Postpartum Followup and Care Coordination

Section 1. Basic Measure Information

1.A. Measure Name
Postpartum Followup and Care Coordination

1.B. Measure Number
0084

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 who were seen for postpartum care within 8 weeks of giving birth who received a breastfeeding evaluation and education, postpartum depression screening, postpartum glucose screening for gestational diabetes patients, and family and contraceptive planning.

This measure was developed by the 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 American Medical Association (AMA).

1.D. Measure Owner
AMA-convened Physician Consortium for Performance Improvement® (PCPI™).

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 receiving the following at a postpartum visit:

- Breastfeeding evaluation and education, including patient-reported breastfeeding.
- Postpartum depression screening.
- Postpartum glucose screening for gestational diabetes patients.
- Family and contraceptive planning.

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

Breastfeeding Evaluation and Education: Patients who were evaluated for breastfeeding before or at 8 weeks postpartum.

Postpartum Depression Screening: Patients who were screened for postpartum depression before or at 8 weeks postpartum. Questions may be asked either directly by a health care provider or in the form of self-completed paper- or computer administered questionnaires; results should be documented in the medical record. Depression screening may include a self-reported validated depression screening tool (e.g., Patient Health Questionnaire-2 [PHQ-2], Beck Depression Inventory, Beck Depression Inventory for Primary Care, Edinburgh Postnatal Depression Scale [EPDS]).

Postpartum Glucose Screening for Gestational Diabetes: Patients who were diagnosed with gestational diabetes during pregnancy who were screened with a glucose screen before or at 8 weeks postpartum.

Family and Contraceptive Planning; Patients who were provided family and contraceptive planning and education (including contraception, if necessary) before or at 8 weeks postpartum.

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 for a postpartum care visit within 8 weeks of giving birth.

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.

Not applicable.

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 (CMS). 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.

The PMCoE Cener of Excellence adopted the PCPI specification process, which places emphasis on developing comprehensive measure specifications for EHRs that can 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 and 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 for providing 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 form 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 specification provides the required information to collect the data needed to calculate the quality measure. The AMA-PCPI, through PMCoE, will make full measure specifications for the measure available for public use in accordance with the terms detailed in the Notice of Grant Award. Please see the attached written statement from AMA-PCPI and PMCoE.

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).**

This measure was developed by the AMA-PCPI, which is a key member of the Pediatric Measurement Center of Excellence (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 developing measures, specifying 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. Today the PCPI portfolio includes measures in more than 46 clinical areas with over 280 individual measures.

The current quality gap in the number of women receiving postpartum followup is well documented and provides a significant area for quality improvement that has implications for mothers, babies, and providers, such as pediatricians.

Managing and ensuring concrete postpartum followup after delivery is a critical challenge to the health care system impacting the quality of care mothers receive after childbirth; followup for depression screening, breastfeeding evaluation, family planning, and glucose screening are important risk factors to evaluate after childbirth. Maternal depression is one of the most common perinatal complications; however, the disorder often remains unrecognized, undiagnosed, and untreated. The various maternal depression disorders are defined by the severity of the depression and the timing and length of the episode. Studies report that 3 to 25 percent of women experience major depression during the year following childbirth. Establishing the diagnosis of gestational diabetes mellitus offers an opportunity not only to improve pregnancy outcome, but also to decrease risk factors associated with the subsequent development of type 2 diabetes. The American College of Obstetricians and Gynecologists' Committee on Obstetric Practice recommends that all women with gestational diabetes mellitus be screened at 6-12 weeks postpartum and managed appropriately. This is a measure of the adequacy of care

provided for those that come for postpartum care, as patients who do not have postpartum visits are excluded from this measure.

A 2008 report summarized results of women's postpartum experiences from two national surveys carried out by Childbirth Connection. The report showed that 6 percent of mothers did not have a postpartum office visit between 3 and 8 weeks after birth. Mothers were asked to rate if physical or emotional problems interfered with their ability to take care of their baby in the first 2 months after birth; 33 percent reported that their postpartum physical health interfered at least "some" with their ability to care for their baby, and 30 percent reported that their postpartum emotional health interfered at least "some."

A 2006 national survey by Childbirth Connection found that almost half of women were not screened for depression in the postpartum period. Survey results showed that only 58 percent of women reported that they were screened for depression during the postpartum visit.

These disparities and quality gaps highlight the importance for this measure in the CHIPRA program.

References

Declercq ER., et al. Listening to Mothers II: Report of the Second National U.S. Survey of Women's Childbearing Experiences. New York: Childbirth Connection; October 2006.

Declercq ER, Sakala C, Corry MP, et al. New mothers speak out: national survey results highlight women's postpartum experiences. New York: Childbirth Connection; August 2008. Available at http://www.childbirthconnection.org/pdfs/new-mothers-speak-out.pdf. Accessed January 5, 2016.

Gaynes BN, Gavin N, Meltzer-Brody S, et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Summary, Evidence Report/Technology Assessment No. 119. (Prepared by the RTI-University of North Carolina Evidence based Practice Center under Contract No. 290-02-0016.). AHRQ Pub. No. 05-E006-1. Rockville, MD: Agency for Healthcare Research and Quality; February 2005.

VanLandeghem, K. Financing strategies for Medicaid reimbursement of maternal depression screening by pediatric providers. National Academy for State Health Policy; April 2006. http://www.library.ca.gov/crb/pmad/docs/VanLandeghem_2006.pdf. Accessed January 5, 2016.

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 during the postpartum period. The measure will provide a mechanism to help assess issues such as breastfeeding use/rates, postpartum depression screening, and planning for contraceptive use. Additionally, it will also help to manage those patients who had gestational diabetes during the pregnancy. This measure is of particular importance for CHIPRA in that it is high impact with Medicaid patients and addresses risks to mothers and babies in the postpartum period.

We encourage the use of this measure by physicians, other health care professionals, and health care systems, or 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 known measures addressing postpartum followup and postpartum depression. We believe that this measure will fill a critical gap.

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:** No.
e. **Service – care for acute conditions:** No.
f. **Service – care for children with acute conditions:** No.
g. **Service – other (please specify):** No.
h. **Measure Topic – duration of enrollment:** Yes.

i. **Measure Topic – clinical quality:** Yes.
j. **Measure Topic – patient safety:** No.
k. **Measure Topic – family experience with care:** No.
l. **Measure Topic – care in the most integrated setting:** No.
m. **Measure Topic other (please specify):** Yes.
n. **Population –women who gave birth:** Seen for postpartum care before or within 8 weeks of giving birth.
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.

The evidence to support postpartum visits and followup is well established and very comprehensive. Research evidence includes clinical practice guidelines and relevant research articles and publications:

Clinical Evidence Base Available for Measure:

Evidence-based clinical practice guidelines from the following organizations were reviewed for this project:

- American College of Obstetricians and Gynecologists.

- American Academy of Family Physicians.

- Centers for Disease Control and Prevention (CDC).

- United States Preventive Services Task Force (USPSTF).

- Veterans Administration/Department of Defense (VA/DoD) Clinical Practice Guideline for Pregnancy Management.

- American Diabetes Association.

- Society of Obstetricians and Gynecologists of Canada.

The following statements are derived from evidence-based guidelines regarding the postpartum visit and followup.

VA/DoD Clinical Practice Guideline for Pregnancy Management; 2009. The following should be included in the postpartum visit:

- Pelvic and breast examinations. [B]

- Cervical smear should be completed as indicated by cervical cancer screening guidelines. [A]

- Initiate or continue the human papilloma virus (HPV) vaccine series for women age < 26 years. [C]

- Screening for postpartum depression. [B]

- Screening for domestic violence. [B]-

- Diabetes testing for patients with pregnancies complicated by gestational diabetes. The 2-hour 75g oral glucose tolerance test (GTT) is recommended, but a fasting glucose can also be done. [B]

- Education about contraception, infant feeding method, sexual activity, weight, exercise and the woman's assessment of her adaptation to motherhood. Pre-existing or chronic medical conditions should be addressed with referral for appropriate followup as indicated.

Breastfeeding

The USPSTF recommends interventions during pregnancy and after birth to promote and support breastfeeding. This recommendation applies to pregnant women, new mothers, and young children. In rare circumstances involving health issues in mothers or infants, such as human immunodeficiency virus (HIV) infection or galactosemia, breastfeeding may be contraindicated,

and interventions to promote breastfeeding may not be appropriate. Interventions to promote and support breastfeeding may also involve a woman's partner, other family members, and friends.

Depression Screening

The 10-question Edinburgh Postnatal Depression Scale (EPDS) is a valuable and efficient way of identifying patients at risk for "perinatal" depression. The EPDS is easy to administer and has proven to be an effective screening tool. Mothers who score above 13 are likely to be suffering from a depressive illness of varying severity. The EPDS score should not override clinical judgment. A careful clinical assessment should be carried out to confirm the diagnosis. The scale indicates how the mother has felt during the previous week. In doubtful cases it may be useful to repeat the tool after 2 weeks.

National Surveys of Postpartum Experiences

A 2008 report summarized results of women's postpartum experiences from two national surveys carried out by the Childbirth Connection, which showed that 6 percent of mothers did not have a postpartum office visit between 3 and 8 weeks after birth. Mothers were asked to rate if physical or emotional problems interfered with their ability to take care of their baby in the first 2 months after birth, and 33 percent reported their postpartum physical health interfered at least "some" with their ability to care for their baby, while 30 percent reported that their postpartum emotional health interfered at least "some." A 2006 national survey by Childbirth Connection found that almost half of women were not screened for depression in the postpartum period. Survey results showed that only 58 percent of women reported that they were screened for depression during the postpartum visit.

References

Declercq ER, et al. Listening to mothers. Report of the Second National U.S. Survey of Women's Childbearing Experiences. New York: Childbirth Connection; October 2006.

U.S. Preventive Services Task Force. Primary care interventions to promote breastfeeding. U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2008 Oct 21; 149(8):560-4.

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 is being derived from the urban, tertiary-care hospital with an EHR system integrating inpatient and outpatient data. The EHR system is certified for the Medicare and Medicaid EHR Incentive Programs. Data being 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, et al., 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 1.00 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 is being 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: (1) at a minimum number of quality reporting events per physician and (2) 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 number of events. Reliability testing results from SNR analysis have been included in support of AMA-PCPI measures submitted for National Quality Forum (NQF) endorsement (National Quality Forum, 2012).

The measure reliability, as assessed using a beta-binomial model of signal to noise ratio, was 0.0. This measure's reliability was tested using 2010 data from the EHR system of an urban, tertiary-care-level hospital. The measure was calculated for the 57 physicians that had a minimum of 10 deliveries in that year.

A second phase of reliability testing on the measure utilized 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—and are used to calculate a kappa statistic and an agreement rate. (See Measure Testing Protocol for PCPI Performance Measures, ama-assn.org/resources/doc/cqi/pcpi-testing-protocol.pdf.) For analyis of the kappa statistics, we used Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33(1):159-74, where:

 0–.20 = slight agreement
.21–.40 = fair agreement
.41–.60 = moderate agreement
.61–.80 = substantial agreement
.81–1 = almost perfect agreement

For measure 10 as a whole, we were unable to calculate a kappa statistic. For the denominator, there is perfect agreement between the automated report and the manual abstraction. Both data modalities identified the 75 eligible patients. For the numerator, the observed agreement between the two modalities is 0.85. There were 11 patients who were found to have met the measure during manual abstraction, while the automated report did not find any patients that met the measure. To meet the measure, there are three or four separate components that need to be completed. We performed a sensitivity analysis to determine the variability in whether or not each of the components was completed. We calculated kappa statistics and agreement percentages for each individual component.

There are separate ways to evaluate family and contraceptive planning. There could be a discussion about family planning, a discussion about contraceptive planning, or there could be an order for a barrier contraceptive device or a hormonal contraceptive device. When we examined at the family planning component, we were unable to calculate a kappa statistic. The automated report did not find the component for any of the 75 eligible patients, but one patient had the component identified during manual abstraction. The observed agreement between the two data modalities is 0.99.

In addition, the kappa statistic could not be calculated for the contraceptive planning component. The automated report did not find the component for any of the 75 eligible patients, but 64 patients had the component identified during manual abstraction. The observed agreement between the two data modalities is 0.15. We then calculated a kappa statistic to assess measure reliability of the measure components. First we analyzed the contraceptive order component. The kappa statistic results show that there is perfect agreement (1.0). The observed agreement between the two data modalities is 1.00.

When we analyzed the gestational diabetes identification component, we were unable to calculate a kappa statistic. The automated report did not find the component for any of the 75 eligible patients, but two patients had the component identified during manual abstraction. The observed agreement between the two modalities is 0.97.

Next, we analyzed the gestational diabetes screening component to identify whether those individuals with gestational diabetes during pregnancy were screened with a glucose screen. We were unable to calculate a kappa statistic, however. The automated report did not find the component for any of the two eligible patients, but one patient had the component identified during manual abstraction. The observed agreement between the two data modalities is 0.50.

Next, we analyzed the postpartum depression screening. The kappa statistic results find that there is slight agreement (0.04). The observed agreement between the two modalities is 0.45. All of the disagreement comes where there is information found by the manual abstraction but not the automated report. This is suggestive of a discrepancy where providers are coding the information from the screening somewhere in the EHR, but it is not where the measure is expecting the information to be found in the EHR.

When we analyzed at the breastfeeding evaluation component, we were unable to calculate a kappa statistic. The automated report did not find the component for any of the 75 eligible patients, but 41 patients had the component identified during manual abstraction. The observed agreement between the two data modalities is 0.45.

When analyzing the breastfeeding evaluation component, we were unable to calculate a kappa statistic. The automated report did not find the component for any of the 75 eligible patients, but 25 patients had the component identified during manual abstraction. The observed agreement between the two modalities is 0.67.

Performance Rate

The performance rate from the automated report is 0 percent. The performance rate from the manual abstraction is 15 percent (11/75).

Data analyses were conducted using SAS/STAT software, version 8.2 (SAS Institute, Cary, NC).

References

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

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

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

National Quality Forum. NQF removes time-limited endorsement for 21 measures; measures now have endorsed status. Press release. Washington, DC: National Quality Forum; January 18, 2012.
qualityforum.org/News_And_Resources/Press_Releases/2012/NQF_Removes_Time - Limited_Endorsement_for_21_Measures;_Measures_Now_Have_Endorsed_Status.aspx.

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.
ncbi.nlm.nih.gov/pmc/articles/PMC2667340/pdf/nihms-99203.pdf.

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 is 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 are reviewed by the expert workgroup, and the measures are adjusted as needed. Other external review groups (e.g., focus groups) may be convened if there are any remaining concerns related to the content validity of the measures.

The expert panel members also assessed the measure 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 Postpartum Followup and Care Coordination measure the mean score was 4.31; 76.9 percent (10/13) of respondents 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 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 this variable to assess disparities and initiate subsequent quality improvement activities.

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.

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

7.D. Rurality/Urbanicity
Future measure testing and implementation will 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.

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 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 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. A series of analyses were subsequently performed in order to characterize the feasibility, integrity, and face validity of the measures being tested.

Feasibility testing was conducted at two sites: an urban, tertiary care hospital; and an urban, public hospital.

One site reported that all nine data elements are captured by their EHR in code or text format. The second site reported that eight of the nine data elements are captured in their EHR as text. This site reported feasibility as "unknown" for one of the measure data elements (gestational diabetes screening).

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."
- "Nonfeasible. 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."

One of two sites rated the feasibility for the measure as "Nonfeasible. Unable to do today." Feedback provided by this site indicates that changes in documentation may be necessary in order to make the measure feasible. The other site rated the feasibility of the measure as "Feasible with workflow modification/changes to EHR." This site currently captures eight of nine of the data elements as free text results. Some workflow modifications or changes to this site's EHR system may be necessary in order to calculate the measure.

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 in 2012, but there have been limited opportunities 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. The testing results 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)**

No.

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

No.

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

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?**

This information is not available.

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), 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 this important measure of appropriateness can help to reduce c-section rates. Inappropriate c-sections can be harmful to mother and child and are a source of rising medical costs. In addition, the work group 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 IT in the collection and calculation of this measure allow 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 is ongoing 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), originally developed by the National Quality Forum with HHS funding, 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 that have been developed 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?

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 programed to give providers prompts to provide 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. 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

Date First Submitted: August 25, 2012

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

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

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

The signed written statement was submitted.