High-Risk Deliveries at Facilities with 24/7 In-House Physician Capable of Safely Managing Labor and Delivery and Performing a Cesarean Section, Including an Emergent Cesarean Section

Section 1. Basic Measure Information

1.A. Measure Name

HROB1: High-Risk Deliveries at Facilities with 24/7 In-House Physician Capable of Safely Managing Labor and Delivery and Performing a Cesarean Section, Including an Emergent Cesarean Section

1.B. Measure Number

0120

1.C. Measure Description

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

Percent of high-risk deliveries that are delivered at a facility with 24/7 in-house physician coverage that is dedicated to obstetrics and includes a physician capable of safely managing labor and delivery and performing a cesarean section, including an emergent cesarean section.

1.D. Measure Owner

Collaboration for Advancing Pediatric Quality Measures (CAPQuaM)

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

Not applicable.

1.F. Measure Hierarchy

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

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

This measure belongs to the Pediatric Quality Measures Program (PQMP) Availability of High-Risk Obstetric Services 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.**

   High-risk obstetrical deliveries at facilities with appropriate high-risk facilities.

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.**

   Structural 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

Number of eligible deliveries that occur in facilities with 24/7 in-house physician coverage that is dedicated to obstetrics, and includes a physician capable of safely managing labor and delivery, and performing a cesarean section, including an emergent cesarean section.

Numerator Elements:

- Number of deliveries.
- Maternal and infant ICD-9-CM codes.
- Response to survey question identified on technical specifications.

1.H. Numerator Exclusions

None.

1.I. Denominator Statement

Overall number of eligible deliveries. Eligible deliveries are identified in two distinct ways. Maternal and infant ICD-9-CM codes are specified in Section 2 Detailed Measure Specifications (see Supporting Documents).

1. Class A: Maternal Diagnoses and Comorbidities
2. Class B: Delivery Complications, Fetal Injury or Compromise, or Suboptimal Infant Diagnoses

a. Maternal Delivery Complication Codes (ICD-9-CM).

b. Maternal Stillbirth or Birth Hypoxia/Asphyxia Codes.
c. Premature or small infant. (Infant codes).

3. Either Class A or Class B

Denominator Elements:

- Number of deliveries.
- Maternal and infant ICD-9-CM codes.
- Maternal DRG, CPT codes, and revenue codes when available.

1.J. Denominator Exclusions

None.

1.K. Data Sources

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

Administrative Data (e.g., claims data); Survey – Health care professional report.

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

Health care professional can be representing a health care facility that delivers babies.

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 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).**

The Collaboration for Advancing Pediatric Quality Measures (CAPQuaM) was assigned the topic of availability of high risk obstetrical services as a PQMP priority by the Agency for Healthcare Research and Quality (AHRQ) and the Centers for Medicare & Medicaid Services (CMS). We developed a set of high-risk obstetric availability measures in close consultation with our expert panel.

Optimal health of children in the United States is fostered by healthy pregnancies, healthy deliveries, and outcomes of pregnancy that include healthy mothers and babies. Appropriate availability of specific aspects of care for pregnant women, in particular those in need of high-risk obstetric services, is necessary to achieve desired outcomes. For example, subspecialty care is necessary for specific high-risk women, and this is the focus of another CAPQuaM measure: Availability of Outpatient Maternal Fetal Medicine Specialty Care for Women with High-Risk Pregnancies.

The current set of measures addresses four critical structures necessary for optimal outcomes among women with high-risk conditions. While there is much interest in obstetrics in classifying levels of obstetric care, we prioritized four specific attributes that others might use to define such levels (Hankins, Clark, Pacheco, et al., 2012).

The proposed availability measures address important gaps in quality and safety and also have the potential to narrow disparities in maternal and neonatal outcomes. These four structural attributes (24-hour in-house physicians covering obstetrics and capable of managing labor and delivery, including performing emergent cesarean sections; 24-hour in house physicians available and capable of providing obstetric anesthesia; 24-hour availability of blood bank/transfusion services; and delivery at a facility with a Level 3 or higher neonatal intensive care unit [NICU]) have the potential to improve both maternal and infant outcomes in the setting of high-risk deliveries. They were chosen to represent a prioritized selection of key structural attributes that impact the timeliness with which a potentially urgent service may be available to women who are delivering in the context of a pregnancy that manifests higher than typical risk. The prioritization process involved our team of stakeholders as well as an expert panel; the clinical and health services judgment of these team members guided the process.

Delivery care provided to pregnant women is critical for the health and well-being of mothers and babies. The burden of chronic illness and risk factors for pregnancy complications (e.g., hypertension, diabetes, advancing maternal age, previous cesarean section) are all rising among women, increasing their risk for morbidity and mortality (Hankins, et al., 2012). Over the past decade, maternal mortality has increased in the United States, and striking racial disparities persist (Berg, Callaghan, Syverson, et al., 2010; Callaghan, Creanga, Kuklina, 2012). Black women are 3 to 4 times as likely as white women to suffer a pregnancy-related death (Berg, et al., 2010). Racial and ethnic disparities are also reflected when considering both the processes and outcomes of neonates (Howell, Hebert, Chatterjee, et al., 2008; Howell, Holzman, Kleinman, et al., 2010; Howell, Stone, Kleinman, et al., 2010).

For every maternal death, 100 or more women suffer severe maternal morbidity, a potentially life-threatening diagnosis, or a life-saving procedure that is associated with pregnancy. Examples include organ failure (e.g., acute renal failure, liver, respiratory), obstetric shock, pulmonary embolism, amniotic embolism, eclampsia, septicemia, cardiac events, mechanical ventilation, transfusion, invasive hemodynamic monitoring, and hysterectomy. Severe maternal morbidity is rising and affects approximately 52,000 women annually in the United States (Callaghan, et al., 2012). Studies using the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS) show the prevalence of at least one severe complication of pregnancy rose 75 percent from 1998-1999 to 2008-2009: renal failure increased by 97 percent, thrombotic embolism by 100 percent, adult respiratory distress syndrome by 75 percent, blood transfusion by 183 percent, and ventilation by 34 percent. Similar to maternal and neonatal mortality, minority women are more likely to experience severe maternal morbidity than white women (Callaghan, et al., 2012).

Severe morbidity is more common at the extremes of reproductive age and for black women as compared with white women. Improving quality and safety of care is an important lever to address these issues, as research suggests that at least one-third to one-half of maternal deaths in the United States may be preventable through improvements in quality of care (Berg, Atrash, Koonin, 1996; Hoyert, Danel, Tully, 2000; Nannini, Weiss, Goldstein, et al., 2002). Additional studies suggest that on the continuum of care to adverse pregnancy outcomes, there are a number of points that can be impacted by improved safety and quality (Joint Commission, 2010). All four structural measures in the HROB measure set are critical to ensure the safety of mothers and

babies in the setting of deliveries that are of higher risk, whether due to maternal comorbidities or complications of pregnancy (American Academy of Pediatrics, 2012; California Maternal Quality Care Collaborative, 2015). We refer to these collectively as high-risk deliveries.

To improve care for women who require high-risk obstetrical services, it is imperative that quality measures address the availability of high-risk obstetrical services by assessing how available key services are at hospitals providing obstetric care. Agencies such as the March of Dimes, American Academy of Pediatrics, American College of Obstetricians and Gynecologists, American Academy of Family Physicians, and American Medical Association have emphasized the need for stratification of facilities based on levels of maternal care, so that the definition of levels of care should be based on the capability to provide more complex care. For example, Table 1 (see Supporting Documents) displays the Indiana Perinatal Network's criteria on when to seek a consultation and when to refer or transport a pregnant woman. Similarly, quality measures can play a critical role in identifying gaps in care delivery and subsequently act to decrease severe maternal morbidity and mortality (Hankins, et al., 2012).

The CAPQuaM measure development process sought to ground availability measures in a definitional framework of what constitutes a high-risk obstetrical service. First, we approached the literature to establish a construct of conditions that potentially can be considered as high-risk, increasing the risk of maternal and/or infant morbidity and mortality. We subsequently convened a multidisciplinary panel of national experts to provide leadership, including helping to establish definitions for both availability and for high-risk obstetrical services. The Panel held a telephone meeting, conducted pre-work via email, and participated in a 2-day face-to-face meeting. By the conclusion of the meeting, the Panel had rated a variety of constructs using this adaptation of the two round RAND/UCLA modified Delphi process.

This is the first measure set that we are aware of that addresses specifically the availability of high-risk obstetrical care. It reflects our perspective that the optimal health of children in the United States is fostered by healthy pregnancies and deliveries. The availability of HROB services is critical for the health of pregnant women who face high-risk deliveries and ultimately for the health of the child they are carrying. An emerging consensus in the literature relates the construct of levels of care for women and newborn services. The American Academy of Pediatrics (AAP) defines special and intensive care newborn services as Levels 2-4 in a specific manner (AAP, 2012), and the field of obstetrics is rapidly moving in that direction. These measures both build on the AAP definition and operationalize components that make up the levels of high-risk obstetrical services. The measures capture the extent to which women in need of HROB services and those who may be at risk for or experiencing a complicated delivery are delivered at hospitals that provide sufficient care.

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).**

Consortium partners at the New York State Department of Health, including the Office of Health Insurance Programs/New York State Medicaid, steering committee, and scientific team have played central roles in the development of these measures. Evidence for a high level of interest in this work in particular was demonstrated by the fact that the CAPQuaM team was asked to present this work in development to the CMS Expert Panel on Improving Maternal and Infant Health Outcomes in Medicaid/CHIP Data, Measurement, and Reporting Workgroup.

More generally, childbirth is the largest category for hospital admissions for commercial payers and Medicaid programs, and the estimated hospital costs associated with childbirth and newborn care are over $80 billion in the United States annually (Andrews, 2008; DeFrances, Cullen, Kozak, 2007). In New York State, 48.6 percent of deliveries in 2011 occurred in women insured by Medicaid (New York State Department of Health, 2013). In our analysis year, 55.6 percent (4,197 neonates) of low birth weight neonates admitted to NICUs across New York State and who were in our study of newborn temperatures (approximately 90 percent of all newborns admitted to level 2 or 3 nurseries) were insured by Medicaid.

Providing high quality care to women with high-risk deliveries has the potential to both improve outcomes and narrow disparities, important national priorities for CMS. In fact, leaders in obstetrics have proposed systematic changes in the delivery of obstetric care to address these issues. Both the peer-reviewed and grey literature propose improved integrated maternal-fetal/neonatal care networks that optimize regionalization of care to improve access to critical 24/7 in-house obstetric services, blood bank/transfusion services, obstetrical anesthesia, and level 3 or 4 NICU services for women with high-risk pregnancies (American Academy of Pediatrics, 2012; California Maternal Quality Care Collaborative, 2015; Hankins, et al., 2012). Therefore, our proposed measures have the potential to have a significant impact on the health of mothers and infants insured by Medicaid. High-risk deliveries disproportionately impact women insured by Medicaid as compared with private insurance. Risk factors that are associated with high-risk deliveries (e.g., hypertension, delivery of low birth weight infants) are all factors that are more prevalent among the Medicaid population. Given the fact that childbirth is the leading category for hospital admissions for Medicaid programs and the fact that high-risk deliveries disproportionately occur among women insured by Medicaid, quality measures targeting high-risk deliveries have the potential to improve quality of care for a sizeable portion of those covered by the Medicaid program.

One key decision made by our expert panel with particular relevance for the vulnerable Medicaid population was establishing that high-risk obstetrical services extend from preconception (e.g., managing the cessation of teratogenic medications) through delivery and the early postpartum period. Risk (and the need for HROB services) could be established at any point in that spectrum, including both premature delivery and an obstetrical emergency, such as a postpartum hemorrhage. The Expert Panel offered definitions regarding which conditions established that a pregnancy required high-risk obstetrical services. They further endorsed constructs important to the assessment of availability of HROB services. Among those constructs, the panel endorsed the

concept of regionalization of care. The panel specifically endorsed the importance of certain services being available 24/7 in the hospital of delivery, among those a qualified obstetrical physician, an obstetrical anesthesiologist, blood banking/transfusion services, and a Level 3 or higher NICU. A working draft of the Panel Summary after the second round of voting can be found in the Supporting Documents. Not specifically incorporated in this summary was the breadth of dialogue regarding what it means to assess availability in this context. The conclusion that guided much of the subsequent conversation was that the role of these availability measures should be to describe availability at a population level even though the unit of analysis that we were to measure directly was an individual pregnancy. There are two key implications – these measures are not intended to assess the quality of care for a given pregnancy. They also are intended to generate a gradient along which availability of HROB services can be assessed. So, while the measures have a concrete interpretation, over time the full nuance of their capacity to describe availability will be enhanced by the establishment of benchmarks in medically and geographically diverse populations and communities.

The co-leads of this measure development, a pediatrician and an obstetrician, collaboratively operationalized these constructs into the measures in the current measure set, working with the CAPQuaM stakeholders, including New York Medicaid, and consulting the expert panelists as appropriate. Using ICD-9-CM codes and a publicly available grouping system, AHRQ's Clinical Classification Software (http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp), the various conditions that could classify a pregnancy as in need of HROB services were specified into those seen in this measure. The four measures in this set incorporate these high-priority conditions and services and address the capacity to have immediately available high-risk services before, during, and after delivery. They describe the proportion of high-risk deliveries that take place in facilities that meet one or more of four structural criteria:

1. 24/7 in-house physician staffing the obstetrical unit who is capable of safely managing labor and delivery, and performing a cesarean section, including an emergent cesarean section.
2. 24/7 in-house obstetrical anesthesia services.
3. 24/7 in-house blood banking/transfusion services.
4. Level 3 or higher NICU services.

The New York State Office of Health Insurance Programs is an active CAPQuaM partner and has been engaged in the conceptualization and development of these measures. Our testing has occurred using Medicaid data and is described below and in Sections 3.C, 6.A, and 6.B.

We assessed measure performance in MAX data for 18 States that had been used for validation activities by another of the CHIPRA Centers of Excellence. The algorithm was modified to use all maternal codes and no infant codes because of the limitations of using anonymous MAX data and the consequent inability to match maternal with infant data. For the proxy OB outcome, self-report of Level 3 OB on the AHA Survey, we were able to match 85 percent of hospitals. We eliminated 3 States with less than 80 percent of deliveries occurring in matched hospitals. In the remaining 15 States, the range of HROB deliveries in hospitals meeting criteria was from 3.11 percent in Wyoming to 52.42 percent in Kansas, with a median of 33.04 percent and an interquartile range of 20.20 percent. The analysis suggests that even this modified approach to the measure using only maternal data is able to capture differences.

Examination of neighboring States (e.g., Arizona and New Mexico) showed similar performance, (35.76 percent and 35.47 percent, respectively), adding face validity to our analysis.

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).

We have developed four related measures based on self-report of whether there is 24-hour, 7-day a week availability of structural characteristics at the facility in which the woman gave birth:

1. Coverage of the obstetrical service by a physician dedicated to the OB service and capable of safely managing labor and delivery, performing a cesarean section, including an emergent cesarean section.
2. In-house coverage dedicated to the obstetrical service by an anesthesiologist who is qualified to provide obstetrical anesthesia.
3. On-site blood banking services/transfusion services that are always available for obstetrical patients, including: testing of blood group and Rh Type; cross-matching; antibody testing; transfusion with on-site and available blood, either ABO specific or 0-Rh-negative; transfusion with fresh frozen plasma; and transfusion with cryoprecipitate.
4. Presence of a Level 3 or higher NICU on campus.

The measures are defined as percent of high-risk deliveries that occur in facilities that meet each characteristic. The selection of these topics is valid and justified by evidence summarized briefly below. All were prioritized during our formal expert process.

An American Hospital Association Chart Book (AHA, 2015) describes OB services as important to be always available: delay can impair maternal and neonatal outcomes. The failure to respond urgently and definitively to fetal distress, maternal hemorrhage or any number of complications during the puerperium can lead to sub-optimal outcomes or death. The Indiana Perinatal Network (Hawkins, et al., 2012) considers 24/7 in house obstetrical services to be part of subspecialty care; our partners in New York State require a maternal-fetal medicine specialist and a neonatologist to be always on-site and available within 20 minutes in order to be designated either a Level 3 Perinatal Center or a Regional Perinatal Center (i.e., Level 4).

Qualifying under the first measure requires coverage of OB by a physician capable of providing the indicated services: physicians may be obstetricians or family physicians qualified to fill those roles. In testing the measure, we used a hospital's self-report of being a Level 3 hospital for Obstetrics on the AHA Survey, supplemented by a New York hospital profiling Web site (New York State Department of Health, 2012) as a proxy. In 2010 in New York State Medicaid, 24.52

percent of Class A deliveries, 27.98 percent of Class B deliveries, and 24.66 percent of Unduplicated deliveries (combined A and B) occurred in hospitals that met the structural measure for OB care. In all measures, Class A and B are reported separately to promote understandability to complement the combined finding (which always will be dominated by Class A).

The integration of OB anesthesia into high-risk care has gained acceptance (Committee on Obstetric Practice, 2011). With increasing complexity of available anesthesia techniques (Bucklin, Hawkins, Anderson, et al., 2005) and an increase in the occurrence of high-risk deliveries, our expert panel chose to operationalize the structural need as a 24/7 anesthesiologist with training in obstetric anesthesia. This is consistent with the literature and reflects the diverse roles of the obstetrical anesthesiologist and his or her ability to manage complications (Committee on Obstetric Practice, 2011, 2009; Barbieri, Camann, 2002). As these measures are intended to assess availability, panelists were not dissuaded by data (Bucklin, et al., 2005) suggesting a shortage of OB anesthesiologists. Such a shortage may motivate the use and elevate the importance of this measure.

The use of general anesthesia in 15-30 percent of emergent c-sections contrasted with less than 5 percent in elective c-sections simultaneously suggests that urgent situations may require different management than do routine ones (arguing for highly skilled specialists) and potentially that a lack of higher level anesthesia care in urgent situations may limit available options for women (Committee on Obstetric Practice, 2011; Bucklin, et al., 2005). Indeed, responses from interviews with clinicians during the CAPQuaM 360 degree process suggest that lack of availability frequently limits women's options.

Of interest, hospitals that deliver between 100 and 500 babies annually represent 36 percent of all hospitals and account for almost 8 percent of births, suggesting the importance of workforce distribution (Bucklin, et al., 2005). These hospitals make up 36 percent of all hospitals, suggesting their critical importance when developing measures of availability. Twenty percent of hospitals that delivered 500-1,500 babies per year (the middle stratum) reported themselves to be regional referral centers for HROB. See the Supporting Documents for Tables 2-4; the data within categories are nationally representative, but final sampling weights are not available to make national estimates across categories.

Variability in anesthesiology staffing is not defined by hospital characteristics. While similar structural characteristics predict obstetrical and anesthesiology coverage, they do not overlap, supporting distinct measures for OB and anesthesiology coverage.

Our transfusion measure incorporates language from the New York State Department of Health criteria to identify Regional Perinatal Centers in New York. The clinical imperative to look at availability of these services is set forth by the California Maternal Quality Care Collaborative (CMQCC; 2015). Hemorrhages occur predictably, in the context of coagulation disorder, somewhat predictably when problems of placentation may be noted before or early in labor, or unpredictably. Large amounts of blood loss may go unnoticed or unappreciated if not monitored, sought, and understood by experienced and meticulous clinicians, often aided by thoughtful protocols. And even in the hands of excellent clinicians, the management of hemorrhage requires

early recognition, proper management to achieve rapid hemostasis, and prompt and sometimes repeated transfusion. Key data from CMQCC are shown in Table 5 (see Supporting Documents).

For our New York State Medicaid data analysis, we used regional perinatal centers (RPC) as a proxy for round-the-clock transfusion services; RPC are required to have these services always available. Considering the three specific categories by which pregnancies are eligible for this measure: 13.38 percent of Class A deliveries, 12.62 percent of Class B deliveries, and 13.46 percent of Class C deliveries occurred in RPC hospitals. We note here another "voltage drop" between OB coverage and blood bank services, validating our decision to include both measures.

Regionalization of perinatal care has been widely accepted in the United States; studies document that delivery at hospitals with Level 3 NICUs is associated with reduced neonatal mortality; the American Academy of Pediatrics encourages regionalization of NICU services (AAP, 2012; McCormick, Shapiro, Starfield, 1985; Phibbs, Bronstein, Buston, et al., 1996) and established Level 3 NICUs as the standard of care for many infants. Our 2010 New York State Medicaid analysis found that the following proportion of deliveries in hospitals that had Level 3 or higher nurseries (identified in this data set by regular submission of Revenue Code 173 or 174): Class A, 34.01 percent; Class B, 37.25 percent; and Unduplicated combined 34.16 percent. Even for Class B, in which the desirability for a NICU is highest and most proximal, nearly two-thirds of women deliver in hospitals that do not have one.

Our literature review, data collection, and data analyses reveal many deliveries in institutions that lack desirable structural characteristics, plus the independent importance of each of these related measures.

Section 4. Measure Categories

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

Does the measure address this category?

a. **Care Setting – ambulatory:** No.
b. **Care Setting – inpatient:** Yes.
c. **Care Setting – other – please specify:** Yes; birthing, delivery.
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:** No.
g. **Service – other (please specify):** Yes; delivery care.
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:** Yes.
m. **Measure Topic other (please specify):** No.
n. **Population – pregnant women:** Yes.
o. **Population – neonates (28 days after birth) (specify age range):** Yes; delivery/newborn.
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):** Yes; pregnant >=10.
s. **Population – adolescents (11 years through 20 years) (specify age range):** Yes; pregnant.
t. **Population – other (specify age range):** Yes; pregnant <=65.
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.

Evidence is discussed throughout this report. A targeted review of the literature is in the Appendix (see Supporting Documents). Further, we interviewed clinicians and engaged clinical societies and accreditors, patient/family groups, New York Medicaid, and others to inform our measure development with the knowledge and experiences of stakeholders, as well as the medical literature. The ratings of the panel along with a brief description of methodology are included (see Supporting Documents). These measures result from the careful conduct of a systematic process.

The availability of high-risk obstetric (HROB) services is a challenging concept, and to develop quality measures that assess the availability of high-risk obstetric services we first needed to define the availability of services and high risk obstetrical services. Specifically we wondered whether the target population could be identified by conditions present in the women, by the clinical services required, or by the clinicians providing the services. Through discussions with our scientific team and steering committee, a review of the literature, and in consultation with our Expert Panel we answered these questions in the following manner. Regarding availability, we expanded on the Anderson and Aday model (Aday, Anderson, 1974), which suggests that utilization of health care is driven by three predisposing characteristics, enabling resources, and need, and that these factors are themselves influenced by the available system of care (Anderson, McCutcheon, Aday, et al., 1983; Kuhlthau, 2011). While their distinction between availability and realized access has blurred over time, we nonetheless chose to respect our assignment by using an availability lens as our framework for these measures.

At a system level, utilization can vary as a result of differences in individual behaviors or system characteristics. The current measures predominantly reflect distribution of system attributes, which may include geography, system design, and/or sufficiency of resources (Kuhlthau, 2011). The definition of HROB services for the purposes of these measures is broad and may include services provided by a variety of clinicians if received by a woman who has an identifiable condition that predisposes her or her baby to an increased risk of morbidity and mortality during the assessment period.

For this measure set we developed two subcategories of high-risk identification. Derived from the literature, Expert Panel ratings, and discussions with our Steering Committee, and from insights drawn from clinician interviews, we include a group of maternal diagnosis codes that place women at increased risk of maternal morbidity and mortality and a group of codes that represent complications of delivery, including low birth weight, that place infants at risk for increased morbidity and mortality. Poor birth outcomes, such as birth asphyxia or stillbirth are included among the latter.

A significant proportion of pregnant women are at higher risk for maternal or infant morbidity and mortality (Hankins, et al, 2012). Professional societies in pediatrics, anesthesia, and obstetrics provide guidance about the need for availability of specific services regarding HROB. A 2009 Joint Statement from the American Society of Anesthesiologists (ASA) and the American College of Obstetricians and Gynecologists (ACOG) called for available OB anesthesia services. Optimal anesthesia care should include credentialed clinicians always available to administer an appropriate anesthetic (Committee on Obstetric Practice, 2009). The Joint Statement also applies to our first measure as it calls for availability of a licensed practitioner who is credentialed to maintain support of vital functions in any OB emergency, including the capacity to start a cesarean delivery within 30 minutes of the decision to perform it (Committee on Obstetric Practice, 2009).

Inadequate physician supervision is an important cause of adverse events around delivery. One review of maternal deaths and near misses, found mismanagement of patient, failure or delay in diagnosis as factors in 90 percent of cases (Geller, Rosenberg, Cox, et al., 2004). Studies focused

on the potential for hemodynamic instability around delivery point out the structures required to manage them (Committee on Obstetric Practice, 2009; Pasupathy, Wood, Pell, et al., 2010). Absence or delay of a physician qualified to deliver OB care has been associated with increased risk and higher rates of intrapartum anoxia (American Hospital Association, 2015). Despite the absence of a randomized trial of 24/7 in house physician coverage of OB, the accumulated evidence supports our expert panel's judgment: this is a critical structural element for HROB.

Although 24-hour in-house anesthesia coverage has not been evaluated in a randomized trial, evidence suggests that inadequate anesthesiologist supervision is associated with maternal death. In a study of 18 years of anesthesia-related deaths in Michigan, more than half were attributed to inadequate supervision by an anesthesiologist (Howell, et al., 2008). Adequate monitoring by an anesthesiologist is vital, as nearly one-third of all births in the United States are cesarean deliveries, an increase of nearly 50 percent since 1996 (Martin, Hamilton, Ventura, 2012). Risk of death for women with an emergency cesarean section is three times as high as for those with a planned cesarean section, suggesting that availability of physicians capable of safely performing an emergent cesarean section is of great concern in obstetric care. Moreover, in a retrospective study examining 1.5 million deliveries from 2000 to 2006, the rate of maternal mortality was 10-fold as high for cesarean delivery as for vaginal delivery (Clark, Belfort, Dildy, et al., 2008). Consistent with well-documented increases in maternal mortality in cesarean versus vaginal deliveries, the risk of severe maternal morbidity also increases (5-10 times as high), which includes hemorrhage and increased blood loss (Shearer, 1993). Pregnancy-related hemorrhage and transfusion rates have increased substantially over the last decade (Callaghan, et al., 2012). Therefore, on-site blood banking/transfusion services are imperative for planned and emergency cesareans, as well as other complications resulting in hemorrhage and extensive maternal blood loss. In their program that designates Regional Perinatal Centers, our partners in the New York State Department of Health operationalize it as: "24-hour capability to provide blood group, Rh Type, cross-matching, antibody testing; Either ABO specific or 0-Rh-negative blood and fresh frozen plasma and cryoprecipitate available at the facility at all times…"

Postpartum hemorrhage remains one of the most significant maternal complications of childbirth in the United States, with peripartum transfusion the most commonly identified morbidity (Ehrenthal, Chichester, Cole, et al., 2012). Given the increased risk for transfusion among women with anemia and placentation disorders, we will assess on-site blood banking for all high-risk deliveries (Rouse, MacPherson, Landon, et al., 2006).

Lastly, our definition of high-risk deliveries includes deliveries of low birth weight infants. An abundance of literature has demonstrated that very small infants delivered in Level 3 nurseries have better outcomes (Chung, Phibbs, Boscardin, et al., 2010; Phibbs, et al., 1996). In the 1970s, regionalization of perinatal care was instituted in the United States, and evaluations have demonstrated that antepartum risk identification and transfer of management of high-risk pregnancies to tertiary centers for delivery resulted in reduced neonatal mortality (McCormick, et al., 1985). Regionalization of perinatal care has been widely accepted in the United States and reaffirmed in a recent American Academy of Pediatrics Policy Statement (AAP, 2012). We include a broader definition to high risk and now propose to measure the proportion of high-risk deliveries that occur in hospitals with Level 3 or higher NICUs.

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

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

This measure has importance as a descriptor of the structural elements necessary for safe maternity care. The rationale can be summarized as follows:

Our expert panel reinforced and prioritized as highly important several structural aspects of high-risk obstetrical care (HROB) that are supported both by the evidence base and by leading clinical societies and other significant actors (see, for example, the New York State Perinatal Designation Matrix in the Appendix [see Supporting Documents], or the California Maternal Quality Care Collaborative OB Hemorrhage Toolkit at http://cmqcc.org/ob_hemorrhage).

There are four topics in the structural measures:

1. Presence of a 24-hour physician dedicated to obstetrics and capable of safely handling routine obstetrical care, as well as obstetrical surgical emergencies, such as emergency cesarean sections.

2. Presence of a 24-hour obstetrical anesthesiologist capable of managing routine and emergent anesthesia care in the labor and delivery setting.

3. Capacity to manage peripartum hemorrhages. Presence of 24-hour blood banking/transfusion services.

4. Presence of a Level 3 or higher neonatal intensive care unit (using American Academy of Pediatrics definitions).

In turn, these four measures represent the capacity to provide critical, often 'life or death' services in a timely way that meets the needs and capacities of these women to obtain them. The significance of these measures raises them to concerns about patient safety. They are specified so as to be able to identify disparities that arise because of socioeconomic, racial/ethnic, and rural/urban considerations. In this regard they address four (Timeliness, Equity, Safety, and Patient-Centeredness) of six characteristics (Efficient and Effective are the other two) of quality care described in the Institute of Medicine's Crossing the Quality Chasm (IOM, 2001). In our review above, we have described the effectiveness of these health care structures on population health. In the absence of a nationwide study evincing evidence that it is cost effective to make services available everywhere in the United States, we have cited evidence that prevention of some of the complications that result from failures will be cost effective or cost saving at least in some circumstances. The proposed measures can provide new measures of availability with which to assess both the outcomes and the cost-effectiveness of future efforts to enhance the availability of HROB services.

We have operationalized the need for HROB services broadly, consistent with the guidance provided by our expert panel. Our definitions borrow from the literature and from AHRQ's own

clinical classification software, and at the margins are defined based upon specific guidance provided by our expert panel.

In so doing, we produce a measure that is more sensitive and less specific, as is desirable for a measure intended to create a gradient at the population level such as we described above. These are not measures designed to assess as good or bad the quality of care for any individual pregnancy. Rather they are designed to provide insight into the availability of HROB services to a population of women who may need them. This approach is consistent with the useful Institute of Medicine definition of quality health care as "The degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" (IOM, 2001) Thus, each of these measures may be said to specify current professional knowledge in a way that produces an index that describes the degree to which specific HROB services (pertaining to delivery) are available to women who are at risk to need them.

The salience and 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 availability of HROB services, described in the next paragraph 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 as 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 (see Supporting Documents).
- 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.

Fortunately, in the case of this measure, we can present both a systematically developed measure and a variety of 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 strengths of this measure derive from its systematic development, its meticulous specification, its careful conceptualization and articulation and its grounding in existing science and consensus.

The data collection and reliability therein depend upon the use of administrative data. These data are used to identify deliveries (our specifications are a slight enhancement of CDC methodologies described by Kuklina, Whiteman, Hillis, et al. [2008]; to Kuklina's work we added Revenue code 722). This was important for our test because the Medicaid MAX data provided by CMS and in which these schemas were tested does not include diagnosis-related groups (DRGs), which are employed in the Kuklina method.

We also tested a variation of the approach to identify deliveries employed by HEDIS in its Timing of Prenatal Care measure in the initial CHIPRA core set. We found that these approaches identified substantially the same population of deliveries in a 16 State subset of the national MAX database. We chose the 18 States to include in an attempt to manifest some standardization of approaches across the seven AHRQ-CMS CHIPRA Centers of Excellence (COEs)—they were recommended to us as a diverse set of States with high data quality by the Children's Hospital of Pennsylvania (CHOP) COE; CHOP has used them extensively in a number of their validation activities. As the different approaches produced 90 percent or more overlap, we decided to specify the measure based upon the Kuklina/CDC approach as both widely used and relevant for the type of population-based approach to measurement proposed in this measure.

In determining which women were to be considered potentially in need of HROB services, our specifications further rely on administrative data. One 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, Kohler, 2001). Further, at the current stage of electronic health record (EHR) development and implementation, chart review is likely to prove infeasible for population-based measures of this

scope. Since this measure is specified to be interpreted at the population and not the individual level, the impact of some of the imperfections of using administrative data will be overcome naturally because of the law of large numbers.

As an illustration of our approach, we provide a case example of our decision to exclude two diagnoses from the inclusion criteria. The expert panel rated valvular heart disease as significant and an indication of the need for HROB. In its deliberations, the panel made clear that conditions such as murmurs or simple mitral valve prolapse, which often are trivial, were not the target of its rating. So in specifying the inclusion specifications, we included Clinical Classifications category 96, Heart Valve Disorders, but specified the removal of three ICD-9-CM codes from that category (4240 Mitral valve disorders, 7852 Undiagnosed cardiac murmurs, 7853 Other abnormal heart sounds). There are two points about this. The CAPQuaM team made the clinical judgment that it was more true to the intention of the panel to accept the error that results from eliminating the rarer more serious isolated mitral valve disorders than the error of including the common and often innocuous mitral valve prolapse in the specified sample. This decision was affirmed when upon their review of our specifications, no panel member questioned that judgment. Secondly, we want to be clear that the three ICD-9-CM codes mentioned were not then used as exclusion criteria if there were other reasons for the pregnancy to be identified as high risk. Rather these codes were removed from the inclusion criteria.

Regarding the assessment of the presence or absence of structural characteristics in this measure set, we have specified this measure to use the results of questionnaires or surveys that we envision as paper, email, or Internet-based. Our feasibility assessment determined that these data are readily available from key individuals at the hospitals. We could imagine that one or more States or health plans have databases that link some or all of these data (especially Level 3 or higher nurseries) to hospitals, and it would be an acceptable approach to use those data. The regular use of Revenue Code 173 or 174 could also be used to identify Level 3 or 4 Nursery care respectively. If challenged, we consider public self-report to be preferable to the use of a database unless there is evidence of deception or fraud.

We have developed our survey questions in accordance with best practices and after studying the American Hospital Association Annual Survey of Hospitals (American Hospital Association, 2008, 2015), which is considered the authoritative survey of hospital structural characteristics in the United States. After careful internal review and revision by the scientific team, appropriate clinicians, and experts, we concluded our development of the four-item questionnaire with one formal cognitive interview. We conducted this interview with the Director of Special Projects in the Office of Patient Excellence at the Mount Sinai Medical Center. This individual is not a clinician and had no previous involvement with the development of these measures. The items were revised and modified in accordance with the findings from that interview and provided back for her review. After a second round of revisions we received confirmation that all relevant issues had been addressed successfully, and that the questions were clear and unambiguous.

We found that of ~119,000 Medicaid deliveries in New York State in 2010, 59,254 were at sufficiently elevated risk to qualify for this measure set: 56,465 (~47 percent) were identified using Class A criteria, 7,800 using Class B (~7 percent), and 59,254 (just under 50 percent) using either Class A or Class B, meaning that 2,789 (or about 5 percent of the overall high-risk

pregnancies) were identified only using Class B. We expected a substantial "voltage drop" between a condition of elevated risk and a complication or an undesirable outcome. Hence, maternal diagnoses codes of Class A will predictably be orders of magnitude larger than the delivery and neonatal codes of Class B. These findings are consistent with our predictions and expectations. Our team had predicted that 40-50 percent of all pregnancies would have elevated risk, and these findings are consistent with the expectations that Medicaid pregnancies would be at least at the higher end of that range. Use of a mother-only algorithm in MAX data in 16 States indicates the proportion of high-risk pregnancies ranges from 31.50 percent in New Jersey to 63.97 percent in Kentucky. The New York MAX finding was 55,379 HROB pregnancies, almost identical to the 56,465 found using internal data bases on the maternal codes, indicating very high reliability across systems.

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 also contains information related to validity. Our definition of high risk obstetrical services results from a formal RAND/UCLA modified Delphi process conducted with a multidisciplinary panel of national experts that included obstetricians, MFM specialists, and a nurse midwife, anesthesiologist and family physician. We carefully operationalized the panel's clinical recommendations by fine tuning AHRQ's Clinical Classification Software. We operationalized panel specifications using data elements that are available in typical administrative data sets. Potential exceptions are elements like race and ethnicity. Our feasibility work confirmed race/ethnicity information is generally available from clinical charts. The CHIPRA legislation (2009) which directs our measures to be capable of identifying disparities and we have specified it to be so, although we are aware of variability in the manner of assignment of race and ethnicity by health care facilities.

Use of administrative data in performance assessment is common. These data contain consistent elements, are available, have information on large numbers of individuals, and are relatively inexpensive. Validity has been established for many administrative databases, their strengths and weaknesses relative to data abstracted from medical records and obtained via survey have been documented, and their use has been encouraged by Federal agencies (Virnig, McBean, 2001). The Centers for Medicare & Medicaid Services has made clear to the participating AHRQ-CMS CHIPRA Centers of Excellence funded to develop measures in the Pediatric Quality Measures Program that it places a premium on feasibility.

Expert panels have been demonstrated to enhance measure development and health care evaluation, including for children (Mangione-Smith, DeCristofaro, Setodji, et al., 2007). Frontline practitioners can assist researchers to create useful measures (Rubio, Berg-Weger, Tebb 2003). CAPQuaM's 360 degree method is highly engaged with collaborators, partners, and the literature. It targets relevant information and perspective, and measures emerge from the

process. Potential measures are tested to the extent that time and resources permit. In developing the HROB availability measures we incorporated:

- Engagement with broadly diverse partnered institutions and senior advisors.
- Detailed literature review.
- Interviews with clinicians from around the country.
- The CAPQuaM scientific team.
- A geographically diverse, multidisciplinary expert panel whose members participated in a two-round RAND/UCLA modified Delphi process, with enhanced follow-up.
- Development of a Boundary Guideline that incorporates 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 using Medicaid data.

Key Aspects of the Validity of HROB Measures

Availability

The construct of availability is complex and can be muddied in the distinction or lack thereof between availability, access, and utilization (Kuhlthau, 2011). For this first PQMP measure set on availability of HROB services, we selected four measures that avoid any potential confusion between availability and access or utilization. In modern medical practice, all women having babies require some form of delivery services. By looking at the rate at which eligible (i.e., high risk) deliveries occur in hospitals that have key structural elements associated with better outcomes, we create an index of the availability of those services. All else equal, we would expect women who live in more medically dense communities to experience greater availability than those in less medically dense communities and those who live in more isolated communities to have less availability. While these measures are challenging to validate definitively, these predictions give us an opportunity to explore construct validity.

High Risk

We have operationalized a systematic expert process informed by a detailed literature review and incorporating a well described and frequently utilized system developed by AHRQ. While we have modified this system, it has been done to be consistent with its use in this context and to remain consistent with the guidance of the expert panel. It is transparent and has high face validity. Our definition that the need for high-risk services extends from preconception to the puerperium implies that maternal conditions (comorbidities), complications of pregnancy, and complications of delivery each may be used to identify the need for high-risk services. For these measures, we classify risk in two ways, one based on maternal diagnoses and another based on delivery, fetal, or infant conditions. We note that all low birth weight infants are products of a

high-risk pregnancy, since premature labor and growth retardation are within our definition of risk. While linking infant and mother charts may occasionally be a challenge for hospitals, it should be less challenging for reporting entities. Our work with the New York State Medicaid data has confirmed the feasibility of such linkage. If linkage is not possible, the Class A portion of the measure can be calculated based on maternal records alone.

Structural Aspects of Care

Data regarding structural aspects are self-reports from health care facilities. We developed a four-item questionnaire with internal review and a single cognitive interview with follow-up review by an individual who could be called on to complete such a survey at Mount Sinai. Our questions are specific and factual, self-report is the current standard for assessing facility characteristics, frequently through the use of the American Hospital Association Survey noted previously. The lack of anonymity for the person completing the survey and the potential verifiability of the questions enhance validity.

We have cited abundant literature that the structural aspects that are the targets of these measures matter. In data from New York State Medicaid for women who met our criteria for high-risk deliveries, we found that these measures vary with a gradient of accessibility of medical services as associated with geographic proximity or metropolitan areas. See Table 6 in the Supporting Documents for results.

We interpret the findings to suggest that these services become less available with increasing rurality, as we had predicted. We designed the measures to identify reduced availability for any reason, including geographic isolation and the observed gradient strongly supports the validity of these as population measures of availability. These measures move in similar directions but not in lock step, confirming that they are measuring related but not identical constructs, as we would hope. The overall availability of these structural components of high-risk obstetrical services is low compared to the identified need. Please see sections 3.C, 5.A, and 6.A, above, for additional evidence of validity.

Section 7. Identification of Disparities

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

7.A. Race/Ethnicity

Our feasibility assessment confirmed that racial and ethnicity data are almost universally available, and that method of assignment of race and ethnicity to the mother varied. It could be based on maternal self-report or assigned by the hospital. National improvement is needed in the methods used to assign race and ethnicity in hospital discharge data. For the purposes of this

measure, we are resigned at this time to using the existing data as recorded in the mothers' medical records.

Testing sites that participated in the CAPQuaM feasibility assessment were asked to determine if maternal race/ethnicity was documented in the maternal chart, the infant chart, or in both charts. Sites were also asked if infant race/ethnicity was documented in the maternal chart, the infant chart, or both charts. Representatives from institutions were asked to determine whether the data source for maternal race/ethnicity was located in an EHR format or a paper format. Institutions were also asked to indicate the difficulty of data abstraction in obtaining maternal race/ethnicity. Responses included very difficult to collect, difficult to collect, not difficult to collect, or unavailable. Virtually all indicated that this was not difficult to collect. The data generally were in the EHR. The New York State Medicaid Program was able to identify race using their information systems; 45 individuals out of nearly 60,000 pregnancies were missing data on race.

We also examined race/ethnicity data in New York State Medicaid files. The following statistics focus on women found to be high risk by at least one of our two approaches identifying high-risk women. Although the scarcity of black women having babies in rural counties limited the scope of our analyses, we were able to see racial differences in the more urban counties. Among women in large metropolitan areas who met our criteria for high-risk deliveries, 44.76 percent of black women, 40.11 percent of Hispanic women, and 30.04 percent of white women with Medicaid delivered in hospitals with Level 3 or higher NICUs. This may reflect housing patterns with increased numbers of minorities in inner cities, more proximate to hospitals with these services. This hypothesis is supported because those living in smaller metropolitan areas (under 250,000), show both lower rates and a different distribution: black women at 33.54 percent, white women at 20.04 percent, and Hispanic women at 13.89 percent.

A different pattern is seen with regional perinatal centers, our proxy for 24/7 blood banking/transfusion centers. For large metro areas among women who met our criteria for high-risk deliveries, 19.25 percent of white women, 13.92 percent of black women, and 13.82 percent of Hispanic women delivered at these institutions. Still a slightly different pattern (black>Hispanic>white) is seen in large metro areas for our OB proxy measure.

We found that our measures are able to identify statistically significant differences in performance across race/ethnicity and poverty and also when stratifying for several of the levels of urbanicity.

7.B. Special Health Care Needs

Not assessed.

7.C. Socioeconomic Status

Institutions participating in feasibility assessments were asked to determine whether sources of payment could be found in patient charts. Payment sources were identified as being in the form of an EHR or a paper record. Representatives from the participating institutions were then asked to assess the difficulty of data abstraction of the payment source. Responses included very difficult to collect, difficult to collect, not difficult to collect, or unavailable. A space was also

provided for institutions to provide an explanation and additional comments that might be insightful. Virtually all indicated that this information was not difficult to collect. The data were generally in the EHRs.

Our feasibility testing demonstrated that we can use Medicaid insurance as a marker for socioeconomic status (SES), and our New York City data demonstrate this to be an important independent predictor of poor maternal and infant 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 which indicate that local disparities in income can serve as an independent predictor of outcomes. It also allows this measure to consider issues of SES while using publicly available data and requiring only the mother's county of residence, a more reliable data point than self-reported income.

Our analysis of U.S. Department of Agriculture (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 7 in the Supporting Documents.

All of New York State lies in the top three quartiles. We would expect to find the largest differences between poorer and other counties, than across the upper end of the spectrum. Nonetheless we conducted the analysis and found statistically significant differences. Quartile 2 was slightly better than the top quartile in performance, but the 3rd quartile, below the median, had less than half the proportion of high-risk women delivering at sites with each of the structural attributes than Quartile 2 (see Table 8 in the Supporting Documents). Interestingly, poor counties performed better than did the most rural counties, confirming that these various approaches to stratification are capturing different information.

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 the results are shown in Table 9 (see Supporting Documents). The population is heavily weighted to metropolitan areas as demonstrated in Table 10 (see Supporting Documents).

As noted, we use Urban Influence Codes (UIC), which have been developed by the USDA based on a number of criteria to describe the levels of urbanicity and rurality. This is intended not only to report within plan differences but to allow for aggregation as appropriate. While each UIC has its own meaningful definition, some researchers choose to aggregate various codes. We recommend consideration of the aggregation schema of Bennett and colleagues at the South Carolina Rural Research Center (Bennett, Olatosi, Probst, 2008). Their aggregation scheme brings together Codes 1 & 2 as Urban; 3,5, & 8 as micropolitan rural; 4,6, & 7 as rural adjacent to a metro area; and 9, 10, 11, & 12 as remote rural. We observe that UIC 5 might as well be aggregated with 4,6,&7 as an adjacent rural area. Further, 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. However, use of such categories is consistent with the ACA's intent that data that are collected for racial and ethnic disparities also look at underserved frontier counties. Frontier health care may be approximated by analysis of the remote rural categories (Hart, 2012). Our judgment was confirmed after 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.

Those interested in care specific to large cities may wish to aggregate rural areas and analyze UIC 1 and 2 separately.

The New York State Medicaid data were sensitive to urbanicity, and this was described above in the Section 6.B Validity.

7.E. Limited English Proficiency (LEP) Populations

LEP was not assessed, but there is nothing intrinsic to the measure to inhibit its use in that population so long as the LEP characteristic can be linked to the pregnancy or delivery data.

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?

The CAPQuaM High-Risk OB measures seek to assess the proportion of high-risk women that deliver at hospitals without certain levels of available services. As such, the data elements of interest include:

- The presence or absence of certain services (24/7 in house laborists, 24/7 in house anesthesia, 24/7 onsite blood banking, Level 3 NICU) for obstetric patients.
- Documentation of conditions that would classify a woman as "high risk."
- The number of deliveries at a given hospital.

For stratification purposes:

- Race and ethnicity.
- Insurance type (Medicaid, private, uninsured).
- Managed care insurance – Yes/No (where applicable).
- Benefit category (for Medicaid and CHIP-eligible cohorts).
- Income level (as recorded for Medicaid and CHIP-eligible cohorts).
- County equivalent and State or zip code of residence.

Several of these data elements are readily available through hospital administrative data. For example, identification of women with "high-risk" conditions, number of deliveries at a given hospital, and presence of a Level 3 (or higher) NICU can be achieved through use of the appropriate ICD9-CM, Clinical Classifications Software (CCS), and/or revenue codes. Additionally, benefit type is typically recorded in health plan, Medicaid, and CHIP administrative data sets.

As part of our feasibility assessment, CAPQuaM partnered with New York State Medicaid to conduct a variety of analyses using their administrative data set. The findings from these analyses indicated that the aforementioned administrative data elements are also readily available at the state-level, and can be abstracted and used for calculating and reporting the CAPQuaM HROB measures. Further, we have specified several variables, for SES, and urbanicity by linking county of residence at the time of delivery to publicly available data sets.

The CAPQuaM feasibility assessment received responses from 9 of 10 sites with obstetrical services around the country. Results from the assessment indicated that, in general, the data elements of interest are available in the medical record system and not difficult to abstract, including race, ethnicity, and zip code or State and county of residence, for those administrative systems that may lack them.

Payment source (insurance type) should be available in a health plan database and is also easily obtained from electronic data at the health care facility. While linking mother and infant records could be challenging at a few sites, this should be less difficult for insurers, and the New York State Medicaid program had no problem doing so for our analyses.

Finally, while data elements are generally available in administrative data sets, we have collected data from health care facilities with obstetrical services and confirmed the availability and accessibility of supplemental data were needed.

Our survey also asked about a number of structural characteristics including 24/7 services, blood banking/transfusion, and NICU services and found these respondents did not have any trouble answering these questions.

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 HROB Structural measures are generally available in the existing data systems, except for the structural attributes themselves. Development of a standardized assessment tool (or adoption of the CAPQuaM Demonstration Survey) and maintenance of a database with results would speed up the use of these measures. Enhancement of collection of patient reported race-ethnicity data into existing administrative systems would also be valuable.

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.

See above.

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?

This is a new measure.

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?**
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?**
Minimum size specified for analysis is 250. Study of HROB deliveries in MAX data in 18 States using slightly less sensitive criteria than those specified herein found range from 1637 (VT) to 55,382 (NY). The Median is 14,500, with 25 percent less than 4,000 deliveries.

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?**
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?**
Minimum size specified for analysis is 250. Study of HROB deliveries in MAX data in 18 States using slightly less sensitive criteria than those specified herein found a range from 1637 (VT) to 55,382 (NY). The Median is 14,500, with 25 percent less than 4,000 deliveries. We specify using urban influence codes which allows for a variety of analyses.

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.

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

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

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

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**
Minimum size specified for analysis is 250. Study of HROB deliveries in MAX data in 18 States using slightly less sensitive criteria than those specified herein found range from 1637 (VT) to 55,382 (NY). The Median is 14,500, with 25 percent less than 4,000 deliveries.

In Use: **Have measure results been reported at this level previously?**
No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**
No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**
None anticipated.

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

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**
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?**
Minimum size specified for analysis is 250. Study of HROB deliveries in MAX data in 18 States using slightly less sensitive criteria than those specified herein found range from 1637 (VT) to 55,382 (NY). The Median is 14,500, with 25 percent less than 4,000 deliveries.

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.

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 specified for this purpose.

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 specified for this purpose.

Provider Level
Hospital: Can compare hospitals

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 specified for this purpose.

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 specified for this purpose.

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)**
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 specified for this purpose.

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 specified for this purpose.

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).

The HROB measures describe the percent of high-risk deliveries that occur in hospitals with the appropriate structural facilities. This measure is straightforward and intuitive, as this represents a desirable clinical practice. Variations at the population level demonstrate differences in the availability of these services for women with high-risk pregnancies and deliveries. These measures are intended for use at the population level and not to assess the quality of care or any individual pregnancy.

We have not tested combining these measures into an index as a 0-4 measure but could envision some States or other entities wanting to do that. We will consider that for our future development work.

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:

- The medical literature (both peer reviewed and gray, including State Web sites).

- Relevant clinicians.
- Organizational stakeholders (our consortium partners, as well as advisory board members,
- see below).
- A multidisciplinary, geographically diverse expert panel, including clinicians and academicians.
- CAPQuaM's scientific team.

Clinical criteria, including consideration of inclusion and exclusion criteria, 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 backgrounds, clinical and regional settings, and expertise. The product of this process was participation by a broad group of experts in the development of clinically detailed scenarios leading to the measures.

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

Alpha testing was performed to assess feasibility, mechanisms of data collection, and operational aspects of collecting and analyzing data for the measure.

Beta testing was performed by the New York State Office of Health Insurance Programs (Medicaid) in close collaboration with the CAPQuaM team, as well as by our team using MAX data from 16 States.

The route to measure specification included development of relevant scenarios and issues for formal processing by our expert panel. Our panel participated in a two-round RAND/UCLA modified Delphi panel that culminated in a 2-day 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.

As health information systems advance, perhaps the administrative data at the heart of this measure could migrate from billing and management systems to the EHR. We are not yet there.

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?

Not at present.

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.

Other than perhaps the race/ethnicity data, the clinical data are a part of routine administrative data systems. The migration of diagnosis data from the EHR directly to administrative systems conceivably could improve the accuracy of the data in the future, although that is not clear. Routine surveys of hospitals regarding key structural attributes could be a part of State health program or State health department administrative or certification activities, and those results could be maintained in a database that would be made available for analysis.

11.D. Health IT Standards

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

No.

If yes, please describe.

Not applicable.

11.E. Health IT Calculation

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

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?

Not applicable.

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).

This measure is based on the self-reported presence of a key structural aspect of hospital care. That is the current state of the art for broad structural measures, absent mechanisms like Accreditation Audits. As these data re not to be collected anonymously, as they are verifiable, as the health care facility has no specific incentive to deceive, and as deception related to performance measurement could be considered fraud, we are confident that this is a mild limitation.

The definition of high-risk obstetrical care is based on a careful, evidence-driven consensus process that was highly engaged and guided by an extraordinary and multidisciplinary panel of national experts. The CAPQuaM team carefully and faithfully operationalized their conclusions and maintained dialogue as we did so. Still, there were infinite combinations of qualifying criteria, and we had to specify one. We are confident that the specifications are strong, the conditions meaningful, and the population at increased risk. But these measures were designed from the outset and explicitly discussed at the expert meeting to be population-based measures. They are intended for the measurement of performance across populations, not for the assessment of the quality of an individual's care. The inevitable noise in the measures was designed to be dwarfed by the signal when applied to large numbers of pregnant women but not for any given individual.

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 innovative set of measures addresses a complex and critical idea: How available are important high-risk obstetrical (HROB) services to women who may need them? We set forth specifications to identify pregnancies that constitute high risk. We assess four critical sets of practices or services and pose the same question for each: what percent of high-risk pregnancies are delivered in facilities that make each of these practices available 24 hours a day, 7 days a week? The practices are: coverage of the OB service by a physician capable of managing labor and delivery and performing an emergent c-section; dedicated coverage of the OB service by an anesthesiologist qualified to provide OB anesthesia; transfusion services; and a Level 3 or higher NICU.

These measures respond to assignment to CAPQuaM, an AHRQ-CMS CHIPRA Center of Excellence in the Pediatric Quality Measurement Program. We have used a rigorous and

systematic process that was highly engaged with clinicians, stakeholders, and experts to develop these measures. We began with the evidence base and the literature. Childbirth accounts for a plurality of hospital admissions for Medicaid programs; our data show that between one- and two- thirds of these deliveries across the country are high risk. Hospital costs for childbirth and neonatal care are large. High-risk women are at risk for increased rates of maternal and/or infant morbidity and mortality. Maternal deaths and near misses are often preventable through improved quality and safety of maternity care. The rapidly rising rate of cesarean sections and associated complications points out the need for OB staffing by physicians. High maternal hemorrhage rates point out the critical importance of transfusion and blood bank services; and the value of NICU care to promote better outcomes for inborn children are well established. These are important measures regarding quality and patient safety in obstetric care. Racial/ethnic disparities in practice are well documented; these four availability measures address important gaps in quality and safety and have the potential to narrow disparities in maternal and neonatal outcomes.

These measures were designed to be population measures, and we have tested them in that regard. As intended, our validation tests showed that more geographically isolated areas show less availability than areas with more dense medical services. We found the measures to be complementary and not duplicative. They were sensitive to differences in socioeconomic status, race, and urbanicity. We found they could be implemented in New York State Medicaid data and in publicly available MAX data, although with some limitations in the latter since mothers and babies could not be linked. Nonetheless, the measures performed well in both data sets (although space limitations curtailed our presentation of the results).

The health of children in the United States is fostered by healthy pregnancies and deliveries that result in healthy mothers and healthy babies. The availability of high-risk obstetric services is critical for the health of pregnant women with high risk-deliveries and ultimately for the health of their unborn infants. It is important to capture the extent to which women with risk factors for a complicated delivery are delivered at hospitals that provide sufficient care for safe monitoring of emergent cesarean sections, obstetrical anesthesia services, and in-house blood banking/transfusion services, as well as a level 3 NICU.

These four measures are particularly relevant to the Medicaid program, as many of the women covered by Medicaid are at increased risk for maternal and infant mortality and morbidity.

References

Aday LA, Andersen R. A framework for the study of access to medical care. Health Serv Res 1974; 9(3):208-20.

American Academy of Pediatrics Committee on Fetus and Newborn. Levels of neonatal care. Pediatrics 2012; 130(3):587-97.

American Hospital Association. Always There, Ready to Care. The 24/7 Role of America's Full-service Hospitals. Washington, DC: American Hospital Association; 2015. Available at http://www.aha.org/content/15/alwaysthere.pdf. Accessed July 17 2017.

American Hospital Association. Data Collection Methods. Health Forum, LLC, Website; 2008, Updated 2017. Available at http://www.ahadataviewer.com/about/data/. Accessed July 18, 2017.

Andersen RM, McCutcheon A, Aday LA, et al. Exploring dimensions of access to medical care. Health Serv Res 1983; 18(1):49-74.

Andrews R. The national hospital bill: the most expensive conditions by payer, 2006. HCUP Statistical Brief 59. Rockville, MD: Agency for Healthcare Research and Quality; 2008.

Barbieri R, Camann WR. A pivotal partnership: Obstetrics and anesthesiology [Editorial]. OB Manage 2002(March):8-9.

Bennett KJ, Olatosi B. Probst J.C. Health disparities: A rural-urban chartbook. Columbia, SC: South Carolina Rural Health Research Center; 2008.

Berg CJ, Atrash HK, Koonin LM, et al. Pregnancy-related mortality in the United States, 1987-1990. Obstet Gynecol 1996; 88(2):161-7.

Berg CJ, Callaghan WM, Syverson C, et al. Pregnancy-related mortality in the United States, 1998 to 2005. Obstet Gynecol 2010; 116(6):1302-9.

Bucklin BA, Hawkins JL, Anderson JR, et al. Obstetric anesthesia workforce survey: Twenty-year update. Anesthesiology 2005; 103(3):645-53.

California Maternal Quality Care Collaborative (CMQC). OB Hemorrhage Toolkit, V.2.0. 2015. Available at http://cmqcc.org/ob_hemorrhage. Accessed July 17, 2017.

Callaghan WM, Creanga AA, Kuklina EV. Severe maternal morbidity among delivery and postpartum hospitalizations in the United States. Obstet Gynecol 2012; 120(5):1029-36.

Chung JH, Phibbs CS, Boscardin WJ, et al. The effect of neonatal intensive care level and hospital volume on mortality of very low birth weight infants. Med Care 2010; 48(7):635-44.

Clark SL, Belfort MA, Dildy GA, et al. Maternal death in the 21st century: Causes, prevention, and relationship to cesarean delivery. Am J Obstet Gynecol 2008; 199(1):36 e1-5; discussion 91-2 e7-11.

Committee on Obstetric Practice. ACOG committee opinion No. 433: Optimal goals for anesthesia care in obstetrics. Obstet Gynecol 2009; 113(5):1197-9.

Committee on Obstetric Practice. Committee Opinion no. 514: Emergent therapy for acute-onset, severe hypertension with preeclampsia or eclampsia. Obstet Gynecol 2011; 118(6):1465-8.

DeFrances CJ, Cullen KA, Kozak LJ. National Hospital Discharge Survey: 2005 annual summary with detailed diagnosis and procedure data. Vital Health Stat 13 2007(165):1-209.

Diamond CRK, 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? Dis Manage 2001; 4(3):121-30.

Ehrenthal DB, Chichester ML, Cole OS, et al. Maternal risk factors for peripartum transfusion. J Womens Health 2012; 21(7):792-7.

Geller SE, Rosenberg D, Cox SM, et al. The continuum of maternal morbidity and mortality: Factors associated with severity. Am J Obstet Gynecol 2004; 191(3):939-44.

Hankins GD, Clark SL, Pacheco LD, et al. Maternal mortality, near misses, and severe morbidity: Lowering rates through designated levels of maternity care. Obstet Gynecol 2012; 120(4):929-34.

Hart G. Frontier/Remote, Island, and Rural Literature Review, Version 2.99. Grand Forks, ND: University of North Dakota; 2012, Available at http://ruralhealth.und.edu/frontier/pdf/lit_review.pdf. Accessed July 18, 2017.

Howell EA, Hebert P, Chatterjee S, et al. Black/white differences in very low birth weight neonatal mortality rates among New York City hospitals. Pediatrics 2008; 121(3):e407-15.

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

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

Hoyert DL, Danel I, Tully P. Maternal mortality, United States and Canada, 1982-1997. Birth 2000; 27(1):4-11.

Institute of Medicine. Crossing the Quality Chasm: The IOM Health Care Quality Initiative. Washington, DC: National Academies Press; 2001.

Joint Commission. Sentinel Event Alert, Issue 44. Preventing maternal death. January 26, 2010. Available at https://www.jointcommission.org/sentinel_event_alert_issue_44_preventing_maternal_death/. Accessed July 17, 2017.

Kuhlthau KA. Measures of availability of health care services for children. Acad Pediatr 2011; 11(3 Suppl):S42-8.

Kuklina EV, Whiteman MK, Hillis SD, et al. An enhanced method for identifying obstetric deliveries: Implications for estimating maternal morbidity. Matern Child Health J 2008; 12(4):469-77.

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.

Martin JA, Hamilton BE, Ventura SJ, et al. Births: Final Data for 2010. Natl Vital Stat Rep 61(1); August 28, 2012. Available at https://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_01.pdf. Accessed July 18, 2017.

McCormick MC, Shapiro S, Starfield BH. The regionalization of perinatal services. Summary of the evaluation of a national demonstration program. JAMA 1985; 253(6):799-804.

Nannini A, Weiss J, Goldstein R, et al. Pregnancy-associated mortality at the end of the twentieth century: Massachusetts, 1990-1999. J Am Med Womens Assoc 2002; 57(3):140-3.

New York State Department of Health. New York State Hospital Profile 2012 (updated 2017). Available at https://profiles.health.ny.gov/hospital/. Accessed July 17, 2017.

New York State Department of Health. Table 4: Live birth summary by mother's race/ethnicity, New York State, 2011. 2013. Available at http://www.health.ny.gov/statistics/vital_statistics/2011/table04.htm. Accessed July 17, 2017.

Pasupathy D, Wood AM, Pell JP, et al. Time of birth and risk of neonatal death at term: Retrospective cohort study. BMJ 2010; 341:c3498.

Phibbs CS, Bronstein JM, Buxton E, et al. The effects of patient volume and level of care at the hospital of birth on neonatal mortality. JAMA 1996; 276(13):1054-9.

Rouse DJ, MacPherson C, Landon M, et al. Blood transfusion and cesarean delivery. Obstet Gynecol 2006; 108(4):891-7.

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

Shearer EL. Cesarean section: Medical benefits and costs. Soc Sci Med 1993; 37(10):1223-31.

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

Section 14: Identifying Information for the Measure Submitter

First Name:	Lawrence
Last Name:	Kleinman
Title:	Director, Mount Sinai CAPQuaM
Organization:	Icahn School of Medicine at Mount Sinai
Mailing Address:	One Gustave L. Levy Place, 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. 17(18)-P008-1-EF
May 2018