BMI Assessment and Recommended Weight Gain

Section 1. Basic Measure Information

1.A. Measure Name

BMI Assessment and Recommended Weight Gain

1.B. Measure Number

0086

1.C. Measure Description

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

Percentage of patients, regardless of age, who gave birth during a 12-month period seen at least once for prenatal care who had a body mass index (BMI) value recorded and were counseled on recommended weight gain during pregnancy at first prenatal care visit. 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: 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 AMA.

1.D. Measure Owner

AMA-convened Physician Consortium for Performance Improvement® (PCPI™) is the measure owner. The AMA has copyright on the measure set.

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.**

   None.

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.**

   Prenatal/Perinatal Performance Measurement Set.

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.**

   None.

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 had a BMI value recorded and were counseled on recommended weight gain during pregnancy at first prenatal care visit

1.H. Numerator Exclusions

None.

1.I. Denominator Statement

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

1.J. Denominator Exclusions

None.

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 attachments for full eSpecifications and coding spreadsheets. Below is an overview of our technical specifications process.

PMCoE adopted the PCPI specification process, which places emphasis on developing comprehensive measure specifications for EHRs, which in turn provide relevant clinical data on patients and actionable feedback to providers. There are several data sources available for collecting performance measures; generally, different data sources require different sets of measure specifications, due to the structure of the systems storing the data. The PCPI recognizes that EHRs are the state of the art for clinical encounters. The PCPI is focusing significant resources and expertise toward specifying and testing measures within EHRs, as they hold the promise of providing the relevant clinical data for measures and feedback to physicians and other health care providers that is timely and actionable.

The type of specifications developed for this measurement set are aligned with the PCPI approach to focus on the development of EHR specifications for new measure development projects. While the PCPI values prospective claims reporting programs and the data these programs can provide, the PCPI is looking to leverage the data in EHRs. This new focus will align the PCPI with national initiatives that highlight the benefits and wealth of data that EHRs bring to health care.

The measure specifications attached with this submission include:

- A text description of the measure.
- 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.
- A visual flow diagram that uses Boolean logic to identify the initial patient population, exclusions, denominator, numerator, and exceptions included in the measure.
- Measure calculation.
- 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.

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).**

4

This measure was developed by the AMA-PCPI, which is a key member of the PMCoE consortium. The AMA-convened Physician Consortium for Performance Improvement® (PCPI™) is a national, physician-led initiative dedicated to improving patient health and safety through the identification and development of evidence-based clinical performance measures and measurement resources that enhance the quality of patient care and foster accountability. The PCPI is nationally recognized for measure development, specification and testing of measures, and enabling use of measures in EHRs. The PCPI's measure development resources include a measure testing protocol, a position statement on the evidence base required for measure development, a composite framework, specification and categorization of measure exceptions, and an outcomes measure framework. The PCPI is made up of over 170 member organizations and individuals including national medical specialty societies, State medical societies, health care professional organizations, Federal agencies, individual members, and other groups interested in improving the quality of health care. Currently, the PCPI portfolio includes measures in more than 46 clinical areas with over 280 individual measures.

The current number of obese women who are pregnant is increasing and provides an important area for quality measurement. Improving the obesity rate in pregnancy will have positive implications for mothers and babies.

Maternal obesity is associated with adverse pregnancy outcomes, including increased risk for gestational diabetes, preeclampsia, cesarean section, and macrosomia. Obese women are more likely to gain in excess of current gestation weight guidelines, which increases the risk for maternal and offspring morbidity. In women of childbearing age, the prevalence of obesity is about 29 percent. To improve outcomes, obstetric providers must effectively evaluate and manage their obese pregnant patients by advising them on the appropriate amount of weight gain in pregnancy, nutritional counseling, and physical activity counseling.

A study looking at practitioner behavior related to managing overweight and obese pregnant patients found that few obstetric providers were fully compliant with clinical practice recommendations, defined obesity correctly, or recommended weight gains concordant with Institute of Medicine (IOM) guidelines (Herring, Platek, Elliott, et al, 2010). Provider personal factors were the strongest correlates of self-reported management practices. A survey was administered to 58 practicing obstetricians, nurse practitioners, and certified nurse midwives at a multispecialty practice in Massachusetts. A 26-item questionnaire that included provider self-reported weight, sociodemographic characteristics, knowledge, attitudes, and management practices with an 8-point score for adherence to eight practices recommended by the American College of Obstetricians and Gynecologists (ACOG) for the management of obese pregnant women. The results showed that among the respondents, 37 percent did not correctly report the minimum BMI for diagnosing obesity, and most reported advising gestational weight gains that were discordant with IOM guidelines, especially for obese women (71 percent). The majority of respondents almost always recommended a range of weight gain (74 percent), advised regular physical activity (74 percent), or discussed diet (64 percent) with obese mothers, but few routinely ordered glucose tolerance testing during the first trimester (26 percent).

Among the Medicaid population, there is more obesity among pregnant women and, hence, an important need for a measure aimed at preventing obesity in pregnancy. Many States have

researched the extent to which women are obese during pregnancy. In Washington State, for example, in 2008, 54 percent of Medicaid-insured women who gave birth had pre-pregnancy weight indicating they were overweight or obese. Among women having their first child (with a parity of 0), one in five is obese before pregnancy; 18 percent of non-Medicaid-insured women and 21.1 percent of Medicaid-insured women were obese before pregnancy.

The proportion classified as obese is higher for low-income women on Medicaid than for higher-income, women not covered by Medicaid at each level of parity. With each successive pregnancy, the proportion classified as obese increases steadily. For Medicaid-insured women, having their fifth (or greater) child, nearly 4 in 10 are obese before pregnancy—the prevalence of obesity in this group (39.2 percent) is nearly double the rate for Medicaid-insured women having their first child (21.1 percent). As weight increases to the levels of overweight and obese, risks for a wide range of medical conditions—such as heart disease, diabetes, high blood pressure, cancer (breast, colon, and endometrial), stroke, and respiratory problems—also increase.

Health consequences of obesity occur both in the general population and among pregnant women. Obese pregnant women have higher rates of delivery by c-section and more complications during and after cesarean delivery. In addition, infants born to obese women have increased risk of stillbirth, prematurity, macrosomia (large for gestational age), and neural tube defects, as well as higher rates of childhood obesity (ACOG, 2005). Overweight and obesity and their associated health problems have a significant impact on the U.S. health care system. One study estimated the annual Medicaid expenditure attributable to obesity for Washington State at $365 million (Finkelstein, 2004). Indirect costs of obesity include absenteeism, decreased productivity, restricted activity, and loss of future income due to early death.

References

American College of Obstetricians and Gynecologists. Counseling and treating obese patient during pregnancy. In ACOG Today, September 2005. Available at https://www.acog.org/-/media/ACOG-Today/acogToday0905.pdf?dmc=1&ts=20160629T1655383566. Accessed June 29, 2016.

Finkelstein EA, Fiebelkorn IC, Wang G. State-level estimates of annual medical expenditures attributable to obesity. Obes Res 2004; 12(1):18-24.

Herring SJ, Platek DN, Elliott P, et al. Addressing obesity in pregnancy: what do obstetric providers recommend? J Womens Health 2010 January; 19(1): 65–70.

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 number of obese pregnant women in the Medicaid population, address educational issues around healthy weight gain, 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 percentage of obese pregnant women is higher in the Medicaid population than in other groups. In addition, there are higher rates of obesity among pregnant minority women, which is both costly to Medicaid and increases risks of undesirable outcomes.

We encourage the use of this measure by physicians, other health care professionals, and health care systems, and health plans, where appropriate. This clinical performance measure is designed for practitioner and/or system level quality improvement to achieve better outcomes for maternity care patients and their babies.

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).

There are no other measures similar to this measure. This measure will fill a significant gap in quality measurement as it addresses a growing problem among pregnant women.

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:** Not applicable.
d. **Service – preventive health, including services to promote healthy birth:** No.
e. **Service – care for acute conditions:** Yes.
f. **Service – care for children with acute conditions:** No.
g. **Service – other (please specify):** No.
h. **Measure Topic – duration of enrollment:** No.
i. **Measure Topic – clinical quality:** Yes.
j. **Measure Topic – patient safety:** Yes.
k. **Measure Topic – family experience with care :** No.
l. **Measure Topic – care in the most integrated setting:** No.
m. **Measure Topic other (please specify):** No.
n. **Population – pregnant women:** 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.

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.

The evidence behind obesity in pregnancy and its link to increased risk of adverse outcomes for mothers and babies is well established and very comprehensive. Research evidence includes clinical practice guidelines, numerous research studies, and data from the Centers for Disease Control and Prevention (CDC) on disparities and quality gaps. Evidence-based clinical practice guidelines that were reviewed for this project include:

- American College of Obstetricians and Gynecologists.
- American Academy of Family Physicians.
- Centers for Disease Control and Prevention.
- United States Preventive Services Task Force.
- Veterans Administration/Department of Defense Clinical Practice Guideline for Pregnancy Management.
- Society of Obstetricians and Gynecologists of Canada.

There is published evidence showing that obesity is rising among pregnant women and is an even greater problem among minority populations. The prevalence of obesity varies by age, gender, race/ethnicity, and socioeconomic status.

The prevalence of obesity (BMI = 30) by race/ethnicity in Washington State among women who gave birth in 2008 was lowest among Asian women (the lowest proportion of obesity, 7.5 percent), while Pacific Islander women had the highest proportion of obesity at 46.3 percent (Washingon State, 2010).

- For four groups—white, Hispanic, Asian, and Pacific Islander—Medicaid-insured women had higher rates of obesity than non-Medicaid-insured women, with the greatest difference among white women.
- For African Americans and American Indians, the rates of obesity were the same, irrespective of Medicaid coverage. This shows the differential effects of poverty and race/ethnicity for different groups of women.

Reference

Washington State. Obesity and pregnancy. First Steps Database. DSHS Research and Data Analysis, April 2010. Report No. 9.99. Olympia, WA.

The following evidence-based clinical practice guideline recommendations were used for this measure:

VA/DoD Clinical Guidelines for Pregnancy Management:

- Recommend assessing and documenting body mass index (BMI) of all pregnant women at the initial visit. Pregnant women found to have a BMI <20 kg/m2 should be referred for nutrition counseling and considered at increased risk for fetal growth restriction.

U.S. Preventive Services Task Force (USPSTF):

- The USPSTF recommends that clinicians screen all adult patients for obesity and offer intensive counseling and behavioral interventions to promote sustained weight loss for obese adults. (B Recommendation, USPSTF, 2003).

Obesity in pregnancy (Society of Obstetricians and Gynecologists of Canada, 2010):

1. Periodic health examinations and other appointments for gynecologic care prior to pregnancy offer ideal opportunities to raise the issue of weight loss before conception. Women should be encouraged to enter pregnancy with a body mass index (BMI) <30 kg/m2, and ideally <25 kg/m2. (III-B).
2. BMI should be calculated from pre-pregnancy height and weight. Those with a pre-pregnancy BMI >30 kg/m2 are considered obese. This information can be helpful in counseling women about pregnancy risks associated with obesity. (II-2B).
3. Obese pregnant women should receive counseling about weight gain, nutrition, and food choices. (II-2B).
4. Obese women should be advised that they are at risk for medical complications such as cardiac disease, pulmonary disease, gestational hypertension, gestational diabetes, and obstructive sleep apnea. Regular exercise during pregnancy may help to reduce some of these risks. (II-2B).
5. Obese women should be advised that their fetus is at an increased risk of congenital abnormalities, and appropriate screening should be done. (II-2B).
6. Obstetric care providers should take BMI into consideration when arranging for fetal anatomic assessment in the second trimester. Anatomic assessment at 20 to 22 weeks may be a better choice for the obese pregnant patient. (II-2B).
7. Obese pregnant women have an increased risk of cessarean section, and the success of vaginal birth after cesarean section is decreased. (II-2B).
8. Antenatal consultation with an anesthesiologist should be considered to review analgesic options and to ensure a plan is in place should a regional anesthetic be chosen. (III-B).
9. The risk of venous thromboembolism for each obese woman should be evaluated. In some clinical situations, consideration for thromboprophylaxis should be individualized. (III-B).

2009 IOM Guidelines for weight gain during pregnancy:

Pre-pregnancy BMI	BMI+ (kg/m2)(WHO)	Total Weight Gain Range (lbs)	Rates of Weight Gain*2nd and 3rd Trimester (Mean Range in lbs/wk)
Underweight	<18.5	28–40	1 (1–1.3)
Normal weight	18.5-24.9	25–35	1(0.8–1)
Overweight	25.0-29.9	15–25	0.6(0.5–0.7)
Obese (includes all classes)	=30.0	11–20	0.5(0.4–0.6)

References

Davies GAL, Maxwell C, McLeod L, et al. Obesity in pregnancy. SOGC Clinical practice guideline. J Obstet Gynaecol Can 2010 Feb; 32(2):165-73.

U.S. Preventive Services Task Force. Screening for and management of obesity in adults (including preconception and postpartum). Grade B. Available at http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/prevention-chronic-care/healthier-pregnancy/documents/adult-obesity-provider-fact-sheet.pdf. Accessed April 18, 2016.

VA/DoD Clinical Practice Guideline for Pregnancy Management. Department of Veterans Affairs, Department of Defense; 2009.

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.

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.

Analytic Method

The study sample for reliability testing was derived from an urban, tertiary-care hospital with an EHR system integrating inpatient and outpatient data. The EHR system is certified for Medicare and Medicaid EHR Incentive Programs. Data used in the analysis are from a patient population of 12,108 for 2010. We carried out an assessment of measure reliability applying a reliability coefficient in the form of the signal to noise ratio (SNR). In SNR analysis, reliability is the measure of confidence in differentiating performance between physicians or other providers (Adams, Mehrotra, McGlynn, 2010; Adams, Mehrotra, Thomas, 2010; Scholle, Roski, Adams, et al, 2008). The signal is the variability in measured performance that can be explained by real differences in physician performance, and the noise is the total variability in measured performance. Reliability is then the ratio of the physician-to-physician variance to the sum of the physician-to-physician variance plus the error variance specific to a physician:

Reliability = Variance (physician-to-physician) / [Variance (physician-to-physician) + Variance (physician-specific-error]

Reliability equal to zero implies that all the variability in a measure is attributable to measurement error. Reliability equal to one implies that all the variability is attributable to real differences in physician performance. Reliability of 0.70 is generally considered a minimum threshold for reliability, and 0.80 is considered very good reliability (Nunnelly, Bernstein, 1994).

The SNR reliability testing was performed using a beta-binomial model. The beta-binomial model assumes the physician performance score is a binomial random variable conditional on the physician's true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates.

Reliability can be estimated at different points. The convention is to estimate reliability at two points: one, at a minimum number of quality reporting events per physician and two, at the average number of quality reporting events per physician. We set the minimum number required as 10 events. Limiting the reliability analysis to only those physicians with a minimum number of events reduces the bias introduced by the inclusion of physicians without a significant numbers of events. Reliability testing results from SNR analysis have been included in support of AMA-PCPI measures submitted for NQF endorsement (National Quality Forum, 2012).

The SNR reliability testing for this measure was undertaken, but results are not available at this time. We produced the automated report from the EHR, but data and analysis of results from reliability testing are not available at this time. We expect that the analysis will provide results on measure reliability, overall measure performance, the distribution of performance rates, and performance stratified by patient race, ethnicity, preferred language, socioeconomic status, and

demographic variables. The structure of the results is the same as that included in our submission of the PMCoE/AMA-PCPI cesarean and episiotomy measures.

A second phase of reliability testing on the measure was undertaken at the same sites where feasibility testing was conducted. This approach utilized parallel forms reliability where measure data elements and performance from an automated report from the EHR were compared to those data from a manual review of the HER—that is, comparison to the gold standard. (See Measure Testing Protocol for PCPI Performance Measures, ama-assn.org/resources/doc/cqi/pcpi-testing-protocol.pdf.)

Adams JL, Mehrotra A, McGlynn EA. Estimating reliability and misclassification in physician profiling. Santa Monica, CA: RAND Corporation; 2010. Available at www.rand.org/pubs/technical_reports/TR863. Accessed April 18, 2016.

Adams JL, Mehrotra A, Thomas JW. Physician cost profiling--reliability and risk of misclassification. N Engl J Med 2010 Mar 18; 362(11):1014-21. Available at http://www.nejm.org/doi/pdf/10.1056/NEJMsa0906323. Accessed April 18, 2016..

National Quality Forum. NQF removes time-limited endorsement for 21 measures; Measures now have endorsed status. Press release, January 18, 2012. Available at http://www.qualityforum.org/News_And_Resources/Press_Releases/2012/NQF_Removes_Time-Limited_Endorsement_for_21_Measures;_Measures_Now_Have_Endorsed_Status.aspx. Accessed April 18, 2016.

Nunnelly J, Bernstein I. Psychometric theory, 3rd ed. New York, NY: McGraw-Hill; 1994.

Scholle SH, Roski J, Adams JL, et al. Benchmarking physician performance: reliability of individual and composite measures. Am J Manag Care 2008; 14:833-8. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2667340/pdf/nihms-99203.pdf. Accessed April 18, 2016.

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 by also 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; 3=Neither Disagree nor Agree; 4 = Agree 5=Strongly Agree).

The survey results show that for the BMI Assessment and Recommended Weight Gain measure the mean score was 4.38; 92.3 percent of respondents (12/13) agree or strongly agree that the scores obtained from the measure as specified will accurately differentiate quality across providers; and no respondents disagree or strongly disagree that the scores obtained from the measure as specified will accurately differentiate quality across providers.

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 include 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 Centers for Medicare & Medicaid (CMS) Blueprint (see https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Blueprint111.pdf for more information about the CMS Blueprint).

Also see Section 8.B.1 and Section 8.B.2.

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 of socioeconomic status data.

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

Also see Section 8.B.1 and Section 8.B.2.

7.D. Rurality/Urbanicity

Future measure testing and implementation will involve collection of 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 these variables 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.

Also see Section 8.B.1 and Section 8.B.2.

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 Tool

The PMCoE Cener 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 metadata 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 DETs collected responses 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 (QA) of the DETs to ensure the data were complete and ready for analysis. Subsequently, a series of analyses were performed to characterize the feasibility, integrity, and face validity of the measures being tested.

Data Element Assessment

All five of the data elements are captured in either code or text format. Feasibility comments provided by one site indicate that the data element for counseling/weight gain recommendation currently is not routinely documented and therefore may not be reliable when extracted from either the inpatient or outpatient EHR.

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 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 test site rated the feasibility of the measure as "Non-feasible. Unable to do today." In their feedback, the test site indicated that this measure will not be feasible if outpatient sites are using paper charts. If the outpatient information is transcribed into the inpatient EHR, the measure could be calculated "fairly easily." Unavailability of the data in an inpatient EHR would not affect feasibility, however, since the measure is specified for ambulatory care settings. There would also need to be routine documentation that pregnancy weight gain counseling was performed. With these adjustments, the feasibility of the measure could be upgraded to "Feasible with workflow modification/changes to EHR."

8.B. Lessons from Use of the Measure

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

The development of the measure was completed, but there has been limited opportunity to have the measure adopted and implemented. Feasibility and reliability testing of the measures have been conducted in EHRs in a variety of settings, including an urban, tertiary care hospital; an urban, public hospital; and a suburban community hospital and provide a description of means of data collection methods and insights into lessons learned. See results presented in Section 6.A. Reliability and Section 8. Feasibility.

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?

Not applicable.

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

Not applicable.

Section 9. Levels of Aggregation

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

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

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

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

State level* Can compare States

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

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

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**
This information is not available.

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?**

This information is not available.

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

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

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

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**
This information is not available.

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?**
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?**
This information is not available.

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?**
This information is not available.

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?**
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?**
This information is not available.

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?**
This information is not available.

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?**

This information is not available.

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?**
This information is not available.

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?**
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?**
This information is not available.

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?**
This information is not available.

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?**
This information is not available.

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?**
This information is not available.

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), comprising representatives from the patient, consumer, and purchaser community. The panel strongly supports this measure and applauds the inclusion of it at the level of the individual clinician. The CPP states this important measure of addressing obesity can help to reduce adverse maternal and neonatal outcomes, as well as reduce medical costs. In addition, the workgroup included member representatives from consumer groups, patient advocacy groups, and a health plan.

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 information technology (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 was carried out at the same sites where feasibility testing was conducted. This approach utilizes parallel forms reliability 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. (See Measure Testing Protocol for PCPI Performance Measures, ama-assn.org/resources/doc/cqi/pcpi-testing-protocol.pdf.)

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/Issues in Implementation for workflow discussion.

11.D. Health IT Standards

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

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

provided with this submission form have not been incorporated into the HQMF eMeasure format, however the information included in the specifications serves 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?

These IT functions could make measure recording in the EHR more feasible and reliable, as well as improve performance on the measure and patient outcomes. For example, computerized decision support with menu drop downs or reminders could be programmed to give providers prompts to provide the appropriate services to patients.

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 deliveries 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 W. Wisconsin Avenue, MS-681

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.