Section I. Basic Measure Information

I.A. Measure Name

CAPQuaM PQMP PERINATAL IV: Thermal condition of low birth weight neonates admitted to Level 2 or higher nurseries in the first 24 hours of life

I.B. Measure Number

0119

I.C. Measure Description

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

Stratifies live-born neonates less than 2500 grams that arrive to a Level 2 or higher nursery on the basis of admission temperature. Strata are cold (<=34.5), very cool (34.51-35.50), cool (35.51-36.50), euthermic (36.51-37.50) and overly warm (> 37.5).

I.D. Measure Owner

CAPQuaM

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

N/A

I.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)'s National Quality Measures Clearinghouse and are available at http://www.qualitymeasures.ahrq.gov/about/hierarchy.aspx:

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 PQMP Inpatient Perinatal Collection #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 Birthweight 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.

   N/A

I.G. Numerator Statement

The numerator has 5 comprehensive and mutually exclusive strata. Each is determined by the number of children whose qualifying temperature (usually the first temperature after arrival to the Level 2 or higher nursery) falls within the criteria for that stratum.

- Stratum 1 “Cold”: All neonates with temperatures less than or equal to 34.5 degrees Celsius.
- Stratum 2 “Very cool”: All neonates with temperatures greater than 34.5 degrees Celsius and less than or equal to 35.5 degrees Celsius;
- Stratum 3 “Cool”:All neonates with temperatures greater than 35.5 degrees Celsius and less than or equal to 36.5 degrees Celsius;
- Stratum 4 “Euthermic”: All neonates with temperatures greater than 36.5 degrees Celsius and less than or equal to 37.5 degrees Celsius;
- Stratum 5: “Overly warm”:All neonates with temperatures greater than 37.5 degrees Celsius.

Data Elements:

- Temperature to first decimal place

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

I.H. Numerator Exclusions

None

I.I. Denominator Statement

All infants born in a medical facility less with birthweights 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 to be included in the denominator may be accomplished through the use of the following ICD-9 codes:

I.J. Denominator Exclusions

- Neonates with comfort care (requires all of the features below): --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]

I.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 II: 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. Example 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 becalculated from a query of an appropriate electronic data source are not requested at this time, the availability of these resources may be a factor indetermining whether a measure can be recommended for use.

A. Description

Divides low birthweight neonates who are admitted to a Level 2 or higher nursery into five strata based upon their admission temperature and calculates the proportion of infants in each stratum based upon their temperature upon arrival to the Level 2 or higher nursery. All temperatures are analyzed using degrees Celsius and reported to one decimal place.

B. Eligible Population

Numerator: Live-born neonates with a birthweight of less than 2500 grams (as identified by ICD-9-CM Principal or Other Diagnosis Codes in Table 1) using the first temperature taken in Level 2 or higher nursery.

The numerator has 5 comprehensive and mutually exclusive strata. Each is determined by the number of children whose qualifying temperature (usually the first temperature after arrival to the Level 2 or higher nursery) falls within the criteria for that stratum.

Stratum 1 “Cold”:All neonates with temperatures less than or equal to 34.5 degrees Celsius.
Stratum 2 “Very cool” All neonates with temperatures greater than 34.5 degrees Celsius and less than or equal to 35.5 degrees Celsius;
Stratum 3 “Cool”:All neonates with temperatures greater than 35.5 degrees Celsius and less than or equal to 36.5 degrees Celsius;
Stratum 4 “Euthermic”:All neonates with temperatures greater than 36.5 degrees Celsius and less than or equal to 37.5 degrees Celsius;

Stratum 5: “Overly warm”: All neonates with temperatures greater than 37.5 degrees Celsius.

Denominator: Live-born neonates with birthweight of less than 2500 grams (as identified from either the medical record or by ICD-9- CM Principal or Other Diagnosis Codes in Table 1) and who were admitted to a level 2 or higher nursery within 24 hours of birth. Exclusions are noted below.

Table 1. Included Populations:ICD-9-CM Principal or Other Diagnosis Code Low Birthweight Diagnosis Codes

76400	LIGHT-FOR-DATES WTNOS	76490	FET GROWTH RETARD WTNOS
76401	LIGHT-FOR-DATES <500G	76491	FET GROWTH RETARD <500G
76402	LT-FOR-DATES 500-749G	76492	FET GROWTH RET 500-749G
76403	LT-FOR-DATES 750-999G	76493	FET GROWTH RET 750-999G
76404	LT-FOR-DATES 1000-1249G	76494	FET GRWTH RET 1000-1249G
76405	LT-FOR-DATES 1250-1499G	76495	FET GRWTH RET 1250-1499G
76406	LT-FOR-DATES 1500-1749G	76496	FET GRWTH RET 1500-1749G
76407	LT-FOR-DATES 1750-1999G	76497	FET GRWTH RET 1750-1999G
76408	LT-FOR-DATES 2000-2499G	76498	FET GRWTH RET 2000-2499G
76410	LT-FOR-DATE W/MAL WTNOS	76500	EXTREME IMMATUR WTNOS
76411	LT-FOR-DATE W/MAL <500G	76501	EXTREME IMMATUR <500G
76412	LT-DATE W/MAL 500-749G	76502	EXTREME IMMATUR 500-749G
76413	LT-DATE W/MAL 750-999G	76503	EXTREME IMMATUR 750-999G
76414	LT-DATE W/MAL 1000-1249G	76504	EXTREME IMMAT 1000-1249G
76415	LT-DATE W/MAL 1250-1499G	76505	EXTREME IMMAT 1250-1499G
76416	LT-DATE W/MAL 1500-1749G	76506	EXTREME IMMAT 1500-1749G
76417	LT-DATE W/MAL 1750-1999G	76507	EXTREME IMMAT 1750-1999G
76418	LT-DATE W/MAL 2000-2499G	76508	EXTREME IMMAT 2000-2499G
76420	FETAL MALNUTRITION WTNOS	76510	PRETERM INFANT NEC WTNOS
76421	FETAL MALNUTRITION <500G	76511	PRETERM NEC <500G
76422	FETAL MALNUTR 500-749G	76512	PRETERM NEC 500-749G
76423	FETAL MAL 750-999G	76513	PRETERM NEC 750-999G
76424	FETAL MAL 1000-1249G	76514	PRETERM NEC 1000-1249G
76425	FETAL MAL 1250-1499G	76515	PRETERM NEC 1250-1499G
76426	FETAL MAL 1500-1749G	76516	PRETERM NEC 1500-1749G
76427	FETAL MALNUTR 1750-1999G	76517	PRETERM NEC 1750-1999G
76428	FETAL MALNUTR 2000-2499G	76518	PRETERM NEC 2000-2499G

EXCLUSIONS

- Neonates who do not survive until the time limit of the measure (15 minutes after arrival to the NICU)
- Neonates with Anencephaly ICD-9-CM 740 Neonates not born in hospital/medical care setting
- 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 had been identified as meeting written institutional criteria for the initiation of therapeutic hypothermia and such therapy was begun or planned {OPTIONAL EXCLUSION}
- Neonates with Comfort care (requires all of the features below): 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)

C.DATA SOURCES

A. Medical record (paper or electronic), may be utilized to identify:
- Date and time of birth
- Date and time of arrival to a Level 2 or higher nursery;
- Date and time of first temperature upon arrival to that nursery;
- Temperature and units of measurement
- Race/ethnicity (preferred data source)
- Home zip code Mother's State and County of Residence and or zip code (preferred data source)
- Born in medical facility or transferred in (preferred data source)
- 5 minute Apgar score
- Birthweight (preferred data source)
- Documentation if child met local criteria for hypothermia and time so identified
- Documentation if hypothermia was planned or initiated before temperature taken
- Insurance type (optional data source)

B. Administrative data with billing and diagnosis codes, utilized to identify:
i. ICD-9 codes to identify low birthweight infants and presence of anencephaly
ii. Revenue codes indicating care in Level 2, 3, or 4 nursery (172, 173, 174)
iii. OPTIONAL source for:
   i. Date of birth
   ii. race/ethnicity
   iii. home zip code
   iv. Whether child was inborn or transferred in
   v. Birthweight range
   vi. Insurance type and benefit plan {Preferred data source}

D. "CALCULATION" and Reporting

Step 1: Identify all live-born neonates with a birthweight less than 2500 grams, using the aforementioned codes or recorded birthweights when practical.

Step 2: Identify all of those neonates from Step 1 who were admitted to Level

2 or higher nursery).

Step 3: Record relevant attributes:

a. Record ICD-9 comorbid diagnoses.
   EXCLUDE those with anencephaly (ICD-9-CM 740xx).
b. Record:
   i. Date and time of birth.
   ii. Birthweight.
   iii. 5 minute Apgar score
   iv. Date and time of arrival to Level 2 or higher nursery.
   v. If child was admitted to a Level 2 or higher nursery from regular newborn care
   vi. If child was inborn or transferred to Level 2 or higher nursery from another facility.
c. If transferred is there documentation that neonate was not born in a medical facility
   EXCLUDE if: child not born in a medical facility

Step 4: Record the following additional data elements for all eligible neonates:

- Race
- Ethnicity
- Insurance type (Medicaid, Commercial, Uninsured)
- Benefit category (HMO, PPO, Medicaid Primary Care Management Plan, Fee for service, Other)
- Zip Code and/ or State and County or equivalent area of Mother's residence. Record FIPS if available
- Evidence child received comfort care only (when appropriate). **EXCLUDE** if so.
- Documentation child was eligible for and received therapeutic hypothermia (when appropriate). **EXCLUDE** if so.

Step 5: Identify and record:

- Time of first temperature taken in the nursery (ARRIVAL TEMPERATURE)
- Value of first temperature taken in the nursery
- Units that temperature was recorded in. If in Fahrenheit calculate Celsius as C=(F-32)*5/9
- Record infant age at time of ARRIVAL temperature

Step 6: If infant's age is > 75 minutes at the time of the initial temperature record the following as the ALTERNATE temperature:

- Time of last temperature taken in the unit where the infant was delivered
- Value of that temperature
- Units that temperature was recorded in. If in Fahrenheit calculate Celsius as C=(F-32)*5/9
- Record infant age at time of that temperature

If infant's age at time of ARRIVAL TEMPERATURE is > 75 minutes AND infant was admitted directly to the Level 2 or higher nursery without transport from another institution OR transfer from the normal newborn nursery, report the lower of the ARRIVAL and the ALTERNATE temperature.

Step 7: Identify which numerator stratum to which the reported temperature should be assigned:

The numerator has 5 comprehensive and mutually exclusive strata. Each is determined by the number of children whose reported temperature falls within the criteria for that stratum.

Stratum 1 "Cold": All neonates with temperatures less than or equal to 34.5 degrees Celsius.

Stratum 2 "Very cool": All neonates with temperatures greater than 34.5 degrees Celsius and less than or equal to 35.5 degrees Celsius; than 37.5 degrees Celsius.

Stratum 3 "Cool": All neonates with temperatures greater than 35.5 degrees Celsius and less than or equal to 36.5 degrees Celsius;

Stratum 4 "Euthermic": All neonates with temperatures greater than 36.5 degrees Celsius and less than or equal to 37.5 degrees Celsius;

Stratum 5: "Overly warm": All neonates with temperatures greater

Step 8. Calculate the percent of neonates who are in each stratum:
= [100*number of children in each stratum] / [total number of infants eligible for the measure]. Percents should be reported to 2 decimal places.

Minimum sample size for reporting overall nursery rates is N=20; Reporting should be further stratified by the application of stratification variables as described below to both the numerator and the denominator. Reporting strata with denominator samples less than N=15 should not be reported.

Step 9. (**Optional**) Calculate 90% confidence intervals around each reported percentage as 1.68 * [square root of the [proportion in the confidence interval * (one minus that percentage) / the eligible N].

Step 10. Using eligible births and qualified temperatures, repeat steps 8 and 9 and report for each stratification category listed below, using the following data elements:

- Birthweight (3 birthweight categories: <999 grams; 1000-1499 grams; 1500-2499 grams)
- Perform stratifications as indicated herein (report for each stratum where denominator >=15):
  - Race and ethnicity (Using White non Hispanic, Black non Hispanic, Hispanic, Asian/Pacific Islander, other)
  - Insurance type (Public/Medicaid, Private/Commercial, None/Other)
  - Admission source (use 3 categories: inborn, transported, transferred from newborn nursery)

Location of delivery
APPLY THESE RULES IN ORDER. STOP WHEN CATEGORIZED

i. Categorize location of delivery as birthing room if:
    1. Location was identified as delivery room on the labor and delivery suite but was not an operating room OR

    2. Location was identified as a birthing room or equivalent OR
    3. Infant was a vaginal delivery other than a multiple gestation AND Operating Room or equivalent (C-section room would be an example of an equivalent to an operating room) is not specified as location.

ii. Otherwise categorize location as OPERATING ROOM if:
    1. Location was identified as an operating room or equivalent, OR
    2. If neonate was delivered by c-section, OR
    3. If infant was a multiple gestation (and location is unspecified) OR
    4. If location is identified as Emergency Department OR other

d. 5 minute Apgar score (Apgar of 5 or less versus 6 or more)

e. Benefit Category (Benefit category (HMO, PPO, Medicaid Primary Care Management Plan, Fee for service, Other)

f. Urban Influence Code(1) or UIC. (2013 urban influence codes available at: http://www.ers.usda.gov/data-products/urban-influence- codes.aspx#.UZUvG2cVoj8 .

   Use mother's place of residence to determine UIC. State and County names can be linked or looked up directly or zip codes can be linked to County indirectly, using the Missouri Census Data Center (http://mcdc.missouri.edu/).

g. Level of Poverty in the County of Residence. The percent of all residents in poverty by county are available from the US Department of Agriculture at http://www.ers.usda.gov/data-products/county-level- data-sets/download-data.aspx Our stratification standards are based on 2011 US population data that we have analyzed with SAS 9.3.

   Using Mother's state and county of residence (or equivalent) or FIPS code, use the variable PCTPOVALL_2011 to categorize into one of 5 Strata:

   i. Lowest Quartile of Poverty if percent in poverty is <=12.5%
   ii. Second Quartile of Poverty if percent in poverty is >12.5% and <=16.5%
   iii. Third Quartile of poverty if percent in poverty is >16.5% and <=20.7%
   iv. First upper quartile (75^{th}-90^{th}) if percent in poverty is >20.7% and <=25.7%

   v. Second upper quartile (>90^{th} percentile) if percent in poverty exceeds 25.7%

iii. Repeat stratifications a-g within birthweight categories (report for all strata for which denominator >=15)

Section III. 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).

III.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 with the active consultation of the Centers for Medicare & Medicaid Services. 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 birthweight 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 (2-8). Modern studies have confirmed and extended these findings, including potential methods to maintain temperature for infants in the Delivery Room (9-11). Laptook et al confirmed the association of temperature loss with poor outcomes in 5277 infants, 401-1499 grams, born at any of 15 academic medical centers participating in the National Institute of Child Health and Development (NICHD) Neonatal Research Network(12). 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(13).

We have collected data from chart review at three diverse hospitals in New York City. All three hospitals had a range of birthweight 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 Figure 1 and 2 on the next page.

Figure 1

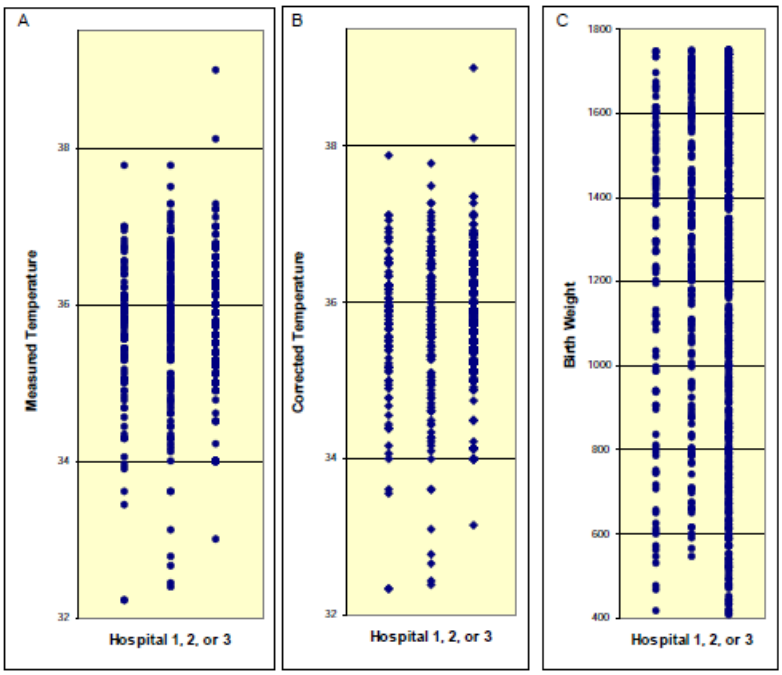

Figure 1 (A-C). Scatter plots of measured temperature (A), temperature adjusted for method of measurement (B), and birth weight in grams (C) for each of the three study hospitals. N=100, 158, and 487 for Hospitals 1, 2, and 3 respectively.

Figure 2

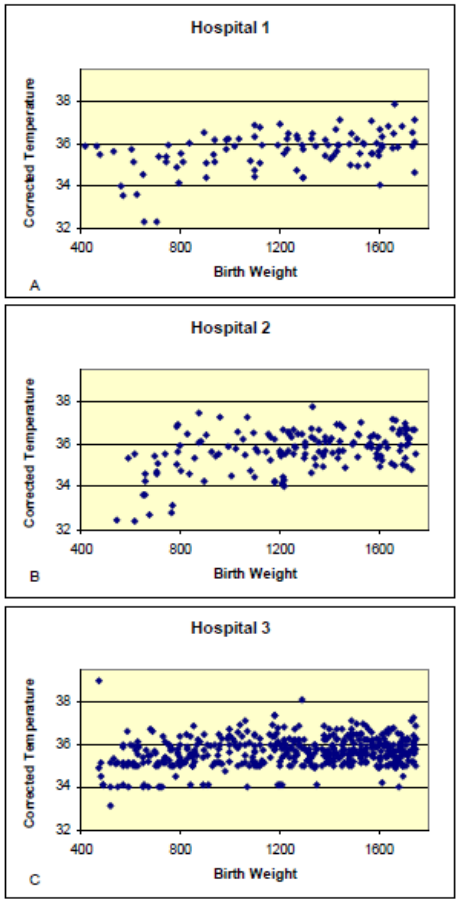

Figure 2 (A-C). Scatter plot of corrected temperature (see text) by birthweight in grams for each of the three study hospitals. Valid measurements on both birthweight and temperature are available for 99 infants in Hospital 1, 147 infants in Hospital 2, and 459 infants in Hospital 3.

Temperature predicted in-hospital mortality after controlling for covariates, whether we dichotomized at the 35.5 degree 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 degrees reduced odds of death more than 35% in the model using a continuous variable (22% for 1° Fahrenheit).** Defining hypothermia as admission temperature below 36.0 would estimate an increase in the odds of mortality of 84%, 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 (13). Regression risk analysis estimates the adjusted risk ratio (ARR) and adjusted risk difference: hypothermia (35.5C) results in an ARR of 1.48 (95% confidence interval 1.03—2.30), indicating a 48% increase in risk, from a baseline risk of 8.9% among those who were euthermic to an exposed risk of 13.1% 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%, from 35.0 to 36.0 by 26%, and from 36.0 to 37.0 by 27%, resulting in absolute risk reductions of 2.8%, 2.4%, and 2.0% respectively. A core body temperature increase from 34.0 to 37.0 is associated with a relative decrease in mortality of 98% and an absolute decrease in mortality of 7.2%, controlling for other factors in the model. The decrease from 36.0 to 35.5 is associated with a 12% increase in the adjusted mortality risk from 9.4% to 10.5%.

Our work confirmed findings in the literature that insurance status and race (14) are associated with outcomes.

Anecdotal reports from among our participating hospitals confirm reports in the literature (15) that attention to thermal management can improve temperature outcomes.

As an appendix, we present 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 despite their submission by the Vermont Oxford Network. We incorporate a highly engaged process to develop an enhanced set of measures.

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 upon the temperature upon admission to the NICU as the initial inpatient perinatal topic in the national Pediatric Quality Measures Program (PQMP). This program is funded by the Child

Health Insurance Reauthorization Act (CHIPRA, 2009). and administered by the Agency for Health Care Research and Quality (AHRQ) in collaboration with the Centers for Medicare & Medicaid Management (CMS). CAPQuaM is one of seven AHRQ-CMS CHIPRA Centers of Excellence across the country. Measures developed under the Pediatric Quality Measures Program will be subject to review by an independent advisory committee.

III.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 birthweight babies are insured by Medicaid. Hypothermia is not only associated with neonatal mortality, but there is evidence (14) 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.

III.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).

VON proposed a measure regarding the adequacy of taking temperatures in low birthweight infants, temperatures taken within an hour of admission to the NICU. This was rejected largely because it was met 98% of the time. While we would hold with VON that 98% 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 on neonatal outcomes (16, 17). We propose a measure that looks at the proportion of low birthweight 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 birthweight 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 would be a larger infant. Hence, this measure is inclusive of all low birthweight 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 assessed 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 36.0 degrees Celsius. It was rejected in part because there is no consensus regarding the desirable threshold. Based on the literature the literature and our own data described above, we believe that temperature provides increasing risk the further it falls below 37 degrees Celsius. Our two temperature measures in this set provide discrete and continuous ways of looking at the distribution of temperatures, stratified by birthweight and reported for various subgroups when sample size is sufficient. Our data demonstrate that optimal thermal management is capable of keeping 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.

Although our data support assertions in the literature(18, 19) that 37 degrees is the euthermic threshold for management of premature infants, we recognize that there are limited data regarding temperature above 36.5 and use this as the pragmatic lower limit for euthermic infants, incorporating explicit criteria developed by our expert panel. There is evidence that an axillary temperature of 36.5 can be well tolerated by full term newborns (19).

Section IV. 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 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	Day 1
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-age 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)

Section V. 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.

V.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 systematically a targeted review of the literature, which is attached as an Appendix. Further we have interviewed clinicians, engaged clinical societies and accreditors, patient/family groups, NY 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 Appendices. A brief summary of the research findings includes that the temperature of low birthweight infants varies based on their management, that every degree below 37 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 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 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 7553 neonates (from 61 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, whether or not we stratified by birthweight category. Deaths were disproportionate among those who were cool, in a graded fashion.

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. 25% of these nurseries had a mean temperature below 36.1. We conclude from this that temperatures do vary across nurseries, further reinforcing our sense that this is an important measure of performance.

See Validity section for further details.

We note above that there is evidence that management can enhance thermal outcomes.

V.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.

The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children (20). 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 (21). The validity of our work has benefited from our use of a formal method, a pragmatic adaptation of the CAPQuaM 360 degree 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, comprised of 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 community. 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 360 degree 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 2 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.

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

Please see discussion and literature summaries presented elsewhere, as well as information in Section VI.

Section VI. 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.

VI.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.

This 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 is more accurate for identifying services with a high likelihood of documentation due to reimbursement (22).
The constructs underlying our measures are:

- Date and time
- 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 age, the optimal approach to measuring the temperature may vary. Measurement approaches that are understood to be valid (articles and specifics of 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 0.5 degrees Celsius warmer than those for whom the site of temperature was not documented to be rectal. Other studies that are in the literature do not find 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 birthweight 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.

VI.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., R^2 for concurrent validity).

The use of electronically available administrative data in healthcare 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 (23). Administrative data are supported, if not encouraged by federal agencies, including NIH, AHRQ, HCFA, and the VA. This measure calls for the use administrative data to identify the universe of low birthweight infants.

The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children (20). 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 (21). The validity of our work has benefited from our use of a formal method, a pragmatic adaptation of the CAPQuaM 360 degree method. The 360 degree 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 2 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.

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 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 0.5 degree “discount” in that our data suggest that optimal outcomes are obtained at 37.0 degrees Celsius, rather than at the 36.5 in the measure (which is still far preferable to cooler). As we noted above, we have data that suggest that this 0.5 degree 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 supports 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 birthweight of 400-2499 grams whose admission temperature was 29 degrees 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=7553. The number of infants ranged from 21 to 370 per hospital and 86.7% 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% of infants were <=34.5 (cold), 9.6% above 34.5 but <= 35.5 (very cool), 48.0% above 35.5 but <= 36.5 (cool), 37.9% above 36.5 but <= 37.5 (Euthermic or Appropriately Warm), and 2.6% above 37.5 or Overly Warm.

There were only 67 newborns that were transferred from another facility. The distributions of temperatures were similar to the inborn infants, with the exception 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 below.

Table 2

Quantile	Estimate
99	37.9
95	37.3
90	37.1
75	36.8
50	36.4
25	36.0
10	35.4
5	35.0
1	34.1

The median was 36.4, mean was 36.3, and 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. 1 infant was euthermic, 1 cool, and 2 were very cool. Nearly 1 percent were transferred from the Newborn

Nursery, of which 48% were euthermic, 44% cool, and only 6% 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 (5254) and the remainder were classified as being born in a labor and delivery room/ birthing room (2245). Of those born in the operating room, 2% were cold, 11% very cool, 72% cool, and 35% euthermic. Those born in the L&D suite were warmer with 2% cold, 7% very cool, 13% cool, 48% euthermic, and 45% too warm ($p<.0001$). This suggests that our categorization of babies born in the OR (while imperfect) does identify a distinct population. Our expert panel recommended that we report by site of delivery.

We found that temperatures varied by birthweight 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% for those under 1000 g (two thirds of all cold babies from a group that was about 12% of all babies). These infants also were least likely to be euthermic, only 25% were so classified, compared to 34% of those in the intermediate weight category and 41% of the larger babies.

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

Section VII. Identification of Disparities

CHIPRA requires that quality measures be able to identify disparities by race, ethnicity, socioeconomic status, and special health care needs. Thus, westrongly 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.

VII.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 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.

VII.B. Special health care needs

Not Assessed

VII.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 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 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 (24). 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 USDA data considering 3142 counties and related geographic units found a mean of 17.2 % of county residents living in poverty, a standard deviation of 6.5%, and an interquartile range of 8.2%. 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 on the next page.

Table 3

Quantile	Percent in Poverty
Maximum	49.9%
99	37.5%
95	28.9%
90	25.7%
75	20.7%
50	16.5%
25	12.5%
10	10.0%
5	8.6%
1	6.1%
Minimum	2.9%

VII.D. Rurality/Urbanicity

As described in the specification 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 3143 county equivalents in the U.S and the results are found in Table 4. See TABLE 4 below.

Table 4

	UIC_2013	
UIC_2013	Frequency	Percent
1	432	13.74
2	735	23.39
3	130	4.14
4	149	4.74
5	242	7.70
6	344	10.94
7	162	5.15
8	269	8.56
9	184	5.85
10	189	6.01
11	125	3.98
12	182	5.79

The population is heavily weighted to metropolitan areas as seen in Table 5. See TABLE 5 below.

Table 5

UIC_2013				
UIC_2013	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1.672E8	55.07	1.672E8	55.07
2	91886000	30.27	2.5909E8	85.34
3	6921700	2.28	2.6601E8	87.62
4	3094100	1.02	2.691E8	88.64
5	10760300	3.54	2.7986E8	92.18
6	7005400	2.31	2.8687E8	94.49
7	1511900	0.50	2.8838E8	94.99
8	8459500	2.79	2.9684E8	97.78
9	2684400	0.88	2.9952E8	98.66
10	1289100	0.42	3.0081E8	99.09
11	1887800	0.62	3.027E8	99.71
12	887700	0.29	3.0359E8	100.00

The set of data shows that 55% of the US population lives in an urban area of greater than 1 million residents (UIC_2013 #1) while 1.33% 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, 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 4877 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 6 per square mile. Further, the literature (26) 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 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.

VII.E. Limited English Proficiency (LEP) Populations

Not assessed

Section VIII. 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.

VIII.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 (birthweight, 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 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 it was 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. birthweight, 5 minute Apgar score) were said to be in all of the EMRs. EMR data was 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% 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 birthweight (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%) classified as Level 2 nurseries submit temperature data, 31 of 36 (86.1%) with a Level 3 designation submit temperature data, and 16 of 18 (88.6%) of Regional Perinatal Centers submit temperature data. These data are virtually complete for those institutions that submit data. These data capture 84.1% of low birthweight 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 electronic medical record and transferred to an analytical database for quality measurement and reporting. A large proportion of these data elements are already captured routinely.

VIII.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 IX. 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 can typically 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. Tested at NY state level.

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 can typically 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?	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 can typically 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 with this population in mind and tested in this population

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

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 can typically 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?	Designed and tested at this level.Some strata may have small sample sizes

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?	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 can typically 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 anticpated at IDS. Not recommended for other options.

Section X. 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 testingwith stakeholders).

This measure was developed and named with lay understanding in mind. Indeed, the CAPQuaM measure process aims to optimize and balance the sometimes competing considerations of validity and meaning, feasibility, usefulness, and understanding. In the current measure set there are three threshold measures that are intended to present useful information in a manner that particularly speaks to ease of interpretation, and this is one of the three.

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 websites)
- 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; and,
- CAPQuaM's scientific team.

Clinical criteria regarding, including consideration of inclusion and exclusion criteria, 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 with a variety of areas 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 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 senior advisory board advised us to use lay terms in naming the measures to the extent that we are able to. This measure uses lay terminology to characterize infants who are “Cold,” “Very cool,” “cool,”, “Euthermic” (about the right temperature), and “Overly warm”.

Section XI. 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.

XI.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. 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 was 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 that 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 and 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.

XI.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?
Not at present. See above. Our feasibility assessment included a survey of 13 hospitals, all with varying degrees of EMR implementation. The potential for this to become an e-measure is clear from the discussion of feasibility above. (See Section 8). We will work with our partners to develop an e-measure as appropriate.

XI.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.

XI.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 http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__standards_ifr/1195)? no

If yes, please describe.

XI.E. Health IT Calculation

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

XI.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 XII. 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 precision and small sample sizes when stratifications are performed. With a sample size of 20, the precision for a 90% confidence interval is +/- 11% for something with a prevalence of 10% and +/- 17% for something with a prevalence of 30%. These are wide but not unreasonable confidence intervals. Those widths decrease to +/-7% and 11% respectively if the sample size increases to 50. Unfortunately, the binomial distribution, which is the most efficient appropriate way to estimate precision, offers only gradual increases with increasing sample size. We recommend that agencies consider asking reporting entities to pool data over multiple years when sample sizes are too small for the desired strata to be analyzed well.

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 XIII. 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 ofthe 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 percent of newborn low birth weight (<2500grams) neonates who fall into each of 5 categories based upon their temperature upon admission to a Level 2 or higher nursery in the first 24 hours of life. Strata are cold (<=34.5), very cool (34.51-35.50), cool (35.51-36.50), euthermic (36.51-37.50) and overly warm (> 37.5). More than 100 years of literature support the ongoing salience of appropriate thermal management of low birthweight infants and,

unfortunately, variable clinical performance persists.

This measure topic was assigned to the CAPQuaM as a PQMP priority by the Agency for Healthcare Quality and Research with the active consultation of the Centers for Medicare & Medicaid Services. In addition to literature 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 birthweight babies are insured by Medicaid. Hypothermia is not only associated with neonatal mortality, but there is evidence (14) 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, Medicaid population is disproportionately black and in our testing data, black infants were disproportionately hypothermic.

In our study of 7553 neonates admitted to Level 2 or higher nurseries in New York State we found that 1.9% of infants were <=34.5 (cold), 9.6% above 34.5 but <= 35.5 (very cool), 48.0% above 35.5 but <= 36.5 (cool), 37.9% above 36.5 but <= 37.5 (Euthermic or Appropriately Warm), and 2.6% above 37.5. Key findings from these analyses were: 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% of Medicaid infants were euthermic, compared to 40% 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 address adequately 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 birthweight 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.

Section XIV: Identifying Information for the Measure Submitter

Complete information about the person submitting the material, including the following:

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:	New York
Postal Code:	10029
Telephone:	2126599556
E-mail:	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.

The signed written statement was NOT submitted