Distribution of Temperatures for Low Birth Weight Neonates

Section 1. Basic Measure Information

1.A. Measure Name

Distribution of Temperatures for Low Birth Weight Neonates Admitted to Level 2 or Higher Nurseries in the First 24 Hours of Life

1.B. Measure Number

0118

1.C. Measure Description

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

Describes the distribution of temperatures of live-born neonates less than 2500 grams that arrive to a Level 2 or higher nursery.

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 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 (LBW) 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

This is a continuous variable. The parameter of interest is the neonate's temperature on admission to the neonatal intensive care unit (NICU). Our standard, assessed in Proposed Measure 2 of the PQMP Inpatient Perinatal Collection No. 1, is that infants admitted to a Level 2 or higher nursery have this temperature taken within 15 minutes of admission.

This measure requests the reporting of the following:

1. Descriptors of the Center of the Distribution (for sample size >=5):
   a. Mean
   b. Median (50th percentile)
   c. 25th percentile
   d. 75th percentile
2. Descriptors of Dispersion:
   a. Interquartile range (for sample size >=5)
   b. Standard deviation (for sample size >=10)
3. Descriptors of the Warm End of the Distribution:
   a. 99th percentile (for sample size >=33)
   b. 95th percentile (for sample size >=20)
   c. 90th percentile (for sample size >=10)
4. Descriptors of the Cool End of the Distribution:
   a. 1st percentile (for sample size >=33)
   b. 5th percentile (for sample size >=20)
   c. 10th percentile (for sample size >=10)

Data Elements:

- Temperature to first decimal place

- Units of temperature (Celsius, Fahrenheit)
- Method of temperature measurement (axillary, rectal, skin, tympanic)

1.H. Numerator Exclusions

None.

1.I. Denominator Statement

All infants born in a medical facility with birth weights less than 2500 grams and admitted to a level 2 or higher nursery within 24 hours of birth. Identification of newborns who may be eligible for inclusion in the denominator may be accomplished through the use of the following ICD-9 codes listed in Table 1 (see supporting documents).

For codes 76400, 76410, 76420, 76490, 76500, birth weights should be verified from the medical record prior to inclusion in measure.

Denominator Elements:

Number of infants less than 2500 grams who are admitted to a Level 2 or higher nursery in the first 24 hours of life

General data elements for measure set:

- Date of birth
- Time of birth
- Time of first temperature taken (any location)
- Date of first temperature taken (any location)
- Value of first temperature taken (any location
- Units of first temperature taken (any location)
- Route of first temperature taken (any location)
- Time of arrival at nursery
- Time of first temperature taken in nursery
- Date of first temperature taken in nursery
- Value of first temperature taken in nursery
- Units of first temperature taken in nursery
- Route of first temperature taken in nursery

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 is referring to a birthing or delivery room on 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

Measure describes a distribution and requires limited calculations. Interquartile range is the absolute difference between the temperature values at the 25th and 75th percentiles (SAS Proc Univariate provides all requested elements).

1.J. Denominator Exclusions

- Neonates with comfort care (requires all of these features: infant died within 48 hours of birth; AND 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.
- Neonates for whom the hospital provides documentation that at the time of arrival to the NICU and before the temperature was taken the infant both met written institutional criteria for therapeutic hypothermia and was managed with hypothermia (this is an optional exclusion).

1.K. Data Sources

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

Administration 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 Pediatric Quality Measures Program (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 (Garrison, 1923; Holt, 1902; Baker, 2000; Pierce, 1875; Currier, 1891; Fischer, 1915; Holt, Macintosh, 1940). 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 have collected data from chart review at three diverse hospitals in New York City. All three hospitals had a range of birth weights and a range of temperatures, both when we considered the actual measured temperature and when we adjusted those that were not taken rectally to create a "corrected" core temperature (see Figures 1 and 2 in the supporting documents).

Temperature predicted in-hospital mortality after controlling for covariates, whether we dichotomized at the 35.5° threshold that our local physicians propose or considered each degree of temperature as a continuous variable. Crossing the threshold into hypothermia more than doubled the odds of death, controlling for other variables in the model. The relationship between temperature and survival is monotonic: an increase of each 1° Celsius up to 37° reduced 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 odds of mortality of 84 percent, p=0.19.

Risk ratio (RR) is a more informative way to express the results than an odds ratio, especially when the underlying risk is large as in this study (Profit, et al., 2011). Regression risk analysis estimates the adjusted risk ratio (ARR) and adjusted risk difference: hypothermia (35.5°C) results in an ARR of 1.48 (95 percent confidence interval 1.03—2.30), indicating a 48 percent increase in risk, from a baseline risk of 8.9 percent among those who were euthermic to an exposed risk of 13.1 percent among those who were hypothermic, controlling for the covariates in the sample. Considering temperature as a continuous variable reveals that increasing the temperature from 34.0° to 35.0° increases the relative chance of survival by 24 percent, from 35.0 to 36.0 by 26 percent, and from 36.0 to 37.0 by 27 percent, resulting in absolute risk reductions of 2.8 percent, 2.4 percent, and 2.0 percent respectively. A core body temperature increase from 34.0° to 37.0° is associated with a relative decrease in mortality of 98 percent and an absolute decrease in mortality of 7.2 percent, controlling for other factors in the model. The

decrease from 36.0° to 35.5° is associated with a 12 percent increase in the adjusted mortality risk from 9.4 percent to 10.5 percent. Our work confirmed findings in the literature that insurance status and race (Miller, Lee, Gould, 2011) are associated with outcomes. Anecdotal reports from among our participating hospitals confirm reports in the literature (Billimoria, Chawla, Bajaj, 2013) that attention to thermal management can improve temperature outcomes. Please see the supporting documents for a more complete literature review. Despite evidence of the importance of temperature on outcomes of neonates, two proposed measures for quality of care – taking the temperature 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 employed a highly engaged process to develop an enhanced set of measures. A distinguished multidisciplinary panel of national experts that included neonatologists, family physicians, nurses, and a pediatric hospitalist specifically requested a measure that incorporates the moments of the distribution into the measure. After beta testing, the first and 99th percentiles were substituted for the panel's preference of highlighting the five highest and lowest temperatures respectively. Two findings motivate this change: first, there are wide variations in sample size, with a number of nurseries having smaller samples sizes, and concern that reporting all these values would invite confusion rather than clarification especially with small samples and further division of the sample when stratifying for reporting. Second, in our pretesting of the measure, we found that the 99th and 1st percentiles were particularly informative. This history, these data, and the absence of currently recommended measures that address adequately this issue all motivated the work of CAPQuaM to develop a measure of quality of care based on temperature upon admission to the NICU as the initial inpatient perinatal topic in the national Pediatric Quality Measures Program (PQMP), funded by the Children's Health Insurance Reauthorization Act of 2009 (CHIPRA, 2009).

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 not only associated with neonatal mortality, but there is evidence (Miller, Lee, Gould, et al., 2011) that intraventricular hemorrhage (IVH) can also be a consequence of hypothermia. IVH is a significant cause of disability, developmental delay, and when serious, 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 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. We note above that there is evidence that management can enhance thermal outcomes.

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 low birth weight infants, temperatures taken within an hour of admission to the NICU. This was rejected largely because it was met 98 percent of the time. While we would hold with VON that 98 percent compliance is inadequate for a quality measure that it is so closely related to patient safety, we have proposed two measures that adopt a slightly different approach. The first hour after life is well known as the "Golden Hour" because of the importance of timely recognition and management of neonatal outcomes (Doyle, Bradshaw, 2012; Reynolds, Pilcher, Ring, et al., 2009).

We propose a measure that looks at the proportion of low birth weight neonates who have a temperature documented within the first hour of life. We consider this a safety measure, as missed hypothermia may lead to shock and death. Those infants who are low birth weight and do not require admission to the advanced care nursery may be at risk for being managed more like full-term infants—that is, 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. Hence, this measure is inclusive of all low birth weight infants. Further, all those infants who require admission to an advanced level of care (a 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. This 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 degrees Celsius. It was rejected in part because there is no consensus regarding the desirable threshold. Based on our read of the data in the literature and our own data described above, we believe that temperature provides increasing risk the further it falls below 37º degrees Celsius. Our two temperature measures in this set provide discrete and continuous ways of looking at the distribution of temperatures, stratified by birth weight and reported for various subgroups when sample size is sufficient. Our data demonstrate that optimal thermal management has the capacity to keep even tiny babies warm. The harmful consequences of cold stress are greater in

smaller babies than in larger ones. Hence, we believe that data should be reported for the entire nursery, as well as stratified when sample size allows.

This 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 are particularly valuable to help guide quality improvement (QI) 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.

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:** No.
d. **Service – preventive health, including services to promote healthy birth:** Yes.
e. **Service – care for acute conditions:** Yes.
f. **Service – care for children with acute 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; first day.
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 and targeted review of the literature, (see 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. As discussed below, our clinical distinctions, including our decision to report ranges and distributions, were informed and shaped by a diverse and superb multidisciplinary panel of national experts. The ratings of the panel along with a brief description of methodology are included as supporting documents.

A brief summary of the research findings includes that the temperatures of low birth weight infants vary, based on their management, that every degree below 37º Celsius adds meaningful risk in a continuous and not only a threshold manner, that consequential outcomes include death and intraventricular hemorrhage, and that hospitals can improve their performance on temperature outcomes.

Further evidence is provided in Section 6.B. Validity. We report on New York State neonatal data. Hospitals use various means to collect data on their high-risk newborns, but they must submit the data using the NICU Module's on-line data entry or import function. To ensure data security and patient confidentiality, hospitals must register their data entry or enter 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 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. Twenty-five 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 is an important measure of performance. See Section 6.B.Validity 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.

Temperature of low birth weight neonates is variable, can be managed at the level of the individual patient and at the level of the unit providing care, and is highly consequential in terms of critical outcomes, such as survival and intraventricular hemorrhage. When viewed at a population level, the lower the temperature, the larger the consequences.

Please see discussion and literature summaries presented in Section 3.A, Evidence for General Importance of the Measure, as well as information in Section 6, Scientific Soundness of the Measure. The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children (Mangione-Smith, DeCristofaro, Setodji, et al., 2007). In addition, 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 (Rubio, Berg-Weger, Tebb, et al., 2003).

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). This presentation is included as an Appendix.
- 2012 Oxford-Mount Sinai Bioethics Consortium (Amsterdam).
- 2012 Child Health Services Research Interest Group at Academy Health (Orlando).

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 who 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 (Diamond, Rask, Kohler, 2004). The constructs underlying our measures are data and time and temperature.

A feasibility study designed to determine the ability and ease of collecting related data showed that date and time are self-evident and that there is mild but manageable variation in how time is reported. This should 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 is required in our measure set. The underlying construct for temperature is the core body temperature of the neonate. For neonates of various sizes and gestational ages, the optimal approach to measuring the temperature may vary. Measurement approaches that are understood to be valid (articles and specifics for this are in our literature review in the Appendix) may include rectal temperatures, axillary temperatures, and when appropriately shielded from a radiant heat source, skin probe temperatures. Our research in New York City hospitals found that neonates who were documented to have a rectal temperature were on average about 0.5º Celsius warmer than those for whom the site of temperature was not documented to be rectal. Other studies reported in the literature do not report such a difference, so this may be thought of as an upper bound regarding potential underestimation of core body temperature.

We understand that it would be a barrier to the wide adoption of this measure were we to specify changes to institutional standards of care regarding how to measure and record the temperature of low birth weight infants or to establish requirements for measurement given the current evidence in the literature. Therefore we do not offer such specification. Instead we ask that reporting agencies record and share the data regarding how each temperature was assessed so that the agencies receiving the data may use that information should they wish to do so. The reliability of modern methods for assessing temperature is very high.

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 reliability section above contains some information related to validity as well. The use of electronically available administrative data in health care research and assessment is becoming increasingly common. Most databases contain consistent elements, are available in a timely manner, provide information about large numbers of individuals, and are relatively inexpensive to obtain and use. Validity has been established, and its strengths and weaknesses relative to data abstracted from medical records and obtained via survey have been documented (Virnig, McBean, 2001). Administrative data are supported, if not encouraged by Federal agencies, including the National Institutes of Health (NIH), AHRQ, the Health Care Financing Administration), and the Veterans Administration (VA). This measure calls for the use administrative data to identify the universe of low birth weight infants. As previously stated, our work was conducted through a formal method, a pragmatic adaptation of the CAPQuaM 360º method.

The measure 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. See Section 5.B, Clinical or Other Rationale Supporting the Focus of the Measure, for additional detail.

Our feasibility work indicates that the time that the temperature is assessed, rather than simply the time that it is documented, is recorded in the medical record, generally an electronic medical record (EMR). This is a critical aspect of the validity of time data. Our underlying construct is core body temperature. Modern temperatures are valid and precise. The core body temperature is the highest of the accurate (legitimate) temperatures that may be obtained, so entities that report this measure will have aligned motivation to estimate temperatures that are as close to the core body temperature as possible. In one sense, the measure was designed with a compromise to pragmatism and can be thought of as having designed in a 0.5º "discount," in that our data suggest that optimal outcomes are obtained at 37.0º Celsius, rather than at the 36.5º in the measure (which is still far preferable to cooler). As we noted above, we have data that suggest that this 0.5º Celsius correction is at least adequate for population-level use. Further, hypothermic infants should be managed clinically using core body temperature, so there is further clinical alignment for the use of a method that approximates core body temperature.

Data from our pretesting support various aspects of this measure. All data are from the New York State neonatal database. 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 (thus excluding potential data errors). Excluded were those with anencephaly or those who expired within 48 hours without receiving respiratory support beyond oxygen in the NICU (N=7,553). The number of infants ranged from 21 to 370 per hospital, and 86.7 percent were admitted to Level 3 or higher hospitals.

For this work we used the first temperature on admission to a level 2 or higher nursery for those admitted within 24 hours of birth. In keeping with the categorical approach applied by the fourth measure in this set, we found that 1.9 percent of infants were <=34.5 (cold), 9.6 percent above 34.5 but <= 35.5 (very cool), 48.0 percent above 35.5 but <= 36.5 (cool), 37.9 percent above 36.5 but <= 37.5 (euthermic or appropriately warm), and 2.6 percent above 37.5 or overly warm.

There were only 67 newborns that were transferred in from another facility. The distributions of temperatures were similar to the inborn infants, with the exception that the transferred infants were slightly more likely to be euthermic. Of the inborn infants, the temperatures ranged from 29.0 to 39.7. See Table 2 in the supporting documents. The median was 36.4, the mean was 36.3, and the standard deviation was 0.7 with an interquartile range of 0.80. Only four infants arrived in the Level 2 or higher nurseries from the emergency department: one infant was euthermic, one was cool, and two were very cool. Nearly 1 percent of the infants were transferred from the Newborn Nursery, of which 48 percent were euthermic, 44 percent cool, and only 6 percent very cool. None were cold.

We did not have delivery location in the dataset and therefore classified neonates born by C-section or deliveries of multiple gestations as being born in the operating room (5,254), and the remainder were classified as being born in a labor and delivery room/birthing room (2,245). Of those born in the operating room, 2 percent were cold, 11 percent very cool, 72 percent cool, and 35 percent euthermic. Those born in the labor and delivery suite were warmer, with 2 percent cold, 7 percent very cool, 13 percent cool, 48 percent euthermic, and 45 percent too warm ($p<.0001$). This suggests that our categorization of babies born in the operating room (while imperfect) does identify a distinct population. Our expert panel recommended that we report by site of delivery.

We found that temperatures varied by birth weight category ($p<.0001$) considering those <1000 grams, 1000-1499 grams, and 1500-2499 grams, as suggested by our expert panel. The percent cold was over 10 percent for those under 1000 g (two-thirds of all cold babies from a group that was about 12 percent of all babies). These infants also were least likely to be euthermic, only 25 percent were so classified, compared to 34 percent of those in the intermediate weight category and 41 percent of the larger babies.

Using the categories defined in Proposed Measure 4 in this set, in-hospital deaths were disproportionately represented among cooler babies. 2.6 percent of babies died before discharge: 24.5 percent of cold; 5.4 percent of very cool, and 2.2 percent of cool babies compared to 1.4 percent of euthermic babies died; 1.6 percent of above normal warmth babies died. Only 20 percent of deaths were among euthermic infants.

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 race and ethnicity data are almost universally available and that method of assignment of race and ethnicity to the babies 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, with non-Hispanic other and Hispanic infants at intermediate values. Race and ethnicity were also independent predictors of temperature in our New York City data.

7.B. Special Health Care Needs

Not assessed.

7.C. Socioeconomic Status

We can use Medicaid insurance as a marker for SES. Our New York City data demonstrate this to be an independent predictor of poor thermal outcomes. We further used 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 geographic 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 of socioeconomic status 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 (USDA) 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 below, shows meaningful dispersion and supports our plan to build off quartiles of distribution with a finer focus in higher areas of poverty (see Table 3 in the supporting documents).

7.D. Rurality/Urbanicity

As described in the specifications, we use urban influence codes 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 with no own town
11 Non-core not adjacent to metro or micro with own town
12 Non-core not adjacent to metro or micro with no own town

We analyzed 3,143 county equivalents in the United States, and our results are listed in Table 4 (see Table 4 in the supporting documents).

The population is heavily weighted to metropolitan areas as shown in Table 5 (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 of the populations lives 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, use of such categories is consistent with the ACA's intent that the Secretary ask that data 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 less than six per square mile. Further, the literature supports (Hart, 2012) 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 for this measure include: date/time of delivery, date/time/value of temperatures after delivery and through the admission to Level 2 or higher nursery, 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.

The 13 hospitals included in the feasibility assessment were geographically and clinically diverse sites and were at varying stages of EMR development. The surveys were completed by the QI 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 in the infant's record, with most elements also available in 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 the 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 can 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 low birth weight 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 not 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 in 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 data base residing on the hospital's secure servers. The planning and development for this implementation is 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 a 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 a 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?**
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 a 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. Small sample size for some categories in our New York State data tests.

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 a 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?**
Not designed for or tested at plan level. Measure is intended to compare clinical units of care or large strata within those units, so health plans with large market share potentially could use this measure.

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?**
Not recommended.

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 a 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 with this population in mind and tested in this population.

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 a 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?**
Not recommended.

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 distribution of temperatures among low birth weight infants. Understanding this measure requires basic numeracy, and better numeracy opens this measure up for more nuanced understanding. As a simple continuous distribution there are many aspects. But the median may be described as the" typical value," the mean as the "average" value, and the interquartile range as the width of the middle part of the distribution. The various percentiles are also easily described. Because these are generally robust descriptors, the stability of most findings are achieved with relatively small sample sizes, again enhancing the capacity for these measures to be understood.

The CAPQuaM measure process aims to optimize and balance the sometimes competing considerations of validity and meaning, feasibility, usefulness, and understanding. Our proposed measures complement this continuous presentation of temperature with a categorical description of the data using lay-friendly terms, such as cold, very cool, and cool.

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° 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 websites)
- 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, including consideration of inclusion and exclusion criteria, reporting approaches, the value of temperature measurement, and specific and meaningful temperature cutoffs, were developed using an enhanced 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 with 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.

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 Delphi 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 regarding the Thermal Condition of Low Birth Weight Neonates Admitted to Level 2 or Higher Nurseries in the First 24 Hours of Life is relevant for implementation in electronic

health records (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 received responses from 12 hospitals on the source record (e.g. Electronic Medical 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 the required data were not difficult to abstract 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 activities. 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?

No.

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

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 part of routine work flow.

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.

11.E. Health IT Calculation

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

Not applicable.

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?

Accurate use of EMR or distinctly created data through a Web portal at the time of admission to the Level 2 or higher nursery offers the potential to create operational run charts and for 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).

The major limitation of this measure relates to small sample sizes when various stratifications are performed. While the mean and standard deviation are highly sensitive to outlying values, the median and interquartile range are particularly robust and therefore valuable with smaller sample sizes.

A minor limitation was discussed above and relates to the variable approaches used for estimating the core body temperature in practice. This is an intuitive proximal outcomes measure that is valid, varies in practice, and can be improved leading to improved outcomes.

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 distribution of temperatures of low birth weight infants upon their admission to a Level 2 or higher nursery in the first 24 hours of life. More than 100 years of literature support the ongoing salience of appropriate thermal management of low birth weight infants and, unfortunately, variable clinical performance persists.

This measure topic was assigned to the CAPQuaM as a PQMP priority by AHRQ with the active consultation of the Centers for Medicare & Medicaid Services (CMS). In addition to literature describing studies conducted in a variety of settings including the NICHD neonatal research network and the Vermont Oxford Network that document 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. The importance of evaluating the spectrum of temperature is evident from our analyses with temperature as a continuous variable. These analyses reveal that each increase in degree of temperature increases the relative chance of survival significantly. 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 (Billimoria, Chawla, Bajaj, et al., 2013) that intraventricular hemorrhage (IVH) can also be a consequence of hypothermia. IVH is a significant cause of disability, developmental delay, and when serious, 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 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.

In our study of 7,553 neonates admitted to Level 2 or higher nurseries in New York State we found that 1.9 percent of infants were <=34.5 (cold), 9.6 percent above 34.5 but <= 35.5 (very cool), 48.0 percent above 35.5 but <= 36.5 (cool), 37.9 percent above 36.5 but <= 37.5 (euthermic or appropriately warm), and 2.6 percent above 37.5. The distribution of mean temperature by nursery ranged from 35.7 to 38.2, with a median of 36.3, a standard error of 0.36, and an interquartile range of 0.4. Twenty-five percent of these nurseries had a mean temperature below 36.1. Key findings from these analyses were: temperature was variable within weight categories; blacks were disproportionately cool compared with Hispanic and 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 infants. 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 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.

References

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

Billimoria Z, Chawla S, Bajaj M, et al. Improving admission temperature in extremely low birth weight infants: a hospital-based multi-intervention quality improvement project. J Perinat Med 2013 41(4):455-60.

Currier A. Diseases of the Newborn. In Ce S, editor. 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.

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

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

Hart G. Frontier/Remote Island, and Rural Literature Review. Rockville, MD: Health Resources and Services Administration; 2012: Available at http://ruralhealth.und.edu/frontier/pdf/lit_review.pdf. Accessed August 25, 2016.

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

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

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

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

Mangione-Smith R, DeCristofaro AH, Setodji CM, et al. The quality of ambulatory care delivered to children in the United States. N Engl J Med 2007; 357(15):1515-23.

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. Social Work Res 2003; 27(2):94-104.

Silverman WA, Fertig 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 36C 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, P.O. Box 1077

City: New York

State: NY

Postal Code: 10029

Telephone: 212-659-9556

Email: Lawrence.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 disclosure (e.g., on the Agency for Healthcare Research and Quality [AHRQ] and/or 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 Publication No. 14(17)-P006-3-EF
October 2016