *(Check all that apply).*

Data used in the measure are:
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other Revenue code data from the UB 04; other EHR for pharmacy

<table>
<thead>
<tr>
<th>Electronic Sources:</th>
<th>H</th>
<th>M</th>
<th>L</th>
<th>I</th>
</tr>
</thead>
</table>

4b. 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 are in a combination of electronic sources

4c. 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:

Since the calculation of the AOI, WAOS and SI rely on administrative data, there can be variation in coding of certain outcomes, both overcoding and undercoding. During the testing and initial operational use, hospital specific reports provide detailed quarterly counts of each adverse event so hospitals can drill down and audit their data. As with any set of measures, auditing data helps improve clinical documentation and coding so the metrics really do reflect the quality of the care provided. Hospitals will often continue case by case review even though they have found consistently high correlation between the audited and the administrative data.

<table>
<thead>
<tr>
<th>Data Collection Strategy/Implementation:</th>
<th>H</th>
<th>M</th>
<th>L</th>
<th>I</th>
</tr>
</thead>
</table>

4d. 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 AOI is both easily measurable and does respond to improvements in care. The administrative data set is standard across the industry and generally very cost effective to collect and analyze. Generally there are some variables that are not included in the administrative set (ie APGAR5) that require a separate file submission but this process requires far less resources than abstracting all the data points. As hospitals continue to upgrade their EHR, all variables showing up on critical quality measures will likely be incorporated into the EHR, making the data "pull" fairly painless. Missing data is rare and while most hospitals/vendors are looking at quarterly data submission and analysis, the time frame for review can be shortened to whatever time period is of interest. The limitation is how quickly the data are available in a particular system and for cross-site comparisons, how quickly the data from all participants can be aggregated and reported.

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 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 NQF-endorsed measure(s): Are the measure specifications completely harmonized?  **No**

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

The list of included codes for #0474 is different from the list for the birth trauma adverse event. The AOI includes the code for Brachial plexus/Erb’s palseym, a serious adverse event that the developers feel is largely preventable. We do not feel there is any impact on interpretability nor does it increase data collection burden as it is a coded trauma.

#### 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):

Brachial plexus/Erb’s palsey is a serious adverse event that can be mitigated through improved obstetric management and therefore should be included in any adverse event rate.

### CONTACT INFORMATION

Co.1 **Measure Steward (Intellectual Property Owner):** Beth Israel Deaconess Medical Center, 330 Brookline Ave., Dept of OB/GYN; Dept of Anesthesia, Boston, Massachusetts, 02215

Co.2 **Point of Contact:** Janet, Muri, MBA, jmuri@npic.org, 401-274-0650-105

Co.3 **Measure Developer if different from Measure Steward:** Beth Israel Deaconess Medical Center, 330 Brookline Ave, Dept of OB/GYN; Dept of Anesthesia, Boston, Massachusetts, 02215

Co.4 **Point of Contact:** Steve, Pratt, MD, spratt@bidmc.harvard.edu, 617-504-0728

Co.5 **Submitter:** Janet, Muri, MBA, jmuri@npic.org, 401-274-0650-105, National Perinatal Information Center

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

Consensus panels were developed for the identification of measures to be included in the composite measure (AOI). This included vetting each of the individual measures.

Peter E. Nielsen, MD; Marlene B. Goldman, ScD; Susan Mann, MD; David E. Shapiro, Ph.D.; Ronald G. Marcus, MB,BCh.; Stephen D. Pratt, MD; Penny Greenberg, RN; Munish Gupta, MD; Patricia McNamee, RN, MS; Mary Salisbury, RN, MSN; David J. Birnbach, MD; Paul A. Gluck, MD; Mark D. Pearlman, MD; Heidi King, MS; David N. Tornburg, MD; Benjamin P. Sachs, MB, BS.; Lauren Bales, MD; Naval Medical Center Camp Pendleton; Ronald Burkman, MD, Baystate Medical Center; Cynthia Brumfield, MD University of Alabama at Birmingham; Peter Cherouny, MD, University of Vermont- Fletcher Allen Health Care; Jack Cooley, MD, National Naval Medical Center; Harold Fox, MD, Johns Hopkins Medical Center; Elizabeth Golladay, MD, Tripler Army Medical Center; Lynn Leventis, MD, Naval Medical Center San Diego; Robert Lorenz, MD, William Beaumont Hospital; William Lucky, MD, Baptist Hospital of Miami; Patrick Nugent, MD, South Shore Hospital; Spike Lipschitz, MD, South Shore Hospital; Chris Stolle, MD, Naval Medical Center of Portsmouth; Cosmas van DeVen, MD, University of Michigan Medical Center; Frank Witter, MD, Johns Hopkins Medical Center, Eileen Hemman, EdD. and Tom Bennettedti, MD, Suznne Walker, RN, MPH and Thomas Strandjord, MD from the University of Washington. In addition, representatives from the American Congress of Obstetrics and Gynecology, the American Society for Obstetric Anesthesia and Perinatology, the American Society of Anesthesiologists, the Association of Women’s Health, Obstetric and Neonatal Nurses, the Armed Forces Institute of Pathology, the U.S. Navy Bureau of
NQF #1769 Adverse Outcome Index

Medicine and Surgery, the Office of the Surgeon General, U.S. Army and TRICARE were present.

National Perinatal Information Center, Providence Rhode Island assisted with translating the AOI into administrative data specifications.

Co.7 Public Contact: Janet, Muri, MBA, jmuri@npic.org, 401-274-0650-105, National Perinatal Information Center

### 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.

Consensus panels were developed for the identification of measures to be included in the composite measure (AOI). This included vetting each of the individual measures.

Peter E. Nielsen, MD; Marlene B. Goldman, ScD; Susan Mann, MD; David E. Shapiro, Ph.D.; Ronald G. Marcus, MB, BCh.; Stephen D. Pratt, MD; Penny Greenberg, RN; Munish Gupta, MD; Patricia McNamee, RN, MS; Mary Salisbury, RN, MSN; David J. Birnbach, MD; Paul A. Gluck, MD; Mark D. Pearlman, MD; Heidi King, MS; David N. Tornburg, MD, MPH; Benjamin P. Sachs, MB, BS.; Lauren Bales, MD; Naval Medical Center Camp Pendleton; Ronald Burkman, MD, Baystate Medical Center; Cynthia Brumfield, MD University of Alabama at Birmingham; Peter Cherouny, MD, University of Vermont- Fletcher Allen Health Care; Jack Cooley, MD, National Naval Medical Center; Harold Fox, MD, Johns Hopkins Medical Center; Elizabeth Golladay, MD, Tripler Army Medical Center; Lynn Leventis, MD, Naval Medical Center San Diego; Robert Lorenz, MD, William Beaumont Hospital; William Lucky, MD, Baptist Hospital of Miami; Patrick Nugent, MD, South Shore Hospital; Spike Lipschitz, MD, South Shore Hospital; Chris Stolle, MD, Naval Medical Center of Portsmouth; Cosmas van Deven, MD, University of Michigan Medical Center; Frank Witter, MD, Johns Hopkins Medical Center, Eileen Hemman, EdD. and Tom Bennedetti, MD, Suzanne Walker, RN, MPH and Thomas Strandjord, MD from the University of Washington. In addition, representatives from the American Congress of Obstetrics and Gynecology, the American Society for Obstetric Anesthesia and Perinatology, the American Society of Anesthesiologists, the Association of Women’s Health, Obstetric and Neonatal Nurses, the Armed Forces Institute of Pathology, the U.S. Navy Bureau of Medicine and Surgery, the Office of the Surgeon General, U.S. Army and TRICARE were present.

National Perinatal Information Center, Providence Rhode Island assisted with translating the AOI into administrative data specifications.

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:

**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: 09, 2011

Ad.5 What is your frequency for review/update of this measure? On going or as required by code changes

Ad.6 When is the next scheduled review/update for this measure? 09, 2013

Ad.7 Copyright statement:

Ad.8 Disclaimers:

Ad.9 Additional Information/Comments:

Date of Submission (MM/DD/YY): 10/14/2011