Prenatal Care Screening

Section 1. Basic Measure Information

1.A. Measure Name

Prenatal Care Screening

1.B. Measure Number

0170

1.C. Measure Description

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

This measure assesses the percentage of patients who received the following screening tests within the specified timeframes: screening for neural tube defects; screening for gestational diabetes; screening for asymptomatic bacteriuria; hepatitis B specific antigen screening; HIV screening; group B streptococcus screening (GBS).

This measure was developed by the American Medical Association (AMA)-convened Physician Consortium for Performance Improvement (PCPI®), which is a key member of the Pediatric Measurement Center of Excellence (PMCoE) consortium. The PMCoE is funded by the Agency for Healthcare Research and Quality (AHRQ) and includes the following consortium members: the American Academy of Pediatrics; American Board of Pediatrics; American Board of Medical Specialties; Northwestern University; Truven Health Analytics (formerly Thomson Reuters); Children's Hospital and Health System, Milwaukee; Medical College of Wisconsin; and the American Medical Association.

1.D. Measure Owner

The AMA-convened PCPI® is the measure owner.

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.**

   Prenatal/Perinatal Performance Measurement Set

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.**

   Not applicable.

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.**

   Not applicable.

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

Patients who received the following screening tests during the prenatal period within the specified timeframes:

- Screening for neural tube defects: screening using maternal serum alpha-fetoprotein (MSAFP) between weeks 15-20 weeks gestation or screening by ultrasound after 16 weeks gestation.
- Screening for gestational diabetes before or at 28 weeks gestation (patients with a diagnosis of diabetes are excluded).
- Screening for asymptomatic bacteriuria before or at 16 weeks gestation.
- Hepatitis B specific antigen screening at first visit (patients with documented immunity to hepatitis B or active hepatitis B are excluded).
- HIV screening at first visit (patients with a diagnosis of HIV are excluded).
- Group B streptococcus screening (GBS) at 35 to 37 weeks gestation (patients with previously diagnosed GBS or a prior baby that was infected are excluded).

Note: To satisfactorily meet the numerator, all components must be performed.

1.H. Numerator Exclusions

None.

1.I. Denominator Statement

All patients, regardless of age, who gave birth during a 12-month period and were seen at least once for prenatal care.

1.J. Denominator Exclusions

Patients who have an active diagnosis of diabetes, hepatitis B, HIV, or AIDS, a positive screening for group b streptococcus, or a previous diagnosis of group b streptococcus infection or GBS infection of infant.

1.K. Data Sources

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

Electronic health record (EHR).

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

Section 2: Detailed Measure Specifications

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

Please see supporting documents for full eSpecifications and coding spreadsheets. The measure specifications include the following components: (1) a text description of the measure; (2) the Data Requirements Table, which outlines the data elements that are required for the measure, including the identification of the clinical vocabularies applicable to a given data element, the NQF Quality Data Model category and State, as well as the timing parameters for each data element; (3) a visual flow diagram that uses Boolean logic to identify the initial patient population, exclusions, denominator, numerator and exceptions included in the measure; (4) measure calculation; and (5) value sets for each of the data elements. The measure specifications provide the required information to collect the data needed to calculate the quality measure.

See Supporting Documents for additional 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).**

Appropriate prenatal care, including timely screening and testing, is an important component for a positive outcome for both mother and baby (ACOG, 2010; Branson, Handsfield, Lampe, et al., 2006; Simonsen, Anderson-Berry, Delair, et al., 2014; U.S. Preventive Services Task Force, 2008). This measure assesses prenatal care screenings where there is a gap in quality of care; therefore, not all prenatal screening tests are included in this measure. The use of ultrasonography to assess for potential fetal abnormalities, confirm the site of pregnancy within the uterus, and determine gestational age is considered the standard of care. Also, the use of ultrasound scanning during the first trimester is correlated with reduced post-term labor induction rates as compared to second trimester ultrasound scanning. Not screening for asymptomatic bacteriuria has been linked to a greater risk for pyelonephritis and for low birth weight (< 2500 g); urine culture can reliably detect asymptomatic bacteriuria. A positive test result is defined as the presence of a single uropathogen in a mid-stream clean-catch specimen of at least 105 colony-forming units per milliliter of urine.

Approximately 7 percent of pregnancies in the United States are complicated by gestational diabetes. Gestational diabetes can lead to neonatal hypoglycemia, respiratory distress syndrome, and fetal macrosomia. Larger infants have increased rates of birth trauma, shoulder dystocia, and cesarean delivery. Women with gestational diabetes who have a higher pre-pregnancy body mass index (BMI) or who gain more weight during pregnancy are more likely to develop type 2 diabetes following pregnancy.

Despite substantial progress in prevention of perinatal group B streptococcal (GBS) disease since the 1990s, GBS remains the leading cause of early-onset neonatal sepsis in the United States. The majority of infections in newborns occur within the first week of life and are designated early-onset disease. Late-onset infections occur in infants aged >1 week, with most infections evident in the first 3 months of life. Young infants with invasive GBS disease usually present with sepsis or pneumonia, and less often contract meningitis, osteomyelitis, or septic arthritis. In pregnant women, GBS can cause clinical infections, but most women have no symptoms associated with genital tract colonization. Urinary tract infections caused by GBS complicate 2 to 4 percent of pregnancies. During pregnancy or the postpartum period, women can contract amnionitis, endometritis, sepsis, or rarely, meningitis caused by GBS (Verani, McGee, Schrag, 2010).

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).**

This measure would fill a gap in the Medicaid and CHIP programs core set of children's health care quality measures aimed at providing services and treatment to promote healthy birth and prevent premature birth. This measure is important to Medicaid and CHIPRA because it expands the core set of measures beyond their current use. The measure will provide a mechanism to help assess the quality of prenatal care screenings among women in the Medicaid population and help to prevent adverse maternal and neonatal outcomes. This measure is of particular importance for CHIPRA in that it is high impact with Medicaid patients and addresses concerns related to both mother and baby. Additionally, the Medicaid population includes a greater than average percentage of obese pregnant women. There are disparities in obesity among minority patients, who are disproportionately represented among the Medicaid population; this is both costly to Medicaid and increases risks of undesirable outcomes.

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

Describe, if known, how this measure complements or improves on an existing measure in this topic area for the child or adult population, or if it is intended to fill a specific gap in an

existing measure category or topic. For example, the proposed measure may enhance an existing measure in the initial core set, it may lower the age range for an existing adult-focused measure, or it may fill a gap in measurement (e.g., for asthma care quality, inpatient care measures).

None.

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:** Yes.
b. **Care Setting – inpatient:** No.
c. **Care Setting – other – please specify:** No.
d. **Service – preventive health, including services to promote healthy birth:** Yes.
e. **Service – care for acute conditions:** No.
f. **Service – care for children with acute conditions:** No.
g. **Service – other (please specify):** Not applicable.
h. **Measure Topic – duration of enrollment:** No.
i. **Measure Topic – clinical quality:** Yes.
j. **Measure Topic – patient safety:** Yes.
k. **Measure Topic – family experience with care:** No.
l. **Measure Topic – care in the most integrated setting:** No.
m. **Measure Topic other (please specify):** No**.**
n. **Population – pregnant women:** Yes.
o. **Population – neonates (28 days after birth) (specify age range):** No.
p. **Population – infants (29 days to 1 year) (specify age range):** No.
q. **Population – pre-school age children (1 year through 5 years) (specify age range):** No.
r. **Population – school-aged children (6 years through 10 years) (specify age range):** No.
s. **Population – adolescents (11 years through 20 years) (specify age range):** No.
t. **Population – other (specify age range):** No.
u. **Other category (please specify):** No.

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.

Studies indicate that many pregnant women are not tested and screened for essential prenatal markers. In 2006, diabetes during pregnancy (diabetes diagnosed both prior to and during pregnancy), was reported at a rate of 42.3 per 1,000 women, (just over 4 percent) compared with 38.5 per 1,000 in 2005. During the 1990s, the diabetes rate increased by an average of 3 percent per year, but between 2000 and 2002, the pace of increase rose to 6 percent per year (American Dietetic Association, 2008).

Evidence is convincing that for pregnant women, detection of and treatment for asymptomatic bacteriuria with antibiotics significantly lowers the incidence of symptomatic urinary tract infections in the mother and low birth weight in the offspring (U.S. Preventive Services Task Force, 2008).

The incidence of invasive GBS infections among pregnant women in the United States declined by 21 percent from 0.29 per 1,000 live births in 1993 to 0.23 in 1998, suggesting that increased use of intrapartum antibiotics also prevented some cases of maternal GBS amnionitis and endometritis. The most robust evaluation comes from a multi-State, population-based analysis of 819,000 live births during 2003--2004 and a similarly designed study of births during 1998--1999. The proportion of infants whose mothers were screened for GBS colonization before delivery increased from 48.1 percent during 1998--1999 to 85.0 percent during 2003--2004; among women screened during 2003--2004, a total of 98.4 percent had a result available at labor. Among screened women, 24.2 percent were documented as GBS-positive, within the range of expected colonization rates. The proportion of mothers with an indication for intrapartum antibiotic prophylaxis who received them also increased substantially, from 73.8 percent during 1998--1999 to 85.1 percent during 2003—2004 (Verani, et al., 2010).

The following evidence statements are quoted verbatim from the referenced clinical guidelines:

Screening for Fetal Chromosomal Abnormalities

First-trimester screening using both nuchal translucency measurement and biochemical markers is an effective screening test for Down syndrome in the general population. At the same false-positive rates, this screening strategy results in a higher Down syndrome detection rate than does the second-trimester maternal serum triple screen and is comparable to the quadruple screen. (Level A)

Women found to have increased risk of aneuploidy with first-trimester screening should be offered genetic counseling and the option of chorionic villus sampling (CVS) or second-trimester amniocentesis. (Level A)

Screening and invasive diagnostic testing for aneuploidy should be available to all women who present for prenatal care before 20 weeks of gestation, regardless of maternal age. Women should be counseled regarding the differences between screening and invasive diagnostic testing. (Level B)

Patients who have a fetal nuchal translucency measurement of 3.5 mm or higher in the first trimester, despite a negative aneuploidy screen, or normal fetal chromosomes, should be offered a targeted ultrasound examination, fetal echocardiogram, or both. (Level B)

American College of Obstetricians and Gynecologists, 2007.

Screening for Asymptomatic Bacteriuria

Pregnant women should have a urine culture to screen for asymptomatic bacteriuria at 12 to 16 weeks' gestation or at the first prenatal visit, if later.

U.S. Preventive Services Task Force (USPSTF), 2008; Grade A recommendation.

Screening for Gestational Diabetes Mellitus

All pregnant women should be screened for gestational diabetes mellitus (GDM) and/or impaired glucose tolerance (IGT); however, depending on level of risk, timing of screening will differ. Research indicates the similarities between GDM and IGT, and both are associated with increased risks of poor maternal/neonatal outcomes if left untreated.

American Dietetic Association, 2008; Strong, Imperative, Grades I and II.

HIV Screening

Clinicians should screen all pregnant women for HIV. There is good evidence that both standard and FDA-approved rapid screening tests accurately detect HIV infection in pregnant women and fair evidence that introduction of universal prenatal counseling and voluntary testing increases the proportion of HIV-infected women who are diagnosed and are treated before delivery.

USPSTF, 2008; Grade A Recommendation.

Universal HIV testing with patient notification should be a routine component of prenatal care; however, this must be in accordance with current State laws.

Joint statement of ACOG/AAP on human immunodeficiency virus screening, reaffirmed 2014.

HIV screening should be a routine part of prenatal care for all women.

HIV screening is recommended after the patient is notified that testing will be performed unless the patient declines (opt-out screening).

Separate written consent for HIV testing should not be required; general consent for medical care should be considered sufficient to encompass consent for HIV testing.

Repeat screening in the third trimester is recommended in certain jurisdictions with elevated rates of HIV infection among pregnant women.

Centers for Disease Control and Prevention (CDC) Web site; One test, two lives. HIV screening for prenatal care. 2011. Available at https://www.cdc.gov/features/1test2lives/

GBS Screening

All pregnant women should be screened at 35--37 weeks' gestation for vaginal and rectal GBS colonization (AII).

At the time of labor or rupture of membranes, intrapartum chemoprophylaxis should be given to all pregnant women identified as GBS carriers (AII).

Colonization during a previous pregnancy is not an indication for intrapartum prophylaxis in subsequent deliveries. Screening to detect GBS colonization in each pregnancy will determine the need for prophylaxis in that pregnancy.

Centers for Disease Control and Prevention; 2010 (Verani, et al., 2010).

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.

See Section 3.

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.

Reliability Testing Results – Parallel Forms Testing Approach

Measure reliability was evaluated for accuracy of the measure denominator and numerator. Parallel forms reliability is used to assess the consistency of the results of two tests constructed in the same way using the same content. Parallel Forms testing data are taken from two modalities (manual abstraction and automated reports) and are used to calculate a kappa statistic and an agreement rate. Data analyses were conducted using SAS/STAT software, version 9.3 (SAS Institute, Cary, NC).

For analysis of kappa, we used the approach described by Landis and Koch (1977):

0.0–.20 = slight agreement
.21–.40 = fair agreement
.41–.60 = moderate agreement
.61–.80 = substantial agreement
.81–1.0 = almost perfect agreement

A kappa less than zero would indicate agreement worse than that expected by chance (Bartko, Carpenter, 1976).

For the measure as a whole, we were unable to calculate a kappa statistic. For the denominator, there was perfect agreement between the automated report and the manual abstraction. Both modalities identified the 72 eligible patients. For the numerator, there was perfect agreement between the two modalities. There were no patients who were found to have met the measure during manual abstraction, while the automated report also did not find any patients that met the measure. To meet the measure, there are six separate screenings to be completed, with four screening components having a specified exclusion. If the patient is excluded from a given screening component, that patient will still need all remaining screening components completed in order to meet the measure. We performed a sensitivity analysis to determine the variability in whether or not each of the screenings was completed. We calculated kappa statistics and agreement percentages for each individual component.

First, we looked at the asymptomatic bacteriuria screening component and were unable to calculate a kappa statistic. The automated report identified the component for 1 of the 72 eligible patients, but none of the patients had the component identified during manual abstraction. There was 99 percent observed agreement between the two modalities.

Next, we looked at the neural tube defect screening component, and the results of inter-rater reliability testing on neural tube defects are presented in Table 1.

Table 1. Inter-rater Reliability Results of Data Abstraction on Neural Tube Defects

Automated Report – Neural Tube Defect	Manual Abstraction - Neural Tube Defect		Total
	NO	YES	
NO	26	44	70
YES	2	0	2
Total	28	44	72

Kappa	-0.06
95% Lower Confidence Limit	-0.13
95% Upper Confidence Limit	0.02
Observed Agreement Percentage	0.36

The kappa statistic results indicate that there is slight disagreement. The observed agreement percentage is fair. Almost all of the disagreement comes where there is information found by the manual abstraction but not by the automated report. This suggests a discrepancy where providers are coding the information from the screening somewhere in the EHR but not where the measure is expecting the information to be found.

Next we looked at the gestational diabetes screening component, and the results of inter-rater reliability testing on the gestational diabetes screening component are presented in Table 2.

Table 2. Inter-rater Reliability Results of Data Abstraction on the Gestational Diabetes Screening Component

Automated Report –Gestational Diabetes	Manual Abstraction – Gestational Diabetes		Total
	NO	YES	
NO	7	59	66
YES	2	4	6
Total	9	63	72

Kappa	-0.04
95% Lower Confidence Limit	-0.12
95% Upper Confidence Limit	0.03
Observed Agreement Percentage	0.15

The kappa statistic results indicate slight disagreement. The observed agreement percentage is slight. Almost all of the disagreement comes where there is information found by the manual abstraction but not by the automated report. This suggests a discrepancy where providers are coding the information from the screening somewhere in the EHR but not where the measure is expecting the information to be found.

Next we looked at the hepatitis B screening component, and the results of inter-rater reliability testing on the hepatitis B screening component are presented in Table 3.

Table 3. Inter-rater Reliability Results of Data Abstraction on the Hepatitis B Screening Component

Automated Report –Hepatitis B	Manual Abstraction – Hepatitis B		Total
	NO	YES	
NO	40	0	40
YES	0	32	32
Total	40	32	72

Kappa	1.00
95% Lower Confidence Limit	1.00
95% Upper Confidence Limit	1.00
Observed Agreement Percentage	1.00

The kappa statistic results show that there is perfect agreement. The observed agreement percentage is perfect.

Next we looked at the HIV screening component. The results of inter-rater reliability testing on the HIV screening component are presented in Table 4.

Table 4. Inter-rater Reliability Results of Data Abstraction on the HIV Screening Component

Automated Report –HIV	Manual Abstraction - HIV		Total
	NO	YES	
NO	41	0	41
YES	0	31	31
Total	41	31	72

Kappa	1.00
95% Lower Confidence Limit	1.00
95% Upper Confidence Limit	1.00
Observed Agreement Percentage	1.00

The kappa statistic results show near-to-perfect agreement. The observed agreement percentage is perfect.

Finally, we looked at the group B streptococcus screening component. The results of inter-rater reliability testing on the group B streptococcus screening component are presented in Table 5.

Table 5. Inter-rater Reliability Results of Data Abstraction on the HIV Screening Component

Automated Report –Group B Streptococcus	Manual Abstraction – Group B Streptococcus		Total
	NO	YES	
NO	15	56	71
YES	0	1	1
Total	15	57	72

Kappa	0.01
95% Lower Confidence Limit	-0.01
95% Upper Confidence Limit	0.02
Observed Agreement Percentage	0.22

The kappa statistic results indicate that there is slight agreement. The observed agreement percentage is fair. All of the disagreement comes where there is information found by the manual abstraction but not by the automated report. This suggests a discrepancy where providers are coding the information from the screening somewhere in the EHR but not where the automated report for the measure is expecting the information to be found.

Performance Rate

The performance rate percentage from the automated report was zero. The performance rate percentage from the manual abstraction was also zero.

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 measure was assessed for content validity and face validity. Evidence of content validity is provided by looking for agreement among subject matter experts. The performance measure was assessed for content validity by a panel of expert workgroup members during the development process. This subject matter expert panel consisted of 24 members, with representation from measure methodologists, patient advocacy groups, and the following clinical specialties: anesthesiology, family practice, geriatric medicine, maternal fetal medicine, neonatology, nurse midwife, obstetrics and gynecology, and perinatal nursing. Additional input on the content validity of draft measures was obtained through a 30-day public comment period and also by soliciting comments from a panel of consumer, purchaser, and patient representatives convened by the PCPI specifically for this purpose. All comments received were reviewed by the expert workgroup, and the measure was adjusted as needed.

The expert panel members also assessed the measure's face validity through an online survey. The survey introduction provided the following definition of face validity: Face validity is the extent to which an empirical measurement appears to reflect that which it is supposed to "at face value." Face validity of an individual measure poses the question of how well the definition and specifications of an individual measure appear to capture the single aspect of care or health care quality as intended. The expert panel was asked to rate their agreement with the following statement: The scores obtained from the measure as specified will accurately differentiate quality across providers. A 5-point Likert scale was used in the survey (1=Strongly Disagree; 2= Disagree; 3=Neither Disagree nor Agree; 4 = Agree 5=Strongly Agree).

The survey results show that for the Prenatal Care Screening measure, the mean score was 4.00; 69.2 percent (9/13) of respondents agreed or strongly agreed that the scores obtained from the measure as specified will accurately differentiate quality across providers. No respondents disagreed or strongly disagreed that the scores obtained from the measure as specified will accurately differentiate quality across providers.

References

American College of Obstetricians and Gynecologists. Screening for fetal chromosomal abnormalities. ACOG Practice Bulletin No. 77. Obstet Gynecol 2007; 109:217-27.

American Dietetic Association. Gestational diabetes mellitus (GDM). Evidence-based nutrition practice guideline. Chicago, IL: American Dietetic Association; 2008.

Bartko JJ, Carpenter WT. On the methods and theory of reliability. J Nerv Ment Dis 1976; 163:307–17).

Branson BM, Handsfield HH, Lampe MA, et al. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health care settings. MMWR Recomm Rep 2006 Sep 22; 55(RR14):1-17.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33(1):159–74.

Simonsen K, Anderson-Berry A, Delair S, et al. Early onset neonatal sepsis. Clin Microbiol Rev 2014; 27(1):21-47.

U.S. Preventive Services Task Force. Screening for asymptomatic bacteriuria in adults: U.S. Preventive Services Task Force Reaffirmation Recommendation Statement. Ann Intern Med 2008; 149:43-7. Available at https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/asymptomatic-bacteriuria-in-adults-screening?ds=1&s=bacteriuria. Accessed February 8, 2017.

Verani JR, McGee L, Schrag SJ. Prevention of Perinatal Group B Streptococcal Disease. Revised Guidelines from CDC, 2010. Recommendations and Reports 2010 Nov 19; 59(RR10):1-32. Available at https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5910a1.htm. Accessed February 8, 2017.

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

We included race and ethnicity as Supplemental Data Elements to collect for each measure to allow for the stratification of measure results by these variables to assess disparities and initiate subsequent quality improvement activities. The CDC value sets for race and ethnicity are referenced in the measure specifications to collect race and ethnicity information, which is the requirement for race and ethnicity outlined in the CMS Blueprint (CMS, 2016).

7.B. Special Health Care Needs

Not applicable for this measure.

7.C. Socioeconomic Status

We include payer as a Supplemental Data Element to collect for each measure to allow for the stratification of measure results by these variables to assess disparities and initiate subsequent quality improvement activities addressing identified disparities, consistent with recent national efforts to standardize the collection socioeconomic status data.

The Payment Typology value set is referenced in the measure specifications to collect payer information, which is the requirement for payer outlined the CMS Blueprint (CMS, 2016).

7.D. Rurality/Urbanicity

Future measure testing and implementation should collect data on the location of the patient and provider populations in order to stratify performance and test for variation by location.

7.E. Limited English Proficiency (LEP) Populations

We include preferred language as a Supplemental Data Element to collect for each measure to allow for the stratification of measure results by this variable to assess disparities and initiate subsequent quality improvement activities addressing identified disparities, consistent with recent national efforts to standardize the collection of preferred language data.

The CDC value set is referenced in the measure specifications to collect preferred language information, which is the requirement for preferred language outlined in the CMS Blueprint (CMS, 2016).

Reference

Centers for Medicare & Medicaid Services. CMS Blueprint. Updated April 2016.

Section 8. Feasibility

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

8.A. Data Availability

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

Data Element Table (DET) Tool

The PMCoE Center of Excellence adopted the AMA-PCPI testing methodology, which uses the Data Element Table (DET) tool to assess the availability of the data and the technical feasibility and implementation feasibility of the measures. The DET is an Excel workbook designed to capture information that will determine whether or not each site can feasibly collect the data for the measures. It is structured to collect meta-data about each data element necessary to construct each measure stored in the EHR. It will also collect information related to integrity and validity of data collection. Specifically, the DET is designed to capture the following information:

- **Data element information:** Whether or not the data element is captured in the EHR, the data source application, primary user interface data location, data type, coding system, unit of measure, frequency of collection, and calculability within the measure context.
- **Measure integrity information:** An assessment by the testing site as to what degree the measure, as specified, retains the originally stated intention of the measure.
- **Measure validity information:** An assessment by the testing site as to what degree the scores obtained from the measure, as specified, will accurately differentiate quality performance across providers.

The DET collected responses that were used to assess technical and implementation feasibility for each measure. Measure technical feasibility was defined as “Can my EHR do this?” and measure implementation feasibility was defined as “Will workflow be used consistently?” The responses were captured in the form of a rating using the following responses:

- “Feasible. Can do today.”
- “Feasible with workflow mod/changes to EHR.”
- “Non-feasible. Unable to do today.”

This information was entered from drop-down options pertaining to the specific criteria and in free text fields for questions related to specific workflow and EHR configurations. The free text fields and specific narrative questions provide qualitative feedback from the sites that can be factored into the overall feasibility grade for the measure. The DET is completed by staff at each testing site. After the completion of the DET by the testing sites, a determination can be made as to which of the measures are relevant for each specific site. For some sites, all of the measures in

the Perinatal/Prenatal Measurement Set may be collected, for others it may be only a few. Once the completed DETs were submitted by the test sites, the PMCoE project team conducted quality assurance of the DETs to ensure the data were complete and ready for analysis. A series of analyses were subsequently performed to characterize the feasibility, integrity, and face validity of the measures being tested.

Feasibility testing was conducted at an urban, tertiary care hospital. The test site reported that their EHR could capture all 43 data elements in code, text, or Boolean format. Of these, 22 elements were captured in code, 16 elements were captured in text, and 4 elements were captured in Boolean format.

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?

Measure Technical Feasibility and Implementation Feasibility

The measure technical feasibility assessment determined how many of the total measure data elements are feasible data elements. A "feasible data element" is one that can be captured by the test site's EHR system. The sites assessed technical feasibility for the measure based on the following rating scale:

- "Feasible. Can do today."
- "Feasible with workflow mod/changes to EHR."
- "Non-feasible. Unable to do today."

The sites also used this scale to assess measure implementation feasibility. Implementation feasibility represents the site's ability to implement the measure using current workflows and EHRs and addresses issues of projected data reliability related to the consistency with which providers document and capture the data elements needed to implement the measure. The technical feasibility and implementation feasibility were rated the same for each of the measures. For example, if the technical feasibility of a measure was rated as "Feasible. Can do today," its implementation feasibility was also rated as "Feasible. Can do today."

The site that evaluated the technical and implementation feasibility for this measure selected the rating of "Feasible with workflow mod/changes to EHR." The information provided by the site indicates that workflow changes would be required, and that they would be able to calculate the measure with their current technical configuration. In their feedback, the test site indicated that maternal serum alpha-fetoprotein (MSAFP) screening, urine culture/asymptomatic bacteriuria screening, and diabetes mellitus screening are not recorded in the EHR unless positive. The site would need to alter their documentation procedures to include recording of screenings that are not positive.

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.

We are not aware of any project or setting where this measure has been used or is currently in use.

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

We are not aware of any project or setting where this measure has been used or is currently in use.

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

We are not aware of any project or setting where this measure has been used or is 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?**
Dependent on specific area.

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?**
Unknown.

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?**
Dependent on specific area.

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?**
Unknown.

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?
Dependent on specific area.

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?**
Unknown.

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?**
Dependent on specific area.

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?**
Unknown.

Provider Level
Individual practitioner: Can compare individual health care professionals

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?**
Dependent on specific area.

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?**
Unknown.

Provider Level
Hospital: Can compare hospitals

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

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

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

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?**
Unknown.

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

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

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

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

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?**
Unknown.

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 AMA-PCPI has worked collaboratively on this measure set with the AMA-PCPI-Consumer Purchaser Panel (CPP), which comprises representatives from the patient, consumer, and purchaser communities. The panel strongly supports this measure and applauds the inclusion of it at the level of the individual clinician. The CPP states that this important measure of addressing prenatal screening can help to reduce adverse maternal and neonatal outcomes, as well as reduce medical costs.

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.

The use of health IT in the collection and calculation of this measure allows for the clinical data to be used to assess measure results. The use of clinical data is more desirable compared to administrative data due to the increased granularity of information that can be collected.

11.B. Health IT Testing

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

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

A second phase of reliability testing on the measure also occurred at the same sites where feasibility testing was conducted. This approach utilizes parallel forms reliability testing where measure data elements and performance from an automated report from the EHR are compared to those data from a manual review of the EHR—that is, comparison to the gold standard.

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.

See Section 8.A.

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)?

Yes.

If yes, please describe.

We use the following standards in the development of our EHR specifications: The Quality Data Model (QDM), developed by the National Quality Forum, the vocabulary recommendations named by the Health IT Standards Committee (of the Office of the National Coordinator for Health IT), (e.g., SNOMED, RxNorm, LOINC), and also referenced in the CMS Blueprint. The vocabulary standards used in the specifications are consistent with those recommendations proposed for Stage II of the CMS EHR incentive program (Meaningful Use). Another available standard is the HL7 Health Quality Measure Format (HQMF), an XML-based structured document to express a quality measure specification. The HQMF is used for specifications included in the Meaningful Use program and also references the QDM. The specifications for this measure have not been incorporated into the HQMF eMeasure format; however, the information included in the specifications serve as the foundation for the HQMF—that is, the PCPI electronic specification outlines the requirements to develop the HQMF.

11.E. Health IT Calculation

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

It is highly likely that missing data or ambiguous information stored in the EHR will lead to calculation errors. The specifications provided for this measure are designed to query the EHR in order to obtain the data required for the measure calculation.

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?

A Clinical Decision Support (CDS) system is an example of how a health IT function may enhance performance on this measure. A CDS template or prompt will facilitate documentation of the data required for the measure at the time the physician is treating the patient. Capturing these data at the point of care ensures the required data are present at the time the retrospective query is performed for measure calculation.

Section 12. Limitations of the Measure

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

The measure may have limited utilization due to the limited adoption of EHRs, particularly among practices treating the Medicaid population. However, the vocabulary standards used in the specifications are as proposed for Stage II of the CMS EHR incentive program (Meaningful Use), so its usability is expected to be enhanced by increased participation in this program. As adoption of EHRs increases, utilization of this measure should also increase.

Section 13. Summary Statement

Provide a summary rationale for why the measure should be selected for use, taking into account a balance among desirable attributes and limitations of the measure. Highlight specific advantages that this measure has over alternative measures on the same topic that were considered by the measure developer or specific advantages that this measure has over existing measures. If there is any information about this measure that is important for the review process but has not been addressed above, include it here.

This measure should be selected because it expands the core set of measures beyond their current use. The measure will provide a mechanism to help assess the appropriateness of prenatal care and prevent adverse neonatal outcomes. This measure is of particular importance for CHIPRA in that it is high impact with Medicaid patients and addresses concerns related to both mother and baby. Additionally, since this measure has full eSpecifications, it can be a candidate for future inclusion in the EHR Incentive Program for Meaningful Use. Our EHR specifications follow the standards in the Quality Data Model (QDM), developed by the National Quality Forum, the

vocabulary recommendations named by the Health IT Standards Committee (of the Office of the National Coordinator for Health IT), (e.g., SNOMED, RxNorm, LOINC), and also referenced in the CMS Blueprint. The vocabulary standards used in the specifications are a part of Stage II of the CMS EHR incentive program (Meaningful Use).

Section 14: Identifying Information for the Measure Submitter

First Name:	Ramesh
Last Name:	Sachdeva, MD, PhD, MBA, FAAP
Title:	Professor of Pediatrics (Critical Care)
Organization:	Medical College of Wisconsin
Mailing Address:	9000 West Wisconsin Avenue
City:	Milwaukee
State:	WI
Postal Code:	53226
Telephone:	414-266-3022
Email:	rsachdeva@chw.org

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

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

Public Disclosure Requirements

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

holder of copyright on each submitted measure or instrument. The authority of the signatory to provide such authorization should be described in the letter.

AHRQ Publication No. 14(17)-P009-14-EF
February 2017