Timely Temperatures for All Low Birth Weight Neonates

Section 1. Basic Measure Information

1.A. Measure Name

Timely Temperatures for All Low Birth Weight Neonates

1.B. Measure Number

0116

1.C. Measure Description

Please provide a non-technical description of the measure that conveys what it measures to a broad audience.

Describes the percent of live-born neonates less than 2500 grams that have a temperature documented within the Golden Hour, from birth to 60 minutes of age.

1.D. Measure Owner

Collaboration for Advancing Pediatric Quality Measures (CAPQuaM)

1.E. National Quality Forum (NQF) ID (if applicable)

Not applicable.

1.F. Measure Hierarchy

Please note here if the measure is part of a measure hierarchy or is part of a measure group or composite measure. The following definitions are used by AHRQ:

1. **Please identify the name of the collection of measures to which the measure belongs (if applicable). A collection is the highest possible level of the measure hierarchy. A collection may contain one or more sets, subsets, composites, and/or individual measures.**

   This measure belongs to the Pediatric Quality Measurement Program (PQMP) Inpatient Perinatal Collection No. 1.

2. **Please identify the name of the measure set to which the measure belongs (if applicable). A set is the second level of the hierarchy. A set may include one or more subsets, composites, and/or individual measures.**

   Thermal Management of Low Birth Weight Infants

3. **Please identify the name of the subset to which the measure belongs (if applicable). A subset is the third level of the hierarchy. A subset may include one or more composites, and/or individual measures.**

   Proximal outcomes subset.

4. **Please identify the name of the composite measure to which the measure belongs (if applicable). A composite is a measure with a score that is an aggregate of scores from other measures. A composite may include one or more other composites and/or individual measures. Composites may comprise component measures that can or cannot be used on their own.**

   Not applicable.

1.G. Numerator Statement

Neonates under 2500 grams who are born in a medical facility.

1.H. Numerator Exclusions

None.

1.I. Denominator Statement

All newborns born in a medical facility with birth weights less than 2500 grams, other than those excluded (see Section 1.J.).

Identification of newborns who may be eligible to be included in the denominator may be accomplished through the use of the following ICD-9 codes listed in Table 1 (see Table 1 in the supporting documents). For codes 76400, 76410, 76420, 76490, and 76500, birth weights should be verified from the medical record prior to inclusion in the measure.

Denominator Elements

Number of infants less than 2500 grams who are born at a hospital or birthing facility.

General Data Elements for the Measure Set

- Date of birth.
- Time of birth.
- Time of first temperature taken.
- Date of first temperature taken.
- Route of first temperature taken.
- Value of first temperature taken.

- Units of first temperature taken.

Other General Data Elements (for stratification and reporting)

- Birth weight.
- 5-minute Apgar.
- Race.
- Ethnicity.
- Insurance type (public, commercial, none, other).
- Benefit category (HMO, PPO, Medicaid Primary Care Management Plan, Fee-for-Service, Other).
- Mother's State and county of residence and/or zip code.
- Medicaid or CHIP benefit/qualifying category.
- Born inside or outside of a medical facility.
- If born in a medical facility:
    1. Location of birth:
        a. Operating Room (e.g., for cesarean section or double set-up delivery).
        b. Birthing Room (Birthing room refers to a birthing or delivery room or a labor and delivery suite that is not an operating room).
        c. Other.
    2. Location of birth unavailable:
        a. If delivery occurred by cesarean section, then put location of birth as operating room.
        b. If this was a twin or multiple gestation delivery, put location of birth as operating room.
        c. Otherwise, put location of birth as birthing room/delivery room.

This measure describes a percent and requires limited calculations. Percent is calculated as 100*number of neonates eligible for the numerator divided by the number of neonates eligible for the denominator.

1.J. Denominator Exclusions

Neonates with comfort care; requires all of the following features:

- Died within 48 hours of birth.
- Received no respiratory support after arrival to the Level 2 or higher nursery other than blow by oxygen (i.e., did not receive CPAP, intubation, or CPR after arrival at Level 2 or higher nursery).
- Neonates with anencephaly (ICD-9-CM 740.0).

1.K. Data Sources

Check all the data sources for which the measure is specified and tested.

Administrative Data (e.g., claims data, Paper Medical Record, Electronic Medical Record).

If other, please list all other data sources in the field below.

Section 2: Detailed Measure Specifications

Provide sufficient detail to describe how a measure would be calculated from the recommended data sources, uploading a separate document (+ Upload attachment) or a link to a URL. Examples of detailed measure specifications can be found in the CHIPRA Initial Core Set Technical Specifications Manual 2011 published by the Centers for Medicare & Medicaid Services. Although submission of formal programming code or algorithms that demonstrate how a measure would be calculated from a query of an appropriate electronic data source are not requested at this time, the availability of these resources may be a factor in determining whether a measure can be recommended for use.

Please see supporting documents for full, detailed measure specifications.

Section 3. Importance of the Measure

In the following sections, provide brief descriptions of how the measure meets one or more of the following criteria for measure importance (general importance, importance to Medicaid and/or CHIP, complements or enhances an existing measure). Include references related to specific points made in your narrative (not a free-form listing of citations).

3.A. Evidence for General Importance of the Measure

Provide evidence for all applicable aspects of general importance:

- **Addresses a known or suspected quality gap and/or disparity in quality (e.g., addresses a socioeconomic disparity, a racial/ethnic disparity, a disparity for Children with Special Health Care Needs (CSHCN), a disparity for limited English proficient (LEP) populations).**
- **Potential for quality improvement (i.e., there are effective approaches to reducing the quality gap or disparity in quality).**
- **Prevalence of condition among children under age 21 and/or among pregnant women**
- **Severity of condition and burden of condition on children, family, and society (unrelated to cost)**

- **Fiscal burden of measure focus (e.g., clinical condition) on patients, families, public and private payers, or society more generally, currently and over the life span of the child.**

- **Association of measure topic with children's future health – for example, a measure addressing childhood obesity may have implications for the subsequent development of cardiovascular diseases.**

- **The extent to which the measure is applicable to changes across developmental stages (e.g., infancy, early childhood, middle childhood, adolescence, young adulthood).**

Inpatient perinatal care was assigned to CAPQuaM as a PQMP priority by the Agency for Healthcare Research and Quality (AHRQ) with the active consultation of the Centers for Medicare & Medicaid Services (CMS). After initial assignment, conversations between CAPQuaM, AHRQ, and CMS resulted in a decision for CAPQuaM to undertake the development of measures related to the temperature of low birth weight neonates. We developed this measure in close consultation with our Consortium partners at the New York State Department of Health, including the Office of Health Insurance Programs/New York State Medicaid.

This measure addresses a key gap in inpatient perinatal care. Evidence that thermal management (such as hot water bottles and incubators) improves survival of newborn and premature infants exists from as early as the late 19th century (Baker, 2000; Currier, 1891; Fischer, 1915; Garrison, 1923; Holt, 1902; Holt, Macintosh, 1940; Pierce, 1875). Modern studies have confirmed and extended these findings, including potential methods to maintain temperature for infants in the Delivery Room (Silverman, Fertig, Berger, 1958; Sinclair, 2007; Watkinson, 2006). Laptook et al confirmed the association of temperature loss with poor outcomes in 5,277 infants, 401-1499 grams, born at any of 15 academic medical centers participating in the National Institute of Child Health and Human Development (NICHD) Neonatal Research Network (Laptook, Salhab, Bhaskar, 2007). A formal item selection process looking at potential measures for infants under 1500 grams identified neonatal temperature as an independent contributor to a composite quality of care measure (Profit, Gould, Zupancic, 2011).

We collected chart review data from three diverse hospitals in New York City. All three hospitals had a range of birth weights and a range of temperatures. Temperature predicted in-hospital mortality after controlling for covariates. The relationship between temperature and survival is monotonic: an increase of each 1° Celsius up to 37° reduced the odds of death more than 35 percent in the model using a continuous variable (22 percent for 1° Fahrenheit). Defining hypothermia as admission temperature below 36.0° would estimate an increase in the risk of mortality of 27 percent, p=0.19.

Our work confirmed findings in the literature that insurance status and race (Reynolds, Pilcher, Ring, 2009) are associated with outcomes. Anecdotal reports from among our participating hospitals confirm reports in the literature (Doyle, Bradshaw, 2012) that attention to thermal management can improve temperature outcomes. As an appendix, we present a more complete literature review; see supporting documents.

Despite evidence of the importance of temperature on outcomes of neonates, two proposed measures for quality of care – taking the temperature within an hour of admission to the neonatal intensive care unit (NICU) and maintaining a temperature of 36.5° at admission to the NICU – were not recommended for endorsement by the National Quality Forum, even though they were submitted by the Vermont Oxford Network. We incorporate a highly engaged process to develop an enhanced set of measures. A distinguished multidisciplinary panel of national experts that included neonatologists, a family physician, nurses, and a pediatric hospitalist articulated that it was a fundamental principle that all low birth weight infants need to have a timely temperature taken, whether sick or healthy, when admitted to a regular nursery or to a special care nursery or NICU. 'Timely' was considered to represent different values by different expert panelists, but in the end none felt it was excusable as a matter of neonatal safety that any low birth weight infant would go their first hour without having a documented temperature.

A Vermont Oxford Network NICU team has migrated the term "The Golden Hour" from field trauma to neonatology to describe the first hour of life (Reynolds, et al., 2009). Prevention of hypothermia was described as the cornerstone of Golden Hour activities and continues as such in more recent writings (Doyle, Bradshaw, 2012). Delay in taking temperatures until after one hour of life is a profound violation of fundamental concepts regarding the management of low birth weight newborns. Since another measure in the measure set considers the timing of temperature assessment in relation to admission to a Level 2 or higher nursery, there was discussion regarding whether this measure should pertain only to those children who are not admitted to a special care nursery; those infants are considered to be at particularly high risk of being ignored or having hypothermia missed because they are seen as being "healthy," even if they are small or premature newborns (Laptook, Jackson, 2006; Laptook, Watkinson, 2008). After discussion, the panel strongly recommended that there not be exclusions. For children who are not admitted to a Level 2 or higher nursery, this is a safety check and, critical as it may be, the only such monitoring for the potential for early or transitional hypothermia.

For those who are admitted to the special or intensive care nursery, there is a complementary measure – percent who had a temperature taken and recorded within the first 15 minutes after admission. For a large subset, that admission temperature would typically fulfill the criteria, and this measure would be satisfied. But for those whose temperature is delayed, this measure within 60 minutes will provide an important safety check. And so the panel's strong recommendation was not to exclude from this population, except for anencephaly or those being managed only for comfort care. Unlike for other measures in this set, we do not exclude neonates being managed under hospital protocols for hypothermia. The relationship of the four measures in this measure set is discussed in the next section.

3.B. Evidence for Importance of the Measure to Medicaid and/or CHIP

Comment on any specific features of this measure important to Medicaid and/or CHIP that are in addition to the evidence of importance described above, including the following:

- **The extent to which the measure is understood to be sensitive to changes in Medicaid or CHIP (e.g., policy changes, quality improvement strategies).**

- **Relevance to the Early and Periodic Screening, Diagnostic and Treatment benefit in Medicaid (EPSDT).**

- **Any other specific relevance to Medicaid/CHIP (please specify).**

In New York State, about half of low birth weight babies are insured by Medicaid. Hypothermia is associated not only with neonatal mortality, but there is also evidence (Miller, Lee, Gould, 2011) that intraventricular hemorrhage (IVH) can also be a consequence of hypothermia. IVH is a significant cause of disability and developmental delay, and when serious, it is a common cause for LBW infants to develop into children with special health care needs. This has broad impact on Medicaid, Medicaid expenses, and early intervention services, including the Early and Periodic Screening, Diagnostic, and Treatment (EPSDT) program services. Hypothermia, through death and disability may have a long tail that impacts families and programs associated with Medicaid. Furthermore, the Medicaid population is disproportionately black, and in our testing data, black infants were disproportionately hypothermic.

We note above that there is evidence that management can enhance thermal outcomes. An overview of the four measures in our measure set is depicted in Figure A in the supporting documents.

3.C. Relationship to Other Measures (if any)

Describe, if known, how this measure complements or improves on an existing measure in this topic area for the child or adult population, or if it is intended to fill a specific gap in an existing measure category or topic. For example, the proposed measure may enhance an existing measure in the initial core set, it may lower the age range for an existing adult-focused measure, or it may fill a gap in measurement (e.g., for asthma care quality, inpatient care measures).

Two excellent measures proposed by The Vermont Oxford Network (VON) are complemented and enhanced by this measure set.

VON proposed a measure regarding the adequacy of taking temperatures in very low birth weight infants, temperatures taken within an hour of admission to the NICU. This was rejected by the National Quality Forum (NQF) largely because it was met almost all of the time. While we would hold with VON that 98 percent compliance is inadequate for a quality measure that is so closely related to patient safety, we have proposed a measure that adopts a slightly different approach. The first hour of life has become known as the “Golden Hour” because of the importance of timely recognition and management on neonatal outcomes (Doyle, Bradshaw, 2012; Reynolds, et al., 2009). This measure includes all LBW infants (not only very low birth weight), thereby including the late preterm infant within our measure. We also look at all infants who are LBW, not only those who are admitted to a special care or intensive care nursery. By so doing, we increase the value of the measure for infants who may be at unrecognized risk for hypothermia.

We propose a measure that looks at the proportion of LBW neonates who have a temperature documented within the first hour of life, regardless of whether they are admitted to an advanced care nursery. We consider this a safety measure, as missed hypothermia may lead to shock and death (Engle, Tomashek, Wallman, 2007; Laptook, Jackson, 2006). Those infants who are low birth weight and do not require admission to the advanced care nursery may be at risk to be managed more like full-term infants without adequate recognition that they are more fragile and in this case more sensitive to severe consequences from cold stress than a larger infant would be.

Indeed, even recently these late preterm infants were called “near term,” leading to confusion that they could be managed safely as are term infants (Engle, et al., 2007; Laptook, Jackson, 2006). These beliefs have not been purged. Hence, this measure is inclusive of all LBW infants. Further, all those infants who require admission to an advanced level of care (Level 2 or higher nursery) have a similar or higher risk of deterioration due to cold stress. Since thermal management is a cornerstone of early care for the sick neonate in the golden hour, our measure set includes a measure that assesses how frequently a temperature is taken and recorded within 15 minutes of arrival to the advanced care nursery. The current measure is for those admitted to the nursery immediately after delivery, as well as those transported or transferred from the newborn nursery within the first day of life.

VON also proposed a measure that reports the proportion of infants cooler than 36.0º Celsius. It also was rejected by NQF, in part because there is no consensus regarding the desirable threshold. Measures 3 and 4 in our proposed measure set address similar issues. Measure 3, the continuous representation of the data, does not focus on judgment of “good” or “bad” and instead provides data that are meaningful and sensitive to change and therefore particularly valuable to help guide quality improvement activities. The measure is organized to describe both ends of the distribution and the central area of the distribution, along with both a sensitive and a robust measure of spread. Measure 4 stratifies temperature as cold, very cool, cool, euthermic, or overly warm.

Section 4. Measure Categories

CHIPRA legislation requires that measures in the initial and improved core set, taken together, cover all settings, services, and topics of health care relevant to children. Moreover, the legislation requires the core set to address the needs of children across all ages, including services to promote healthy birth. Regardless of the eventual use of the measure, we are interested in knowing all settings, services, measure topics, and populations that this measure addresses. These categories are not exclusive of one another, so please indicate "Yes" to all that apply.

Does the measure address this category?

a. **Care Setting – ambulatory:** No.
b. **Care Setting – inpatient: Yes.**
c. **Care Setting – other – please specify:** Yes. Newborn birthing care.
d. **Service – preventive health, including services to promote healthy birth:** Yes.
e. **Service – care for acute conditions:** Yes.
f. **Service – care for children with special health care needs/chronic conditions:** Yes.
g. **Service – other (please specify):** No.
h. **Measure Topic – duration of enrollment:** No.
i. **Measure Topic – clinical quality:** Yes.
j. **Measure Topic – patient safety:** Yes.
k. **Measure Topic – family experience with care:** No.
l. **Measure Topic – care in the most integrated setting:** No.
m. **Measure Topic other (please specify):** No.
n. **Population – pregnant women:** No.
o. **Population – neonates (28 days after birth) (specify age range):** Yes.

p. **Population – infants (29 days to 1 year) (specify age range):** No.
q. **Population – pre-school age children (1 year through 5 years) (specify age range):** No.
r. **Population – school-aged children (6 years through 10 years) (specify age range):** No.
s. **Population – adolescents (11 years through 20 years) (specify age range):** No.
t. **Population – other (specify age range):** No.
u. **Other category (please specify):** Not applicable.

Section 5. Evidence or Other Justification for the Focus of the Measure

The evidence base for the focus of the measures will be made explicit and transparent as part of the public release of CHIPRA deliberations; thus, it is critical for submitters to specify the scientific evidence or other basis for the focus of the measure in the following sections.

5.A. Research Evidence

Research evidence should include a brief description of the evidence base for valid relationship(s) among the structure, process, and/or outcome of health care that is the focus of the measure. For example, evidence exists for the relationship between immunizing a child or adolescent (process of care) and improved outcomes for the child and the public. If sufficient evidence existed for the use of immunization registries in practice or at the State level and the provision of immunizations to children and adolescents, such evidence would support the focus of a measure on immunization registries (a structural measure).

Describe the nature of the evidence, including study design, and provide relevant citations for statements made. Evidence may include rigorous systematic reviews of research literature and high-quality research studies.

Please see evidence and references discussed in section 3 above. In addition, we have conducted a systematic, targeted review of the literature (see Appendix in the supporting documents). Further we have interviewed clinicians and engaged clinical societies and accreditors, patient/family groups, New York Medicaid, and others to inform our measure development with the intelligence and experiences of stakeholders as well as the medical literature. Further, our measure operationalizes constructs defined by a diverse and superb multidisciplinary panel of national experts who participated in a RAND/UCLA Modified Delphi process. The ratings of the panel along and a brief description of methodology can be found in the supporting documents. Further evidence is provided in the validity section, below.

We report on New York State neonatal data. Hospitals use various means to collect the data on their high risk newborns, but they must submit the data using the NICU Module's online data entry or import function. To ensure data security and patient confidentiality, hospitals must register their data entry or importing users through the NYSDOH Health Commerce System before they are granted controlled access to the Web-based NICU Module.

Key findings from our study of 7,553 neonates (from 61 nurseries) in New York State are: temperature was variable within weight categories; blacks were disproportionately cool compared with Hispanic and non-Hispanic others who, in turn, were disproportionately cool compared with non-Hispanic whites, whether or not we stratified by birth weight category. Deaths were disproportionate among those who were cool, in a graded fashion.

The distribution of mean temperature by nursery ranged from 35.7º to 38.2º, with a median of 36.3º, a standard error of 0.36, and an interquartile range of 0.4; 25 percent of these nurseries had a mean temperature below 36.1º. We conclude from this that temperatures do vary across nurseries, further reinforcing our sense that this topic is an important measure of performance.

Using the Mount Sinai Data Warehouse, which is linked to Mount Sinai's Epic electronic health record (EHR), we looked at the time of the first recorded temperature for low birth weight newborns for a 1-year period (446 infants for whom the time of the first temperature could be identified electronically) across the weight spectrum and found that on average there were several infants each month for whom temperature-taking was delayed, and that these infants were across all weight categories. These data confirm that while infants generally have their temperature taken within 60 minutes, even at a teaching hospital with Level 4 care such as Mount Sinai, which has succeeded in raising admission temperatures for LBW infants because of sustained attention to the issue, it is not universal. See Validity section for further details.

5.B. Clinical or Other Rationale Supporting the Focus of the Measure (optional)

Provide documentation of the clinical or other rationale for the focus of this measure, including citations as appropriate and available.

This is discussed in detail above in the Importance of the Measure section. This measure has importance as a descriptor of the process of care and as a patient safety measure. Delay greater than 1 hour is outside of the standard of acceptable care for low birth weight infants.

The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children (Diamond, Rask, Kohler, 2004); and practitioners have been identified as a resource for researchers in developing and revising measures, since they are on the frontlines working with the populations who often become research participants. Involving practitioners can assist researchers in the creation of measures that are appropriate and easily administered (Howell, Stone, Kleinman, 2010).

The validity of our work has benefited from our use of a formal method, a pragmatic adaptation of the CAPQuaM 360º method. The method as adapted to the perinatal measures was specifically designed to develop valid and reliable measures in the face of pragmatic epistemological uncertainty. That is, recognizing that practice extends well beyond the research base, we designed this method to allow us to develop reliable and valid state-of-the-science measures, in part by explicitly modeling and accounting for uncertainties in the measure development, in part by the conceptualization and implementation of a Boundary Guideline (see below). We have shared and refined this approach in a number of venues including within the PQMP, which comprises the various PQMP AHRQ-CMS CHIPRA Centers of Excellence, the State PQMP participants, and AHRQ and CMS participants. All presentations have invited dialogue and feedback. This work has been similarly presented at a number of Grand Rounds/weekly

conferences in the New York-New Jersey area, as well was to national/international audiences including the Bioethics and children's health services communities. These latter venues include:

- 2012 Pediatric Academic Societies State of the Science Plenary (Boston), 2012. This presentation can be found in the supporting documents.
- Oxford-Mount Sinai Bioethics Consortium (Amsterdam), 2012.
- Child Health Services Research Interest Group, Academy Health (Orlando), 2012.

Feedback from these presentations has been extremely positive. The Boundary Guideline construct has generated particular enthusiasm. We asked the Bioethics Consortium to extrapolate the primum non nocere (First, do no harm) principle to apply regarding this aspect of performance measurement. We received strong feedback that not only is it ethical to measure using systematically developed measures (even in the context of some uncertainty), but that it is ethically preferable to use such measures compared with the alternative of providing care that is not assessed (and perhaps not assessable) because of residual uncertainty.

The 360º method is highly engaged with collaborators, partners, and the literature. It seeks to target relevant information and perspective and to have measures emerge from the process. The potential measures are then tested to the extent that time and resources permit. In developing the perinatal measures we incorporate:

- A high level of engagement with partnered institutions and senior advisors that bring into the process a wide diversity of stakeholders.
- A detailed literature review that is updated and supplemented as needed.
- Interviews with clinicians.
- The CAPQuaM scientific team (professionals qualified in neonatology, pediatrics, obstetrics and gynecology, epidemiology, quality measurement and improvement, patient safety, and public health).
- A geographically diverse, multidisciplinary expert panel whose members participated in a two- Round RAND/UCLA modified Delphi process, with enhanced follow up.
- Development of a Boundary Guideline that takes a multi-vectorial approach to incorporate simultaneously a variety of gradients, including gradients of importance, relevance, and certainty, as appropriate to the construct being represented;
- Specification and review of measures and approaches to measurement by stakeholders and experts.
- Testing and assessment of measure performance to the extent feasible, given resources and available time.

Fortunately, in the case of this proposed measure we can present both a systematically developed measure and strong evidence to support its use.

Section 6. Scientific Soundness of the Measure

Explain the methods used to determine the scientific soundness of the measure itself. Include results of all tests of validity and reliability, including description(s) of the study

sample(s) and methods used to arrive at the results. Note how characteristics of other data systems, data sources, or eligible populations may affect reliability and validity.

6.A. Reliability

Reliability of the measure is the extent to which the measure results are reproducible when conditions remain the same. The method for establishing the reliability of a measure will depend on the type of measure, data source, and other factors.

Explain your rationale for selecting the methods you have chosen, show how you used the methods chosen, and provide information on the results (e.g., the Kappa statistic). Provide appropriate citations to justify methods.

The basis for the scientific soundness of this measure lies in the use of a hybrid of administrative/encounter and medical records data. Though they have their limitations, these data types have been shown in multiple studies to be a reliable source of information for population-level quality measurement. One such study found that quality measures that could be calculated using administrative data showed higher rates of performance than indicated by a review of the medical record alone, and that claims data are more accurate for identifying services with a high likelihood of documentation due to reimbursement (Howell, Holzman, Kleinman, et al., 2010).

A feasibility study of diverse hospitals from across the country and in different stages of EHR development was conducted. Our feasibility study was designed to determine the ability and ease of collecting related data. The results from this study show that date and time are self-evident, and that there is mild but manageable variation in how time is reported. This limited variation will not impair the calculation of a neonate's age or the relationship of the time of measurement to the time of birth or of arrival to the NICU as may be required in our measure set. Twelve of 15 respondents were clear that the data would be in the infant record, and three others thought it would be in the mother's chart. Nine of 10 who responded to the question indicated the data would be available in the EMR, while one thought that it was more likely in the paper record. None thought the data would be very difficult to obtain.

In our team's work studying processes and outcomes of neonatal care in three New York City hospitals, we found that chart abstractors could be readily trained to collect valid and reliable data regarding the thermal management of children (and other processes of care) using a simple portable electronic data abstraction tool (Rubio, Berg-Weger, Tebb, et al., 2003; Virnig, McBean, 2001).

6.B. Validity

Validity of the measure is the extent to which the measure meaningfully represents the concept being evaluated. The method for establishing the validity of a measure will depend on the type of measure, data source, and other factors.

Explain your rationale for selecting the methods you have chosen, show how you used the methods chosen, and provide information on the results (e.g., R2 for concurrent validity).

The validity of our measure stems not only from the use of a formal process that was highly engaged with stakeholders and the literature in order to generate potential measures, but from empirical data analysis of both the Mount Sinai Data Warehouse and the New York State

Department of Health Inpatient Neonatal database, which has data on virtually all children admitted to Level 2 or higher nurseries in the State.

Our testing (using Mount Sinai data) of ICD-9 codes as a way to identify low birth weight infants found that 99 infants out of 677 who were identified with our ICD-9 screen (which was recommended to us by The Joint Commission), had birth weights of over 2500. The distribution of ICD-9 codes for this cohort that were 2500 grams or above LBW is reported in Table A (see supporting documents).

Of the 99 infants, five had recorded birth weights of 2500 grams, consistent with the ICD-9 codes used. We have indicated in our specifications that the various ICD-9 codes, such as 764.00, 764.10, and 765.10 that represent poor fetal growth without a specified weight need to have their eligibility for the measure confirmed with an actual birth weight.

The key constructs underlying our measures are: date and time of birth, arrival to the Level 2 or higher nursery, and taking of the first temperature.

Testing with data from the New York State Neonatal database supports various aspects of this measure. Our data include reports from 20 Level 2 nurseries, 27 Level 3 nurseries, and 14 Regional Perinatal Centers that contributed 20 or more infants for the reporting year assessed. In our data, we included all inborn infants from these hospitals with a birth weight of 400-2499 grams whose admission temperature was 29º Celsius or higher. Excluded were those with anencephaly or those who expired within 48 hours without receiving respiratory support beyond oxygen in the NICU (N=7553). The number of infants ranged from 21 to 370 per hospital, and 86.7 percent were admitted to Level 3 or higher hospitals.

We investigated time of first temperature among infants admitted to the neonatal intensive care unit within 24 hours of birth. Overall, we found that temperatures taken later than 15 minutes after arrival were significantly more likely to be euthermic and less likely to be cool or cold, consistent with our expected findings.

Our data analysis confirms that there is variability in the time at which temperatures are taken. Statewide, 86.8 percent of LBW infants have their temperature taken within 15 minutes of arrival to the nursery. We also investigated age of the neonate at time that the first temperature was taken. We found that 10.8 percent of LBW infants (n=815) did not have documentation of a temperature within the first hour of life. The systematic variation, including the racial differences noted above and the apparent structural variation seen across the level 2, 3, and 4 nurseries reinforce our decision to prioritize these proposed measures of timing as important process of care measures, with failure of the 60-minute measure representing a meaningful failure that jeopardizes patient safety. Data regarding age of neonate and temperature can be seen in Table B (see supporting documents).

Temperatures measured after 60 minutes of life were higher than those measured within the first hour ($p<.0001$). Our findings have important implications. The temperature difference reminds us that temperature in LBW infants is largely a factor of environment, and that the potentially chaotic environment surrounding delivery and transport immediately following delivery is very different from the potentially more controlled environment of the nursery an hour or more after birth. So the earlier and later temperatures are actually measuring different constructs. Failure to measure a timely temperature after birth misses the opportunity to identify and manage early cold stress. Further, if temperature is a quality indicator as we propose, the higher later

temperatures may become an incentive to not enter early cool temperatures into the permanent medical record.

We also employed a multitude of experts and diverse stakeholders – clinicians, scientists, payers, purchasers, and consumers – as another means of establishing validity. We believe this to be central to validity in the context of measuring quality amidst uncertainty. We obtained feedback on the face validity of the constructs, the development of the boundary guidelines, and the measure's testing. The use of Expert Panels has been demonstrated to be useful in measure development and evaluation, and practitioners have been identified as a resource for researchers in developing and revising measures, since they are on the frontlines working with the populations who often become research participants. Involving practitioners can assist researchers in the creation of measures that are appropriate and easily administered (Howell, et al., 2010).

Throughout development, CAPQuaM brought together stakeholders to ensure their iterative engagement in advancing quality measures that are understandable, salient and actionable. CAPQuaM employed a 360º method, designed to involve key stakeholders in meaningful ways. Our development process for this measure cultivated formal input from:

- Medical literature (both peer reviewed and gray, including State Web sites).
- Relevant clinicians.
- Organizational stakeholders (our consortium partners, as well as advisory board members, see below).
- Multidisciplinary, geographically diverse expert panel including clinicians and academicians.
- CAPQuaM's scientific team.

Clinical criteria regarding reporting approaches, including consideration of inclusion and exclusion criteria, the value of temperature measurement, and specific and meaningful temperature cutoffs were developed using a modified version of the RAND/UCLA modified Delphi Panels. CAPQuaM sought recommendations from major clinical societies and other stakeholders to identify academic and clinician expert panel participants from a variety of areas and backgrounds, clinical and regional settings, and expertise. The product of this process was participation by a broad group of experts in the development of clinically detailed scenarios leading to the measures.

The route to measure specification included development of relevant scenarios and issues for formal processing by our expert panel, whose members participated in a two-round RAND/UCLA modified Delhi panel that culminated in a day-long, in-person meeting hosted at the Joint Commission and moderated by a pediatrician and an obstetrician-gynecologist. The output from that panel meeting was summarized in the form of a boundary guideline that was then used to guide the measure specification and prioritization.

Our feasibility work indicates that the time the temperature is assessed, rather than simply the time that it is recorded, is documented in the medical record, generally an EMR. This is a critical aspect of the validity of time data.

Section 7. Identification of Disparities

CHIPRA requires that quality measures be able to identify disparities by race, ethnicity, socioeconomic status, and special health care needs. Thus, we strongly encourage nominators to have tested measures in diverse populations. Such testing provides evidence for assessing measure's performance for disparities identification. In the sections below, describe the results of efforts to demonstrate the capacity of this measure to produce results that can be stratified by the characteristics noted and retain the scientific soundness (reliability and validity) within and across the relevant subgroups.

7.A. Race/Ethnicity

Our feasibility assessment confirmed that racial and ethnicity data are almost universally available, and that method of assignment of race and ethnicity to the baby varied. Assignment could be based on maternal self-report or assigned by the hospital, most typically as the mother's race and ethnicity. National improvement is needed in the methods used to assign race and ethnicity to newborns in the hospital. For the purposes of this measure, we are resigned at this time to using the existing data as recorded in the infants' medical records.

Racial differences were seen in our New York State neonatal data analysis, with black babies most likely to be cold, very cool, or cool and least likely to be euthermic or above normal. ($p<.001$). Whites were least likely to be cool, and non-Hispanic other and Hispanic infants were at intermediate values. Race and ethnicity were also independent predictors of temperature in our New York City data. We also saw racial differences in the timing of temperatures among those admitted to Level 2 or higher nurseries. Table C (see supporting documents) depicts the percent of infants admitted to a Level 2 or higher nursery who had their temperatures taken within 15 minutes of arrival.

While the precise meaning of these data might be argued, it is clear that the timing of temperature taking has systematic variations that certainly include the birth weight and may include race/ethnicity. The uncertainty of the meaning of these data speaks to the importance of monitoring for systematic racial/ethnic differences in performance within and across health care organizations.

7.B. Special Health Care Needs

Not assessed.

7.C. Socioeconomic Status

We can use Medicaid insurance as a marker for socioeconomic status (SES). Our New York City data demonstrate this to be an independent predictor of poor thermal outcomes.

We further use the national distribution of percent of individuals in poverty to establish five categories that reflect the counties' level of poverty. We considered other data such as county median income or county unemployment but felt that the percent of individuals in poverty was a more integrative measure. The use of a county-level rather than an individual measure is consistent with recent applications of hierarchical methods to study the impact of poverty and

also with data that indicate that local disparities in income are an independent predictor of outcomes (Kawachi, Berkman, 2003). It also allows this measure to consider issues related to SES while using publicly available data and requiring only the mother's county of residence, a more reliable data point than self-reported income.

Our analysis of U.S. Department of Agriculture data considering 3,142 counties and related geographic units found a mean of 17.2 percent of county residents living in poverty, a standard deviation of 6.5 percent, and an interquartile range of 8.2 percent. The distribution illustrated in Table 3 (see supporting documents), shows meaningful dispersion and supports our plan to build off quartiles of distribution with a finer focus in higher areas of poverty.

7.D. Rurality/Urbanicity

As described in the specifications, we use urban influence codes (Hall, Kaufman, Ricketts, 2006) to describe the level of rurality or urbanicity.

Metropolitan

1 In large metro area of 1+ million residents

2 In small metro area of less than 1 million residents

Non-metropolitan

3 Micropolitan adjacent to large metro

4 Non-core adjacent to large metro

5 Micropolitan adjacent to small metro

6 Non-core adjacent to small metro with own town

7 Non-core adjacent to small metro no own town

8 Micropolitan not adjacent to a metro area

9 Non-core adjacent to micro with own town

10 Non-core adjacent to micro without own town

11 Non-core not adjacent to metro or micro with own town

12 Non-core not adjacent to metro or micro without own town

We analyzed 3,143 county equivalents in the United States, and the results can be seen in Table 4 (see supporting documents).

The population is heavily weighted to metropolitan areas (see Table 5 in the supporting documents). The data show that 55 percent of the U.S. population lives in an urban area of greater than 1 million residents (UIC_2013 #1), while 1.33 percent live in a county that does not contain a town of at least 2,500 residents (UIC_2013 #10-12). While this approach to rurality does not map exactly to the population density-based definition of frontier (< 6 persons per square mile) as articulated in the Affordable Care Act (ACA), use of such categories is consistent

with the ACA's intent that the Secretary ask that data that are collected for racial and ethnic disparities also look at underserved frontier counties. For example we notice that the total population in UIC=12 is 887,700, spread over 182 counties for a density of 4,877 per county. In other words, if the typical UIC=12 county were about 30*30 miles in size, the average density across these counties would be fewer than six per square mile. Further, the literature (Hart, 2012) supports the aggregation of UIC 9-12 as a specific approach to approximating frontier areas based upon county level data. CAPQuaM consulted with Gary Hart, Director of the Center for Rural Health at the University of North Dakota School of Medicine & Health Sciences, who is heading a Health Resources and Services Administration (HRSA)-funded project to develop new methods to analyze frontier health. We clarified that his work suggests that UIC 9-12 is the best overall approach to using county level data to study frontier health. Inclusion of UIC 8 would make the analysis more sensitive to including frontier areas but at a meaningful cost in sensitivity.

7.E. Limited English Proficiency (LEP) Populations

Not assessed.

Section 8. Feasibility

Feasibility is the extent to which the data required for the measure are readily available, retrievable without undue burden, and can be implemented for performance measurement. Using the following sections, explain the methods used to determine the feasibility of implementing the measure.

8.A. Data Availability

1. What is the availability of data in existing data systems? How readily are the data available?

Data elements necessary for this measure include: date/time of delivery, date/time/value of first temperature after delivery, infant characteristics (birth weight, Apgar), delivery characteristics (e.g., location of delivery, nursery level, delivery type), and demographics (e.g., race, ethnicity, insurance, zip code). To determine the availability and ease of collecting these data elements, CAPQuaM used three primary sources: a feasibility survey of 13 hospitals conducted by The Joint Commission under contract to CAPQuaM, analysis of the Mount Sinai Data Warehouse, and a New York Statewide neonatal database that is a part of a voluntary statewide effort championed by the New York State Department of Health.

We included 13 geographically and clinically diverse hospitals at varying stages of EMR development in the feasibility assessments. The surveys were completed by the quality improvement team at each hospital. Results of these surveys revealed that the data elements required for these measures (or the information required to calculate the data element (e.g., age of neonate at time of temperature) are available at the hospital level within existing medical record systems and are not difficult to abstract.

For delivery characteristics, respondents indicated that information would be available on the infant's record, with most elements also available on the mother's record. The EMR was the

preferred source of such data elements. For all other items, 12 hospitals indicated that the data were not difficult to collect, and none said that data were unavailable. A similar pattern of responses was seen regarding questions about identifying the date and time of delivery and of arrival to the intensive care nursery. Times at which the measurement was taken (rather than the time of documentation) were universally described as present. In general, the required data elements were reported to be not difficult to collect (12/13). Data on the infant (e.g. birth weight, 5 minute Apgar score) were said to be in all of the EMRs. EMR data were seen as available to identify those managed for comfort care only, and 12 hospitals indicated that such data would not be difficult to collect. Depending upon the data element, 11-13 of the sites said that race and ethnicity data and payment source would be available from the EMR. Two sites indicated that there would be a challenge to linking an infant's chart to the mother's chart, with more than 80 percent of the others indicating that such linkages could be performed electronically.

Analysis of the Mount Sinai Data Warehouse found that temperatures and time of temperature are often available in the Epic EMR. We found our ICD-9 schema was capable of identifying LBW infants. Some of the codes not specifically associated with a birth weight (e.g., growth retardation) were less specific for identifying LBW neonates. Details are discussed in the validation section. Of the hospitals that participate in the New York State neonatal database and using New York State designations, 23 of 25 (92.0 percent) classified as Level 2 nurseries submit temperature data, 31 of 36 (86.1 percent) with a Level 3 designation submit temperature data, and 16 of 18 (88.6 percent) of Regional Perinatal Centers submit temperature data. These data are virtually complete for those institutions that submit data. These data capture 84.1 percent of LBW admissions to Level 2 or higher nurseries in one year. Medicaid represents nearly half of babies entered into the database. We conclude that the necessary data are available at the level of the hospital, and that such data could be collected by health plans or Medicaid programs or other entities with contractual arrangements with the providing hospitals.

2. If data are not available in existing data systems or would be better collected from future data systems, what is the potential for modifying current data systems or creating new data systems to enhance the feasibility of the measure and facilitate implementation?

The data required for the CAPQuaM perinatal measures are generally available in the existing data systems. We cannot comment on the readiness of systems to provide routine output into a database suitable for analysis and generation of these measures, but there are no fundamental barriers to such being accomplished. We are in the process of developing an intranet-based interface for the collection of relevant data at the time of admission to the NICU at the Mount Sinai Medical Center to serve as a demonstration site for the efficient implementation of these data and these measures for quality measurement.

As indicated above, much if not all of the needed data could be captured in the EMR and transferred to an analytical database for quality measurement and reporting. A large proportion of these data elements are already captured routinely.

8.B. Lessons from Use of the Measure

1. Describe the extent to which the measure has been used or is in use, including the types of settings in which it has been used, and purposes for which it has been used.

The measure is being implemented for routine quality measurement at the Mount Sinai Medical Center.

2. If the measure has been used or is in use, what methods, if any, have already been used to collect data for this measure?

We plan to use the Epic EMR to the extent possible and supplement with an electronic data entry system that is algorithmic and efficient, with a database residing on the hospital's secure servers. The planning and development for this implementation are ongoing.

3. What lessons are available from the current or prior use of the measure?

The measure is not currently in use.

Section 9. Levels of Aggregation

CHIPRA states that data used in quality measures must be collected and reported in a standard format that permits comparison (at minimum) at State, health plan, and provider levels. Use the following table to provide information about this measure's use for reporting at the levels of aggregation in the table.

For the purpose of this section, please refer to the definitions for provider, practice site, medical group, and network in the Glossary of Terms.

If there is no information about whether the measure could be meaningfully reported at a specific level of aggregation, please write "Not available" in the text field before progressing to the next section.

Level of aggregation (Unit) for reporting on the quality of care for children covered by Medicaid/ CHIP†:

State level* Can compare States

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**

Yes.

Data Sources: **Are data sources available to support reporting at this level?**

No.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**

One hospital typically can provide meaningful sample size. Stratified analysis will benefit from aggregation of multiple facilities. Sample size of 15-20 per stratum is adequate to provide useful information.

In Use: **Have measure results been reported at this level previously?**

No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**

No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**

None anticipated.

Other geographic level: Can compare other geographic regions (e.g., MSA, HRR)

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**

Yes.

Data Sources: **Are data sources available to support reporting at this level?**
No.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**

One hospital typically can provide meaningful sample size. Stratified analysis will benefit from aggregation of multiple facilities. Sample size of 15-20 per stratum is adequate to provide useful information.

In Use: **Have measure results been reported at this level previously?**

No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**

No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**

None anticipated. Measure is specified using Urban Influence Codes. Because Zip codes or counties are requested, other geographic aggregations are feasible.

Medicaid or CHIP Payment model: Can compare payment models (e.g., managed care, primary care case management, FFS, and other models)

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**

Yes.

Data Sources: **Are data sources available to support reporting at this level?**

Yes.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**

One hospital typically can provide meaningful sample size. Stratified analysis will benefit from aggregation of multiple facilities. Sample size of 15-20 per stratum is adequate to provide useful information.

In Use: **Have measure results been reported at this level previously?**

No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**

No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**

None anticipated.

Health plan*: Can compare quality of care among health plans.

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**

No.

Data Sources: **Are data sources available to support reporting at this level?**

No.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**

One hospital typically can provide meaningful sample size. Stratified analysis will benefit from aggregation of multiple facilities. Sample size of 15-20 per stratum is adequate to provide useful information.

In Use: **Have measure results been reported at this level previously?**

No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**

No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**

Although the unit of analysis is the neonate, this measure was designed to aggregate at the level of the hospital or other higher aggregation of hospitals. Since plans manage a fraction of the patients within a hospital, we have no experience looking at it by health plan. We did look at Medicaid vs. commercial statewide in our testing.

Provider Level

Individual practitioner: Can compare individual health care professionals

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**

No.

Data Sources: **Are data sources available to support reporting at this level?**

No.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**

Specified for hospitals but not for individual clinicians.

In Use: **Have measure results been reported at this level previously?**

No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**

No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**

Not recommended.

Provider Level

Hospital: Can compare hospitals

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**

Yes.

Data Sources: **Are data sources available to support reporting at this level?**

Yes.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**

One hospital typically can provide meaningful sample size. Stratified analysis will benefit from aggregation of multiple facilities. Sample size of 15-20 per stratum is adequate to provide useful information.

In Use: **Have measure results been reported at this level previously?**

No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**

No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**

None anticipated. Designed to be used as part of a larger measure set.

Provider Level

Practice, group, or facility:** Can compare: (i) practice sites; (ii) medical or other professional groups; or (iii) integrated or other delivery networks

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**

Yes.

Data Sources: **Are data sources available to support reporting at this level?**

Yes.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**

One hospital typically can provide meaningful sample size. Stratified analysis will benefit from aggregation of multiple facilities. Sample size of 15-20 per stratum is adequate to provide useful information.

In Use: **Have measure results been reported at this level previously?**

No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**

No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**

None anticipated.

Section 10. Understandability

CHIPRA states that the core set should allow purchasers, families, and health care providers to understand the quality of care for children. Please describe the usefulness of this measure toward achieving this goal. Describe efforts to assess the understandability of this measure (e.g., focus group testing with stakeholders).

This measure describes the percent of eligible LBW neonates who have had a temperature taken within the first hour of birth. It is a process measure and a patient safety measure, and its meaning is self-evident.

This measure is intuitive and as a simple percentage, it is readily understood. Understandability is at the heart of CAPQuaM's measure development process. Throughout development, CAPQuaM brought together diverse stakeholders – clinicians, scientists, payers, purchasers, consumer organizations, and others – to ensure their iterative engagement in advancing quality

measures that are understandable, salient and actionable. CAPQuaM employed a 360-degree method, designed to involve key stakeholders in meaningful ways.

Our development process for this measure cultivated formal input from:

- Medical literature (both peer reviewed and gray, including State Web sites).
- Relevant clinicians.
- Organizational stakeholders (our consortium partners, as well as advisory board members,
- see below).
- Multi-disciplinary, geographically diverse expert panel including clinicians and academicians.
- CAPQuaM's scientific team.

Clinical criteria, including consideration of inclusion and exclusion criteria, regarding reporting approaches, the value of temperature measurement, and specific and meaningful temperature cutoffs were developed using a modified version of the RAND/UCLA modified Delphi Panels.

CAPQuaM sought recommendations from major clinical societies and other stakeholders to identify academic and clinician expert panel participants from a variety of backgrounds, clinical and regional settings, and expertise. The product of this process was participation by a broad group of experts in the development of clinically detailed scenarios leading to the measures.

CAPQuaM integrated perspectives from a national consortium, Steering Committee, and Senior Advisory Board at each step of the process, in addition to a continuing collaboration with AHRQ. Our team far exceeded the required minimums for expertise outside of the mainstream medical system, ensuring understandability at various levels and by a variety of audiences.

Alpha testing was performed to assess feasibility, mechanisms of data collection, and operational aspects of collecting and analyzing data for the measure. Beta testing included analysis of Mount Sinai and statewide data.

The route to measure specification included development of relevant scenarios and issues for formal processing by our expert panel who participated in a two-round RAND/UCLA modified Delhi panel that culminated in a day-long, in-person meeting hosted at the Joint Commission and moderated by a pediatrician and an obstetrician-gynecologist. The output from that panel meeting was summarized in the form of a boundary guideline that was then used to guide the measure specification and prioritization.

Section 11. Health Information Technology

Please respond to the following questions in terms of any health information technology (health IT) that has been or could be incorporated into the measure calculation.

11.A. Health IT Enhancement

Please describe how health IT may enhance the use of this measure.

Our measure is relevant for implementation in EHRs. The use of health IT will mitigate onerous data collection and data mining, as electronic querying enables efficient searching for relevant ICD-9 and CCS codes for this measure. Additionally, institutional use of EHR facilitates downstream clinical decision support that will prompt appropriate measurement and documentation of neonatal thermal management. In assessing the feasibility of capturing necessary data elements for the measure, we queried hospitals on the source record (e.g. Electronic Health Record, Paper Medical Record, Infant Record, Maternal Record) for measure numerator and denominator elements, and found consistency across all 12 respondents. This included characteristics such as time of arrival to the NICU as well as infant temperature in the delivery room and upon admission to the NICU. Additionally, the feasibility assessment also assessed ease of capturing necessary data elements on the part of the hospital site, and most sites responded that it was not difficult to abstract the required data from the chart. There were, however, discrepancies in the format for reporting date and time in the medical record, suggesting that the fields required to calculate the measure are not currently standardized. The lack of standardization of required fields suggests that these data fields need to be incorporated into EHR technical standards, so as to increase feasibility and reliability of measure reporting based on EHR data.

We are working with Mount Sinai Medical Center's NICU, which has decided to implement this measure as a routine part of its quality measurement. We are designing an intranet portal and data collection system to sit within the medical center's firewall that will collect the necessary data elements at the time of admission to the NICU. We are exploring the capacity for this system to handshake and collect or distribute information via the EPIC API.

11.B. Health IT Testing

Has the measure been tested as part of an electronic health record (EHR) or other health IT system?

Yes.

If so, in what health IT system was it tested and what were the results of testing?

Yes, please see section above.

11.C. Health IT Workflow

Please describe how the information needed to calculate the measure may be captured as part of routine clinical or administrative workflow.

These data are already captured as a part of routine work flow. The limitation is that they are not always captured in searchable fields or form. See above for discussion.

11.D. Health IT Standards

Are the data elements in this measure supported explicitly by the Office of the National Coordinator for Health IT Standards and Certification criteria (see healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__standards_ifr/1195)?

No.

If yes, please describe.

Not applicable.

11.E. Health IT Calculation

Please assess the likelihood that missing or ambiguous information will lead to calculation errors.

The query of the Mount Sinai Data Warehouse data, which included Epic EMR data, had no challenge in calculating the age that first temperature was obtained or in reporting temperature in degrees Fahrenheit once the temperature was found. However, our review of the results leads us to doubt that the first temperature is universally recorded in its intended field (many small neonates had no temperatures recorded in their charts at all). Certain flow sheets and other items are scanned and entered into the EMR and this represents a challenge to the national migration to health IT in the current environment. The use of fixed fields and searchable text fields in preference to scanned documents is a part of the evolution of health IT that is still in process. Advancement of these practices would improve the capacity to perform this as an e-measure. In theory, this measure is ideally suited for electronic measurement, using EMR data.

11.F. Health IT Other Functions

If the measure is implemented in an EHR or other health IT system, how might implementation of other health IT functions (e.g., computerized decision support systems in an EHR) enhance performance characteristics on the measure?

Optimization of the EHR to improve data capture would offer the potential to create operational run charts and the use of statistical process control and QI approaches to improve performance and clinical outcomes.

Section 12. Limitations of the Measure

Describe any limitations of the measure related to the attributes included in this CPCF (i.e., availability of measure specifications, importance of the measure, evidence for the focus of the measure, scientific soundness of the measure, identification of disparities, feasibility, levels of aggregation, understandability, health information technology).

As noted above, limitations on the capacity to calculate the measure from the current IT infrastructure is a meaningful limitation.

The measure is a simple intuitive measure. It would be complemented by measures of temperature or more complex measures of thermal management. Our measure set is a meaningful first step in that process.

Section 13. Summary Statement

Provide a summary rationale for why the measure should be selected for use, taking into account a balance among desirable attributes and limitations of the measure. Highlight specific advantages that this measure has over alternative measures on the same topic that

were considered by the measure developer or specific advantages that this measure has over existing measures. If there is any information about this measure that is important for the review process but has not been addressed above, include it here.

This measure describes the proportion of infants who are less than 2500 grams and have

documentation that their first temperature was taken in the first hour of life or the "golden hour." More than 100 years of literature supports the ongoing salience of appropriate thermal management of low birth weight infants and, unfortunately, variable clinical performance persists. Our proposed measure is a process measure that is also a patient safety indicator.

This measure topic was assigned to the CAPQuaM as a PQMP priority by the Agency for Health Care Quality and Research with the active consultation of the Centers for Medicare & Medicaid Services. In addition to literature reporting studies conducted in a variety of settings, including the NICHD neonatal research network and the Vermont Oxford Network, that documents this problem, we have found performance concerns in New York City and New York State. Our chart review data from three diverse hospitals in New York City showed variation in temperatures recorded across the weight spectrum within and between hospitals. These differences were meaningful with cooler babies more likely to die.

In New York State, about half of low birth weight babies are insured by Medicaid. Hypothermia is not only associated with neonatal mortality, but there is evidence (Reynolds, et al., 2009) that intraventricular hemorrhage (IVH) can also be a consequence of hypothermia. IVH is a significant cause of disability and developmental delay, and when serious, it is a common cause for LBW infants to develop into children with special health care needs. This has broad implications for Medicaid, Medicaid expenses, and early intervention services, including EPSDT services. Hypothermia, through death and disability may have a long tail that impacts families and programs associated with Medicaid. Furthermore, the Medicaid population is disproportionately black, and in our testing data, black infants were disproportionately hypothermic.

Key findings from our study of 7,553 neonates admitted to Level 2 or higher nurseries in New York State are: temperature was variable within weight categories; blacks were disproportionately cool compared with Hispanic or non-Hispanic others who were disproportionately cool compared with non-Hispanic whites; and deaths were disproportionate among those who were cool, in a graded fashion. Only 36 percent of Medicaid infants were euthermic, compared to 40 percent of commercially insured babies. We also found systematic differences in the timing of when the temperatures were taken.

This history, these data, and the absence of currently recommended measures that adequately address this issue all motivate the work of the CAPQuaM to develop this measure as part of the initial set of inpatient perinatal measures developed in the PQMP. Clinically, we have demonstrated that the temperature of low birth weight neonates is variable, and it is highly consequential in terms of critical outcomes like survival and intraventricular hemorrhage. Institutional anecdotal evidence supports literature observations that thermal management can be managed and improved at the unit level with improved outcomes. Despite limitations, our data from Mount Sinai demonstrate that a substantial number of LBW neonates do not have documentation of temperatures in the first hour. Since temperature is critical for early clinical management, even a single failure is too many.

This measure is both a patient safety indicator and an independent metric related to a desirable process of care that was put forth and endorsed by the systematic CAPQuaM process, including the recommendations of a multidisciplinary national expert panel using a RAND/UCLA modified Delphi process. We are proud to nominate this measure for consideration as a core measure in the CHIPRA/Medicaid pediatric set.

References

Baker JP. The incubator and the medical discovery of the premature infant. J Perinatol 2000; 20(5):321-8.

Currier A. Diseases of the Newborn. In Ce S, ed. Philadelphia, PA: F.A. Davis; 1891.

Diamond CC, Rask KJ, Kohler SA. . Use of Paper Medical Records Versus Administrative Data for Measuring and Improving Health Care Quality: Are We Still Searching for a Gold Standard? Disease Management 2004; 4(3):121-30.

Doyle KJ, Bradshaw WT. Sixty golden minutes. Neonatal Netw 2012; 31(5):289-94.

Engle WA, Tomashek KM, Wallman C. "Late-preterm" infants: a population at risk. Pediatrics 2007; 120(6):1390-401.

Fischer L. Diseases of Infancy and Childhood. Philadelphia, PA: F.A. Davis; 1915.

Garrison F. History of Pediatrics. In Abt IA, Philadelphia, PA: W.B. Saunders; 1923.

Hall SA, Kaufman JS, Ricketts TC. Defining urban and rural areas in U.S. epidemiologic studies. J Urban Health 2006; 83(2):162-75.

Hart G. Frontier/Remote Island, and Rural Literature Review, 2012. Available at http://ruralhealth.und.edu/frontier/pdf/lit_review.pdf. Accessed August 31, 2016.

Holt LE. The Care of Premature and Delicate Infants. In The Diseases of Infancy and Childhood, 2nd Ed, Rev. New York, NY: D. Appleton; 1902.

Holt LE, Macintosh, R. Holt's Diseases of Infants and Children, 11th ed. New York, NY: Appleton; 1940.

Howell EA, Holzman I, Kleinman LC, et al. Surfactant use for premature infants with respiratory distress syndrome in three New York City hospitals: discordance of practice from a community clinician consensus standard. J Perinatol 2010; 30(9):590-5.

Howell EA, Stone J, Kleinman LC, et al. Approaching NIH guideline-recommended care for maternal-infant health: clinical failures to use recommended antenatal corticosteroids. Matern Child Health J 2010;14(3):430-6.

Kawachi I, Berkman, LF. Neighborhoods and Health. New York, NY: Oxford University Press; 2003.

Laptook A, Jackson GL. Cold stress and hypoglycemia in the late preterm ("near-term") infant: impact on nursery of admission. Semin Perinatol 2006; 30(1):24-7.

Laptook AR, Salhab W, Bhaskar B. Admission temperature of low birth weight infants: predictors and associated morbidities. Pediatrics 2007; 119(3):e643-9.

Laptook AR, Watkinson M. Temperature management in the delivery room. Semin Fetal Neonatal Med 2008; 13(6):383-91.

Miller SS, Lee HC, Gould JB. Hypothermia in very low birth weight infants: distribution, risk factors and outcomes. J Perinatol 2011; 31(Suppl 1):S49-56.

Pierce RV. The People's Common Sense Medical Adviser in Plain English. Buffalo, NY: World's Dispensary Printing Office; 1875.

Profit J, Gould JB, Zupancic JA, et al. Formal selection of measures for a composite index of NICU quality of care: Baby-MONITOR. J Perinatol 2011; 31(11):702-10.

Reynolds RD, Pilcher J, Ring A, et al. The Golden Hour: care of the LBW infant during the first hour of life; one unit's experience. Neonatal Netw 2009; 28(4):211-9; quiz 55-8.

Rubio D, Berg-Weger M, Tebb SS, et al. Objectifying content validity: conducting a content validity study in social work research. Soc Work Res 2003; 27(2):94-104.

Silverman WA, Feritg JW,, Berger AP. The influence of the thermal environment upon the survival of newly born premature babies. Pediatrics 1958; 22:876-86.

Sinclair JC. Servo-control for maintaining abdominal skin temperature at 36 C in low birth weight infants. Cochrane Database Syst Rev 2007; (1):CD001074.

Virnig BA, McBean M. Administrative data for public health surveillance and planning. Annu Rev Public Health 2001; 22:213-30.

Watkinson M. Temperature control of premature infants in the delivery room. Clin Perinatol 2006; 33(1):43-53, vi.

Section 14: Identifying Information for the Measure Submitter

First Name:	Lawrence
Last Name:	Kleinman
Title:	Director, Mount Sinai CAPQuaM
Organization:	Collaboration for Advancing Pediatric Quality Measures
Mailing Address:	One Gustave L. Levy Place, Box 1077
City:	New York
State:	NY
Postal Code:	10029
Telephone:	212-659-9556
Email:	**l**awrence.kleinman@mssm.edu

The CHIPRA Pediatric Quality Measures Program (PQMP) Candidate Measure Submission Form (CPCF) was approved by the Office of Management and Budget (OMB) in accordance with the Paperwork Reduction Act.

The OMB Control Number is 0935-0205 and the Expiration Date is December 31, 2015.

Public Disclosure Requirements

Each submission must include a written statement agreeing that, should U.S. Department of Health and Human Services accept the measure for the 2014 and/or 2015 Improved Core Measure Sets, full measure specifications for the accepted measure will be subject to public Centers for Medicare & Medicaid Services [CMS] websites), except that potential measure users will not be permitted to use the measure for commercial use. In addition, AHRQ expects that measures and full measure specifications will be made reasonably available to all interested parties. "Full measure specifications" is defined as all information that any potential measure implementer will need to use and analyze the measure, including use and analysis within an electronic health record or other health information technology. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure. This statement must be signed by an individual authorized to act for any holder of copyright on each submitted measure or instrument. The authority of the signatory to provide such authorization should be described in the letter.

AHRQ Pub. No. 14(17)-P006-4-EF

November 2016