NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1341 **NQF Project:** Child Health Quality Measures 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Autism Screening
De.2 **Brief description of measure***:* The percentage of children who turned 2 years old during the measurement year who had an autism screening and proper follow up performed between 6 months and 2 years of age.
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** This measure is included in the NCQA composite measure: Comprehensive Well Care for Children by Age 2 Years
De.4 **National Priority Partners Priority Area:** Patient and family engagement, Care coordination, Population health De.5 **IOM Quality Domain**: Effectiveness, Timeliness De.6 **Consumer Care Need:** Staying healthy

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** Proprietary measure A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission A.4 **Measure Steward Agreement attached:**	**A** **Y☐** **N☐**

B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Accountability	**C** Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☐ N☐
(for NQF staff use**)** **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** High resource use, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Autism, or autistic spectrum disorder (ASD), is a developmental disorder. Children with ASD demonstrate deficits in social interaction, verbal and nonverbal communication, and repetitive behaviors or interests. Many ASD children are highly attuned or even painfully sensitive to certain sounds, textures, tastes, and smells, and can be oblivious to extreme cold or pain(NIMH, 2008). Many children with ASD have some degree of mental impairment, and one in four develop seizures (NIMH, 2008). Early intervention can improve long-term outcomes. Estimates of the prevalence of ASD vary widely. The Centers for Disease Control and Prevention's (CDC) Autism and Developmental Disabilities Monitoring Network released data in 2007 that showed about one in 150 eight-year-old children in multiple areas of the U.S. had an ASD (CDC, 2007). The National Institute of Mental Health (NIMH) estimates the prevalence to be one in every 500 children. Younger ages at diagnosis, migration, changes in diagnostic criteria, and inclusion of milder cases is partially responsible; to what extent is not certain. However, according to the NIMH, recent reports suggest that the incidence of autism may be substantially increasing (NIMH, 2008).	**1a** C☐ P☐ M☐ N☐

Comment [KP1]: 1a. The measure focus addresses:
•a specific national health goal/priority identified by NQF's National Priorities Partners; OR
•a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

Each individual with autism accrues about $3.2 million in costs to society over his or her lifetime, with lost productivity and adult care being the most expensive components (Leslie, 2007). In total, autism costs society more than $35 billion in direct and indirect expenses each year.

1a.4 Citations for Evidence of High Impact: American Academy of Pediatrics, Section on Developmental and Behavioral Pediatrics Committee on Coding and Nomenclature. Guidance on reporting developmental screening, testing. AAP News. 2005;26 :34

American Academy of Pediatrics Committee on Children with Disabilities. The Pediatrician's Role in the Diagnosis and Management of Autistic Spectrum Disorder in Children. Pediatrics 2001; 107 No5.

Centers for Disease Control and Prevention. Autism Information Center – Overview. http://www.cdc.gov/ncbddd/autism/overview.htm. Updated 2007.

Douglas L. Leslie, PhD; Andrés Martin, MD, MPH . Health Care Expenditures Associated With Autism Spectrum Disorders. Arch Pediatr Adolesc Med. 2007;161(4):350-355.

National Institute of Mental Health. Autism Spectrum Disorders (Pervasive Developmental Disorders) http://www.nimh.nih.gov/health/publications/autism/complete-publication.shtml. Updated 2008.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Early intervention for autism is important. Although the age of diagnosis has been decreasing in recent years, children still do not receive the proper diagnosis until 3½ to 4 years old (Gupta, 2007). It is estimated that 16 percent of children have some type of developmental and/or behavioral disorder; however, only 30 percent of these cases are identified before a child begins school. This measure will encourage autism screening using a standardized tool.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
From 1999 to 2000, approximately two percent of children under three years of age received early intervention services in the U.S. under the Individuals with Disabilities Education Act Part C, whereas an estimated five percent of preschoolers were served under Part B (DOE, 2001). Moreover, one study found that in a sample of 121 pediatricians, where more than 60 percent reported using a developmental screening test, only 15 to 20 percent of these physicians screened more than ten percent of their patients.

Many clinicians hesitate to discuss the possibility of a diagnosis of autism with parents of young children, even when some symptoms are present, due to concerns about family distress, the possible adverse effects of labeling a child, the possibility of being incorrect, or the hope that the symptoms will reverse over time. However, the positive outcomes of accurate diagnosis may far outweigh the negative effects, and families universally express the desire to be informed as early as possible.

1b.3 Citations for data on performance gap:
Gupta VB, et al. Identifying Children With Autism Early? Pediatrics 2007;119;152-153

Marcus, L. M., & Stone, W. L. (1993). Assessment of the young autistic child. In E. Schopler & G. B. Mesibov (Eds.), Preschool issues in autism? New York: Plenum Press. (From: Pauline A. Filipek et al).

Palfrey JS; Singer JD; Walker DK; Butler JA. Early identification of children´s special needs: a study in five metropolitan communities. J Pediatr 1987 Nov;111(5):651-9.

Pauline A. Filipek, Pasquale J. Accardo, Grace T. Baranek, Edwin H. Cook, Jr., Geraldine Dawson, Barry Gordon, Judith S. Gravel, Chris P. Johnson, Ronald J. Kallen, Susan E. Levy, Nancy J. Minshew, Barry M. Prizant, Isabelle Rapin, Sally J. Rogers, Wendy L.

Smith RD. The use of developmental screening tests by primary-care pediatricians. J Pediatr. 1978; 93:524–527 (From Sices et al).

1b
C☐
P☐
M☐
N☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Stone, Stuart Teplin, Roberto F. Tuchman, and Fred R. Volkmar. The Screening and Diagnosis of Autistic Spectrum Disorders. Journal of Autism and Developmental Disorders Vol. 29, No. 6, 1999

U.S. Department of Education. Twenty-Third Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act. Washington, D.C.; 2001. (From Sices et al)

1b.4 Summary of Data on disparities by population group:
ASD occurs in all racial/ethnic, and socioeconomic groups. Males are more often affected by the disorder, with one in 94 boys diagnosed, and they are four times more likely than females to be diagnosed (ASA, 2008).

ASDs tend to occur more often than expected among people who have certain other medical conditions, including Fragile X syndrome, tuberous sclerosis, congenital rubella syndrome, and untreated phenylketonuria (PKU) (CDC, 2007). Some drugs taken during pregnancy also have been linked with a higher risk of autism, specifically the prescription drug thalidomide (CDC, 2007).

1b.5 Citations for data on Disparities:
Autism Society of America. http://www.autism-society.org/site/PageServer?pagename=about_home. Updated 2008.

Centers for Disease Control and Prevention. Autism Information Center – Overview. http://www.cdc.gov/ncbddd/autism/overview.htm. Updated 2007.

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Early intervention services have been shown to be associated with improved long-term outcomes (AAP, 2001) and an easing of parental anxiety (Gupta et al 2007). According to the AAP, currently accepted strategies are to "improve the overall functional status of the child by enrolling the child in an appropriate and intensive early intervention program that promotes development of communication, social, adaptive, behavioral, and academic skills; decrease maladaptive and repetitive behaviors through use of behavioral and sometimes pharmacologic strategies; and help the family manage the stress associated with raising a child with autism, particularly by providing information about community resources, respite care, and parent support organizations (AAP 2001)."

Although there is growing agreement among experts that early and sustained intensive behavioral and educational interventions may improve overall outcomes, there is less agreement regarding the relative effectiveness of specific intervention strategies or the degree to which they should be delivered (AAP, 2001). Intervention strategies should be tailored to the child´s needs; although the menu of services may vary among children, all children with ASD should be cared for in the context of the medical home (AAP 2001).

1c.2-3. Type of Evidence: Evidence-based guideline, Expert opinion

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Major pediatric health organizations recommend autism screening based on scientific evidence. The American Academy of Pediatrics has recommended administering autism-specific screening tools at the 18-month preventive care visit (in addition to a general developmental screening tool) (Bright futures, 2006). The policy statement recommends surveillance for developmental problems at the 9-, 18-, and 30-month visits, plus screening with an autism-specific tool at the age of 18 months. Screening with an autism-specific screening tool should be repeated at the age of 24 months or at any encounter when a parent raises concerns (Gupta VB, 2007). The American Academy of Neurology recommends that developmental surveillance be performed at all well-child visits from infancy through school-age, and at any age thereafter if concerns are raised about social acceptance, learning, or behavior. The CDC recommends that screening tests used solely for identifying children with developmental disabilities should be given to all children during the 9-month, 18 month, and 24- or 30- month well-child visits.

1c
C☐
P☐
M☐
N☐

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
oAccess – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
oEfficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
Fair to Good

1c.6 Method for rating evidence: Expert Consensus

1c.7 Summary of Controversy/Contradictory Evidence: The USPSTF concluded that the evidence was insufficient to recommend for or against the use of brief, formal screening instruments in primary care to detect speech and language delay in children. However, it is important to note that this recommendation did NOT examine ASD specifically. The USPSTF recommendation statement for speech and language delay and accompanying explanation are below.

The USPSTF concludes that the evidence is insufficient to recommend for or against routine use of brief, formal screening instruments in primary care to detect speech and language delay in children up to 5 years of age.

Speech and language delay affects 5 to 8 percent of preschool children, often persists into the school years, and may be associated with lowered school performance and psychosocial problems. The USPSTF found insufficient evidence that brief, formal screening instruments that are suitable for use in primary care for assessing speech and language development can accurately identify children who would benefit from further evaluation and intervention. Fair evidence suggests that interventions can improve the results of short-term assessments of speech and language skills; however, no studies have assessed long-term outcomes. Furthermore, no studies have assessed any additional benefits that may be gained by treating children identified through brief, formal screening who would not be identified by addressing clinical or parental concerns. No studies have addressed the potential harms of screening or interventions for speech and language delays, such as labeling, parental anxiety, or unnecessary evaluation and intervention. Thus, the USPSTF could not determine the balance of benefits and harms of using brief, formal screening instruments to screen for speech and language delay in the primary care setting.

1c.8 Citations for Evidence (*other than guidelines***):** Council on Children with Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee and Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006; 118: 405-420

Gupta VB, Hyman SL, Johnson CP, et al. Identifying children with autism early? Pediatrics. 2007;119; 152-153

Hagan, JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove, IL: American Academy of Pediatrics

American Academy of Pediatrics. The Pediatrician´s Role in the Diagnosis and Management of Autistic Spectrum Disorder in Children. PEDIATRICS Vol. 107 No. 5 May 2001, pp. 1221-1226

Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. Practice parameter: Screening and diagnosis of autism. December 2008.

Center for Disease Control. Autism Information Center. Screening and Diagnosis. Update April 2008.
U.S. Department of Health & Health Services http://www.ahrq.gov/clinic/uspstf06/speech/speechrs.htm

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
American Academy of Pediatrics (2007): The AAP recommends autism screening at the 18-month and 24-month well-baby examinations. Before 18 months of age, screening tools that evaluate social and communication skills may assist in systematic detection of early signs of ASD.

Common, classic presentations of ASD are lack of speech, scripted speech, parroting without communicative intent, and pop-up and giant words. Earlier prespeech deficits are often present and, if recognized, may allow earlier diagnosis. These deficits may include lack of appropriate gaze or of warm,

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

joyful expressions with gaze; lack of alternating to-and-fro pattern of vocalizations between infant and parent; lack of recognition of parent's voice; disregard for vocalizations (e.g., own name) with keen awareness for environmental sounds; lack of expressions such as "oh-oh" or "huh." Based on a review of these results and his or her own observations, the pediatrician may make a negative or positive determination.

If ASD is not ruled out:
No action is taken when ASD is ruled out, but 3 immediate responses are triggered for positive cases, including: a referral to an autism diagnostic clinic for a definitive evaluation, a prescription for an early intervention program for treatment, and a referral to an audiologist to rule out hearing problems.

Grade: Expert Consensus Policy Statement

American Academy of Neurology (2008): Developmental surveillance should be performed at all well-child visits from infancy through school-age, and at any age thereafter if concerns are raised about social acceptance, learning, or behavior.Screening should be performed not only for autism-related symptoms but also for language delays, learning difficulties, social problems, and anxiety or depressive symptoms.

Recommended developmental screening tools include the Ages and Stages Questionnaire, the BRIGANCE(R) Screens, the Child Development Inventories, and the Parents´ Evaluations of Developmental Status.
Because of the lack of sensitivity and specificity, the Denver-II (DDST-II) and the Revised Denver Pre-Screening Developmental Questionnaire (R-DPDQ) are not recommended for appropriate primary-care developmental surveillance.
Screening specifically for autism should be performed on all children failing routine developmental surveillance procedures using one of the validated instruments: the Checklist for Autism in Toddlers (CHAT) or the Autism Screening Questionnaire.

Further developmental evaluation is required whenever a child fails to meet any milestones (babbling; gesturing; single words by 16 months; two-word spontaneous phrases by 24 months; loss of any language or social skills at any age.
- Siblings of children with autism should be carefully monitored for acquisition of social, communication, and play skills, and the occurrence of maladaptive behaviors.
- Laboratory investigations recommended for any child with developmental delay and/or autism include audiologic assessment and lead screening
- Early referral for a formal audiologic assessment should include behavioral audiometric measures, assessment of middle ear function, and electrophysiologic procedures using experienced pediatric audiologists with current audiologic testing methods and technologies
- Lead screening should be performed in any child with developmental delay and pica.

Additional periodic screening should be considered if the pica persists
Grade: A recommendation for patient management that reflects moderate clinical certainty (usually requires one or more Class II studies or a strong consensus of Class III evidence).

Centers for Disease Control and Prevention (2008): The CDC recommends all children be screened for ASD using Screening tests used solely for identifying children with developmental disabilities should be given to all children during the 9-month, 18-month, and 24- or 30-month well-child visits. Thorough evaluation may include clinical observations, parent interviews, developmental histories, psychological testing, speech and language assessments, and possibly the use of one or more autism diagnostic scales. Because ASDs are complex disorders, a comprehensive evaluation may also include physical, neurological, and genetic testing. Many tools have been designed to assess ASDs in young children, but no single tool should be used as the only basis for diagnosing autism. Diagnostic tools usually rely on two main sources of information—parents' or caregivers' descriptions of their child's development and direct observation of behavior.

If a parent or doctor thinks there could be a problem, there should be a referral to see a developmental pediatrician or other specialist. Parents can also call local early intervention agency (for children under 3) or public school (for children 3 and older).
Grade: Expert Consensus

1c.10 Clinical Practice Guideline Citation: Hagan, JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures:

Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove, IL: American Academy of Pediatrics American Academy of Pediatrics. The Pediatrician´s Role in the Diagnosis and Management of Autistic Spectrum Disorder in Children. PEDIATRICS Vol. 107 No. 5 May 2001, pp. 1221-1226 Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. Practice parameter: Screening and diagnosis of autism. December 2008. Centers for Disease Control and Prevention. Autism Information Center. Screening and Diagnosis. Update April 2008. U.S. Preventive Services Task Force. Screening for speech and language delay in preschool children: recommendation statement. Pediatrics. 2006;117(2):497-501 **1c.11 National Guideline Clearinghouse or other URL:** Assessment, diagnosis and clinical interventions for children and young people with autism spectrum disorders. A national clinical guideline. **1c.12 Rating of strength of recommendation (***also provide narrative description of the rating and by whom***):** AAP: Expert Consensus Policy Statement; AAN: A recommendation; CDC: Expert Consensus **1c.13 Method for rating strength of recommendation** (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:** Expert consensus with evidence review **1c.14 Rationale for using this guideline over others:** NCQA convened a multistakeholder panel of experts to review evidence and guidelines for child health care. The Child Health Measurement Advisory Panel reviewed these guidelines together with the health importance and field test results of this measure. The MAP concluded that the health importance, evidence and feasibility supports this measure.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report*?	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** Children who had documentation in the medical record of an autism screening between 6 months and 2 years of life. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** 2 years **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):**	**2a-specs** **C☐** **P☐** **M☐** **N☐**

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

Documentation must include a note indicating the date of screening and the following.
The type of standardized tool used
A result of normal, abnormal or indeterminate
For abnormal or indeterminate results, evidence of cconfirmatory testing, referral or treatment

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
Children who turned 2 years of age between January 1 of the measurement year and December 31 of the measurement year and who had documentation of a face-to-face visit between the clinician and the child that predates the child's birthday by at least 12 months.

2a.5 Target population gender: Female, Male
2a.6 Target population age range: 6 months to 2 years old

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
1 year

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Children who turned 2 years of age between January 1 of the measurement year and December 31 of the measurement year and who had documentation of a face-to-face visit between the clinician and the child that predates the child's birthday by at least 12 months.

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** None

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
NA

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
None

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
NA

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
Step 1: Determine the denominator
Children who turned age 2 years in the measurement year, AND
Who had a visit that predates the child´s birthday by 12 months
Step 2: Determine the numerator
Children who had documentation in the medical record of an autism screening between 6 months and 2 years of life.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
For this physician-level measure, we anticipate the entire population will be used in the denominator. If a sample is used, a random sample is ideal. NCQA's work has indicated that a sample size of 30-50 patients would be necessary for a typical practice size of 2000 patients.

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Paper medical record/flow-sheet, Electronic clinical data, Electronic Health/Medical Record

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
Medical record

2a.26-28 Data source/data collection instrument reference web page URL or attachment:

2a.29-31 Data dictionary/code table web page URL or attachment:

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Clinicians: Individual, Clinicians: Group, Population: national, Population: regional/network

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*
Ambulatory Care: Office, Ambulatory Care: Clinic, Ambulatory Care: Hospital Outpatient, Behavioral health/psychiatric unit

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Behavioral Health: Mental Health, Clinicians: Nurses, Clinicians: Physicians (MD/DO), Clinicians: PT/OT/Speech

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)***:** NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure).

2b.2 Analytic Method *(type of reliability & rationale, method for testing)***:**
We did not conduct reliability testing for this measure.

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
We did not conduct reliability testing for this measure.

2b
C☐ P☐ M☐ N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)***:** NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure).

2c.2 Analytic Method *(type of validity & rationale, method for testing)***:**
NCQA tested the measure for face validity using a panel of stakeholders with specific expertise in measurement and child health care. This panel included representatives from key stakeholder groups, including pediatricians, family physicians, health plans, state Medicaid agencies and researchers. Experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area. This measure was deemed valid by the expert panel. In addition, this measure does not utilize administrative data sources; data recorded in the chart is considered the gold standard.

For autism screening, the expert panel concluded that the most important aspect of care was whether screening was documented using a scientifically sound standardized instrument and whether or not follow-up of abnormal or indeterminate results were documented in the medical chart.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)***:**
Eligible: 180
Needed and Received Follow Up: 1/1 (100%)

2c
C☐ P☐ M☐ N☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Screening Documented: 39% Results Documented: 38% Standardized Tool Documented: 38% Results and Proper Follow Up Documented: 38%	
2d. Exclusions Justified **2d.1 Summary of Evidence supporting exclusion(s):** Upon reviewing the measure, the expert panel suggested adding an exclusion for children already diagnosed or in treatment. Note, this exclusion is not evidence dependent but rather a specification issue. **2d.2 Citations for Evidence:** NA **2d.3 Data/sample** *(description of data/sample and size)*: NA **2d.4 Analytic Method** *(type analysis & rationale)*: NA **2d.5 Testing Results** *(e.g., frequency, variability, sensitivity analyses)*: NA	**2d** C☐ P☐ M☐ N☐ NA☐
2e. Risk Adjustment for Outcomes/ Resource Use Measures **2e.1 Data/sample** *(description of data/sample and size)*: NA **2e.2 Analytic Method** *(type of risk adjustment, analysis, & rationale)*: NA **2e.3 Testing Results** *(risk model performance metrics)*: NA **2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:** The measure assesses prevention and wellness in a general population; risk adjustment is not indicated.	**2e** C☐ P☐ M☐ N☐ NA☐
2f. Identification of Meaningful Differences in Performance **2f.1 Data/sample from Testing or Current Use** *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure). **2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*: Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance **2f.3 Provide Measure Scores from Testing or Current Use** *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*: Eligible: 180 Needed and Received Follow Up: 1/1 (100%) Screening Documented: 39% Results Documented: 38% Standardized Tool Documented: 38% Results and Proper Follow Up Documented: 38%	**2f** C☐ P☐ M☐ N☐
2g. Comparability of Multiple Data Sources/Methods **2g.1 Data/sample** *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)	**2g** C☐ P☐ M☐ N☐

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
AND
•precisely defined and specified:
–if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;Error! Bookmark not defined. OR ... [1]

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by ra ... [2]

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically ... [3]

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

2g.2 Analytic Method *(type of analysis & rationale)*: This measure is chart review only; no other sources were identified by the expert panel; this measure does not utilize administrative data. **2g.3 Testing Results** *(e.g., correlation statistics, comparison of rankings)*: NA	**NA**☐
2h. Disparities in Care **2h.1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: This measure is not stratified by disparities. **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** NA	**2h** **C**☐ **P**☐ **M**☐ **N**☐ **NA**☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* **C**☐ **P**☐ **M**☐ **N**☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	Eval Rating
3a. Meaningful, Understandable, and Useful Information **3a.1 Current Use:** Not in use but testing completed **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*): This measure is not currently publicly reported. NCQA is exploring the feasibility of adding this measure and its related measures into a physician-level program and/or the HEDIS® measurement set as appropriate. **3a.3 If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years*): This measure is not currently used in QI. NCQA is exploring the feasibility of adding this measure and its related measures into a physician-level program and/or the HEDIS® measurement set as appropriate. NCQA anticipates that after we release these measures, they will become widely used, as all our measures do. **Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) **3a.4 Data/sample** *(description of data/sample and size)*: NA **3a.5 Methods** *(e.g., focus group, survey, QI project)*: NCQA vetted the measures with its expert panel. In addition, throughout the development process, NCQA vetted the measure concepts and specifications with other stakeholder groups, including the National Association of State Medicaid Directors, NCQA's Health Plan Advisory Council, NCQA's Committee on Performance Measurement, and the American Academy of Pediatrician's Quality Improvement Innovation Network. After field testing, NCQA also conducted a debrief call with field test participants. In the form of a group interview, NCQA systematically sought feedback on whether the measures were understandable, feasible, important, and had face validity.	**3a** **C**☐ **P**☐ **M**☐ **N**☐

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

3a.6 Results *(qualitative and/or quantitative results and conclusions)*: NCQA received feedback that the measure is understandable, feasible, important and valid.	
3b/3c. Relation to other NQF-endorsed measures **3b.1 NQF # and Title of similar or related measures:** NA	
(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): **3b.2 Are the measure specifications harmonized? If not, why?** NA	**3b** C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** NA **5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:** NA	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes **4a.1-2 How are the data elements that are needed to compute measure scores generated?** Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition), Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) No **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.** NCQA plans to eventually specify this measure for electronic health records.	**4b** C☐ P☐ M☐ N☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No	**4c** C☐ P☐ M☐ N☐

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

4c.2 **If yes, provide justification.**	**NA**☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences 4d.1 **Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** During the measure development process the Child Health MAP and measure development team worked with NCQA's certified auditors and audit department to ensure that the measure specifications were clear and auditable. The denominator, numerator and optional exclusions are concisely specified and align with our audit standards.	**4d** **C**☐ **P**☐ **M**☐ **N**☐
4e. Data Collection Strategy/Implementation 4e.1 **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** Based on field test results, we have specified the measure to assess whether screening was documented and whether use of a standardized tool was documented. Our field test results showed that these data elements are available in the medical record. In addition, our field test participants noted that many were able to program these requirements into their electronic health record systems, and several implemented point-of-service physician reminders for this measure. 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** Collecting measures from medical charts is time-consuming and can be burdensome. Adapting this measure in electronic health records may relieve some of this burden. 4e.3 **Evidence for costs:** Based on field test participant feedback and other stakeholder input 4e.4 **Business case documentation:** NA	**4e** **C**☐ **P**☐ **M**☐ **N**☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** **C**☐ **P**☐ **M**☐ **N**☐
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited** ☐
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y**☐ **N**☐ **A**☐
CONTACT INFORMATION	
Co.1 **Measure Steward (Intellectual Property Owner)** Co.1 **Organization** National Committee for Quality Assurance, 1100 13th Street NW, Suite 1000, Washington, District Of Columbia, 20005 Co.2 **Point of Contact** Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-	

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

Measure Developer If different from Measure Steward
Co.3 **Organization**
National Committee for Quality Assurance, 1100 13th Street NW, Suite 1000, Washington, District Of Columbia, 20005

Co.4 **Point of Contact**
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-

Co.5 **Submitter If different from Measure Steward POC**
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-, National Committee for Quality Assurance

Co.6 **Additional organizations that sponsored/participated in measure development**

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 **Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.**
Child Health Measurement Advisory Panel:
Jeanne Alicandro
Barbara Dailey
Denise Dougherty, PhD
Ted Ganiats, MD
Foster Gesten, MD
Nikki Highsmith, MPA
Charlie Homer, MD, MPH
Jeff Kamil, MD
Elizabeth Siteman
Mary McIntyre, MD, MPH
Virginia Moyer, MD, MPH, FAAP
Lee Partridge
Xavier Sevilla, MD, FAAP
Michael Siegal
Jessie Sullivan

Ad.2 **If adapted, provide name of original measure:** NA
Ad.3-5 **If adapted, provide original specifications URL or attachment**

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 **Year the measure was first released:**
Ad.7 **Month and Year of most recent revision:**
Ad.8 **What is your frequency for review/update of this measure?**
Ad.9 **When is the next scheduled review/update for this measure?**

Ad.10 **Copyright statement/disclaimers:**

Ad.11 -13 **Additional Information web page URL or attachment:**

Date of Submission (*MM/DD/YY*)**:** 08/30/2010

Page 10: [1] Comment [KP16] **Karen Pace** **10/5/2009 8:59:00 AM**

2e. For outcome measures and other measures (e.g., resource use) when indicated:

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;**Error! Bookmark not defined.** OR

rationale/data support no risk adjustment.

Page 10: [2] Comment [k17] **Karen Pace** **10/5/2009 8:59:00 AM**

13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Page 10: [3] Comment [k19] **Karen Pace** **10/5/2009 8:59:00 AM**

14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1417 **NQF Project:** Child Health Quality Measures 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Screening for hyperbilirubinemia in term and near term neonates
De.2 **Brief description of measure***:* Percentage of newborn infants > 2500g birthweight who receive either serum or transcutaneous bilirubin screening prior to hospital discharge
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure**
De.4 **National Priority Partners Priority Area:** Safety De.5 **IOM Quality Domain**: Safety De.6 **Consumer Care Need:** Staying healthy

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission A.4 **Measure Steward Agreement attached:**	**A** **Y**☐ **N**☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least	**B** **Y**☐

every 3 years. Yes, information provided in contact section	**N**☐
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ▶**Purpose:** Public reporting, Internal quality improvement	**C** **Y**☐ **N**☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** **Y**☐ **N**☐
(for NQF staff use**) Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** **Y**☐ **N**☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:** **Staff Reviewer Name(s):**	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Severity of illness 1a.2 1a.3 **Summary of Evidence of High Impact:** Bilirubin encephalopathy results in major lifelong morbidity and cost and is generally preventable if hyperbilirubinemia is identified and treated in a timely manner. 1a.4 **Citations for Evidence of High Impact:** 1. Bhutani VK et al. Predictive ability of a pre-discharge hour-specific serum bilirubin for subsequent significant hyperbilirubinemia in healthy term and near term newborns. Pediatrics 1999:103:6-14 2. Mah MP et al. Reduction in severe hyperbilirubinemia after institution of predischarge bilirubin screening Pediatrics 2010125 e 1143-8 3. American Acadamy of Pediatrics Clincal Practice Guidelines. Management of hyperbilirubinemia in the newborn infant 35 weeks or more gestation. 2004 4.Eggert LD et al. The effect of instituting a pre-hospital discharge newborn bilirubin screening program in a 16 hospital health system Pediatrics 2006;1176:e855	**1a** **C**☐ **P**☐ **M**☐ **N**☐
1b. Opportunity for Improvement 1b.1 **Benefits (improvements in quality) envisioned by use of this measure:** The AAP has emphasized the difficulty in judging early stages of clinical jaundice from physicial exam alone, particulary in infants of color, and well as the ongoing problem with bilirubin encephalopathy in the term newborn.	**1b** **C**☐ **P**☐ **M**☐ **N**☐

Comment [KP1]: 1a. The measure focus addresses:
•a specific national health goal/priority identified by NQF's National Priorities Partners; OR
•a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
The AAP has emphasized the difficulty in judging early stages of clinical jaundice from physicial exam alone, particulary in infants of color, and well as the ongoing problem with bilirubin encephalopathy in the term newborn.

1b.3 Citations for data on performance gap:
Mah MP et al. Reduction in severe hyperbilirubinemia after institution of predischarge bilirubin screening Pediatrics 2010125 e 1143-8
American Acadamy of Pediatrics Clincal Practice Guidelines. Management of hyperbilirubinemia in the newborn infant 35 weeks or more gestation. 2004

1b.4 Summary of Data on disparities by population group:
The AAP has emphasized the difficulty in assessing clinical jaundice, and that this problem is especially common in newborns of color.

1b.5 Citations for data on Disparities:
American Acadamy of Pediatrics Clincal Practice Guidelines. Management of hyperbilirubinemia in the newborn infant 35 weeks or more gestation. 2004

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Bilirubin encephalopathy does not occur in term and near term infants without significant hyperbilirubinemia. Risk thresholds have been quantitatively defined

1c.2-3. Type of Evidence: Observational study, Evidence-based guideline, Expert opinion

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Prevention of severe hyperbilirubinemia (> 25mg%) will reliably prevent bilirubin encephalopathy in term and near term newborns. Predischarge screening and use of the Bhutani nomogram allows accuate identification, appropriate follow up and early treatment (phototherapy) in infants at risk for pathologic hyperbilirubinemia. Severe hyperbilirubinemia (>25mg%) may be almost entirely prevented by universal predischarge screening

References:
1. Bhutani VK et al. Predictive ability of a pre-discharge hour-specific serum bilirubin for subsequent significant hyperbilirubinemia in healthy term and near term newborns. Pediatrics 1999:103:6-14
2. Mah MP et al. Reduction in severe hyperbilirubinemia after institution of predischarge bilirubin screening Pediatrics 2010125 e 1143-8
3. American Acadamy of Pediatrics Clincal Practice Guidelines. Management of hyperbilirubinemia in the newborn infant 35 weeks or more gestation. 2004
4. Eggert LD et al. The effect of instituting a pre-hospital discharge newborn bilirubin screening program in a 16 hospital health system Pediatrics 2006;1176:e855

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom*)**:**
II

1c.6 Method for rating evidence: II

1c.7 Summary of Controversy/Contradictory Evidence: Hypothetically, a skilled clinician may be able to use physicil observation to detect early jaundice in white infants, thus avoiding the need for acutual bilirubin quantitation. However, while good data exists to document the efficacy of transcutaneous or serum screening, no evidence exists to document the efficacy of clinical observation across broad populations. Further the continued occurence of bilirubin encephalopathy in unscreened term and near term newborns is well documented and suggests the inefficacy of clincial observation among the general

1c
C☐
P☐
M☐
N☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and th ... [1]

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a ... [2]

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

pediatrician population in the U.S. **1c.8 Citations for Evidence (***other than guidelines***):** 1. Bhutani VK et al. Predictive ability of a pre-discharge hour-specific serum bilirubin for subsequent significant hyperbilirubinemia in healthy term and near term newborns. Pediatrics 1999:103:6-14 2. Mah MP et al. Reduction in severe hyperbilirubinemia after institution of predischarge bilirubin screening Pediatrics 2010125 e 1143-8 3. American Acadamy of Pediatrics Clincal Practice Guidelines. Management of hyperbilirubinemia in the newborn infant 35 weeks or more gestation. 2004 4. Eggert LD et al. The effect of instituting a pre-hospital discharge newborn bilirubin screening program in a 16 hospital health system Pediatrics 2006;1176:e855 **1c.9 Quote the Specific guideline recommendation (***including guideline number and/or page number***):** "The best documented method for assessing the risk of subsequent hyperbilirubinemia is to measure the TSB or TcB level and plot the results on a nomogram" AAP (see above citation) **1c.10 Clinical Practice Guideline Citation:** "The best documented method for assessing the risk of subsequent hyperbilirubinemia is to measure the TSB or TcB level and plot the results on a nomogram" AAP (see above citation) **1c.11 National Guideline Clearinghouse or other URL:** na **1c.12 Rating of strength of recommendation (***also provide narrative description of the rating and by whom***):** II **1c.13 Method for rating strength of recommendation** (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:** na **1c.14 Rationale for using this guideline over others:** see above. no NQF metrics currently address this issue.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** Number of neonates with birthweight >2500g who receive either serum or transcutaneous bilirubin screening prior to hospital discharge **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the*	**2a-specs** **C☐** **P☐** **M☐** **N☐**

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

numerator)**:**
Birth to hospital discharge

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions*)**:**
Birth weight > 2500g
Serum or transcutaneous bilirubin test performed

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured*)**:**
All newborns > 2500g

2a.5 Target population gender: Female, Male
2a.6 Target population age range: Neonates

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator*)**:**
Birth to hospital discharge

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions*)**:**
Birth, with birthweight > 2500g

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population*)**:** none

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions*)**:**
na

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions*)**:**
na

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method*)**:**
na

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps*)**:**
Neonates screened/total neonates

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
chi square with Yates correction

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
na

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Electronic administrative data/claims, Lab data

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.*)**:**
na

2a.26-28 Data source/data collection instrument reference web page URL or attachment:

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

2a.29-31 Data dictionary/code table web page URL or attachment:

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency, Population: national

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*
Hospital

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Laboratory

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: Measure has been tested in approximately 1 million infants (see reference Mah et al)over 21 states

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
cohort studies

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Application of such screening eliminated pathologic levels of hyperbilirubinemia in normal term and near term neonates whose caregivers were compliant with recommended care. see references, Mah et al and Eggert et al.

2b C☐ P☐ M☐ N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: Over 1 million infants

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
Cohort observational studies of rates of pathologic hyperbilirubinemia. Both studies, conducted in different, large populations, demonstrated similar results. (see Mah et al and Eggert et al)

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
Validity demonstrated over large and diverse populations, see Mah et al. Universal newborn screening correlates well with subsequent risk of hyperbilirubinemia, see Bhutani et al and Mah et al.

2c C☐ P☐ M☐ N☐

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
none

2d.2 Citations for Evidence:
na

2d.3 Data/sample *(description of data/sample and size)*: na

2d.4 Analytic Method *(type analysis & rationale)*:
na

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
na

2d C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: na

2e C☐ P☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
AND
•precisely defined and specified:
–if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of ca ... [3]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;Error! Bookmark not defined. OR ... [4]

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*: na **2e.3 Testing Results** *(risk model performance metrics)*: na **2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:** no risk adjustment necessary since this measure applied primarily to normal, term and near term newborns.	M☐ N☐ NA☐
2f. Identification of Meaningful Differences in Performance **2f.1 Data/sample from Testing or Current Use** *(description of data/sample and size)*: Testing in approximately 1 million newborns demonstrates ease of assessment of % infants screened. **2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*: Chi square with Yates correction using 2 tailed P values. **2f.3 Provide Measure Scores from Testing or Current Use** *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*: Distribution by % newborns screened suggests rates approaching 100% can be achieved across a large, diverse population	**2f** C☐ P☐ M☐ N☐
2g. Comparability of Multiple Data Sources/Methods **2g.1 Data/sample** *(description of data/sample and size)*: Only 2 different data source exist - serum or transcutaneous assessment. Both have been shown to be equivalent. see Bhutani et al, Mah et al, Eggert et al. Administrative claims data used to collect statistics. **2g.2 Analytic Method** *(type of analysis & rationale)*: Analysis of administrative claims data **2g.3 Testing Results** *(e.g., correlation statistics, comparison of rankings)*: Chi square with Yates correction	**2g** C☐ P☐ M☐ N☐ NA☐
2h. Disparities in Care **2h.1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: na **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** na	**2h** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* C☐ P☐ M☐ N☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	Eval Rating
3a. Meaningful, Understandable, and Useful Information **3a.1 Current Use:** In use	**3a** C☐ P☐

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large) (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** public reporting expected to follow potential NQF approval. **3a.3 If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years*)**:** http://www.hcahealthcare.com/CustomPage.asp?guidCustomContentID=8838FE94-377C-4AE4-BE74-FFA58C708791 **Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) **3a.4 Data/sample** *(description of data/sample and size)***:** A simple % in a large population is easily understood **3a.5 Methods** *(e.g., focus group, survey, QI project)***:** na **3a.6 Results** *(qualitative and/or quantitative results and conclusions)***:** na	**M☐** **N☐**
3b/3c. Relation to other NQF-endorsed measures **3b.1 NQF # and Title of similar or related measures:** none	
(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): **3b.2 Are the measure specifications harmonized? If not, why?** na	**3b** **C☐** **P☐** **M☐** **N☐** **NA☐**
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** na **5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:** na	**3c** **C☐** **P☐** **M☐** **N☐** **NA☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** **C☐** **P☐** **M☐** **N☐**
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes	**4a** **C☐**

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

4a.1-2 How are the data elements that are needed to compute measure scores generated? Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition)	**P☐ M☐ N☐**
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	**4b** **C☐ P☐ M☐ N☐**
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** **C☐ P☐ M☐ N☐ NA☐**
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences **4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** Simplicity of measure (using single lab analysis without exclusions and simple % calculation minimized chance of error.	**4d** **C☐ P☐ M☐ N☐**
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** Data is easily collected electronically. It is being reported quarterly in HCA´s population of 220,000 delivieries annually. No significant difficulties in collection or understanding of data have been encountered. **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** none **4e.3 Evidence for costs:** na **4e.4 Business case documentation:** na	**4e** **C☐ P☐ M☐ N☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** **C☐ P☐ M☐ N☐**
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited ☐**

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

Steering Committee: Do you recommend for endorsement? **Comments:**	**Y☐** **N☐** **A☐**

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
Hospital Corporation of America, 1 Park Plaza, Building 2-W4, Nashville, Tennessee, 37202

Co.2 Point of Contact
Steven, Clark, steven.clark1@hcahealthcare.com, 801-440-1630-

Measure Developer If different from Measure Steward
Co.3 Organization
Hospital Corporation of America, 1 Park Plaza, Building 2-W4, Nashville, Tennessee, 37202

Co.4 Point of Contact
Steven, Clark, steven.clark1@hcahealthcare.com, 801-440-1630-

Co.5 Submitter If different from Measure Steward POC
Steven, Clark, steven.clark1@hcahealthcare.com, 801-440-1630-, Hospital Corporation of America

Co.6 Additional organizations that sponsored/participated in measure development

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
na

Ad.2 If adapted, provide name of original measure: na
Ad.3-5 If adapted, provide original specifications URL or attachment

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 Year the measure was first released: 2006
Ad.7 Month and Year of most recent revision: 01, 2006
Ad.8 What is your frequency for review/update of this measure? annually
Ad.9 When is the next scheduled review/update for this measure? 01, 2010

Ad.10 Copyright statement/disclaimers:

Ad.11 -13 Additional Information web page URL or attachment: URL
http://www.hcahealthcare.com/CustomPage.asp?guidCustomContentID=8838FE94-377C-4AE4-BE74-FFA58C708791

Date of Submission (*MM/DD/YY*)**:** 08/30/2010

Page 3: [1] Comment [k4] **Karen Pace** **10/5/2009 8:59:00 AM**

1c. The measure focus is:

- an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;

OR

- if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
  - Intermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
  - Process – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
    if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
  - Structure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
  - Patient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
  - Access – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
  - Efficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Page 3: [2] Comment [k5] **Karen Pace** **10/5/2009 8:59:00 AM**

4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Page 6: [3] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;

AND

- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

AND

- precisely defined and specified:
  - if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);

if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

Page 6: [4] Comment [KP16] **Karen Pace** **10/5/2009 8:59:00 AM**

2e. For outcome measures and other measures (e.g., resource use) when indicated:

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;**Error! Bookmark not defined.** OR

rationale/data support no risk adjustment.

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1403 **NQF Project:** Child Health Quality Measures 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Newborn Blood Spot Screening
De.2 **Brief description of measure***:* The percentage of children who turned 6 months old during the measurement year who had documentation of a newborn metabolic screening test results by 6 months of age.
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** This measure appears in the composite Comprehensive Well Care by Age 6 Months.
De.4 **National Priority Partners Priority Area:** Care coordination, Population health De.5 **IOM Quality Domain**: Effectiveness, Timeliness De.6 **Consumer Care Need:** Staying healthy

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** Proprietary measure A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission A.4 **Measure Steward Agreement attached:**	**A** Y☐ N☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and	**B**

update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Accountability	C Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	D Y☐ N☐
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:** **Staff Reviewer Name(s):**	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Frequently performed procedure, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Annually an estimated 4.1 million infants are screened for genetic and metabolic disorders. Of these, 4,000 infants are diagnosed with a genetic and metabolic disorder. On average, an additional 1,000 infants have a genetic and metabolic disorders that go undetected. (Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children, 2004). The genetic metabolic diseases are caused either by an abnormality in a person´s genes or by the presence/absence of key proteins whose production is directed by specific genes. The three most common genetic disorders are phenylketonuria (PKU), galactosemia (a sickle-cell disorder) and congenital hypothyroidism. Hyperphenylalaninemia is an abnormal increase in the concentration of the amino acid phenylalanine (Phe) in the blood. When the concentration of Phe is very high (_20 mg/dL or 1210 _mol/L) and there is accumulation of phenylketones, the condition is called classic phenylketonuria (PKU). (National Center for Biotechnology Information. 2006) The reported incidence ranges from 1 in 19 000 to 1 in 13 500 newborn infants. For non-PKU hyperphenylalaninemia, the estimated incidence is 1 in 48 000 newborn infants. (NIH, 2000) PKU is rarely diagnosed before 6 months of age without newborn screening, because the most common manifestation without treatment is developmental delay followed by mental retardation. Untreated	**1a** C☐ P☐ M☐ N☐

Comment [KP1]: 1a. The measure focus addresses:
•a specific national health goal/priority identified by NQF's National Priorities Partners; OR
•a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

individuals may also develop microcephaly, delayed or absent speech, seizures, eczema, and behavioral abnormalities. (Celia I. Kaye, 2006) Galactosemia is an increased concentration of galactose in the blood. The genetic disorders that cause galactosemia vary in severity from a benign condition to a life-threatening disorder of early infancy. Early diagnosis and treatment of the latter condition can be life saving. (Celia I. Kaye, 2006)

Thyroid hormone deficiency at birth is one of the most common treatable causes of mental retardation. There are multiple etiologies of this disorder, both heritable and sporadic, varying in severity. Congenital hypothyroidism (CH) occurs in 1 in 4000 to 1 in 3000 newborns. Programs reporting a higher incidence may include some transient cases. (Celia I. Kaye, 2006)

1a.4 **Citations for Evidence of High Impact:** Overview of NBS Programs: State of the States. Briefing presented at: the first meeting of the Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children; June 7-8, 2004; Washington, DC.

National Center for Biotechnology Information. OMIM: Online Mendelian Inheritance in Man [database]. Available at: www.ncbi.nlm.nih.gov/entrez/query.fcgi?db_OMIM. Accessed March 1, 2006

National Institutes of Health. Consensus Development Conference on Phenylketonuria (PKU): Screening and Management. Bethesda, MD: US Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development; 2000

Celia I. Kaye, MD, PhD, and the Committee on Genetics. American Academy of Pediatrics: Newborn Screening Fact Sheets. 2006 PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

1b. Opportunity for Improvement

1b.1 **Benefits (improvements in quality) envisioned by use of this measure:** Newborn screening is a recognized preventive measure for the early detection of disorders that can cause severe mental retardation, chronic disability or death. Early detection of these abnormalities can prevent morbidity and mortality. The Newborn Screening Authoring Committee (2008) stated that an important goal of newborn screening is to identify infants with treatable congenital conditions before they become symptomatic. Pediatricians and emergency care physicians are often among the first health care professionals to encounter symptomatic infants, so they should be knowledgeable about the newborn screening program, ACT sheets for suspected conditions, and local or regional pediatric medical subspecialists to whom infants can be referred.The state newborn screening program usually can provide information about suspected conditions and expedite the newborn's follow-up confirmatory testing and care.

This measure encourages pediatricians and primary care physicians to ensure results of hospital-based newborn screenings are in the medical chart and to perform needed follow up.

1b.2 **Summary of data demonstrating performance gap (variation or overall poor performance) across providers:**
While infants are screened in the hospital, national recommendations suggest primary care physicians should receive notification of positive newborn screens within 5 to 7 days after testing. Despite this recommendation, one study showed that only slightly more than half received results within 2 weeks; others not at all. The majority of clinicians reported rarely attempting to obtain written copies of screening results if they were not readily available (Oyeku et al., 2010).

In a study focusing on the likelihood of primary care clinician's follow-up of positive newborn screening results for Sickle Cell Disease, nearly 84 percent (71 of 85) reported that they hardly ever attempted to obtain a written copy of newborn screening results when reports were not readily available during a clinic visit. For their patients with positive or abnormal newborn screening results, only 50 percent received results within two weeks of birth (Oyeku et al, 2010).

In addition, overall, clinicians' knowledge of newborn screening management is poor (Oyeku et al, 2010). In 2006, a national survey found that most primary care physicians thought that they were responsible for newborn screening follow-up care. Unfortunately, many felt unprepared to manage follow-up care for a child with a positive newborn screen. For example, nearly 20% of the pediatricians and half of the family

1b
C ☐
P ☐
M ☐
N ☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

physicians reported that they were not competent to discuss PKU (Kemper et al, 2006).

These gaps in coordination of care represent a missed opportunity to treat patients and educate families about these conditions.

1b.3 Citations for data on performance gap:
Kemper, Uren, Moseley & Clark. Primary Care Physicians' Attitudes Regarding Follow-up Care for Children with positive Newborn Screening Results. Pediatrics 2006;118;1836-1841.

Oyeku, Feldman, Ryan, Muret-Wagstaff, Neufeld. Primary Care Clinicians' Knowledge and Confidence About Newborn Screening for Sickle Cell Disease: Randomized Assessment of Educational Strategies. JAMA. VOL. 102, NO. 8, AUGUST 2010.

1b.4 Summary of Data on disparities by population group:
There are large variations in the incidence of PKU by ethnic and cultural groups, with individuals of Northern European ancestry and American Indian/Alaska Native individuals having a higher incidence than black, Hispanic, and Asian individuals.(NIH, 2000)

Congenital hypothyroidism (CH) seems to occur more commonly in Hispanic and American Indian/ Alaska Native people (1 in 2000 to 1 in 700 newborns) and less commonly in black people (1 in 3200 to 1 in 17 000 newborns). Programs report a consistent 2:1 female/male ratio, which is unexplained but speculated to be related to an autoimmune risk factor. (Celia I. Kaye, 2006)

1b.5 Citations for data on Disparities:
National Institutes of Health. Consensus Development Conference on Phenylketonuria (PKU): Screening and Management. Bethesda, MD: US Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development; 2000

Celia I. Kaye, MD, PhD, and the Committee on Genetics. American Academy of Pediatrics: Newborn Screening Fact Sheets. 2006 PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Many metabolic diseases, if detected and treated early, can lead to improved outcomes. For example, early treatment of PKU is associated with improved intellectual outcome. There is an inverse relationship between age at diagnosis of congenital hypothyroidism and neurodevelopmental outcome; the later treatment is started, the lower the IQ will be.

1c.2-3. Type of Evidence: Evidence-based guideline, Expert opinion

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
There is evidence that early detection of metabolic diseases can lead to improved outcomes. Furthermore, comprehensive state newborn screening programs involve more than the initial screening. Diagnosis, follow-up, treatment and evaluation are also vital components to ensure that children with potentially life threatening conditions receive necessary care (Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children, 2004). Children with PKU who are treated appropriately after positive newborn screening have average intelligence as measured by IQ tests; on average their intelligence is slightly lower when compared with parent and sibling IQs. There is an inverse relationship between the age at which treatment is begun and the IQ level, even in PKU that is treated early (Hellekson, 2001). Adolescents and young adults who are treated early and continuously seem to have no increased incidence of psychiatric, emotional, or functional disorders, and there is no increase in problems of self-concept (Landolt, 2002; Sullivan, 2001). With early detection of galactosemia, parents can exclude galactose from their child's diet. The exclusion of galactose can improve the life-threatening complications of classic galactosemia. This treatment has only limited efficacy in the prevention of long-term complications from galactosemia. Complications include impaired cognitive development, with mean IQ in the range of 70 to 90; verbal dyspraxia, a speech disorder attributable to a sensorimotor disturbance of articulation; growth delay,

1c
C☐
P☐
M☐
N☐

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
oAccess – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
oEfficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

with ultimate height in the normal range; neurologic findings, including tremor and ataxia beginning in midchildhood to middle age; and ovarian failure, manifesting as delayed puberty, primary amenorrhea, secondary amenorrhea, or oligomenorrhea. (Berry, 2001) Prepubertal children with GALT deficiency are also at increased risk of having decreased bone mineral density despite normal calcium intake. (Panis, 2004). For congenital hypothyroidism, most newborn screening programs report no difference in global IQ score compared with sibling or classmate controls, whereas some report a reduction in IQ ranging from 6 to 15 points. Recent data suggest that a starting dose of 10 to 15 _g/kg per day normalized serum thyrotropin by 1 month and resulted in a higher IQ as compared with infants started on a lower treatment dose (Salerno, 2002).

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):** Good

1c.6 Method for rating evidence: Expert consensus

1c.7 Summary of Controversy/Contradictory Evidence: There is general agreement that newborn blood spot testing is an important practice. The current national controversy concerning newborn screening involves the discrepancy in the number of genetic screenings mandated by each state. Each state (and the District of Columbia) determines its own list of diseases and methods for screening. All states test for a core group of disorders including PKU, hypothyroidism and galactosemia. However, each state´s mandated newborn screening tests vary tremendously despite identical World Health Organization criteria for disorder screening. State screening laws vary based on disorder prevalence, detectability, treatment availability, outcome and overall cost effectiveness. For instance, North Carolina mandates 32 tests, while Arkansas only screens for four conditions.

However, this measure does not specify which screening tests are done but rather ensures that the results of any screening tests mandated by the state are documented in the medical record and transferred to primary care. The intent of this measure is to assess care coordination.

1c.8 Citations for Evidence (*other than guidelines***):** Berry GT, Leslie N, Reynolds R, Yager CT, Segal S. Evidence for alternate galactose oxidation in a patient with deletion of the galactose-1-phosphate uridyltransferase gene. Mol Genet Metab. 2001;72:316–321

Hagan JF, Shaw Js, Ducan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove Village, IL: American Academy of Pediatrics.

Hellekson KL; National Institutes of Health. NIH consensus statement on phenylketonuria. Am Fam Physician. 2001;63: 1430–1432

Celia I. Kaye, MD, PhD, and the Committee on Genetics. American Academy of Pediatrics: Newborn Screening Fact Sheets. 2006 PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Kilpatrick NM, Awang H, Wilcken B, Christodoulou J. The implication of phenylketonuria on oral health. Pediatr Dent. 1999;21:433–437

Landolt MA, Nuoffer JM, Steinmann B, Superti-Furga A. Quality of life and psychologic adjustment in children and adolescents with early treated phenylketonuria can be normal. J Pediatr. 2002;140:516–521

Newborn Screening Authoring Committee. Newborn Screening Expands: Recommendations for Pediatricians and Medical Homes—Implications for the System. 2008. www.pediatrics.org/cgi/doi/10.1542/ peds.2007-3021
doi:10.1542/peds.2007-3021

Panis B, Forget PP, van Kroonenburgh MJ, et al. Bone metabolism in galactosemia. Bone. 2004;35:982–987

Perez-Duenas B, Valls-Sole J, Fernandez-Alvarez E, et al. Characterization of tremor in phenylketonuric patients. J Neurol. 2005;252:1328–1334

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

Salerno M, Militerni R, Bravaccio C, et al. Effect of different starting doses of levothyroxine on growth and intellectual outcome at four years of age in congenital hypothyroidism. Thyroid. 2002;12:45–52

Sullivan JE. Emotional outcome of adolescents and young adults with early and continuously treated phenylketonuria. J Pediatr Psychol. 2001;26:477–484

Overview of NBS Programs: State of the States. Briefing presented at: the first meeting of the Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children; June 7-8, 2004; Washington, DC.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
Newborn screening programs are state-based, so the number of tests performed, retesting guidelines, and other important issues vary from state to state. All states and U.S. territories screen newborns for phenylketonuria (PKU), hypothyroidism, galactosemia and sickle cell disease.

In 2005, the American Academy of Pediatrics (AAP) endorsed a report from the American College of Medical Genetics (ACMG), which recommended that all states screen newborn infants for a core panel of 29 treatable congenital conditions and an additional 25 conditions that may be detected by screening.

The Secretary of Health and Human Services' Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children (ACHDGDNC)† also adopted that report. Some states are now screening for more than 50 congenital conditions, many of which are rare and unfamiliar to pediatricians and other primary health care professionals. In the foreseeable future, screening programs will likely adopt screening technologies that will further expand the number of conditions screened and tests offered.

In 2004, the Maternal and Child Health Bureau of the Health Resources and Services Administration called on states to adopt a uniform panel of 29 newborn screening tests performed using tandem mass spectrometry, which requires blood from only a single heel-stick.

1c.10 Clinical Practice Guideline Citation: Newborn Screening Authoring Committee. Newborn Screening Expands: Recommendations for Pediatricians
and Medical Homes—Implications for the System. 2008. www.pediatrics.org/cgi/doi/10.1542/ peds.2007-3021
doi:10.1542/peds.2007-3021

http://www.aap.org/healthtopics/newbornscreening.cfm

http://www.aafp.org/online/etc/medialib/aafp_org/documents/policy/state/newborn.Par.0001.File.tmp/stateadv_newbornscreening.pdf

1c.11 National Guideline Clearinghouse or other URL: Follow-up testing for metabolic diseases identified by expanded newborn screening using tandem mass spectrometry.
http://www.guideline.gov/content.aspx?id=14282&search=newborn+screening

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
Good

1c.13 Method for rating strength of recommendation (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:**
State mandates

1c.14 Rationale for using this guideline over others:
This measure is based on the body of guidelines and literature as evaluated by an expert panel.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report*?	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐**

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

	N☐
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** Children who had documentation in the medical record of a newborn blood spot screening and results by age 6 months. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** 2 years **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Documentation must include a note indicating both of the following. • A blood spot or metabolic screening test result of normal, abnormal or indeterminate The blood spot or metabolic test is any test required by the state.	
2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):** Children who turned 6 months of age between January 1 of the measurement year and December 31 of the measurement year and who had documentation of a face-to-face visit between the clinician and the child that predates the child's birthday by at least 6 months. **2a.5 Target population gender:** Female, Male **2a.6 Target population age range:** 0 – 6 months **2a.7 Denominator Time Window (***The time period in which cases are eligible for inclusion in the denominator***):** 1 year **2a.8 Denominator Details (***All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):** See 2a4; chart review only	
2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** None **2a.10 Denominator Exclusion Details (***All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):** NA	
2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):** The measure is not stratified	
2a.12-13 Risk Adjustment Type: No risk adjustment necessary **2a.14 Risk Adjustment Methodology/Variables (***List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**	**2a-specs** **C☐** **P☐** **M☐** **N☐**

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

NA

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
Step 1: Determine the denominator
Children who turned the requisite age in the measurement year, AND
Who had a visit within the past 12 months of the child´s birthday
Step 2: Determine the numerator
Children who had documentation in the medical record of the screening or service during the measurement year or the year previous to the measurement year.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
For this physician-level measure, we anticipate the entire population will be used in the denominator. If a sample is used, a random sample is ideal. NCQA's work has indicated that a sample size of 30-50 patients would be necessary for a typical practice size of 2000 patients.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Paper medical record/flow-sheet, Electronic clinical data, Electronic Health/Medical Record

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
Medical Record

2a.26-28 Data source/data collection instrument reference web page URL or attachment:

2a.29-31 Data dictionary/code table web page URL or attachment:

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Clinicians: Individual, Clinicians: Group, Population: national, Population: regional/network

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*
Ambulatory Care: Office, Ambulatory Care: Clinic, Ambulatory Care: Hospital Outpatient

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)***:** NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2b.2 Analytic Method *(type of reliability & rationale, method for testing)***:**
We did not conduct reliability testing for this measure.

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
NA

2b
C☐
P☐
M☐
N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)***:** NCQA received data from 19 physician practices

2c
C☐
P☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

who submitted 10 records per measure (total 190 records per measure)

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
NCQA tested the measure for face validity using a panel of stakeholders with specific expertise in measurement and child health care. This panel included representatives from key stakeholder groups, including pediatricians, family physicians, health plans, state Medicaid agencies and researchers. Experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
This measure was deemed valid by the expert panel. In addition, this measure does not utilize administrative data sources; data recorded in the chart is considered the gold standard.

M☐ N☐

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
No exclusions

2d.2 Citations for Evidence:
NA

2d.3 Data/sample *(description of data/sample and size)*: NA

2d.4 Analytic Method *(type analysis & rationale)*:
NA

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
NA

2d C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: NA

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
NA

2e.3 Testing Results *(risk model performance metrics)*:
NA

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: The measure assesses prevention and wellness in a general population; risk adjustment is not indicated.

2e C☐ P☐ M☐ N☐ NA☐

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance

2f.3 Provide Measure Scores from Testing or Current Use *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
Elig Population: 180
Performance Rates
Results Documented: 87%

2f C☐ P☐ M☐ N☐

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
AND ... [1]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;Error! Bookmark not defined. OR ... [2]

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by ra ... [3]

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically ... [4]

2g. Comparability of Multiple Data Sources/Methods

2g.1 **Data/sample** *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2g.2 **Analytic Method** *(type of analysis & rationale)*:
This measure is chart review only; no other sources were identified by the expert panel; this measure does not utilize administrative data.

2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)*:
NA

2g
C☐ P☐ M☐ N☐ NA☐

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

2h. Disparities in Care

2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: The measure is not stratified to detect disparities.

2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:**
NA

2h
C☐ P☐ M☐ N☐ NA☐

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*

2

Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?**
Rationale:

2
C☐ P☐ M☐ N☐

3. USABILITY

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)

Eval Rating

3a. Meaningful, Understandable, and Useful Information

3a.1 **Current Use:** Not in use but testing completed

3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*):
This measure is not currently publicly reported. NCQA is exploring the feasibility of adding this measure and its related measures into a physician-level program and/or the HEDIS® measurement set as appropriate

3a.3 **If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years*):
This measure is not currently used in QI. NCQA is exploring the feasibility of adding this measure and its related measures into a physician-level program and/or the HEDIS® measurement set as appropriate. NCQA anticipates that after we release these measures, they will become widely used, as all our measures do.

Testing of Interpretability (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*)

3a.4 **Data/sample** *(description of data/sample and size)*: Expert panel, other stakeholders, and 19 physician field test participants

3a.5 **Methods** *(e.g., focus group, survey, QI project)*:
NCQA vetted the measures with its expert panel. In addition, throughout the development process, NCQA vetted the measure concepts and specifications with other stakeholder groups, including the National Association of State Medicaid Directors, NCQA's Health Plan Advisory Council, NCQA's Committee on

3a
C☐ P☐ M☐ N☐

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

Performance Measurement, and the American Academy of Pediatrician's Quality Improvement Innovation Network. After field testing, NCQA also conducted a debrief call with field test participants. In the form of a group interview, NCQA systematically sought feedback on whether the measures were understandable, feasible, important, and had face validity. 3a.6 **Results** *(qualitative and/or quantitative results and conclusions)*: NCQA received feedback that the measure is understandable, feasible, important and valid.	
3b/3c. Relation to other NQF-endorsed measures 3b.1 **NQF # and Title of similar or related measures:**	
(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): 3b.2 **Are the measure specifications harmonized? If not, why?**	**3b** C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value 3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** 5.1 **If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:** The Centers for Disease Control and Prevention (CDC), the HRSA Maternal and Child Health Bureau (MCHB) and the National Committee for Quality Assurance (NCQA) have submitted 2010 Child Health Quality Measures to NQF that relate to the topic of newborn screening. However the measures target different care settings and data sources. CDC, MCHB, and NCQA are collaborating to ensure the measure specifications have distinctive additive value and are harmonized. Please note this applies to both Newborn Blood Spot Screening (the current measure) as well as NCQA´s Newborn Hearing Screening measure submission.	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes 4a.1-2 **How are the data elements that are needed to compute measure scores generated?** Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition), Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources	**4b** C☐

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

4b.1 Are all the data elements available electronically? (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) No **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.** NCQA plans to eventually adapt this measure for use in electronic health records.	P☐ M☐ N☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences **4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** During the measure development process the Child Health MAP and measure development team worked with NCQA's certified auditors and audit department to ensure that the measure specifications were clear and auditable. The denominator, numerator and any exclusions are concisely specified and align with our audit standards.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** Based on field test results, we have specified the measure to assess whether screening was documented and whether results were also documented in the medical record. Our field test results showed that these data elements are available in the medical record. In addition, our field test participants noted that many were able to program these requirements into their electronic health record systems, and several implemented point-of-service physician reminders for this measure. **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** Collecting measures from medical charts is time-consuming and can be burdensome. Adapting this measure in electronic health records may relieve some of this burden. **4e.3 Evidence for costs:** Based on field test participant feedback and other stakeholder input. **4e.4 Business case documentation:**	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, met? **Rationale:**	**4** C☐ P☐ M☐ N☐
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited** ☐
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y**☐ **N**☐

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

A☐

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
National Committee for Quality Assurance, 1100 13th Street, NW Suite 1000, Washington, District Of Columbia, 20005

Co.2 Point of Contact
Sepheen, Byron, byron@ncqa.org, 202-955-3573-

Measure Developer If different from Measure Steward
Co.3 Organization
National Committee for Quality Assurance, 1100 13th Street, NW Suite 1000, Washington, District Of Columbia, 20005

Co.4 Point of Contact
Sepheen, Byron, byron@ncqa.org, 202-955-3573-

Co.5 Submitter If different from Measure Steward POC
Sepheen, Byron, byron@ncqa.org, 202-955-3573-, National Committee for Quality Assurance

Co.6 Additional organizations that sponsored/participated in measure development

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Child Health Measurement Advisory Panel:
Jeanne Alicandro
Barbara Dailey
Denise Dougherty, PhD
Ted Ganiats, MD
Foster Gesten, MD
Nikki Highsmith, MPA
Charlie Homer, MD, MPH
Jeff Kamil, MD
Elizabeth Siteman
Mary McIntyre, MD, MPH
Virginia Moyer, MD, MPH, FAAP
Lee Partridge
Xavier Sevilla, MD, FAAP
Michael Siegal
Jessie Sullivan

Ad.2 If adapted, provide name of original measure:
Ad.3-5 If adapted, provide original specifications URL or attachment

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 Year the measure was first released:
Ad.7 Month and Year of most recent revision:
Ad.8 What is your frequency for review/update of this measure?
Ad.9 When is the next scheduled review/update for this measure?

Ad.10 Copyright statement/disclaimers: © 2009 by the National Committee for Quality Assurance
1100 13th Street, NW, Suite 1000

Ad.11 -13 Additional Information web page URL or attachment:

Date of Submission (*MM/DD/YY*)**:** 09/02/2010

Page 9: [1] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;

AND

- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

AND

- precisely defined and specified:
  - if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);

if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

Page 9: [2] Comment [KP16] **Karen Pace** **10/5/2009 8:59:00 AM**

2e. For outcome measures and other measures (e.g., resource use) when indicated:

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR

rationale/data support no risk adjustment.

Page 9: [3] Comment [k17] **Karen Pace** **10/5/2009 8:59:00 AM**

13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Page 9: [4] Comment [k19] **Karen Pace** **10/5/2009 8:59:00 AM**

14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1401 **NQF Project:** Child Health Quality Measures 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Maternal Depression Screening
De.2 **Brief description of measure***:* The percentage of children who turned 6 months during the measurement year who had documentation of a maternal depression screening and proper follow-up performed between 0 and 6 months of life.
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** This measure is included in the NCQA composite measure: Comprehensive Well Care for Children by Age 6 Months
De.4 **National Priority Partners Priority Area:** Care coordination, Population health De.5 **IOM Quality Domain**: Effectiveness, Timeliness De.6 **Consumer Care Need:** Staying healthy

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** Proprietary measure A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission A.4 **Measure Steward Agreement attached:**	**A** **Y☐** **N☐**

B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	**C** Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☐ N☐
(for NQF staff use**) Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** High resource use, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Maternal depression is one of the most common perinatal complications; however, the disorder often remains unrecognized, undiagnosed, and untreated (VanLandeghem, 2006). The various maternal depression disorders are defined by the severity of the depression and the timing and length of the episode. Studies report that three to 25 percent of women experience major depression during the year following childbirth (Gaynes BN, 2005 ; Kessler RC, 1994)). Maternal depression should be distinguished from the "baby blues", which is much more common but lasts only a few days and has little effect on functioning (Hagan, JF, 2008). The incidence of depression may be higher in women who already have young children (VanLandeghem, 2006; Gaynes BN, 2005). Maternal depression can greatly affect mothers, their baby, and their family's well-being. Postpartum depression can have lasting effects on a mother's self-esteem and confidence as a mother (Epperson, 1999). 1a.4 **Citations for Evidence of High Impact:** Epperson, C Neill, MD. Postpartum Major Depression: Detection and Treatment. American Family Physician. April 15, 1999.	**1a** C☐ P☐ M☐ N☐

Comment [KP1]: 1a. The measure focus addresses:
- a specific national health goal/priority identified by NQF's National Priorities Partners; OR
- a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

Gaynes BN, G. 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 Publication No. 05-E006-1. Rockville, MD: Agency for Healthcare Research and Quality. February 2005.

Hagan, JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove, IL: American Academy of Pediatrics.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch Gen Psychiatry 1994;51:8-19.

VanLandeghem, Karen, MPH. National Academy for State Health Policy. Financing Strategies for Medicaid Reimbursement of Maternal Depression Screening by Pediatric Providers. April 2006.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: This measure encourages health care providers to screen new mothers for maternal depression. Periodic screening for maternal depression has been recommended and found to be feasible during an infant health supervision visits. Pediatricians have an opportunity to screen and intervene during well child visits.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
Screening is important, as mothers with post-partum depression who are not treated can have symptoms that carry over into the second year post-partum. Mothers that have had post-partum depression are also more likely to have a recurrence with subsequent children. (Epperson, C Neill, 1999).

More than 10% of mothers experience depression six weeks after giving birth, whether it is minor of major post partum depression (PPD). There are clinically and cost-effective treatments available for PPD, but unfortunately less than half of PPD cases are ever diagnosed (Gibson, 2010). Less than 50% of mothers with an infant child are currently being screened for postpartum depression (Gjerdingen, Crow, McGovern, Miner, Center, 2009).

1b.3 Citations for data on performance gap:
Epperson, C Neill, MD. Postpartum Major Depression: Detection and Treatment. American Family Physician. April 15, 1999.

Jennifer Gibson. Screening for Postpartum Depression Not Worth the Time or Money. March 27, 2010.

Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Postpartum Depression Screening at Well-Child Visits: Validity of a 2-Question Screen and the PHQ-9. Annals of Family Medicine 7:63-70 (2009).

1b.4 Summary of Data on disparities by population group:
Risk factors that increase the likelihood of depression include poverty, chronic maternal health conditions, domestic violence, substance abuse, and marital discord. Parents of children with special health care needs also should be closely monitored for depression symptoms (Hagan, JF, 2008).

1b.5 Citations for data on Disparities:
Hagan, JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove, IL: American Academy of Pediatrics.

1b
C☐ P☐ M☐ N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** There is strong evidence that children whose mothers are depressed experience developmental delays, including delays in expressive language development, cognitive skills, and emotional development (Epperson, 1999). Studies have shown a

1c
C☐ P☐ M☐ N☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
oAccess – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
oEfficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., … [1]

possible link between maternal depression and a decreased likelihood that a child's health and environment are safeguarded.

1c.2-3. **Type of Evidence:** Evidence-based guideline, Expert opinion

1c.4 **Summary of Evidence** (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
A number of adverse health concerns could develop in children of mothers suffering from untreated postpartum depression including delayed psychological, cognitive, neurological and motor development. They are also at higher risk of developing habits of avoidance and distressed behavior. Compared with non-depressed mothers, women who have suffered through post partum depression are 3 times more likely to see serious emotional problems in their children and are 10 times more likely to have a poor relationship with the child (Gjerdingen, Yawn, 2007). Timely identification of PPD in mothers could potentially interrupt this cycle, hopefully before damage to mother, child, and family becomes irreparable.

1c.5 **Rating of strength/quality of evidence** (*also provide narrative description of the rating and by whom*)**:**
Fair to good

1c.6 **Method for rating evidence:** Expert consensus

1c.7 **Summary of Controversy/Contradictory Evidence:** There has been evidence of controversy related to whose responsibility it is to provide postpartum depression screenings. While postpartum depression is not always easily identified and oftentimes can be misdiagnosed (i.e. false positives) there is substantial benefit in urging pedestrians to provide postpartum depression screenings for mothers while at their offices. Pediatricians have several options if the mother of one of their patients screens positive for depression. They can refer the mother back to her primary physician or at least help educate her about postpartum depression and its effects on children (Levin, 2007).

1c.8 **Citations for Evidence** (*other than guidelines*)**:** Barbara P. Yawn. Postpartum Depression: Prevalence and Considerations in Screening. February 2010.

U.S. Preventive Services Task Force. Screening for Depression, May 2002.
Hagan, JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove, IL: American Academy of Pediatrics

Gjerdingen DK, Yawn BP. Postpartum Depression Screening: Importance, Methods, Barriers, and Recommendations for Practice. The Journal of the American Board of Family Medicine 20 (3): 280-288 (2007).

Jane Collingwood. The Efficacy of Postpartum Depression Screening. August 17, 2010.

Aaron Levin. Postpartum-Depression Questions Should Be Routine for Pediatricians. Psychiatric News, February 2, 2007,
Volume 42, Number 3, Page 27.

1c.9 **Quote the Specific guideline recommendation** (*including guideline number and/or page number*)**:**
U.S. Preventive Services Task Force (2002)
The USPSTF recommends screening for depression in clinical practices that have systems in place to assure accurate diagnosis, effective treatment, and follow up for the general adult population*
Grade: B Recommendation
*NOTE: General adult population (not specific to mothers of newborns)

Bright Futures (2008)
Health care professionals should screen mothers on the following topics:

Mothers of one week old infants:
Discuss health and depression, family stress, uninvited advice, parent role.
Differentiate between short-term "baby blues" and postpartum depression, and counsel and refer was

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

appropriate:
It may be helpful to advise women that the ´postpartum blues´ are a different entity from depression. The ´blues´, with characteristic tearfulness, anxiety and low mood, are relatively common but are transient, peaking at 3–5 days after birth and resolving by 10–14 days.

Mothers of one month old infants:
Discuss maternal health (postpartum, checkup, depression, substance abuse)

Mothers of two month old children:
Discuss maternal health (maternal postpartum, checkup and resumption of activities, depression)
Grade: Expert Consensus

1c.10 Clinical Practice Guideline Citation: U.S. Preventive Services Task Force. Screening for Depression, May 2002.

Hagan, JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove, IL: American Academy of Pediatrics

1c.11 National Guideline Clearinghouse or other URL:
http://www.guideline.gov/content.aspx?id=12994&search=maternal+depression+screening+and+maternal+depression+and+maternal+depression

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
Fair to good

1c.13 Method for rating strength of recommendation (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:**
USPSTF-based and expert consensus with evidence review

1c.14 Rationale for using this guideline over others:
This measure is based on a review of the body of evidence and guidelines as a whole.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report*?

1

Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?**
Rationale:

1
Y☐
N☐

2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)

Eval Rating

2a. MEASURE SPECIFICATIONS

S.1 Do you have a web page where current detailed measure specifications can be obtained?
S.2 If yes, provide web page URL:

2a. Precisely Specified

2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):**
Children who had documentation in the medical record of a maternal depression screening by age 6 months

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator***):**
2 years

2a-specs
C☐
P☐
M☐
N☐

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions***):**
Documentation must include a note indicating the date and evidence of screening the mother for maternal depression.
or
A note indicating evidence of at least one of the following
– Mother currently in treatment for any behavioral condition
– Mother currently on medication for depression
Note: Evidence of maternal depression screening may come from the child's or mother's medical chart.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
Children who turned 6 months of age between January 1 of the measurement year and December 31 of the measurement year and who had documentation of a face-to-face visit between the clinician and the child that predates the child's birthday by at least 12 months.

2a.5 Target population gender: Female, Male
2a.6 Target population age range: 0-6 months

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
1 year

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
See 2a4; chart review only

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** None

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
None

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
None

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
NA

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
Step 1: Determine the denominator
Children who turned the requisite age in the measurement year, AND
Who had a visit within the past 12 months of the child´s birthday
Step 2: Determine the numerator
Children who had documentation in the medical record of the screening or service during the measurement year or the year previous to the measurement year.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
For this physician-level measure, we anticipate the entire population will be used in the denominator. If a

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

sample is used, a random sample is ideal. NCQA's work has indicated that a sample size of 30-50 patients would be necessary for a typical practice size of 2000 patients.

2a.24 **Data Source (***Check the source(s) for which the measure is specified and tested)*
Paper medical record/flow-sheet, Electronic clinical data, Electronic Health/Medical Record

2a.25 **Data source/data collection instrument (***Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
Medical Record

2a.26-28 **Data source/data collection instrument reference web page URL or attachment:**

2a.29-31 **Data dictionary/code table web page URL or attachment:**

2a.32-35 **Level of Measurement/Analysis** (*Check the level(s) for which the measure is specified and tested)*
Clinicians: Individual, Clinicians: Group, Population: national, Population: regional/network

2a.36-37 **Care Settings (***Check the setting(s) for which the measure is specified and tested)*
Ambulatory Care: Amb Surgery Center, Ambulatory Care: Office, Ambulatory Care: Clinic, Ambulatory Care: Hospital Outpatient, Behavioral health/psychiatric unit

2a.38-41 **Clinical Services** (*Healthcare services being measured, check all that apply*)
Behavioral Health: Mental Health, Clinicians: Nurses, Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 **Data/sample** *(description of data/sample and size)***:** NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2b.2 **Analytic Method** *(type of reliability & rationale, method for testing)***:**
We did not conduct reliability testing for this measure.

2b.3 **Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
NA

2b
C☐
P☐
M☐
N☐

2c. Validity testing

2c.1 **Data/sample** *(description of data/sample and size)***:** NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2c.2 **Analytic Method** *(type of validity & rationale, method for testing)***:**
NCQA tested the measure for face validity using a panel of stakeholders with specific expertise in measurement and child health care. This panel included representatives from key stakeholder groups, including pediatricians, family physicians, health plans, state Medicaid agencies and researchers. Experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)***:**
This measure was deemed valid by the expert panel. In addition, this measure does not utilize administrative data sources; data recorded in the chart is considered the gold standard.

2c
C☐
P☐
M☐
N☐

2d. Exclusions Justified

2d.1 **Summary of Evidence supporting exclusion(s):**
No Exclusions

2d
C☐
P☐
M☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
AND
•precisely defined and specified:
–if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category … [2]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

2d.2 Citations for Evidence:
NA

2d.3 Data/sample *(description of data/sample and size)*: NA

2d.4 Analytic Method *(type analysis & rationale)*:
NA

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
NA

N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: NA

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
NA

2e.3 Testing Results *(risk model performance metrics)*:
NA

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: The measure assesses prevention and wellness in a general population; risk adjustment is not indicated.

2e C☐ P☐ M☐ N☐ NA☐

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance

2f.3 Provide Measure Scores from Testing or Current Use *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
Elig Population: 180
Performance Rate:
Screening documented in the Medical Chart: 30%

2f C☐ P☐ M☐ N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2g.2 Analytic Method *(type of analysis & rationale)*:
This measure is chart review only; no other sources were identified by the expert panel; this measure does not utilize administrative data.

2g.3 Testing Results *(e.g., correlation statistics, comparison of rankings)*:
NA

2g C☐ P☐ M☐ N☐ NA☐

2h. Disparities in Care

2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)*: The measure is not stratified to detect disparities.

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities,

2h C☐ P☐ M☐ N☐ NA☐

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR rationale/data support no risk adjustment.

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

provide follow-up plans:
NA

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*

2

Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?**
Rationale:

2
C☐
P☐
M☐
N☐

3. USABILITY

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)

Eval Rating

3a. Meaningful, Understandable, and Useful Information

3a.1 **Current Use:** Not in use but testing completed

3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:**
This measure is not currently publicly reported. NCQA is exploring the feasibility of adding this measure and its related measures into a physician-level program and/or the HEDIS® measurement set as appropriate.

3a.3 **If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years***):**
This measure is not currently used in QI. NCQA is exploring the feasibility of adding this measure and its related measures into a physician-level program and/or the HEDIS® measurement set as appropriate. NCQA anticipates that after we release these measures, they will become widely used, as all our measures do.

Testing of Interpretability (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*)
3a.4 **Data/sample** *(description of data/sample and size)***:** Expert panel, other stakeholders, and 19 physician field test participants

3a.5 **Methods** *(e.g., focus group, survey, QI project)***:**
NCQA vetted the measures with its expert panel. In addition, throughout the development process, NCQA vetted the measure concepts and specifications with other stakeholder groups, including the National Association of State Medicaid Directors, NCQA's Health Plan Advisory Council, NCQA's Committee on Performance Measurement, and the American Academy of Pediatrician's Quality Improvement Innovation Network.

After field testing, NCQA also conducted a debrief call with field test participants. In the form of a group interview, NCQA systematically sought feedback on whether the measures were understandable, feasible, important, and had face validity.

3a.6 **Results** *(qualitative and/or quantitative results and conclusions)***:**
NCQA received feedback that the measure is understandable, feasible, important and valid.

3a
C☐
P☐
M☐
N☐

3b/3c. Relation to other NQF-endorsed measures

3b.1 **NQF # and Title of similar or related measures:**

(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**

3b. Harmonization

3b

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): **3b.2 Are the measure specifications harmonized? If not, why?**	C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** **5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:** NA	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, met? **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes **4a.1-2 How are the data elements that are needed to compute measure scores generated?** Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition), Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) No **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.** NCQA plans to eventually adapt this measure for use in electronic health records.	**4b** C☐ P☐ M☐ N☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences **4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** During the measure development process the Child Health MAP and measure development team worked with NCQA's certified auditors and audit department to ensure that the measure specifications were clear and auditable. The denominator, numerator and any exclusions are concisely specified and align with our audit standards.	**4d** C☐ P☐ M☐ N☐

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

4e. Data Collection Strategy/Implementation

4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:
Based on field test results, we have specified the measure to assess whether screening was documented and whether use of a standardized tool was documented. Our field test results showed that these data elements are available in the medical record. In addition, our field test participants noted that many were able to program these requirements into their electronic health record systems, and several implemented point-of-service physician reminders for this measure.

4e.2 Costs to implement the measure (*costs of data collection, fees associated with proprietary measures*)**:**
Collecting measures from medical charts is time-consuming and can be burdensome. Adapting this measure in electronic health records may relieve some of this burden.

4e.3 Evidence for costs:
Based on field test participant feedback and other stakeholder input

4e.4 Business case documentation:

4e
C☐ P☐ M☐ N☐

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?

4

Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?**
Rationale:

4
C☐ P☐ M☐ N☐

RECOMMENDATION

(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**

Time-limited ☐

Steering Committee: Do you recommend for endorsement?
Comments:

Y☐ N☐ A☐

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
National Committee for Quality Assurance, 1100 13th Street NW, Suite 1000, Washington, District Of Columbia, 20005

Co.2 Point of Contact
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-

Measure Developer If different from Measure Steward
Co.3 Organization
National Committee for Quality Assurance, 1100 13th Street NW, Suite 1000, Washington, District Of Columbia, 20005

Co.4 Point of Contact
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-

Co.5 Submitter If different from Measure Steward POC
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-, National Committee for Quality Assurance

Co.6 Additional organizations that sponsored/participated in measure development

ADDITIONAL INFORMATION
Workgroup/Expert Panel involved in measure development Ad.1 **Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.** Child Health Measurement Advisory Panel: Jeanne Alicandro Barbara Dailey Denise Dougherty, PhD Ted Ganiats, MD Foster Gesten, MD Nikki Highsmith, MPA Charlie Homer, MD, MPH Jeff Kamil, MD Elizabeth Siteman Mary McIntyre, MD, MPH Virginia Moyer, MD, MPH, FAAP Lee Partridge Xavier Sevilla, MD, FAAP Michael Siegal Jessie Sullivan
Ad.2 **If adapted, provide name of original measure:** Ad.3-5 **If adapted, provide original specifications URL or attachment**
Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 **Year the measure was first released:** Ad.7 **Month and Year of most recent revision:** Ad.8 **What is your frequency for review/update of this measure?** Ad.9 **When is the next scheduled review/update for this measure?**
Ad.10 **Copyright statement/disclaimers:** 1100 13th Street, NW, Suite 1000
Ad.11 -13 **Additional Information web page URL or attachment:**
Date of Submission (*MM/DD/YY*)**:** 09/01/2010

Page 3: [1] Comment [k5] **Karen Pace** **10/5/2009 8:59:00 AM**

4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Page 7: [2] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;

AND

- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

AND

- precisely defined and specified:
  - if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);

if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1399 **NQF Project:** Child Health Quality Measures 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Developmental Screening by 2 Years of Age
De.2 **Brief description of measure***:* The percentage of children who turned 2 years old during the measurement year who had a developmental screening and proper follow-up performed between 6 months and 2 years of age.
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** This measure appears in the composite Comprehensive Well Care by Age 2 Years.
De.4 **National Priority Partners Priority Area:** Population health De.5 **IOM Quality Domain**: Effectiveness, Timeliness De.6 **Consumer Care Need:** Staying healthy

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** Proprietary measure A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission A.4 **Measure Steward Agreement attached:**	**A** Y☐ N☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and	**B**

update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Accountability	C Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	D Y☐ N☐
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** The American Academy of Pediatrics (AAP) defines a developmental delay as a "condition in which a child is not developing and/or achieving skills according to the expected time frame." A child that is developmentally challenged may face many barriers throughout life; these barriers are even more severe if a delay in development is not detected early. Delayed or disordered development can lead to further health and behavior problems, including failure in school and social and emotional problems.(Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee, 2006) Approximately 12 to 18 percent of U.S. children may have a developmental and behavioral problem. However, only about two percent of children from birth to two years old receive the necessary early intervention services.(Hix-Small, Hollie, PhD, et al., 2007) A child who is identified as having a delay in development by the time he starts school and participates in early intervention programs is more likely to graduate high school, hold a job, live independently, and avoid teen pregnancy, delinquency and violent crimes -- representing a saved cost to society of between $30,000 and $100,000 per child.(Glascoe FP, PhD, et al., 2007) Studies have shown that developmental surveillance based on non-standardized clinical judgment and	1a C☐ P☐ M☐ N☐

Comment [KP1]: 1a. The measure focus addresses:
•a specific national health goal/priority identified by NQF's National Priorities Partners; OR
•a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

observation alone does not accurately identify children with delays. Therefore, national recommendations call for routine, standardized screening of children three times in the first three years (at the 9, 18 and 24- or 30-month well-visit).

1a.4 Citations for Evidence of High Impact: Hagan JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children and Adolescent, Third Edition, Elk Grove Village IL. American Academy of Pediatrics.

Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006;118(1):405-420

Hix-Small, Hollie, PhD, et al. Impact of Implementing Developmental Screening at 12 and 24 Months in a Pediatric Practice Pediatrics Vol. 120 No. 2 August 2007, pp. 381-389

Glascoe FP, PhD and Shapiro, HL, MD. Introduction to Developmental and Behavioral Screening. 2007. http://www.dbpeds.org/articles/detail.cfm?TextID=5

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Pediatricians are not usually successful in identifying children with developmental delays without use of a standardized tool (Hix-Small, 2007). This measure will encourage the use of standardized tools for developmental screening, as delineated by guidelines. Children who are identified earlier are more likely to have developmental promotion activities, that can further improve the likihood that they will be able to start school ready to learn. Demonstrated quality improvement activities such as the ABCD Screening Academy have shown that providers can feasibly and sustainably implement standardized screening, and when done so, more children are refereed to Early Intervention and other services and that the kinds and types of referrals performed are more appropriate than was previously done without standardized screening.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
Findings from the National Survey of Children Health show that only 19.5% of children are screened in the first five years of life. Despite the evidence, the use of standardized developmental screening tools is uncommon; only about 20 percent of physicians routinely use developmental screening tests (The Commonwealth Fund, 2008). One study found that pediatricians failed to identify and refer 60 to 80 percent of children with developmental delays in a timely manner. Another study found that 68 percent of children with delays were not detected by pediatricians. Though many significant delays occur before school age, less than 50 percent of children with delays are identified before starting school -- leading to missed opportunities for treatment (Hix-Small, 2007).

1b.3 Citations for data on performance gap:
http://www.nschdata.org

Commonwealth Fund. Quality Matters, May 6 2008.

Hix-Small, Hollie, PhD, et al. Impact of Implementing Developmental Screening at 12 and 24 Months in a Pediatric Practice Pediatrics Vol. 120 No. 2 August 2007, pp. 381-389

Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006;118(1):405-420

The American Academy of Pediatrics, Council on Children With Disabilities, Section on Developmental and Behavioral Pediatrics, Bright Futures Steering Committee, and Medical Home Initiatives for Children With Special Needs. Identifying infants and young children with developmental disorder in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006. 118(1): 405-420.

1b
C☐
P☐
M☐
N☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Bethell, CD, Reuland, C, Halfon, N, Olsen, L, Schor, E., Measuring the Quality of Preventive and Developmental Services for Young Children: National Estimates and Patterns of Clinicians' Performance. Pediatrics. June 2004.

Pinto-martin, J, Dunkle M, Earls M, Fliedner D, Cynthia L. Developmental States of Developmental Screening: Steps to Implementation of a Successful Program. American Journal of Public Health. 95, 11: 1928-1932.

King T., Trandon, D, Macias, M, et al. Implementing developmental screening and referrals: Lessons learned from a national project. Pediatrics, V 125, No 2, Feb 2010.

Sand N, Silverstein M, Glascoe FP, et al. Pediatrician's reported practices regarding developmental screening: do guidelines work? Do they help? Pediatrics 2005; V116 (1): 174-179

Smith RD. The use of developmental screening tests by primary-care pediatricians. J Pediatrics. 1978; 93(3): 524-527.

Zuckerman KE, Boudreau AA, Lipstein EA, Kuhlthau KA, and Perrin JM. Household Language, Parent Developmental Concerns, and Child Risk for Developmental Disorder. Academic Pediatrics. 2009; 9(2): 97-105.

1b.4 Summary of Data on disparities by population group:

Studies suggest income disparities exist for developmental screening. One study found that only 23 percent of low-income children receive recommended preventive and developmental services (Bethell et al, 2002). The Early Intervention Periodic Screening, Diagnosis and Treatment (EPSDT) benefit for Medicaid children includes screening at each visit, however, as of 2007, 28 states were engaged in lawsuits due to a failure to properly deliver this service (Glascoe et al, 2007). Another study found that children most at risk for school difficulty were those whose mothers had less than a high school education, those who came from single-mother families, those who had received public assistance, and those who lived in families in which the primary language was not English (High, 2008)." Specific ally related to screening, the National Survey of Children's Health found that while improvements were needed in increasing screening for all children, significant variations existed in the rates of screening by race-ethnicity and insurance status.

1b.5 Citations for data on Disparities:

Bethell at al. Partnering with parents to promote the healthy development of young children enrolled in Medicaid. New York NY: The commonwealth Fund, 2002.
Glascoe FP, PhD and Shapiro, HL, MD. Introduction to Developmental and Behavioral Screening. 2007. http://www.dbpeds.org/articles/detail.cfm?TextID=5
High, Pamela C. and the Committee on Early Childhood, Adoption, and Dependent Care and Council on School Health. School Readiness. Pediatrics 2008;121;e1008-e1015
http://www.nschdata.org

Pinto-martin, J, Dunkle M, Earls M, Fliedner D, Cynthia L. Developmental States of Developmental Screening: Steps to Implementation of a Successful Program. American Journal of Public Health. 95, 11: 1928-1932.

King T., Trandon, D, Macias, M, et al. Implementing developmental screening and referrals: Lessons learned from a national project. Pediatrics, V 125, No 2, Feb 2010.

Sand N, Silverstein M, Glascoe FP, et al. Pediatrician's reported practices regarding developmental screening: do guidelines work? Do they help? Pediatrics 2005; V116 (1): 174-179

Smith RD. The use of developmental screening tests by primary-care pediatricians. J Pediatrics. 1978; 93(3): 524-527.

Zuckerman KE, Boudreau AA, Lipstein EA, Kuhlthau KA, and Perrin JM. Household Language, Parent Developmental Concerns, and Child Risk for Developmental Disorder. Academic Pediatrics. 2009; 9(2): 97-

105.

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Early identification of developmental disabilities through surveillance and screening can lead to timely evaluation, diagnosis and appropriate treatment, including developmental intervention.

1c.2-3. Type of Evidence: Evidence-based guideline, Expert opinion

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Developmental surveillance should be a component of every preventive care visit. Standardized developmental screening tools should be used when such surveillance identifies concerns about a child´s development. Furthermore, it is recommended that standardized screening for developmental, behavioral and social delays occur at the 9-, 18-, and 24-month OR 30-month well visits.
When a child has a positive screening result for a developmental problem, developmental and medical evaluations to identify the specific developmental disorders and related medical problems are warranted. Children diagnosed with developmental disorders should be identified as children with special health care needs; chronic-condition management for these children should be initiated.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom*)**:**
Good

1c.6 Method for rating evidence: Expert consensus with evidence review

1c.7 Summary of Controversy/Contradictory Evidence: The USPSTF did not review developmental screening generally. Rather, the Task Force reviewed the routine use of brief, formal screening instruments in primary care to detect speech and language delay in children. This recommendation received an "I Statement":

The USPSTF concludes that the evidence is insufficient to recommend for or against routine use of brief, formal screening instruments in primary care to detect speech and language delay in children up to 5 years of age.

Speech and language delay affects 5 to 8 percent of preschool children, often persists into the school years, and may be associated with lowered school performance and psychosocial problems. The USPSTF found insufficient evidence that brief, formal screening instruments that are suitable for use in primary care for assessing speech and language development can accurately identify children who would benefit from further evaluation and intervention. Fair evidence suggests that interventions can improve the results of short-term assessments of speech and language skills; however, no studies have assessed long-term outcomes. Furthermore, no studies have assessed any additional benefits that may be gained by treating children identified through brief, formal screening who would not be identified by addressing clinical or parental concerns. No studies have addressed the potential harms of screening or interventions for speech and language delays, such as labeling, parental anxiety, or unnecessary evaluation and intervention. Thus, the USPSTF could not determine the balance of benefits and harms of using brief, formal screening instruments to screen for speech and language delay in the primary care setting.

Secondly, It is important to note that is measure does not included standardized screening for a specific domain of development (e.g. social emotional screening via the ASQ-SE, autism screening) as it is anchored to recommendations focused on global developmental screening using tools that focus on identifying risk for developmental, behavioral and social delays. National recommendations also call for autism screening at the 18-month and 24-month well-visit and future, separate measures may specified and build off the data collection efforts used for this measure to capture domain-specific screening. Additionally, many of the ABCD states included a distinct focus on complementary, but separate, screening specifically focused on social-emotional development (using tools such as the ASQ-SE). Similarly, future efforts may maximize the data collection efforts for this measure to include additional specifications focused specifically on

1c
C☐
P☐
M☐
N☐

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
oAccess – evidence that an association exists between access to a health service and the outcomes of, or experience with, care. ... [1]

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

social-emotional screening so that a separate measure may be calculated.

1c.8 Citations for Evidence (*other than guidelines***):** Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics 2006 Jul;118(1):405-20.

Hagan JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children and Adolescent, Third Edition, Elk Grove Village IL. American Academy of Pediatrics.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
Institute for Clinical Systems Improvement:
Providers should perform the following on infants: Developmental assessment of: motor skills, language development and social development.
ICSI: Level III

Michigan Quality Improvement Consortium (2007):
From Birth to 24 months, developmental assessments should be performed.
Grade: Consensus and ICSI-Based

American Academy of Pediatrics (2006):
Medical Professionals should use standardized developmental screening tools to screen children and 9 months, 18 months:
• Developmental and medical evaluations to identify the specific developmental disorders and related medical problems
• Referred to early developmental intervention and early childhood services and scheduled for earlier return visits to increase developmental surveillance.
• Identified as children with special health care needs; chronic-condition management for these children should be initiated.
Grade: Consensus and Guideline-Based

Bright Futures (2008):
At 9, 18 and 30 Month Visits, health care providers should perform structured developmental screens. Referral should be made to an appropriate early intervention program or developmental specialist for evaluation.
Grade: Consensus and Guideline-Based

1c.10 Clinical Practice Guideline Citation: Hagan, JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove, IL: American Academy of Pediatrics

Institute for Clinical Systems Improvement. Preventive Services for Children and Adolescents Thirteenth Edition. October 2007
[AAP] Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics 2006 Jul;118(1):405-20.
Michigan Quality Improvement Consortium. Routine preventive services for children and adolescents (ages 2-18). Southfield (MI): Michigan Quality Improvement Consortium; 2007 May. 1 p.

1c.11 National Guideline Clearinghouse or other URL:
http://www.guideline.gov/search/search.aspx?term=developmental+screening

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
Consensus and Guideline-Based

1c.13 Method for rating strength of recommendation (*If different from* USPSTF system*, also describe*

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

rating and how it relates to USPSTF): Expert consensus with evidence review **1c.14 Rationale for using this guideline over others:** NCQA convened a multistakeholder panel of experts to review evidence and guidelines for child health care. The Child Health Measurement Advisory Panel reviewed these guidelines together with the health importance and field test results of this measure. The MAP concluded that the health importance, evidence and feasibility supports this measure. In addition, NCQA collaborated with CAHMI in order to understand the state perspectives regarding implementation of such a measure. States indicated that the measures concept used in this measure and the joint NCQA-CAHMI state-level measure are important, scientifically acceptable and feasible.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report*, **met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** Children who had documentation in the medical record of a screening for risk of developmental, behavioral and social delays by 2 years of age. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** 2 years **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Documentation must include a note indicating the date of screening, the standardized developmental screening tool used, and evidence that tool was completed and scored. Tools must meet the following criteria: . 1) Developmental domains: The following domains must be included in the standardized developmental screening tool: motor, language, cognitive, and social-emotional. 2) Established Reliability: Reliability scores of approximately 0.70 or above. 3)Established Findings Regarding the Validity: •Concurrent validity: This compares screening results with outcomes derived from a reliable and valid diagnostic assessment usually performed 7-10 days after the screening test. The validity coefficient reports the agreement between the two tests (Meisels & Atkins-Burnett, 2005). Predictive validity: This compares the screening results with measures of children's performance obtained 9-12 months later (Meisels & Atkins-Burnett, 2005). Validity scores for the tool must be approximately 0.70 or above. Measures of validity must be conducted on a significant number of children and using an appropriate standardized developmental or social-emotional assessment instrument(s).	**2a-specs** **C☐** **P☐** **M☐** **N☐**

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

4)Established Sensitivity/Specificity: Sensitivity and specificity scores of approximately 0.70 or above.

Current recommended tools that meet these criteria:
Ages and Stages Questionnaire (ASQ) - 2 months–5 years
Battelle Developmental Inventory Screening Tool (BDI-ST) - Birth–95 months
Bayley Infant Neuro-developmental Screen (BINS) - 3 months–2 years
Brigance Screens-II - Birth–90 months
Child Development Inventory (CDI) - 18 months–6 years
Child Development Review-Parent Questionnaire (CDR-PQ) - 18 months–5 years
Infant Development Inventory - Birth–18 months
Parents' Evaluation of Developmental Status (PEDS) - Birth–8 years

Tools NOT Included in This Measure: It is important to note that standardized tools specifically focused on one domain of development [e.g. child's socio-emotional development (ASQ-SE) or autism (M-CHAT)] are not included in the list above as this measure is anchored to recommendations focused on global developmental screening using tools that focus on identifying risk for developmental, behavioral and social delays.

2a.4 **Denominator Statement (***Brief, text description of the denominator - target population being measured***):**
Children who turned 2 years of age between January 1 of the measurement year and December 31 of the measurement year and who had documentation of a face-to-face visit between the clinician and the child that predates the child's birthday by at least 12 months.

2a.5 **Target population gender:** Female, Male
2a.6 **Target population age range:** 6 months to 2 years old

2a.7 **Denominator Time Window (***The time period in which cases are eligible for inclusion in the denominator***):**
1 year

2a.8 **Denominator Details (***All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Children who turned 2 years of age between January 1 of the measurement year and December 31 of the measurement year and who had documentation of a face-to-face visit between the clinician and the child that predates the child's birthday by at least 12 months.

2a.9 **Denominator Exclusions (***Brief text description of exclusions from the target population***):** None

2a.10 **Denominator Exclusion Details (***All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
NA

2a.11 **Stratification Details/Variables (***All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
None

2a.12-13 **Risk Adjustment Type:** No risk adjustment necessary

2a.14 **Risk Adjustment Methodology/Variables (***List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
NA

2a.15-17 **Detailed risk model available Web page URL or attachment:**

2a.18-19 **Type of Score:** Rate/proportion
2a.20 **Interpretation of Score:** Better quality = Higher score
2a.21 **Calculation Algorithm (***Describe the calculation of the measure as a flowchart or series of steps***):**
Step 1: Determine the denominator
Children who turned the 2 years of age in the measurement year, AND

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Who had a visit within the past 12 months of the child´s birthday

Step 2: Determine the numerator
Children who had documentation in the medical record of developmental screening using a standardized tool during the measurement year.
Documentation must include a note indicating the standardized tool that was used, the date of screening and
evidence that the tool was completed and scored.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)*:
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
For this physician-level measure, we anticipate the entire population will be used in the denominator. If a sample is used, a random sample is ideal. NCQA's work has indicated that a sample size of 30-50 patients would be necessary for a typical practice size of 2000 patients.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Paper medical record/flow-sheet, Electronic clinical data, Electronic Health/Medical Record

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
Medical Record

2a.26-28 Data source/data collection instrument reference web page URL or attachment:

2a.29-31 Data dictionary/code table web page URL or attachment:

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Clinicians: Individual, Clinicians: Group, Population: national, Population: regional/network

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*
Ambulatory Care: Office, Ambulatory Care: Clinic, Ambulatory Care: Hospital Outpatient

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: We did not conduct reliability testing for this measure.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
NA

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
NA

2b
C☐
P☐
M☐
N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
NCQA tested the measure for face validity using a panel of stakeholders with specific expertise in

2c
C☐
P☐
M☐
N☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

measurement and child health care. This panel included representatives from key stakeholder groups, including pediatricians, family physicians, health plans, state Medicaid agencies and researchers. Experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
This measure was deemed valid by the expert panel.

In addition, this measure does not utilize administrative data sources; data recorded in the chart is considered the gold standard.

2d. Exclusions Justified

2d.1 **Summary of Evidence supporting exclusion(s):**
No Exclusions

2d.2 **Citations for Evidence:**
NA

2d.3 **Data/sample** *(description of data/sample and size)*: NA

2d.4 **Analytic Method** *(type analysis & rationale)*:
NA

2d.5 **Testing Results** *(e.g., frequency, variability, sensitivity analyses)*:
NA

2d
C☐
P☐
M☐
N☐
NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 **Data/sample** *(description of data/sample and size)*: No risk adjustment

2e.2 **Analytic Method** *(type of risk adjustment, analysis, & rationale)*:
NA

2e.3 **Testing Results** *(risk model performance metrics)*:
NA

2e.4 **If outcome or resource use measure is not risk adjusted, provide rationale:** The measure assesses prevention and wellness in a general population; risk adjustment is not indicated.

2e
C☐
P☐
M☐
N☐
NA☐

2f. Identification of Meaningful Differences in Performance

2f.1 **Data/sample from Testing or Current Use** *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2f.2 **Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*:
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance

2f.3 **Provide Measure Scores from Testing or Current Use** *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
Eligible N: 180
Screening documented: 88%
Results documented: 87%
Standardized tool documented: 71%

2f
C☐
P☐
M☐
N☐

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
AND
•precisely defined and specified:
–if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category ... [2]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR rationale/data support no risk adjustment.

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for a ... [3]

2g. Comparability of Multiple Data Sources/Methods

2g.1 **Data/sample** *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2g.2 **Analytic Method** *(type of analysis & rationale)*:
This measure is chart review only; no other sources were identified by the expert panel; this measure does not utilize administrative data.

2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)*:
NA

2g
C☐ P☐ M☐ N☐ NA☐

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

2h. Disparities in Care

2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: The measure is not stratified to detect disparities.

2h.**2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:**
NA

2h
C☐ P☐ M☐ N☐ NA☐

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*

2

Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?**
Rationale:

2
C☐ P☐ M☐ N☐

3. USABILITY

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)

Eval Rating

3a. Meaningful, Understandable, and Useful Information

3a.1 **Current Use:** Not in use but testing completed

3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*):
This measure is not currently publicly reported. NCQA is exploring the feasibility of adding this measure and its related measures into a physician-level program and/or the HEDIS® measurement set as appropriate.

3a.3 **If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years*):
This measure is not currently used in QI. NCQA is exploring the feasibility of adding this measure and its related measures into a physician-level program and/or the HEDIS® measurement set as appropriate. NCQA anticipates that after we release these measures, they will become widely used, as all our measures do.

Testing of Interpretability (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*)
3a.4 **Data/sample** *(description of data/sample and size)*: Expert panel, 19 physician field test participants, Medicaid State Directors, ABCD Academy

3a.5 **Methods** *(e.g., focus group, survey, QI project)*:
NCQA vetted the measures with its expert panel. In addition, throughout the development process, NCQA vetted the measure concepts and specifications with other stakeholder groups, including the National Association of State Medicaid Directors, NCQA's Health Plan Advisory Council, NCQA's Committee on

3a
C☐ P☐ M☐ N☐

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

Performance Measurement, and the American Academy of Pediatrician's Quality Improvement Innovation Network.

After field testing, NCQA also conducted a debrief call with field test participants. In the form of a group interview, NCQA systematically sought feedback on whether the measures were understandable, feasible, important, and had face validity.

3a.6 **Results** *(qualitative and/or quantitative results and conclusions)*:
NCQA received feedback that the measure is understandable, feasible, important and valid.

3b/3c. Relation to other NQF-endorsed measures

3b.1 **NQF # and Title of similar or related measures:**
11

(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**

3b. Harmonization
If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population):
3b.2 **Are the measure specifications harmonized? If not, why?**
The National Quality Form has endorsed the Promoting Healthy Development Survey (PHDS) [NQF # 0011], which includes a measure of screening for risk of developmental, behavioral and social delays based on family surveys.

In addition, NCQA and CAHMI are jointly submitting a developmental screening measure specified for state-level measurement.

The measure of screening based on the PHDS is complementary, but different from this measure in the following ways:
• Data source: The screening measure in the PHDS is based on parental report
• Denominator: The PHDS sampling is anchored to children who have had 1 or more well-child visit

The state-level measures that we are submitting jointly with CAHMI is harmonized with this measure.

3b
C☐ P☐ M☐ N☐ NA☐

3c. Distinctive or Additive Value
3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:**
This measure complements the state-level measure of screening submitted by NCQA and CAHMI.

5.1 **If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:**
NA

3c
C☐ P☐ M☐ N☐ NA☐

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?

3

Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?**
Rationale:

3
C☐ P☐ M☐ N☐

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)

Eval Rating

4a. Data Generated as a Byproduct of Care Processes

4a.1-2 **How are the data elements that are needed to compute measure scores generated?**
Data generated as byproduct of care processes during care delivery (Data are generated and used by

4a
C☐ P☐ M☐

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition), Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	N☐
4b. Electronic Sources 4b.1 **Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) No 4b.2 **If not, specify the near-term path to achieve electronic capture by most providers.** NCQA plans to eventually adapt this measure for use in electronic health records.	**4b** C☐ P☐ M☐ N☐
4c. Exclusions 4c.1 **Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No 4c.2 **If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences 4d.1 **Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** During the measure development process the Child Health MAP and measure development team worked with NCQA's certified auditors and audit department to ensure that the measure specifications were clear and auditable. The denominator, numerator and any exclusions are concisely specified and align with our audit standards.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation 4e.1 **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** Based on field test results, we have specified the measure to assess whether screening with a standardized tool was documented. Our field test results showed that these data elements are available in the medical record. In addition, our field test participants noted that many were able to program these requirements into their electronic health record systems, and several implemented point-of-service physician reminders for this measure. In working with CAHMI on the state-level measure, we worked to ensure our age ranges and requirements for a standardized tool were consistent in responding to state experiences. 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** Collecting measures from medical charts is time-consuming and can be burdensome. Adapting this measure in electronic health records may relieve some of this burden. 4e.3 **Evidence for costs:** Based on field test participant feedback and other stakeholder input 4e.4 **Business case documentation:**	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** C☐

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

	P☐ M☐ N☐
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	Time-limited ☐
Steering Committee: Do you recommend for endorsement? **Comments:**	Y☐ N☐ A☐

CONTACT INFORMATION

Co.1 **Measure Steward (Intellectual Property Owner)**
Co.1 **Organization**
NCQA, 1100 13th St, NW, Suite 1000, Washington, District Of Columbia, 20005

Co.2 **Point of Contact**
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-

Measure Developer If different from Measure Steward
Co.3 **Organization**
NCQA, 1100 13th St, NW, Suite 1000, Washington, District Of Columbia, 20005

Co.4 **Point of Contact**
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-

Co.5 **Submitter If different from Measure Steward POC**
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-, NCQA

Co.6 **Additional organizations that sponsored/participated in measure development**
None

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 **Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.**
Child Health Measurement Advisory Panel:
Jeanne Alicandro
Barbara Dailey
Denise Dougherty, PhD
Ted Ganiats, MD
Foster Gesten, MD
Nikki Highsmith, MPA
Charlie Homer, MD, MPH
Jeff Kamil, MD
Elizabeth Siteman
Mary McIntyre, MD, MPH
Virginia Moyer, MD, MPH, FAAP
Lee Partridge
Xavier Sevilla, MD, FAAP
Michael Siegal
Jessie Sullivan

The NCQA Child Health MAP advised NCQA during measure development. They evaluated the way staff specified measures, assessed the content validity of measures, and reviewed field test results. As you can see from the list, the MAP consisted of a balanced group of experts, including representatives from pediatricians, family physicians, researchers, Medicaid CHIP offices and health plans.

Note that, in addition to the MAP, we also vetted these measures with a host of other stakeholders, as is our process. Thus, our measures are the result of consensus from a broad and diverse group of stakeholders, in addition to the Child Health MAP.
Ad.2 **If adapted, provide name of original measure:** NA Ad.3-5 **If adapted, provide original specifications URL or attachment**
Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 **Year the measure was first released:** Ad.7 **Month and Year of most recent revision:** Ad.8 **What is your frequency for review/update of this measure?** Ad.9 **When is the next scheduled review/update for this measure?**
Ad.10 **Copyright statement/disclaimers:**
Ad.11 -13 **Additional Information web page URL or attachment:**
Date of Submission (*MM/DD/YY*)**:** 09/24/2010

Page 5: [1] Comment [k4] **Karen Pace** **10/5/2009 8:59:00 AM**

1c. The measure focus is:

- an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;

OR

- if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
  - Intermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
  - Process – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
    if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
  - Structure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
  - Patient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
  - Access – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
  - Efficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Page 10: [2] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;

AND

- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

AND

- precisely defined and specified:
  - if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);

if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

Page 10: [3] Comment [k19] **Karen Pace** **10/5/2009 8:59:00 AM**

14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1397 **NQF Project:** Child Health Quality Measures 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Sudden Infant Death Syndrome Counseling
De.2 **Brief description of measure***:* The percentage of children who tured 6 months old during the measurement year and who had Suddent Infant Death Syndrome (SIDS) counseling and proper follow-up.
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** This measure appears in the composite Comprehensive Well Care by Age 6 Months.
De.4 **National Priority Partners Priority Area:** Patient and family engagement, Population health, Safety De.5 **IOM Quality Domain**: Effectiveness, Safety, Timeliness De.6 **Consumer Care Need:** Staying healthy

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** Proprietary measure A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission A.4 **Measure Steward Agreement attached:**	**A** Y☐ N☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and	**B**

update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Accountability	C Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	D Y☐ N☐
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers, Leading cause of morbidity/mortality, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Sudden Infant Death Syndrome (SIDS) is the most common cause of deaths among infants age one month to one year old; in the U.S. alone, 2,500 infants die from SIDS a year (The Nemours Foundation, 2005; AAP, 2005). The accepted definition of SIDS is "The sudden death of an infant under 1 year of age, which remains unexplained after a thorough case investigation, including performance of a complete autopsy, examination of the death scene, and review of the clinical history" (AAP, 2005). A SIDS death is rare in the first month of life; the occurrence peaks between two and three months of age and continues to decline until it is no longer a threat at age one. Organizations, including the AAP and the SIDS Global Strategy Task Force, concluded the risk of SIDS outweighs any benefits of stomach sleeping (Pollack and Frohna, 2002). 1a.4 **Citations for Evidence of High Impact:** The Nemours Foundation. Sudden Infant Death Syndrome (SIDS). http://kidshealth.org/parent/general/sleep/sids.html. Updated: September 2005. American Academy of Pediatrics, Task Force on Infant Positioning and SIDS. Positioning and SIDS. Pediatrics. 1992;89:1120–1126.	**1a** C☐ P☐ M☐ N☐

Comment [KP1]: 1a. The measure focus addresses:
•a specific national health goal/priority identified by NQF's National Priorities Partners; OR
•a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

American Academy of Pediatrics. Task Force on Sudden Infant Death Syndrome. The Changing Concept of Sudden Infant Death Syndrome: Diagnostic Coding Shifts, Controversies Regarding the Sleeping Environment, and New Variables to Consider in Reducing Risk. Pediatrics Vol. 116 No. 5 November 2005, pp. 1245-1255.

Pollack HA and Frohna JG. Infant Sleep Placement After the Back to Sleep Campaign. Pediatrics Vol. 109(4) April 2002.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: In 1992, the American Academy of Pediatrics (AAP) issued a recommendation that healthy term infants be placed on their backs (supine) to sleep and to avoid the prone sleeping position. This measure encourages health care providers to counsel mothers and caregivers on the importance of placing infants in the supine sleeping position, which could prevent sudden infant death syndrome.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
As a result of the "Back to Sleep" campaign of 1994, parents have been encouraged to place infants in the supine sleep position opposed to the prone sleep position, meaning laying babies on their back instead of their stomachs when putting them down for the night of just a nap. The prone sleep position more than triples the risk of sudden infant death syndrome among infants less than a year old (Kinney, Thach, 2009).

In the ten years spanning from 1992 to 2002, evidence showed the rate in which infants were placed in the prone sleep position decreased by 63.7%, from 75% to 11.3%. Unfortunately, in 2008 this rate had increased to 14.5%. In 2006, the National Center for Health Statistics reported a total of 2,323 SIDS deaths across the nation resulting in a SIDS rate of 0.54 per 1000 live births (Carolan, 2009). This is clear evidence that there is still an important need for counseling parents about the potential dangers of placing their babies in the prone sleep position.

1b.3 Citations for data on performance gap:
Kinney HC, Thach BT. The Sudden Infant Death Syndrome. N Engl J Med 2009; 361:795-805, August 20, 2009.

Patrick L Carolan, MD. Sudden Infant Death Syndrome at http://emedicine.medscape.com/article/1004238-overview. Oct 1, 2009.

1b.4 Summary of Data on disparities by population group:
The rate of SIDS among Hispanic/Latino infants are the lowest compared to white infants and African American infants. African American infants have SIDS incidences twice that of white infants. The AAP (2005) noted that campaigns to eradicate SIDS should especially concentrate on the black and American Indian/Alaska Native populations. Low birth weight infants also have higher incidences of SIDS. (Pollack and Frohna, 2002)
Low socioeconomic status and low educational attainment of mothers also affects adherence to recommendations to place infants in the supine sleeping position. Other risk factors include poor prenatal care, smoking or drinking during pregnancy or after birth, and young age of the mother (The Nemours Foundation, 2005; Hagan, JF, 2008; AAP, 2005).

1b.5 Citations for data on Disparities:
American Academy of Pediatrics. Task Force on Sudden Infant Death Syndrome. The Changing Concept of Sudden Infant Death Syndrome: Diagnostic Coding Shifts, Controversies Regarding the Sleeping Environment, and New Variables to Consider in Reducing Risk. Pediatrics Vol. 116 No. 5 November 2005, pp. 1245-1255

Pollack HA and Frohna, JG. Infant Sleep Placement After the Back to Sleep Campaign. Pediatrics Vol. 109 No. 4 April 2002

The Nemours Foundation. Sudden Infant Death Syndrome (SIDS). http://kidshealth.org/parent/general/sleep/sids.html. Updated: September 2005.

1b
C☐
P☐
M☐
N☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Hagan, JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove, IL: American Academy of Pediatrics.

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Continuous education and counseling of caregivers to place infants on their backs to sleep is associated with positive health outcomes. Since the American Academy of Pediatrics (AAP) released its recommendation in 1992 that infants be placed in a non-prone sleeping position, there has been a major decrease in the incidence of SIDS. According to one study, the SIDS rate for the U.S. was 1.20 deaths per 1000 live births in 1992. In 2001, the SIDS rate was reported at 0.56 deaths per 1000 live births, (Mathews, 2003), representing a decrease of 53 percent over 10 years.

1c.2-3. Type of Evidence: Evidence-based guideline, Expert opinion

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
The AAP identifies any nonprone position (i.e. side or supine) as being optimal for reducing SIDS risk. In 2000, on the basis of new evidence, the AAP advised that placing infants on their backs confers the lowest risk and is the preferred position. However, the risk of side position was reported as less risky than prone, and the AAP advised that if the side position is used, caregivers should be advised to bring the dependent arm forward to lessen the likelihood of the infant rolling to the prone position. The AAP guideline is endorsed by several other organizations, including the National Institute of Child Health and Human Development , the Association of SIDS and Infant Mortality Programs, and the SIDS Alliance.

In large part due to the employment of the supine sleeping position, the incidence of SIDS decreased by 56% in the United Stated from 1992 to 2003 (Coleman-Phox, Odouli, Li, 2008). Generally thought of as common knowledge and practice, placing infants in the supine sleep position is the safest way to lay them down for either naps or bed to prevent SIDS.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom*)**:**
Good

1c.6 Method for rating evidence: Expert Consensus

1c.7 Summary of Controversy/Contradictory Evidence: None

1c.8 Citations for Evidence (*other than guidelines*)**:** Kattwinkel J, Brooks J, Myerberg D; American Academy of Pediatrics, Task Force on Infant Positioning and SIDS. Positioning and SIDS. Pediatrics. 1992;89:1120-1126

Kimberly Coleman-Phox; Roxana Odouli; De-Kun Li. Use of a Fan During Sleep and the Risk of Sudden Infant Death Syndrome. Arch Pediatr Adolesc Med. 2008;162(10):963-968.

Mathews TJ, Menacker F, MacDorman MF. Infant mortality statistics from the 2001 period linked birth/infant death data set. Natl Vital Stat Rep. 2003;52(2):1-28

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number*)**:**
American Academy of Pediatrics (2005) and other organizations*
Infants should be placed for sleep in a supine position (wholly on the back) for every sleep. Side sleeping is not as safe as supine sleeping and is not advised. Also discuss the relationship of SIDS and sleep surface; objects in crib; smoking; location and temperature of sleep environment; bed sharing; pacifier use while sleeping; not using commercial monitors to reduce SIDS; avoid development of positional plagiocephaly ("tummy time, etc")
* The AAP guideline is endorsed by the Association of SIDS and Infant Mortality Programs - Professional Association; National Institute of Child Health and Human Development - Federal Government Agency

1c
C☐
P☐
M☐
N☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
oAccess – evidence that an association exists between access to a health service and the outcomes of, or experience with, care. ... [1]

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

[U.S.]; SIDS Alliance - Professional Association

Bright Futures (2008)
Bright Futures recommends that health care providers provide anticipatory guidance for parents and caregivers of newborns up to age four months. Health care providers should discuss sleep positions and topics of back to sleep, location, and crib safety.

ICSI (2007)
ICSI recommends that parents of infants from birth to two years of age be asked how the child is positioned for sleep. Parents should be informed of the importance of back-sleeping position. Health care providers should also demonstrate the appropriate sleeping position when the patient is under medical care.

- Infants should be placed on their back for sleep. Side sleeping is no longer recognized as an alternative position. Advise about the appropriate sleeping position starting in the newborn nursery
- Infant sleep surfaces should be firm and there should be no loose bedding or soft objects around the infant.
- Parents should be encouraged not to smoke, as this has many important health benefits. Smoking during pregnancy has been shown to be associated with increased risk of SIDS.
- A proximate but separate sleeping environment and the use of pacifiers have been recommended. These should be discussed with parents in the context of fully supporting breastfeeding

Level II (Good Evidence)

Michigan Quality Improvement Consortium (2007)
The Michigan Quality Improvement Consortium recommends parental education and counseling for newborns up to to 24 months of age to place infants on their back to sleep.
B Level of Evidence

1c.10 Clinical Practice Guideline Citation: American Academy of Pediatrics. Task Force on Sudden Infant Death Syndrome. The Changing Concept of Sudden Infant Death Syndrome: Diagnostic Coding Shifts, Controversies Regarding the Sleeping Environment, and New Variables to Consider in Reducing Risk. Pediatrics Vol. 116 No. 5 November 2005, pp. 1245-1255
Hagan, JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove, IL: American Academy of Pediatrics
Institute for Clinical Systems Improvement. Preventive Services for Children and Adolescents Thirteenth Edition. October 2007
Michigan Quality Improvement Consortium. Routine preventive services for infants and children (birth-24 months). Southfield (MI): Michigan Quality Improvement Consortium; 2007 May. 1

1c.11 National Guideline Clearinghouse or other URL:
http://www.guideline.gov/content.aspx?id=15116&search=sids and
http://www.guideline.gov/content.aspx?id=13314&search=sids+infant

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
ICSI Level III Evidence (good)

1c.13 Method for rating strength of recommendation (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:**
Evidence Review

1c.14 Rationale for using this guideline over others:
There is broad guideline support for this measure.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report*?	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)

Eval Rating

2a. MEASURE SPECIFICATIONS

S.1 Do you have a web page where current detailed measure specifications can be obtained?
S.2 If yes, provide web page URL:

2a. Precisely Specified

2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):**
Children who had documentation in the medical record of a SIDS counseing by age 6 months

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator***):**
2 years

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions***):**
Documentation must include a note indicating at least one of the following.
• Engagement in discussion about placing infants on their backs to sleep or the risks of Sudden Infant Death Syndrome (SIDS)
• Checklist indicating that SIDS was addressed
• Counseling or referral for SIDS education
• Member received educational materials on SIDS
• Anticipatory guidance for SIDS

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
Children who turned 6 months of age between January 1 of the measurement year and December 31 of the measurement year and who had documentation of a face-to-face visit between the clinician and the child that predates the child's birthday by at least 12 months.

2a.5 Target population gender: Female, Male
2a.6 Target population age range: 0-6 months

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
1 year

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
See above; chart review only

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** None

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
NA

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
None

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual*

2a-specs
C☐
P☐
M☐
N☐

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

models, statistical models, or other aspects of model or method)**:**
NA

2a.15-17 **Detailed risk model available Web page URL or attachment:**

2a.18-19 **Type of Score:** Rate/proportion
2a.20 **Interpretation of Score:** Better quality = Higher score
2a.21 **Calculation Algorithm (***Describe the calculation of the measure as a flowchart or series of steps***):**
Step 1: Determine the denominator
Children who turned the requisite age in the measurement year, AND
Who had a visit within the past 12 months of the child´s birthday
Step 2: Determine the numerator
Children who had documentation in the medical record of the screening or service during the measurement year or the year previous to the measurement year.

2a.22 **Describe the method for discriminating performance** *(e.g., significance testing)***:**
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance

2a.23 **Sampling (Survey) Methodology** *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
For this physician-level measure, we anticipate the entire population will be used in the denominator. If a sample is used, a random sample is ideal. NCQA's work has indicated that a sample size of 30-50 patients would be necessary for a typical practice size of 2000 patients.

2a.24 **Data Source (***Check the source(s) for which the measure is specified and tested)*
Paper medical record/flow-sheet, Electronic clinical data, Electronic Health/Medical Record

2a.25 **Data source/data collection instrument (***Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
Medical Record

2a.26-28 **Data source/data collection instrument reference web page URL or attachment:**

2a.29-31 **Data dictionary/code table web page URL or attachment:**

2a.32-35 **Level of Measurement/Analysis** (*Check the level(s) for which the measure is specified and tested)*
Clinicians: Individual, Clinicians: Group, Population: national, Population: regional/network

2a.36-37 **Care Settings (***Check the setting(s) for which the measure is specified and tested)*
Ambulatory Care: Office, Ambulatory Care: Clinic, Ambulatory Care: Hospital Outpatient

2a.38-41 **Clinical Services** (*Healthcare services being measured, check all that apply*)
Clinicians: Nurses, Clinicians: PA/NP/Advanced Practice Nurse, Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 **Data/sample** *(description of data/sample and size)***:** NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2b.2 **Analytic Method** *(type of reliability & rationale, method for testing)***:**
We did not conduct reliability testing for this measure.

2b.3 **Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
We did not conduct reliability testing for this measure.

2b
C☐
P☐
M☐
N☐

2c. Validity testing

2c

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

2c.1 Data/sample *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
NCQA tested the measure for face validity using a panel of stakeholders with specific expertise in measurement and child health care. This panel included representatives from key stakeholder groups, including pediatricians, family physicians, health plans, state Medicaid agencies and researchers. Experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area. This measure was deemed valid by the expert panel. In addition, this measure does not utilize administrative data sources; data recorded in the chart is considered the gold standard.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
NA

C☐ P☐ M☐ N☐

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
No exclusions

2d.2 Citations for Evidence:
NA

2d.3 Data/sample *(description of data/sample and size)*: NA

2d.4 Analytic Method *(type analysis & rationale)*:
NA

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
NA

2d
C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: NA

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
NA

2e.3 Testing Results *(risk model performance metrics)*:
NA

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: The measure assesses prevention and wellness in a general population; risk adjustment is not indicated.

2e
C☐ P☐ M☐ N☐ NA☐

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure)

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance

2f.3 Provide Measure Scores from Testing or Current Use *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
Elig Population: 180

2f
C☐ P☐ M☐ N☐

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
AND ... [2]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;Error! Bookmark not defined. OR ... [3]

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by ra ... [4]

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically ... [5]

Documentation of counseling for SIDS: 76%	
2g. Comparability of Multiple Data Sources/Methods 2g.1 **Data/sample** *(description of data/sample and size)*: NCQA received data from 19 physician practices who submitted 10 records per measure (total 190 records per measure) 2g.2 **Analytic Method** *(type of analysis & rationale)*: This measure is chart review only; no other sources were identified by the expert panel; this measure does not utilize administrative data 2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)*: NA	**2g** C☐ P☐ M☐ N☐ NA☐
2h. Disparities in Care 2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: The measure is not stratified to detect disparities. 2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** NA	**2h** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* C☐ P☐ M☐ N☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	Eval Rating
3a. Meaningful, Understandable, and Useful Information 3a.1 **Current Use:** Not in use but testing completed 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*): This measure is not currently publicly reported. NCQA is exploring the feasibility of adding this measure and its related measures into a physician-level program and/or the HEDIS® measurement set as appropriate. 3a.3 **If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years*): This measure is not currently used in QI. NCQA is exploring the feasibility of adding this measure and its related measures into a physician-level program and/or the HEDIS® measurement set as appropriate. NCQA anticipates that after we release these measures, they will become widely used, as all our measures do. **Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) 3a.4 **Data/sample** *(description of data/sample and size)*: NA 3a.5 **Methods** *(e.g., focus group, survey, QI project)*: NCQA vetted the measures with its expert panel. In addition, throughout the development process, NCQA vetted the measure concepts and specifications with other stakeholder groups, including the National	**3a** C☐ P☐ M☐ N☐

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

Association of State Medicaid Directors, NCQA's Health Plan Advisory Council, NCQA's Committee on Performance Measurement, and the American Academy of Pediatrician's Quality Improvement Innovation Network.

After field testing, NCQA also conducted a debrief call with field test participants. In the form of a group interview, NCQA systematically sought feedback on whether the measures were understandable, feasible, important, and had face validity.

3a.6 **Results** *(qualitative and/or quantitative results and conclusions)*:
NCQA received feedback that the measure is understandable, feasible, important and valid.

3b/3c. Relation to other NQF-endorsed measures

3b.1 **NQF # and Title of similar or related measures:**

(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**

3b. Harmonization
If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population):
3b.2 **Are the measure specifications harmonized? If not, why?**

3b
C☐
P☐
M☐
N☐
NA☐

3c. Distinctive or Additive Value
3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:**

5.1 **If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:**
NA

3c
C☐
P☐
M☐
N☐
NA☐

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?

3

Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?**
Rationale:

3
C☐
P☐
M☐
N☐

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)

Eval Rating

4a. Data Generated as a Byproduct of Care Processes

4a.1-2 **How are the data elements that are needed to compute measure scores generated?**
Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition), Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)

4a
C☐
P☐
M☐
N☐

4b. Electronic Sources

4b.1 **Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*)
No

4b.2 **If not, specify the near-term path to achieve electronic capture by most providers.**

4b
C☐
P☐
M☐
N☐

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

NCQA plans to eventually specify this measure for electronic health records.	
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences **4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** During the measure development process the Child Health MAP and measure development team worked with NCQA's certified auditors and audit department to ensure that the measure specifications were clear and auditable. The denominator, numerator and optional exclusions are concisely specified and align with our audit standards.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** Based on field test results, we have specified the measure to assess whether screening was documented and whether use of a standardized tool was documented. Our field test results showed that these data elements are available in the medical record. In addition, our field test participants noted that many were able to program these requirements into their electronic health record systems, and several implemented point-of-service physician reminders for this measure. **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** Collecting measures from medical charts is time-consuming and can be burdensome. Adapting this measure in electronic health records may relieve some of this burden. **4e.3 Evidence for costs:** Based on field test participant feedback and other stakeholder input **4e.4 Business case documentation:**	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, met? **Rationale:**	**4** C☐ P☐ M☐ N☐
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited** ☐
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y**☐ **N**☐ **A**☐
CONTACT INFORMATION	

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
National Committee for Qualtiy Assurance, 1100 13th Street NW, Suite 1000, Washington, District Of Columbia, 20005

Co.2 Point of Contact
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-

Measure Developer If different from Measure Steward
Co.3 Organization
National Committee for Qualtiy Assurance, 1100 13th Street NW, Suite 1000, Washington, District Of Columbia, 20005

Co.4 Point of Contact
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-

Co.5 Submitter If different from Measure Steward POC
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-, National Committee for Qualtiy Assurance

Co.6 Additional organizations that sponsored/participated in measure development

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Child Health Measurement Advisory Panel:
Jeanne Alicandro
Barbara Dailey
Denise Dougherty, PhD
Ted Ganiats, MD
Foster Gesten, MD
Nikki Highsmith, MPA
Charlie Homer, MD, MPH
Jeff Kamil, MD
Elizabeth Siteman
Mary McIntyre, MD, MPH
Virginia Moyer, MD, MPH, FAAP
Lee Partridge
Xavier Sevilla, MD, FAAP
Michael Siegal
Jessie Sullivan

Ad.2 If adapted, provide name of original measure: NA
Ad.3-5 If adapted, provide original specifications URL or attachment

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 Year the measure was first released:
Ad.7 Month and Year of most recent revision:
Ad.8 What is your frequency for review/update of this measure?
Ad.9 When is the next scheduled review/update for this measure?

Ad.10 Copyright statement/disclaimers:

Ad.11 -13 Additional Information web page URL or attachment:

Date of Submission (*MM/DD/YY*)**:** 08/30/2010

Page 4: [1] Comment [k4] **Karen Pace** **10/5/2009 8:59:00 AM**

1c. The measure focus is:

- an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;

OR

- if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
  - Intermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
  - Process – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
    if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
  - Structure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
  - Patient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
  - Access – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
  - Efficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Page 8: [2] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;

AND

- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

AND

- precisely defined and specified:
  - if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);

if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

Page 8: [3] Comment [KP16] **Karen Pace** **10/5/2009 8:59:00 AM**

2e. For outcome measures and other measures (e.g., resource use) when indicated:

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR

rationale/data support no risk adjustment.

Page 8: [4] Comment [k17] **Karen Pace** **10/5/2009 8:59:00 AM**

13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Page 8: [5] Comment [k19] **Karen Pace** **10/5/2009 8:59:00 AM**

14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1391 **NQF Project:** Child Health Quality Measures 2010

MEASURE DESCRIPTIVE INFORMATION

De.1 **Measure Title:** Perinatal Care: Measure 1: Frequency of Ongoing Prenatal Care, Measure 2: Prenatal and Postpartum Care

De.2 **Brief description of measure**: We are combining 2 measures into one form because measure features and evidence are the same or similar.
Measure 1: Frequency of Ongoing Prenatal Care (FPC): The percentage of Medicaid deliveries between November 6 of the year prior to the measurement year and November 5 of the measurement year that received the following number of expected prenatal visits.
•<21 percent of expected visits
•21 percent–40 percent of expected visits
•41 percent–60 percent of expected visits
•61 percent–80 percent of expected visits
•=81 percent of expected visits
This measure uses the same denominator as the Prenatal and Postpartum Care measure.

Measure 2: Prenatal & Postpartum Care (PPC): The percentage of deliveries of live births between November 6 of the year prior to the measurement year and November 5 of the measurement year. For these women, the measure assesses the following facets of prenatal and postpartum care.
• Rate 1: Timeliness of Prenatal Care. The percentage of deliveries that received a prenatal care visit as a member of the organization in the first trimester or within 42 days of enrollment in the organization.
• Rate 2: Postpartum Care. The percentage of deliveries that had a postpartum visit on or between 21 and 56 days after delivery.

1.1-2 **Type of Measure:** Access
De.3 **If included in a composite or paired with another measure, please identify composite or paired measure**
None

De.4 **National Priority Partners Priority Area:** Care coordination, Population health
De.5 **IOM Quality Domain**: Effectiveness, Timeliness

De.6 Consumer Care Need: Staying healthy

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* **A.1 Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes **A.2 Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** Proprietary measure, Proprietary complex measure with fees **A.3 Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission **A.4 Measure Steward Agreement attached:**	**A** **Y☐** **N☐**
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** **Y☐** **N☐**
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	**C** **Y☐** **N☐**
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. **D.1Testing:** Yes, fully developed and tested **D.2 Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** **Y☐** **N☐**
(for NQF staff use**) Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:** **Staff Notes to Reviewers** (*issues or questions regarding any criteria*)**:**	**Met** **Y☐** **N☐**
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, Frequently performed	**1a**

Comment [KP1]: 1a. The measure focus addresses:
- a specific national health goal/priority identified by NQF's National Priorities Partners; OR
- a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

procedure, Leading cause of morbidity/mortality, High resource use, Severity of illness, Patient/societal consequences of poor quality

C☐ P☐ M☐ N☐

1a.2

1a.3 Summary of Evidence of High Impact: Each year, about four million women give birth in the United States. While many women experience normal pregnancies without problems, about one million women have one or more complications during pregnancy, labor and delivery, or postpartum period. Studies indicate that as many as half of all deaths from pregnancy complications could be prevented if women had better access to health care, better quality of care, and changed their health and lifestyle habits (CDC, 2002). Women who receive prenatal care late in their pregnancy or who do not receive any care are at increased risk of bearing infants who are low birth weight, stillborn or who die within the first year of life (National Center for Health Statistics, 2010).

The impact of pregnancy complications on health care costs is considerable. Pregnancy complications before delivery account for more than two million hospital days of care and over one billion dollars each year in the U.S. (CDC, 2002). One driver of excessive maternity costs is premature babies, or babies born before the 37th week. Preterm/low birth weight infants in the United States account for half of infant hospitalization costs and one-quarter of pediatric costs. Costs for these preterm/low birth weight admissions totaled $5.8 billion, representing 47 percent of the costs for all infant hospitalizations and 27 percent of the cost for all pediatric stays. The average cost associated with preterm/low birth weight infant hospital stays is $15,100 and the averagelength of stay is 12.9 days Conversely, for uncomplicated newborns, the averages hospital stay is approximately $600 and 1.9 days. The study lead by Russel found that costs were highest for extremely preterm infants (<28 weeks' gestation/birth weight <1000 g), on average $65,600. Often times the costs are associated with respiratory-related complications. (Rebecca B. Russell, 2007)

1a.4 Citations for Evidence of High Impact: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. Safe Motherhood: Promoting Health for Women Before, During and After Pregnancy. March 2002. www.cdc.gov/nccdphp/drh/mh_ataglance.htm

National Center for Health Statistics, Division of Vital Statistics, DHS; Summary Health Statistic for U.S. Children: National Health Interview Survey, 2009.Vintzileos, A, Ananth, C, Smulian, JC, Scorza, WE, Knuppel, RA. The impact of prenatal care on postneonatal deaths in the presence and absence of antenatal high-risk conditions. Am J Obstet Gynecol November 2002; 187(5):1258-1262.

American Academy of Pediatrics and The American College of Obstetricians and Gynecologists. Guidelines for Perinatal Care (5th Edition). October 2002.

Pennsylvania Health Care Cost Containment Council. Promoting Maternal Health in the Workplace. PHC4 FYI, Issue No. 21. December 2003.

Rebecca B. Russell, Nancy S. Green, Claudia A. Steniner, Susan Meikle. Cost of Hospitalization for Preterm and Low Birth Weight Infants in the United States. PEDIATRICS Vol. 120 No. 1 July 2007, pp. e1-e9 (doi:10.1542/peds.2006-2386)

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Research indicates that early, comprehensive prenatal care and consistent visits throughout pregnancy can promote healthier pregnancies and reduce the risk of costly, adverse birth outcomes. This measure ensures that perinatal care occurs and in a timely and consistent fashion.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:

Despite great national wealth, the U.S. continues to rank poorly relative to other industrialized nations on infant mortality and other birth outcomes, and with wide inequities by race/ethnicity. Across all years, about 60% of community health centers (CHC) mothers received first-trimester prenatal care and more than 70% received postpartum and newborn care.

In a study on the effects of pregnancy and childbirth, mothers´ lack of knowledge about postpartum health

1b
C☐ P☐ M☐ N☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

was the main finding (Kline, Martin, & Deyo, 1998).

In addition, NCQA's HEDIS measure has shown that performance among health plans is low. The rate for timeliness of prenatal care was 81.37% in 2007; and the rate for postpartum care was just 58.6%.

1b.3 Citations for data on performance gap:

Kline C. R, Martin D. P, Deyo R. A. Health consequences of pregnancy and childbirth as perceived by women and clinicians. Obstetrics and Gynecology. 1998;92:842–848.

Gaynes B. N, Gavin N, Meltzer-Brody S, Lohr K. N, Swinson T, Gartlehner G. 2005. Perinatal depression: Prevalence, screening, accuracy, and screening outcomes (AHRQ Publication No. 05-E006-1). et al. Rockville, MD: Agency for Healthcare Research Quality.

1b.4 Summary of Data on disparities by population group:

In the United States, substantial racial/ethnic disparities exist in birth outcomes. As of 2002, the infant mortality rate for blacks (13.5 per 1,000 live births) was more than 2.5 times that of whites (5.7 per 1,000), Hispanics (5.4 per 1,000), and Asians (4.7 per 1,000) (Arias et al. 2003). Black infants were about twice as likely to be delivered low birth weight (LBW) (13.3%) as whites (6.9%) and Hispanics (6.5%); and black infants (17.5%) were more likely to be delivered preterm than either Hispanics (11.6%) or whites (11.0%). Both LBW and preterm birth have been associated with increased risks of infant mortality, and developmental disabilities such as mental retardation and cerebral palsy.

Substantial racial/ethnic disparities also persist in the receipt of prenatal care that has been associated with better birth outcomes (Ickovics et al. 2003;). In 2002, blacks (75%) and Hispanics (77%) were less likely than whites (89%) and Asian/Pacific Islanders (85%) to receive prenatal care in the first trimester (Martin et al. 2003). Similarly, receipt of adequate prenatal care (defined by the Revised-Graduated Index of Prenatal Care Utilization) was reported by 57%of whites and 51% of blacks (Alexander, Kogan, and Nabukera 2002). Despite these differences, other studies have challenged the effectiveness of prenatal care in reducing disparities in birth outcomes due to the strength of other, more difficult to address, factors such as social class and hereditary risks (Lu and Halfon 2003; Lu et al. 2003;).

Maximizing access to prenatal care is a key element of public health strategy to improve the early initiation and appropriate utilization of prenatal care to improve pregnancy outcomes. Utilization of prenatal care is known to vary cross-sectionally by sociodemographic characteristics, notably race/ethnicity, education, age, and marital status (Braverman P,2000).

Contemporary policy thinking about access to health care typically focuses on gaps in health insurance, other economic and transportation barriers, and lack of information as impediments to utilizing care (Frisbie WP, 2001). While some of these factors are persistent over a woman's life, others such as familiarity with prenatal services change in regular or random patterns.
Psychosocial factors may also delay initiation of care, undermine adherence to the standard schedule of visits, or both (Sarnoff R, 2001). For example, women in some sociodemographic groups may be more inclined to find the organization of services to be impersonal or threatening, and the content of services to be unresponsive to their concerns and ordinary mode of life (Pagnini DL, 2000). Some of these attitudinal factors may have a consistent impact on prenatal care throughout the lifetimes of such women. Others may, however, be responses to experience from earlier pregnancies.

1b.5 Citations for data on Disparities:

Alexander,G. R., M. D. Kogan, and S. Nabukera. 2002. ''Racial Differences in Prenatal Care Use in the United States: Are Disparities Decreasing?'' American Journal of Public Health 92 (12): 1970–5.

Arias, E., M. F. MacDorman, D. M. Strobino, and B. Guyer. 2003. ''Annual Summary of Vital Statistics—2002.'' Pediatrics 112 (6, Part 1): 1215–30.

Ickovics, J. R., T. S. Kershaw, C. Westdahl, S. S. Rising, C. Klima, H. Reynolds, and U. Magriples. 2003. ''Group Prenatal Care and Preterm Birth Weight: Results from a Matched Cohort Study at Public Clinics.'' Obstetrics and Gynecology 102 (5, Part 1): 1051–57.

Lu, M. C., and N. Halfon. 2003. ''Racial and Ethnic Disparities in Birth Outcomes: A Life-Course Perspective.'' Maternal and Child Health Journal 7 (1): 13–30.

Martin, J. A., B. E. Hamilton, P. D. Sutton, S. J. Ventura, F. Menacker, and M. L. Munson. 2003. ''Births: Final Data for 2002.'' National Vital Statistics Reports 52 (10): 1–113.

Lu, M. C., V. Tache, G. R. Alexander, M. Kotelchuck, and N. Halfon. 2003. ''Preventing Low Birth Weight: Is Prenatal Care the Answer?'' Journal of Maternal- Fetal and Neonatal Medicine 13 (6): 362–80.

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Proper perinatal care is associated with improved birth outcomes. For example, one study found that 25.6 percent of women who did not receive prenatal care delivered preterm infants compared to 9.2 percent of women who received even a minimum amount of prenatal care (Vintzileos et al., 2002).
In 2001, infants of mothers who received no prenatal care had an infant mortality rate of 34.8 per 1,000 live births, compared to an infant mortality rate of only 6.2 per 1,000 when prenatal care was initiated in the first trimester of pregnancy (Matthews et al., 2003). Observational studies have consistently shown that groups having more post-delivery visits have lower maternal, fetal and neonatal illness and mortality.

Regarding postpartum visits, not only do many women experience some degree of emotional liability in the postpartum period, which warrants a follow-up visit, but they will also need personalized care during this time to hasten the development of a healthy mother-infant relationship and a sense of maternal confidence (ACOG, 2002). Should the pregnancy have an abnormal outcome, the postpartum visit is an advantageous time to discuss implications of such conditions as diabetes mellitus, intrauterine growth restriction, preterm birth, hypertension or other conditions that may recur in any future pregnancies (ACOG, 2002). The postpartum visit is also an ideal time to begin preconceptional counseling for patients who may wish to have future pregnancies.

1c.2-3. Type of Evidence: Evidence-based guideline, Expert opinion

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
The goal of the prenatal contact is to exchange information and identify existing risk factors that may impact the pregnancy (DoD/VA, 2002). Lack of prenatal care can be considered a high-risk factor for postneonatal death. A study that sought to determine the association between prenatal care (defined as one visit) and postneonatal death rates (defined as the number of deaths of infants between 28 and 365 days of life) found that the postneonatal deaths among women who had prenatal care was 2.1 per 1,000 women, whereas the rate among women without prenatal care was 5.9 per 1,000 (Vintzileos, A, et al., 2002). These rates applied to women without high-risk conditions. Women whose prenatal care fails to meet established standards are at a greater risk for pregnancy complications and negative birth outcomes (National Center for Health Statistics, 1997).

The goal of postpartum care is to assess the physical and psychosocial status of the mother after the mother's discharge. The majority of maternal and neonatal deaths, as well as a significant burden of long-term morbidity, occur during the postpartum period (WHO, 1998). The postpartum visit should include obtaining an interval history and performing a physical exam to evaluate the patient's current status. Additionally, the emotional status of a woman whose pregnancy had an abnormal outcome should be reviewed, as many women experience some degree of emotional liability during the postpartum period. It is also an advisable time to begin preconceptional counseling for patients who may wish to have future pregnancies (ACOG Guidelines, 2002).

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
Good

1c.6 Method for rating evidence: Expert Consensus

1c
C☐
P☐
M☐
N☐

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
oAccess – evidence that an association exists between access to a health service and the outcomes of, or experience with, care. ... [1]

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

1c.7 Summary of Controversy/Contradictory Evidence: None

1c.8 Citations for Evidence (*other than guidelines***):** Veterans Health Administration, Department of Defense. DoD/VA clinical practice guideline for the management of uncomplicated pregnancy. Washington (DC): Department of Veterans Affairs; 2002 October.

Department of Reproductive Health and Research (RHR), World Health Organization (WHO). Postpartum care of the mother and newborn: a practical guide; 1998.

Vintzileos, A, Ananth, C, Smulian, JC, Scorza, WE, Knuppel, RA. The impact of prenatal care on postneonatal deaths in the presence and absence of antenatal high-risk conditions. Am J Obstet Gynecol November 2002; 187(5):1258-1262

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
ACOG
Timeliness of Prenatal Care: Once pregnancy occurs, patients should have early contact with an obstetrician to begin counseling about prenatal testing and to develop a management plan. The frequency of subsequent antepartum office visits is determined by the individual needs of the woman and the assessment of her risks. According to ACOG guidelines, for an uncomplicated pregnancy, the ACOG guidelines suggest the following frequency of office visits: monthly office visits from the initial prenatal visit until 29 weeks of pregnancy; weekly office visits from 36 weeks until delivery; office visits every two to three weeks from 29 weeks to 36 weeks of pregnancy. (ACOG guidelines, 2010)
Postpartum Care: The timing of the postpartum visit has been a topic for debate. According to the ACOG, the mother should visit her physician for a postpartum review and examination approximately 4 to 6 weeks after delivery. This interval may be modified according to the needs of the patient with medical, obstetric, or intercurrent complications (ACOG Guidelines, 2002).

The DoD/VA clinical practice guideline for management of uncomplicated pregnancy
Spports the recommended 8 weeks after delivery postpartum visit. Evidence suggests that eight weeks is the optimal time to decrease the rate of false positive cervical smears, though consideration of the mother's schedule should also be taken into account (2002).
A visit within 7-14 days of delivery may be advisable after a cesarean delivery or a complicated gestation, primarily to assess the surgical wounds and healing. The standard postpartum care visit is recommended in follow-up to this initial visit (ACOG Guidelines) to ensure the woman's uterus has reduced to its normal size, and to conduct a depression screening and family planning counseling.

1c.10 Clinical Practice Guideline Citation: American Academy of Pediatrics and The American College of Obstetricians and Gynecologists. Guidelines for Perinatal Care (5th Edition). October 2002.
1c.11 National Guideline Clearinghouse or other URL: Routine prenatal and postnatal care. http://www.guideline.gov/content.aspx?id=13174&search=prenatal+and+postpartum+care

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
Good

1c.13 Method for rating strength of recommendation (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:**
USPSTF

1c.14 Rationale for using this guideline over others:
The measures are access and use of service measures that are based on the body of evidence and guidelines regarding perinatal care.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report*?	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** Measure 1: FPC Received the following number of expected prenatal visits. • <21 percent of expected visits • 21 percent–40 percent of expected visits • 41 percent–60 percent of expected visits • 61 percent–80 percent of expected visits • =81 percent of expected visits Measure 2: PPC Rate 1: Received a prenatal care visit as a member of the organization in the first trimester or within 42 days of enrollment in the organization. Rate 2: Had a postpartum visit on or between 21 and 56 days after delivery. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** 2 years **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Measure 1: FPC Administrative Specification Women who had an unduplicated count of <21 percent, 21 percent–40 percent, 41 percent–60 percent, 61 percent–80 percent or =81 percent of the number of expected visits, adjusted for the month of pregnancy at time of enrollment and gestational age. For each delivery, follow the steps below to calculate each woman's ratio of observed-to-expected prenatal care visits. Medical Record Specification: Women who had an unduplicated count of the number of expected visits that was <21 percent, 21 percent–40 percent, 41 percent–60 percent, 61 percent–80 percent or =81 percent of the number of expected visits, adjusted for the month of pregnancy at time of enrollment and gestational age. The visits may be identified through either administrative data or medical record review. The numerator is calculated retroactively from date of delivery or EDD. Measure 2: PPC Administrative Specification Rate 1 A prenatal visit in the first trimester or within 42 days of enrollment, depending on the date of enrollment in the organization and the gaps in enrollment during the pregnancy. Include only visits that occur while the member was enrolled. Markers for Early Prenatal Care Obtainable From Administrative Data • CPT: 59400*, 59425*, 59426*, 59510*, 59610*, 59618* • CPT Category II: 0500F, 0501F, 0502F Rate 2:	**2a-specs** C☐ P☐ M☐ N☐

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

A postpartum visit (Table PPC-E) to an OB/GYN practitioner or midwife, family practitioner or other PCP for a pelvic exam or postpartum care on or between 21 and 56 days after delivery.
Codes to Identify Postpartum Visits
57170, 58300, 59400*, 59410*, 59430, 59510*, 59515*, 59610*, 59614*, 59618*, 59622*, 88141-88143, 88147, 88148, 88150, 88152-88155, 88164-88167, 88174, 88175 , 99501
0503F
G0101, G0123, G0124, G0141, G0143-G0145, G0147, G0148, P3000, P3001, Q0091
V24.1, V24.2, V25.1, V72.3, V76.2
89.26, 91.46
0923
10524-7, 18500-9, 19762-4, 19764-0, 19765-7, 19766-5, 19774-9, 33717-0, 47527-7, 47528-5

Medical Record Specification
Rate 1:
Prenatal care visit to an OB/GYN practitioner or midwife, family practitioner or other PCP. For visits to a family practitioner or PCP, a diagnosis of pregnancy must be present. Documentation in the medical record must include a note indicating the date when the prenatal care visit occurred, and evidence of one of the following.
- A basic physical obstetrical examination that includes auscultation for fetal heart tone, or pelvic exam with obstetric observations, or measurement of fundus height (a standardized prenatal flow sheet may be used)
- Evidence that a prenatal care procedure was performed, such as:
  - Screening test in the form of an obstetric panel (e.g., hematocrit, differential WBC count, platelet count, hepatitis B surface antigen, rubella antibody, syphilis test, RBC antibody screen, Rh[D] and ABO blood typing), or
  - TORCH antibody panel alone or
  - A rubella antibody test/titer with an Rh incompatibility (ABO/Rh) blood typing, or
  - Echography of a pregnant uterus
- Documentation of LMP or EDD in conjunction with either of the following.
  - Prenatal risk assessment and counseling/education, or
  - Complete obstetrical history

Note: For members whose last enrollment segment was after 219 days prior to delivery (i.e., between 219 days prior to delivery and the day of delivery), count documentation of a visit to an OB/GYN, family practitioner or other PCP with a principal diagnosis of pregnancy.

Rate 2:
Postpartum visit to an OB/GYN practitioner or midwife, family practitioner or other PCP on or between 21 and 56 days after delivery. Documentation in the medical record must include a note indicating the date when a postpartum visit occurred and one of the following.
- Pelvic exam, or
- Evaluation of weight, BP, breasts and abdomen, or
  - Notation of "breastfeeding" is acceptable for the "evaluation of breasts" component
- Notation of postpartum care, including but not limited to the following:
  - Notation of "postpartum care," "PP care," "PP check," "6-week check"
  - A preprinted "Postpartum Care" form in which information was documented during the visit.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**

2a.5 Target population gender: Female
2a.6 Target population age range: Women of childbearing years

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
1 year

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**

Measure 1: FPC
Product line Medicaid.
Age None specified.
Continuous enrollment 43 days prior to delivery through 56 days after delivery.
Allowable gap No allowable gap during the continuous enrollment period.
Anchor date Date of delivery.
Benefit Medical.
Event/diagnosis Delivered a live birth on or between November 6 of the year prior to the measurement year and November 5 of the measurement year. Women who delivered in a birthing center should be included in this measure. Refer to Table PPC-A and Table PPC-B.
Multiple births. Women who had two separate deliveries (different dates of service) between November 6 of the year prior to the measurement year and November 5 of the measurement year should count twice. Women who have multiple live births during one pregnancy should be counted once in the measure.
The organization must exclude members for whom a prenatal visit is not indicated. These exclusions are indicated by a dash (–) in Table FPC-A.

Measure 2: PPC
Product lines Commercial, Medicaid (report each product line separately).
Age None specified.
Continuous enrollment 43 days prior to delivery through 56 days after delivery.
Allowable gap No allowable gap during the continuous enrollment period.
Anchor date Date of delivery.
Benefit Medical.
Event/ diagnosis Delivered a live birth on or between November 6 of the year prior to the measurement year and November 5 of the measurement year. Women who delivered in a birthing center should be included in this measure. Refer to Tables PPC-A and PPC-B for codes to identify live births.
Multiple births. Women who had two separate deliveries (different dates of service) between November 6 of the year prior to the measurement year and November 5 of the measurement year should be counted twice. Women who had multiple live births during one pregnancy should be counted once in the measure.

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** None

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
NA

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
None

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
NA

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
Measure 1: FPC
Step 1 Identify the delivery date using hospital discharge data.
Step 2 Identify the date when the member enrolled in the organization and determine the stage of pregnancy at time of enrollment. If the member has gaps in enrollment during pregnancy, use the last enrollment segment to determine continuous enrollment in the organization. For members with a gap in enrollment any time during pregnancy (including a gap in the first trimester), the last enrollment segment is the enrollment start date during the pregnancy that is closest to the delivery date.
Use the following approach (or an equivalent method) to calculate the stage of pregnancy at time of enrollment. If gestational age is not available, assume a gestational age of 280 days (40 weeks).

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

- Convert gestational age into days.
- Subtract gestational age (in days) from the date of delivery (step 1).
- Subtract the date obtained above from the date when the member enrolled in the organization to determine the stage of pregnancy at time of enrollment.
- Divide the numbers of days the member was pregnant at enrollment (step 3) by 30. Round the resulting number according to the .5 rule to a whole number.

For example, delivery date is August 8, 2010; gestational age is 33 weeks; date of enrollment is May 6, 2010. Given these variables, the process is:

- Gestational age in days is 231 days (33 weeks ? 7 days/week).
- Date of delivery – gestational age (in days) is December 22, 2009 (August 8, 2009 – 231 days).
- Date when the member enrolled in the organization – date obtained in step 2 is 135 days (May 6, 2010 – December 22, 2009).
- Month in which prenatal care began is 4.5 months (135 days/30 days) and then round up to 5 months using the 0.5 rule.

This member's stage of pregnancy at time of enrollment is 5 months.

Step 3 Use Table FPC-A to find the number of recommended prenatal visits by gestational age and stage of pregnancy at time of enrollment per the American College of Obstetricians and Gynecologists (ACOG). The chart subtracts the number of missed visits prior to the date the member enrolled from the number of recommended visits for a given gestational age.

ACOG recommends that women with an uncomplicated pregnancy receive visits every 4 weeks for the first 28 weeks of pregnancy, every 2–3 weeks until 36 weeks of pregnancy, and weekly thereafter. For example, ACOG recommends 14 visits for a 40-week pregnancy. If the member enrolled during her fourth month (3 missed visits prior to enrollment in the organization), the expected number of visits is 14 – 3 = 11.

For deliveries with a gestational age <28 weeks or >42 weeks, calculate the expected number of prenatal care visits using the date when the member enrolled and ACOG's recommended schedule of visits. For example, if gestational age is 26 weeks and the member enrolled during her second month of pregnancy, the expected number of prenatal care visits is 5 (6 expected visits [1 visit every 4 weeks or 6 visits in 24 weeks], less 1 visit missed in the first month).

If gestational age is 43 weeks and the member enrolled during her third month of pregnancy, the expected number of prenatal care visits is 15 (14 expected visits for a 40-week gestation plus 1 visit each additional week [17 total expected prenatal care visits], less 2 visits missed in the first and second months).

Step 4 Identify the number of prenatal care visits the member received during the course of her pregnancy and while enrolled in the organization using claims and encounter data. Use Table PPC-C to identify prenatal visits that occurred during the first trimester. The organization may use any of the four rules presented in the table to search for evidence of prenatal care; a woman's record only needs to satisfy one rule.

Use Table PPC-D to identify prenatal visits that occurred during the second and third trimester. Visits that occur on the date of delivery and meet the prenatal visit criteria count toward the measure.

Count as a single visit, a HCPCS code that falls on the same date of service as a CPT or UB Revenue code. Using Table PPC-C, Decision Rule 2 as an example, count as a single visit, HCPCS H1004, CPT 99201 and ICD-9-CM Diagnosis code 651.03 that fall on the same date of service.

If the member had a gap in enrollment, count only the visits received during the last enrollment segment.

Step 5 Calculate the ratio of observed visits (step 4) over expected visits (step 3).

Step 6 Report each woman in the appropriate category.

- <21 percent
- 21 percent–40 percent
- 41 percent–60 percent
- 61 percent–80 percent
- =81 percent of expected visits

Note: Ultrasound and lab results alone should not be considered a visit; they must be linked to an office visit with an appropriate practitioner in order to count for this measure.

Measure 2: PPC

Step 1 Identify live births. Use Method A and Method B below to identify all women with a live birth between November 6 of the year prior to the measurement year and November 5 of the measurement year. The organization must use both methods to identify the eligible population, but a member only needs to be

identified by one to be included in the measure.
Step 2 Identify continuous enrollment. For women identified in step 1, determine if enrollment was continuous between 43 days prior to delivery and 56 days after delivery, with no gaps.
Step 3 Determine enrollment status during the first trimester. Determine if women identified in step 2 were enrolled on or before 280 days prior to delivery (or estimated date of delivery [EDD]). For these women, go to step 4. For women not enrolled on or before 280 days prior to delivery (or EDD), who were therefore pregnant at the time of enrollment, proceed to step 6.
Step 4 Determine continuous enrollment for the first trimester. Determine if women identified in step 3 were continuously enrolled during the first trimester (176–280 days prior to delivery [or EDD]) with no gaps in enrollment. For these women, use one of the four decision rules in Table PPC-C to determine if there was a prenatal visit during the first trimester.4 For women who were not continuously enrolled during the first trimester, proceed to step 5.
Step 5 For women who had a gap between 176 and 280 days before delivery, proceed to step 6.
Step 6 For women identified in step 3 and step 5, determine the start date of the last enrollment segment.5 For women not enrolled in the organization on or before 280 days before delivery (or EDD) and for women who had a gap between 176 and 280 days before delivery (step 5), determine the start date of the last enrollment segment.
For women whose last enrollment started on or between 219 and 279 days before delivery, proceed to step 7. For women whose last enrollment started less than 219 days before delivery proceed to step 8.
Step 7 Determine numerator compliance if enrollment started on or between 219 and 279 days before delivery. If the last enrollment segment started on or between 219 and 279 days before delivery, determine numerator compliance using the numerator criteria in Table PPC-D and find a visit between the last enrollment start date and 176 days before delivery.6
Step 8 Determine numerator compliance if enrollment started less than 219 days before delivery (i.e., between 219 days before delivery and the day of delivery). If the last enrollment segment started less than 219 days before delivery, determine numerator compliance using Table PPC-D numerator criteria for a visit within 42 days after enrollment.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
A systematic sample of members drawn from the eligible population. Frequency of Ongoing Prenatal Care and Prenatal and Postpartum Care measures must use the same systematic sample for both. The organization may reduce the sample size using the current year's lowest product-line-specific administrative rate for the rate of women who received >=81 percent of expected prenatal care visits and the two rates from Prenatal and Postpartum Care. It may also use the prior year's lowest audited product-line-specific rates for the rate of women who received >=81 percent of expected prenatal care visits and the two rates from Prenatal and Postpartum Care.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Paper medical record/flow-sheet, Electronic administrative data/claims

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
Medical Record

2a.26-28 Data source/data collection instrument reference web page URL or attachment:

2a.29-31 Data dictionary/code table web page URL or attachment:

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Health Plan, Integrated delivery system, Population: national, Population: regional/network

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*
Ambulatory Care: Office, Ambulatory Care: Clinic, Ambulatory Care: Hospital Outpatient

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Clinicians: Nurses, Clinicians: PA/NP/Advanced Practice Nurse, Clinicians: Physicians (MD/DO), Other midwife

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: We did not conduct reliability testing for this measure.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
NA

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
NA

2b C☐ P☐ M☐ N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: expert panel

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
NCQA tested the measure for face validity using a panel of stakeholders with specific expertise in measurement and child health care. This panel included representatives from key stakeholder groups, including pediatricians, family physicians, health plans, state Medicaid agencies and researchers. Experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
This measure was deemed valid by the expert panel.

2c C☐ P☐ M☐ N☐

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
None

2d.2 Citations for Evidence:
NA

2d.3 Data/sample *(description of data/sample and size)*: NA

2d.4 Analytic Method *(type analysis & rationale)*:
NA

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
NA

2d C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: NA

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
NA

2e.3 Testing Results *(risk model performance metrics)*:
NA

2e C☐ P☐ M☐ N☐ NA☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus; ... [2]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outc ... [3]

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and w ... [4]

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: The measure assesses utilization and access in a general population; risk adjustment is not indicated.

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)***:** HEDIS National Data

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)***:**
Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance

2f.3 Provide Measure Scores from Testing or Current Use *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)***:**
Frequency of Ongoing Prenatal Care
National Means
HEDIS 2006 Data
<21%
13.52
HEDIS 2007 Data
12.36
>= 81%
HEDIS 2006 Data
58.6
HEDIS 2007 Data
59.59
21-40%
HEDIS 2006 Data
6.04
HEDIS 2007 Data
6.63
41-60%
HEDIS 2006 Data
7.84
HEDIS 2007 Data
7.74
61-80%
HEDIS 2006 Data
14.1
HEDIS 2007 Data
13.85

Prenatal and Postpartum Care
Postpartum Care
HEDIS 2006
59.08
HEDIS 2007
58.6

Timeliness of Prenatal Care
HEDIS 2006
81.24
HEDIS 2007
81.37

2f
C☐
P☐
M☐
N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample *(description of data/sample and size)***:**

2g
C☐
P☐

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

2g.2 Analytic Method *(type of analysis & rationale)*:	**M☐ N☐ NA☐**
2g.3 Testing Results *(e.g., correlation statistics, comparison of rankings)*:	
2h. Disparities in Care	
2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)*: The measure is not stratified to detect disparities.	**2h C☐ P☐ M☐ N☐ NA☐**
2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: NA	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* **C☐ P☐ M☐ N☐**
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	Eval Rating
3a. Meaningful, Understandable, and Useful Information	
3a.1 Current Use: In use	
3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large) (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*): This measure is used in public reporting.	
3a.3 If used in other programs/initiatives (*If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years***):** This measure is a measure in the Healthcare Effectiveness Data and Information Set (HEDIS)	
Testing of Interpretability (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) **3a.4 Data/sample** *(description of data/sample and size)*: General public and other stakeholder groups (i.e. HEDIS users)	
3a.5 Methods *(e.g., focus group, survey, QI project)*: For the health plan measure, we released the measure for public comment and reviewed all results with the NCQA Committee on Performance Measurement (CPM). We also reviewed first-year results with the CPM.	
3a.6 Results *(qualitative and/or quantitative results and conclusions)*: NCQA received feedback that the measure is understandable, feasible, important and valid. Upon review of public comment results, the Committee on Performance Measurement approved the NCQA staff recommendation to add the measure to HEDIS. After reviewing first-year analysis results, the CPM approved the staff recommendation to publicly report the measure. The measure was deemed usable and feasible.	**3a C☐ P☐ M☐ N☐**
3b/3c. Relation to other NQF-endorsed measures	

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

3b.1 NQF # and Title of similar or related measures:	
(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): **3b.2 Are the measure specifications harmonized? If not, why?**	**3b** C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** **5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:** NA	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes **4a.1-2 How are the data elements that are needed to compute measure scores generated?** Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition), Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) No **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.** NCQA may eventually specify this measure for electronic health records.	**4b** C☐ P☐ M☐ N☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences	**4d** C☐

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

4d.1 **Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** All measures that are used in NCQA programs are audited.	**P☐ M☐ N☐**
4e. Data Collection Strategy/Implementation 4e.1 **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** Based on field test results, we have specified the measure to assess whether visit occurred. 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** This measure appears in HEDIS and is subject to HEDIS costs. 4e.3 **Evidence for costs:** Based on user feedback and other stakeholder input. 4e.4 **Business case documentation:**	**4e** **C☐ P☐ M☐ N☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** **C☐ P☐ M☐ N☐**
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited ☐**
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y☐ N☐ A☐**

CONTACT INFORMATION

Co.1 **Measure Steward (Intellectual Property Owner)**
Co.1 **Organization**
National Committee for Quality Assurance, 1100 13th Street NW, Suite 1000, Washington, District Of Columbia, 20005

Co.2 **Point of Contact**
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-

Measure Developer If different from Measure Steward
Co.3 **Organization**
National Committee for Quality Assurance, 1100 13th Street NW, Suite 1000, Washington, District Of Columbia, 20005

Co.4 **Point of Contact**
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-

Co.5 **Submitter If different from Measure Steward POC**
Sepheen, Byron, MHS, byron@ncqa.org, 202-955-3573-, National Committee for Quality Assurance

Co.6 **Additional organizations that sponsored/participated in measure development**

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development

Ad.1 **Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.**

Over the years, the following expert panel has contributed to many of the measures in the HEDIS set that apply to women and children.

David Archer, MD
Eastern Virginia Medical School
Grant P. Bagley, MD, JD
Arnold & Porter
Thomas J. Benedetti, MD
University of Washington Medical Center
Denis Dougherty
Agency for Healthcare Research and Quality (AHRQ)
Christopher B. Forrest, MD, PhD
The Children's Hospital of Philadelphia
Shirley Girouard, PhD, RN
Southern Connecticut State University
Bill Heuston, MD
Medical University of South Carolina
Mary Kay Holleran
Highmark Caring Foundation
Charles Homer MD, MPH
National Initiative for Children's Healthcare Quality
Marilyn C. Jones, MD
Children's Hospital
Milton Kotelchuck, PhD, MPH
Boston University School of Public Health
Mark Mandell, MD
Partners Community Health Care, Inc.
Dorothy Mann, PhD, MPH
Consultant
Robert H. Pantell, MD
University of California, San Francisco
Lee Partridge

Ad.2 **If adapted, provide name of original measure:** NA
Ad.3-5 **If adapted, provide original specifications URL or attachment**

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.6 **Year the measure was first released:** 1997
Ad.7 **Month and Year of most recent revision:** 07, 2010
Ad.8 **What is your frequency for review/update of this measure?** Annual
Ad.9 **When is the next scheduled review/update for this measure?** 07, 2011

Ad.10 **Copyright statement/disclaimers:** Frequency of Ongoing Prenatal Care:

Prenatal and Postpartum Care:

Ad.11 -13 **Additional Information web page URL or attachment:**

Date of Submission (*MM/DD/YY*)**:** 09/02/2010

Page 5: [1] Comment [k4] **Karen Pace** **10/5/2009 8:59:00 AM**

1c. The measure focus is:

- an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;

OR

- if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
  - Intermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
  - Process – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
    if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
  - Structure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
  - Patient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
  - Access – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
  - Efficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Page 12: [2] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;

AND

- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

AND

- precisely defined and specified:
  - if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);

if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

Page 12: [3] Comment [KP16] **Karen Pace** **10/5/2009 8:59:00 AM**

2e. For outcome measures and other measures (e.g., resource use) when indicated:

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;**Error! Bookmark not defined.** OR

rationale/data support no risk adjustment.

Page 12: [4] Comment [k17] **Karen Pace** **10/5/2009 8:59:00 AM**

13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1351 **NQF Project:** Child Health Quality Measures 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Proportion of infants covered by Newborn Bloodspot Screening (NBS)
De.2 **Brief description of measure***:* What percentage of infants had bloodspot newborn screening performed as mandated by state of birth?
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** This measure is paired with other measures relevant to the monitoring and measurement of the early screening evaluation and intervention process.
De.4 **National Priority Partners Priority Area:** Population health De.5 **IOM Quality Domain**: Effectiveness De.6 **Consumer Care Need:** Living with illness

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary A.4 **Measure Steward Agreement attached:**	**A** **Y**☐ **N**☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and	**B**

update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Accountability, Accreditation	C Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** No, testing will be completed within 12 months D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	D Y☐ N☐
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers, Frequently performed procedure, High resource use, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** One in 800 infants born each year has a newborn screening detectable disorder, all of which can cause death or morbidity unless treated. U.S. Preventive Services Task Force. The USPSTF recommends screening for PKU, congenital hypothyroidism and sickle cell disease in all newborn infants. There is good evidence that NBS testing is highly accurate and leads to earlier identification and treatment of infants with these disorders. Good-quality evidence shows that early detection improves developmental and overall health outcomes. http://www.uspreventiveservicestaskforce.org/uspstf/uspsspku.htm http://www.uspreventiveservicestaskforce.org/uspstf/uspscghy.htm http://www.uspreventiveservicestaskforce.org/uspstf/uspshemo.htm 1a.4 **Citations for Evidence of High Impact:** ACMG Report- Newborn Screening Toward a Uniform Panel PEDIATRICS Vol. 106 No. 3 September 2000, pp. 595 - "The Importance of Newborn Screening" American Academy of Pediatrics NBS Task Force http://www.uspreventiveservicestaskforce.org/uspstf/uspsspku.htm http://www.uspreventiveservicestaskforce.org/uspstf/uspscghy.htm	1a C☐ P☐ M☐ N☐

Comment [KP1]: 1a. The measure focus addresses:
•a specific national health goal/priority identified by NQF's National Priorities Partners; OR
•a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

http://www.uspreventiveservicestaskforce.org/uspstf/uspshemo.htm

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: By using this measure, we will be able to document the proportion of infants screened by NBS, identify state programs with problems and assist them in making sure NBS is universally available and documented.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
More than 10 states are unable to assess this measure, others range from 80% and some states have >100% screened, which indicates lack of uniformity and reliability of current reporting.

1b.3 Citations for data on performance gap:
http://nnsis.uthscsa.edu/xreports.aspx?XREPORTID=17&FORMID=44&FCLR=1

1b.4 Summary of Data on disparities by population group:
This is a national program and at this time, there is no specific information on disparities by population group.

1b.5 Citations for data on Disparities:

1b
C☐ P☐ M☐ N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** If we cannot measure what percentage of infants are screened, we cannot address gaps in coverage of this public health program and infants could be dying from preventable disorders that have a reliable test and should be universally available.

1c.2-3. Type of Evidence: Cohort study, Evidence-based guideline, Expert opinion, Systematic synthesis of research

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
95% of infants with PKU (detectable on NBS) will have mental retardation, seizures and brain changes if not started on treatment in first 2 weeks of life.
25% of infants with MCAD (NBS disorder) die with their first illness if not prospectively followed.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
Grade A: based on US Preventive Services Task Force

1c.6 Method for rating evidence: Based on USPSTF

1c.7 Summary of Controversy/Contradictory Evidence: There is limited evidence about the harms of screening, with conflicting research findings regarding anxiety associated with falsepositive test results. There is limited information about the harms of treatment

1c.8 Citations for Evidence (*other than guidelines***):** Boles RG, Buck EA, Blitzer MG, Platt MS, Cowan TM, Martin SK, Yoon H, Madsen JA, Reyes-Mugica M, Rinaldo P. "Retrospective biochemical screening of fatty acid oxidation disorders in postmortem livers of 418 cases of sudden death in the first year of life", J Pediatr 1998; 132(6):924-33.
Paine, RS. The variability in manifestations of untreated patients with phenylketonuria aciduria). Pediatrics 20:290-302, 1957.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
Screening for Phenylketonuria (PKU)- USPSTF

1c
C☐ P☐ M☐ N☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure … [1]

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve … [2]

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system … [3]

Screening for Congenital Hypothyroidism - USPSTF
Screening for Sickle Cell Disease in Newborns - USPSTF

1c.10 Clinical Practice Guideline Citation: Serving the Family From Birth to the Medical Home
Newborn Screening: A Blueprint for the Future A Call for a National Agenda on State Newborn Screening Programs
1c.11 National Guideline Clearinghouse or other URL:

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
The USPSTF recommends screening for phenylketonuria (PKU) in newborns; Congenital Hypothyroidism; Sickle Cell Disease. Grade: A Recommendation

1c.13 Method for rating strength of recommendation (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:**
Rationale for PKU screening:
Importance: PKU is an inborn error of phenylalanine metabolism that occurs in from 1 per 13,500 to 1 per 19,000 newborns in the United States. In the absence of treatment during infancy, most persons with this disorder will develop severe mental retardation.1,2

Detection: Two approaches, fluorometry and tandem mass spectrometry, are in common use. The sensitivity and specificity of fluorometry are 100% and 51%, respectively, and of tandem mass spectrometry, 100% and 98%, respectively.3

Benefits of Detection and Early Treatment: There is good evidence that detection by neonatal screening and early treatment of PKU substantially improve neurodevelopmental outcomes for affected persons.

Harms of Detection and Early Treatment: False-positive tests could generate considerable parental anxiety
Sickle Cell:
Rationale
Importance: Sickle cell anemia (hemoglobin SS) affects 1 in 375 African American newborns born in the United States and smaller proportions of children in other ethnic groups. Without prompt diagnosis and the initiation of prophylactic antibiotics and pneumococcal conjugate vaccination by 2 months of age, children with sickle cell anemia are vulnerable to life-threatening pneumococcal infections.1

Detection: In the United States, most state-based screening programs utilize thin-layer isoelectric focusing (IEF) or high performance liquid chromatography (HPLC) techniques performed on capillary blood collected from a heel stick and absorbed onto filter paper. The sensitivity and specificity of each of these tests approaches 100%.
Rationale for Congenital Hypothyroid:
Importance: Primary congenital hypothyroidism occurs in approximately 1 of every 3,000-4,000 newborns in the United States. In the absence of prompt diagnosis and treatment, most persons with this disorder will develop various degrees of neurological, motor and growth deficits, including irreversible mental retardation.

Detection: In the U.S., most state-based screening programs utilize serum thyroxine (T4) and/or thyroid-stimulating hormone (TSH) performed on capillary blood collected from a heel stick and adsorbed onto filter paper.

Benefits of Detection and Early Treatment: Early detection of CH by neonatal screening and appropriate treatment substantially improves neurodevelopmental outcomes for affected persons.

Harms of Detection and Early Treatment: Positive test results, whether true positive or false positive, cause anxiety in parents. For some parents, this anxiety may be considerable.

USPSTF Assessment: The USPSTF concludes that there is high certainty that the net benefit is substantial.

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

Benefits of detection and early intervention: There is good evidence that early detection of sickle cell anemia followed by prophylactic oral penicillin substantially reduces the risk of serious infections during the first few years of life. Additional benefits result from pneumococcal conjugate vaccination and parental education about early warning signs of infection. Finally, detection of sickle cell disease permits counseling for family members about disease management and future reproductive decisions. Harms of detection and early treatment: Incidental detection of sickle cell carrier status and hemoglobin disorders of questionable clinical significance has the potential to cause psychosocial harms, which may include exposure of non-paternity, stigma and discrimination, negative impact on self-esteem, and anxiety about future health. The USPSTF concludes that there is high certainty that the net benefit of screening for sickle cell disease in newborns is substantial 1c.14 **Rationale for using this guideline over others:** These are the only USPSTF guideline related to newborn bloodspot screening and the only nationally recognized guidelines. Recently the Secretary of Health and Human Services endorsed the Uniform Screening Panel as put forward by the Secretary´s Advisory Committee on Heritable Disorders in Newborns and Children.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report*, **met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 **Do you have a web page where current detailed measure specifications can be obtained?** S.2 **If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 **Numerator Statement (***Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** The number of infants born in a state who have a valid newborn screen performed- in accordance with the state of birth mandated program specifications 2a.2 **Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** The time period varies upon needs of the particular user (e.g. calendar year, quarterly, monthly) but must be the same for both the numerator and denominator. 2a.3 **Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Number of infants with newborn bloodspot screen performed as documented/collected by the state newborn screening program.	
2a.4 **Denominator Statement (***Brief, text description of the denominator - target population being measured***):** number of infants born in a state during the time period used in the numerator (same area used for numerator) 2a.5 **Target population gender:** Female, Male	**2a-specs** **C☐** **P☐** **M☐** **N☐**

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

2a.6 Target population age range: birth to 2 weeks

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
The time period varies upon needs of the particular user (e.g. calendar year, quarterly, monthly) but must be the same for both the numerator and denominator.

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
This should be information gathered by the state public health department by birth certificates or hospital birth records for matching with the numerator.

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** infants who die prior to normal time frame for collection of newborn screen or infants who have a formal waiver signed by the parents/guardians refusing the state newborn screen

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
Joint Commission Discharge Disposition - Death Value Set (86986.v1) 1.3.6.1.4.1.33895.1.3.0.12. "Patient Deceased": Patient has expired.

LOINC# 54108-6 LA6644-4 C0580717 "Parental refusal"

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
None because state mandates apply to all infants and do not stratify by NICU status, prematurity, geographic location, or insurance coverage. In the future we might explore health disparities, but current measures will be applied to all infants born in a state.

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
(1) The time period for births included in the estimate is specified (see 2a.2, 2a.7).
(2) All live births that occurred in a state during the time period are selected.
(3) Result of step 2 is filtered to remove children who died prior to discharge (see 2a.9, 2a.10) or parental waiver. This result is saved
The numerator is calculated using the following step:
(4) Result of step 3 is further filtered to be limited to the subset with a NBS performed (see 2a.3) This result is saved as the numerator (see 2a.1).
The denominator is:
(5) Result of step (3)

Porportion calculated using the following step:
(6) HRSA NBS measure is calculated by dividing the numerator (result of step 4) by the denominator (result of step 5).

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
chi squared comparison of porportions- Method to discriminate performance is based upon jurisdictionally based statistical measurement reflecting local and national variability.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
not applicable

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Paper medical record/flow-sheet, Electronic administrative data/claims, Public health data/vital statistics, Electronic Health/Medical Record, Lab data

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
Number of infants screened- NNSIS national newborn screening information system, collects the number of NBS performed in each state, will work to distinguish exact number of infants screened via tracking and linkage- need to distinguish between initial screens and repeat screens.
number of infants born- state birth certificates and hospital discharge records

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL
http://nnsis.uthscsa.edu/xreports.aspx?XREPORTID=5

2a.29-31 Data dictionary/code table web page URL or attachment: URL
http://nnsis.uthscsa.edu/NNSIS_User_Manual_V2.pdf

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency, Population: states, Program: Other state newborn screening program

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*
Hospital, Ambulatory Care: Clinic

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Laboratory, Other public health agency

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)***:** This is a population wide collection of data, gathered and assessed at the state level in the various state newborn screening programs. All states have a NBS program and collect data related to this program.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)***:**
N/A - in the process of testing the methodology and determining reliability

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
This should be a simple matter of matching births to newborn screening results. However, the numbers need to be verified and the matching needs to be done.

2b
C☐
P☐
M☐
N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)***:** This is a population wide collection of data, gathered and assessed at the state level in the various state newborn screening programs.

2c.2 Analytic Method *(type of validity & rationale, method for testing)***:**
N/A - in the process of testing the methodology and determining reliability

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)***:**
This should be a simple matter of matching births to newborn screening results. However, the numbers need to be verified and the matching needs to be done.

2c
C☐
P☐
M☐
N☐

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
The exclusions only apply where legally mandated at the state level. Some states allow for exclusion based

2d
C☐
P☐
M☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
AND
•precisely defined and specified:
–if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category ... [4]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

on a waiver, others do not. Some states all collect bloodspot screens on infant that expire prior to the mandated time frame for further evaluation if there is a question of a heritable disorder as causative. 2d.2 **Citations for Evidence:** http://genes-r-us.uthscsa.edu/resources/consumer/statemap.htm This website can direct people to various states and their rules. 2d.3 **Data/sample** *(description of data/sample and size)*: 2d.4 **Analytic Method** *(type analysis & rationale)*: 2d.5 **Testing Results** *(e.g., frequency, variability, sensitivity analyses)*:	**N☐** **NA☐**
2e. Risk Adjustment for Outcomes/ Resource Use Measures 2e.1 **Data/sample** *(description of data/sample and size)*: N/A- at this time there is no risk adjustment needed 2e.2 **Analytic Method** *(type of risk adjustment, analysis, & rationale)*: 2e.3 **Testing Results** *(risk model performance metrics)*: 2e.4 **If outcome or resource use measure is not risk adjusted, provide rationale:**	**2e** **C☐** **P☐** **M☐** **N☐** **NA☐**
2f. Identification of Meaningful Differences in Performance 2f.1 **Data/sample from Testing or Current Use** *(description of data/sample and size)*: At this time, cannot be determined 2f.2 **Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*: Will need to survey with good, reliable numbers in order to establish a baseline and then differentiate what are difference with impact. This is a state public health surveillance system, not a sampling sysitem. 2f.3 **Provide Measure Scores from Testing or Current Use** *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:	**2f** **C☐** **P☐** **M☐** **N☐**
2g. Comparability of Multiple Data Sources/Methods 2g.1 **Data/sample** *(description of data/sample and size)*: Will need to develop a method for linking and tracking in order to get reliable numbers of infants screened and infants born in the geographical area of interest. 2g.2 **Analytic Method** *(type of analysis & rationale)*: 2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)*:	**2g** **C☐** **P☐** **M☐** **N☐** **NA☐**
2h. Disparities in Care 2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: N/A at this time	**2h** **C☐** **P☐** **M☐** **N☐**

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR rationale/data support no risk adjustment.

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:**	**NA**☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* **C**☐ **P**☐ **M**☐ **N**☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	Eval Rating
3a. Meaningful, Understandable, and Useful Information 3a.1 **Current Use:** Testing not yet completed 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** At this time, a facsimile of the data is reported out by state on the http://genes-r-us.uthscsa.edu website. However, at this time it is not verified and linked to insure that infants are not double entered does not occur, nor does linking to birth records or birth certificates. In the future, will be able to provide accurate aggregate data but the state specifics will be password protected and disseminated with their discretion. This was previously a Title V Block Grant Performance Measure and was tracked by the states. The decision was made to change the emphasis in the Block Grant to follow-up and so they changed the specific measure. 3a.3 **If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years***):** In the process of updating the information system this is reported in, working with various health informatics programs to provide automated linking and messaging systems, which will allow for less time intensive data entry and more reliable numbers. This will be incorporated into a NBS QI system at a national level with breakdowns by state. The Newborn Screening Saves Lives Act of 2008 also mandates reporting of quality indicators for newborn screening programs. A related measure proposed for Healthy People 2020 is: HP2020 Objective Text: MICH HP2020-22: Increase appropriate newborn blood-spot screening and follow-up testing a. Increase the number of states that verify, through linkage with vital records, that all newborns are screened shortly after birth for conditions mandated by their State-sponsored screening program. **Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) 3a.4 **Data/sample** *(description of data/sample and size)***:** This is aggregate data to look at screening rates and evaluate the needs for further funding or programmatic assistance. 3a.5 **Methods** *(e.g., focus group, survey, QI project)***:** QI project to cover full NBS system evaluation. 3a.6 **Results** *(qualitative and/or quantitative results and conclusions)***:**	**3a** **C**☐ **P**☐ **M**☐ **N**☐
3b/3c. Relation to other NQF-endorsed measures 3b.1 **NQF # and Title of similar or related measures:**	

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

Proposed measures from NCQA for physician documentation of NBS results in patient record.	
(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): **3b.2 Are the measure specifications harmonized? If not, why?** Yes.	**3b** C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** This measure is aimed at the state NBS programs and accuracy of their ability to track the screened population. The NCQA measure is for physician documentation and records tracking. **5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:** There is no competing measure that is population based rather than practice or hospital based and population data is required to measure this dimension of quality.	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes **4a.1-2 How are the data elements that are needed to compute measure scores generated?** Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition)	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	**4b** C☐ P☐ M☐ N☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences **4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.**	**4d** C☐ P☐ M☐

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

Babies born in one state may get another screen in a second state that could result in double counting, however with good records linking and tracking this can be eliminated.	N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** Requires an accurate standardized denominator and numerator to successfully determine that all infants have been accounted for and received necessary screen. The limitation has been that states only report the number of screens performed, not tracking by individual infant. **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** This is a measure calculated by public health and based on NBS lab reporting and matching with birth records and certificates. Public health information systems must be capable of having a specific NBS record on each infant and be capable of differentiating initial vs repeat screen. Such systems are in use in States. For other public health programs infrastructure may need to be strengthened and there will be a cost to this additional data collection. **4e.3 Evidence for costs:** **4e.4 Business case documentation:**	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** C☐ P☐ M☐ N☐
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited** ☐
Steering Committee: Do you recommend for endorsement? **Comments:**	Y☐ N☐ A☐

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

CONTACT INFORMATION
Co.1 Measure Steward (Intellectual Property Owner) **Co.1 Organization** HRSA, 5600 Fishers Ln Rm 18A-19, Rockville, Maryland, 20857 **Co.2 Point of Contact** Sara, Copeland, MD, scopeland@hrsa.gov, 301-443-8860-
Measure Developer If different from Measure Steward **Co.3 Organization** HRSA, 5600 Fishers Ln Rm 18A-19, Rockville, Maryland, 20857 **Co.4 Point of Contact** Sara, Copeland, MD, scopeland@hrsa.gov, 301-443-8860-
Co.5 Submitter If different from Measure Steward POC

Sara, Copeland, MD, scopeland@hrsa.gov, 301-443-8860-, HRSA
Co.6 Additional organizations that sponsored/participated in measure development
ADDITIONAL INFORMATION
Workgroup/Expert Panel involved in measure development **Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.** CDC Newborn Screening Quality Assessment Program, National Newborn Screening and Genetics Resource Center
Ad.2 If adapted, provide name of original measure: **Ad.3-5 If adapted, provide original specifications URL or attachment**
Measure Developer/Steward Updates and Ongoing Maintenance **Ad.6 Year the measure was first released:** 2010 **Ad.7 Month and Year of most recent revision:** 2010 **Ad.8 What is your frequency for review/update of this measure?** This is a new measure that will be released in Fall, 2010 and an annual review/update is planned **Ad.9 When is the next scheduled review/update for this measure?** 2011
Ad.10 Copyright statement/disclaimers:
Ad.11 -13 Additional Information web page URL or attachment: URL http://genes-r-us.uthscsa.edu
Date of Submission (*MM/DD/YY*)**:** 08/30/2010

Page 3: [1] Comment [k4] **Karen Pace** **10/5/2009 8:59:00 AM**

1c. The measure focus is:

- an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;

OR

- if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
  - Intermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
  - Process – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
    if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
  - Structure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
  - Patient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
  - Access – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
  - Efficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Page 3: [2] Comment [k5] **Karen Pace** **10/5/2009 8:59:00 AM**

4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Page 3: [3] Comment [k6] **Karen Pace** **10/5/2009 8:59:00 AM**

3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

Page 7: [4] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;

AND

- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

AND

- precisely defined and specified:
  - if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);

if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1448 **NQF Project:** Child Health Quality Measures 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 Measure Title: Developmental Screening in the First Three Years of Life
De.2 Brief description of measure*:* The percentage of children screened for risk of developmental, behavioral and social delays using a standardized screening tool in the first three years of life. This is a measure of screening in the first three years of life that includes three, age-specific indicators assessing whether children are screened by 12 months of age, by 24 months of age and by 36 months of age.
1.1-2 Type of Measure: Process **De.3 If included in a composite or paired with another measure, please identify composite or paired measure** NA
De.4 National Priority Partners Priority Area: Care coordination, Population health **De.5 IOM Quality Domain**: Effectiveness, Timeliness **De.6 Consumer Care Need:** Staying healthy

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* **A.1 Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes **A.2 Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** **A.3 Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission **A.4 Measure Steward Agreement attached:**	**A** **Y** ☐ **N** ☐

B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Other Program evaluation.	**C** Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** No, testing will be completed within 12 months D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☐ N☐
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** The American Academy of Pediatrics (AAP) defines a developmental delay as a "condition in which a child is not developing and/or achieving skills according to the expected time frame." A child that is developmentally challenged may face many barriers throughout life; these barriers are even more severe if a delay in development is not detected early. Delayed or disordered development can lead to further health and behavior problems, including failure in school and social and emotional problems.(Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee, 2006) Approximately 12 to 18 percent of U.S. children may have a developmental and behavioral problem. However, only about two percent of children from birth to two years old receive the necessary early intervention services.(Hix-Small, Hollie, PhD, et al., 2007) A child who is identified as having a delay in development by the time he starts school and participates in early intervention programs is more likely to graduate high school, hold a job, live independently, and avoid teen pregnancy, delinquency and violent crimes -- representing a saved cost to society of between $30,000 and $100,000 per child.(Glascoe FP, PhD, et al., 2007)	**1a** C☐ P☐ M☐ N☐

Studies have shown that developmental surveillance based on non-standardized clinical judgment and observation alone does not accurately identify children with delays. Therefore, national recommendations call for routine, standardized screening of children three times in the first three years (at the 9, 18 and 24- or 30-month well-visit).

1a.4 Citations for Evidence of High Impact: Hagan JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children and Adolescent, Third Edition, Elk Grove Village IL. American Academy of Pediatrics.

Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006;118(1):405-420

Hix-Small, Hollie, PhD, et al. Impact of Implementing Developmental Screening at 12 and 24 Months in a Pediatric Practice Pediatrics Vol. 120 No. 2 August 2007, pp. 381-389

Glascoe FP, PhD and Shapiro, HL, MD. Introduction to Developmental and Behavioral Screening. 2007. http://www.dbpeds.org/articles/detail.cfm?TextID=5

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Pediatricians are not usually successful in identifying children with developmental delays without use of a standardized tool (Hix-Small, 2007). This measure will encourage the use of standardized tools for developmental screening, as delineated by guidelines. Children who are identified earlier are more likely to have developmental promotion activities, that can further improve the likihood that they will be able to start school ready to learn. Demonstrated quality improvement activities such as the ABCD Screening Academy have shown that providers can feasibly and sustainably implement standardized screening, and when done so, more children are refereed to Early Intervention and other services and that the kinds and types of referrals performed are more appropriate than was previously done without standardized screening

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
Findings from the National Survey of Children Health show that only 19.5% of children are screened in the first five years of life. Despite the evidence, the use of standardized developmental screening tools is uncommon; only about 20 percent of physicians routinely use developmental screening tests (The Commonwealth Fund, 2008). One study found that pediatricians failed to identify and refer 60 to 80 percent of children with developmental delays in a timely manner. Another study found that 68 percent of children with delays were not detected by pediatricians. Though many significant delays occur before school age, less than 50 percent of children with delays are identified before starting school -- leading to missed opportunities for treatment (Hix-Small, 2007).

1b.3 Citations for data on performance gap:
http://www.nschdata.org

Commonwealth Fund. Quality Matters, May 6 2008.

Hix-Small, Hollie, PhD, et al. Impact of Implementing Developmental Screening at 12 and 24 Months in a Pediatric Practice Pediatrics Vol. 120 No. 2 August 2007, pp. 381-389

Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006;118(1):405-420

The American Academy of Pediatrics, Council on Children With Disabilities, Section on Developmental and Behavioral Pediatrics, Bright Futures Steering Committee, and Medical Home Initiatives for Children With Special Needs. Identifying infants and young children with developmental disorder in the medical home: an

1b
C☐
P☐
M☐
N☐

algorithm for developmental surveillance and screening. Pediatrics. 2006. 118(1): 405-420.

Bethell, CD, Reuland, C, Halfon, N, Olsen, L, Schor, E., Measuring the Quality of Preventive and Developmental Services for Young Children: National Estimates and Patterns of Clinicians' Performance. Pediatrics. June 2004.

Pinto-martin, J, Dunkle M, Earls M, Fliedner D, Cynthia L. Developmental States of Developmental Screening: Steps to Implementation of a Successful Program. American Journal of Public Health. 95, 11: 1928-1932.

King T., Trandon, D, Macias, M, et al. Implementing developmental screening and referrals: Lessons learned from a national project. Pediatrics, V 125, No 2, Feb 2010.

Sand N, Silverstein M, Glascoe FP, et al. Pediatrician's reported practices regarding developmental screening: do guidelines work? Do they help? Pediatrics 2005; V116 (1): 174-179

Smith RD. The use of developmental screening tests by primary-care pediatricians. J Pediatrics. 1978; 93(3): 524-527.

Zuckerman KE, Boudreau AA, Lipstein EA, Kuhlthau KA, and Perrin JM. Household Language, Parent Developmental Concerns, and Child Risk for Developmental Disorder. Academic Pediatrics. 2009; 9(2): 97-105.

1b.4 Summary of Data on disparities by population group:

Studies suggest income disparities exist for developmental screening. One study found that only 23 percent of low-income children receive recommended preventive and developmental services (Bethell et al, 2002). The Early Intervention Periodic Screening, Diagnosis and Treatment (EPSDT) benefit for Medicaid children includes screening at each visit, however, as of 2007, 28 states were engaged in lawsuits due to a failure to properly deliver this service (Glascoe et al, 2007). Another study found that children most at risk for school difficulty were those whose mothers had less than a high school education, those who came from single-mother families, those who had received public assistance, and those who lived in families in which the primary language was not English (High, 2008)." Specifically related to screening, the National Survey of Children's Health found that while improvements were needed in increasing screening for all children, significant variations existed in the rates of screening by race-ethnicity and insurance status.

1b.5 Citations for data on Disparities:

Bethell at al. Partnering with parents to promote the healthy development of young children enrolled in Medicaid. New York NY: The commonwealth Fund, 2002.
Glascoe FP, PhD and Shapiro, HL, MD. Introduction to Developmental and Behavioral Screening. 2007. http://www.dbpeds.org/articles/detail.cfm?TextID=5
High, Pamela C. and the Committee on Early Childhood, Adoption, and Dependent Care and Council on School Health. School Readiness. Pediatrics 2008;121;e1008-e1015
http://www.nschdata.org

Pinto-martin, J, Dunkle M, Earls M, Fliedner D, Cynthia L. Developmental States of Developmental Screening: Steps to Implementation of a Successful Program. American Journal of Public Health. 95, 11: 1928-1932.

King T., Trandon, D, Macias, M, et al. Implementing developmental screening and referrals: Lessons learned from a national project. Pediatrics, V 125, No 2, Feb 2010.

Sand N, Silverstein M, Glascoe FP, et al. Pediatrician's reported practices regarding developmental screening: do guidelines work? Do they help? Pediatrics 2005; V116 (1): 174-179

Smith RD. The use of developmental screening tests by primary-care pediatricians. J Pediatrics. 1978; 93(3): 524-527.

Zuckerman KE, Boudreau AA, Lipstein EA, Kuhlthau KA, and Perrin JM. Household Language, Parent

Developmental Concerns, and Child Risk for Developmental Disorder. Academic Pediatrics. 2009; 9(2): 97-105.

1c. Outcome or Evidence to Support Measure Focus

1c.1 **Relationship to Outcomes** (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Early identification of developmental disabilities through surveillance and screening can lead to timely evaluation, diagnosis and appropriate treatment, including developmental intervention.

1c.2-3. **Type of Evidence:** Evidence-based guideline, Expert opinion

1c.4 **Summary of Evidence** (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Developmental surveillance should be a component of every preventive care visit. Standardized developmental screening tools should be used when such surveillance identifies concerns about a child´s development. Furthermore, it is recommended that standardized screening for developmental, behavioral and social delays occur at the 9-, 18-, and 24-month OR 30-month well visits.
When a child has a positive screening result for a developmental problem, developmental and medical evaluations to identify the specific developmental disorders and related medical problems are warranted. Children diagnosed with developmental disorders should be identified as children with special health care needs; chronic-condition management for these children should be initiated.

1c.5 **Rating of strength/quality of evidence (***also provide narrative description of the rating and by whom***):**
Good

1c.6 **Method for rating evidence:** Expert consensus with evidence review.

1c.7 **Summary of Controversy/Contradictory Evidence:** The USPSTF did not review developmental screening generally. Rather, the Task Force reviewed the routine use of brief, formal screening instruments in primary care to detect speech and language delay in children. This recommendation received an "I Statement":

The USPSTF concludes that the evidence is insufficient to recommend for or against routine use of brief, formal screening instruments in primary care to detect speech and language delay in children up to 5 years of age.

Speech and language delay affects 5 to 8 percent of preschool children, often persists into the school years, and may be associated with lowered school performance and psychosocial problems. The USPSTF found insufficient evidence that brief, formal screening instruments that are suitable for use in primary care for assessing speech and language development can accurately identify children who would benefit from further evaluation and intervention. Fair evidence suggests that interventions can improve the results of short-term assessments of speech and language skills; however, no studies have assessed long-term outcomes. Furthermore, no studies have assessed any additional benefits that may be gained by treating children identified through brief, formal screening who would not be identified by addressing clinical or parental concerns. No studies have addressed the potential harms of screening or interventions for speech and language delays, such as labeling, parental anxiety, or unnecessary evaluation and intervention. Thus, the USPSTF could not determine the balance of benefits and harms of using brief, formal screening instruments to screen for speech and language delay in the primary care setting.

Secondly, It is important to note that is measure does not included standardized screening for a specific domain of development (e.g. social emotional screening via the ASQ-SE, autism screening) as it is anchored to recommendations focused on global developmental screening using tools that focus on identifying risk for developmental, behavioral and social delays. National recommendations also call for autism screening at the 18-month and 24-month well-visit and future, separate measures may specified and build off the data collection efforts used for this measure to capture domain-specific screening. Additionally, many of the ABCD states included a distinct focus on complementary, but separate, screening specifically focused

1c
C ☐
P ☐
M ☐
N ☐

on social-emotional development (using tools such as the ASQ-SE). Similarly, future efforts may maximize the data collection efforts for this measure to include additional specifications focused specifically on social-emotional screening so that a separate measure may be calculated.

1c.8 Citations for Evidence (*other than guidelines***):** Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics 2006 Jul;118(1):405-20.

Hagan JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children and Adolescent, Third Edition, Elk Grove Village IL. American Academy of Pediatrics.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
Institute for Clinical Systems Improvement:
Providers should perform the following on infants: Developmental assessment of: motor skills, language development and social development.
ICSI: Level III

Michigan Quality Improvement Consortium (2007):
From Birth to 24 months, developmental assessments should be performed.
Grade: Consensus and ICSI-Based

American Academy of Pediatrics (2006):
Medical Professionals should use standardized developmental screening tools to screen children and 9 months, 18 months:
• Developmental and medical evaluations to identify the specific developmental disorders and related medical problems
• Referred to early developmental intervention and early childhood services and scheduled for earlier return visits to increase developmental surveillance.
• Identified as children with special health care needs; chronic-condition management for these children should be initiated.
Grade: Consensus and Guideline-Based

Bright Futures (2008):
At 9, 18 and 30 Month Visits, health care providers should perform structured developmental screens. Referral should be made to an appropriate early intervention program or developmental specialist for evaluation.
Grade: Consensus and Guideline-Based

1c.10 Clinical Practice Guideline Citation: Hagan, JF, Shaw JS, Duncan PM, eds. 2008. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Elk Grove, IL: American Academy of Pediatrics

Institute for Clinical Systems Improvement. Preventive Services for Children and Adolescents Thirteenth Edition. October 2007
[AAP] Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics 2006 Jul;118(1):405-20.
Michigan Quality Improvement Consortium. Routine preventive services for children and adolescents (ages 2-18). Southfield (MI): Michigan Quality Improvement Consortium; 2007 May. 1 p.

1c.11 National Guideline Clearinghouse or other URL:
http://www.guideline.gov/search/search.aspx?term=developmental+screening

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
Consensus and Guideline-Based

1c.13 Method for rating strength of recommendation (*If different from* USPSTF system, *also describe rating and how it relates to USPSTF)*: Expert consensus with evidence review **1c.14 Rationale for using this guideline over others:** This measure represents the shared experiences of NCQA in operationalizing a feasible, meaningful measure for Medicaid Managed Care Organizations and physician practices and the learnings that the CAHMI gathered in providing measurement technical assistance to State Medicaid agencies and pediatric health care providers participating in the ABCD Screening Academy. This measure represents areas of synergy in the work conducted by both groups to yield feasible, valuable measures. As part of this effort, NCQA and CAHMI convened a group of multi-stakeholder panel of users and experts to review the specifications, evidence and guidelines for developmental screening for children. These stakeholders included persons from State Medicaid agencies in the states who participated in the Assuring Better Child Development program.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report*, **met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** The numerator identifies children who were screened for risk of developmental, behavioral and social delays using a standardized tool. National recommendations call for children to be screened at the 9, 18, and 24- OR 30-month well visits to ensure periodic screening over the first three years. The measure is based on three, age-specific indicators. Indicator 1: Children who had screening for risk of developmental, behavioral and social delays using a standardized screening tool that was documented by 12 months of age Indicator 2: Children who had screening for risk of developmental, behavioral and social delays using a standardized screening tool that was documented by 24 months of age Indicator 3: Children who screening for risk of developmental, behavioral and social delays using a standardized screening tool that was documented by 36 months of age **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** Twelve months – 1 year. **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Claims data: CPT codes 96110 (Developmental testing, with interpretation and report) Claims NOT Included in This Measure: It is important to note that 96110 claims that include modifiers indicating standardized screening for a specific domain of development (e.g. social emotional screening via the ASQ-SE, autism screening) should not be included as this measure is anchored to recommendations	**2a-specs** **C☐** **P☐** **M☐** **N☐**

focused on global developmental screening using tools that focus on identifying risk for developmental, behavioral and social delays.

Medical Chart:
Documentation must include a note indicating the date of screening, the standardized developmental screening tool used, and evidence that tool was completed and scored.

Tools must meet the following criteria: .
1) Developmental domains: The following domains must be included in the standardized developmental screening tool: motor, language, cognitive, and social-emotional.
2) Established Reliability: Reliability scores of approximately 0.70 or above.
3)Established Findings Regarding the Validity- Concurrent validity: This compares screening results with outcomes derived from a reliable and valid diagnostic assessment usually performed 7-10 days after the screening test. The validity coefficient reports the agreement between the two tests (Meisels & Atkins-Burnett, 2005). Predictive validity: This compares the screening results with measures of children's performance obtained 9-12 months later (Meisels & Atkins-Burnett, 2005).

Validity scores for the tool must be approximately 0.70 or above. Measures of validity must be conducted on a significant number of children and using an appropriate standardized developmental or social-emotional assessment instrument(s).

4)Established Sensitivity/Specificity: Sensitivity and specificity scores of approximately 0.70 or above.

Current recommended tools that meet these criteria:
Ages and Stages Questionnaire (ASQ) - 2 months–5 years
Battelle Developmental Inventory Screening Tool (BDI-ST) - Birth–95 months
Bayley Infant Neuro-developmental Screen (BINS) - 3 months–2 years
Brigance Screens-II - Birth–90 months
Child Development Inventory (CDI) - 18 months–6 years
Child Development Review-Parent Questionnaire (CDR-PQ) - 18 months–5 years
Infant Development Inventory - Birth–18 months
Parents' Evaluation of Developmental Status (PEDS) - Birth–8 years

Tools NOT Included in This Measure: It is important to note that standardized tools specifically focused on one domain of development [e.g. child's socio-emotional development (ASQ-SE) or autism (M-CHAT)] are not included in the list above as this measure is anchored to recommendations focused on global developmental screening using tools that focus on identifying risk for developmental, behavioral and social delays.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
Indicator 1: Members who turn 12 months of age between January 1 of the measurement year and December 31 of the measurement year
Indicator 2: Members who turn 24 months of age between January 1 of the measurement year and December 31 of the measurement year
Indicator 3: Members who turn 36 months of age between January 1 of the measurement year and December 31 of the measurement year

2a.5 Target population gender: Female, Male
2a.6 Target population age range: First three years of life.

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
One year

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
See 2a4

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** None.

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
NA

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
The measure is stratified by the following ages:
By 12 months (Indicator 1)
By 24 months (Indicator 2)
By 36 months (Indicator 3)

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
NA

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
Step 1: Determine the denominator
Identify the denominator for each age-specific indicator:
Indicator 1: Members who turn 12 months of age between January 1 of the measurement year and December 31 of the measurement year
Indicator 2: Members who turn 24 months of age between January 1 of the measurement year and December 31 of the measurement year
Indicator 3: Members who turn 36 months of age between January 1 of the measurement year and December 31 of the measurement year

Step 2: Determine the numerator
Claims Data:
Children for whom a claim of 96110 was submitted during the measurement year.
Medical Chart:
Children who had documentation in the medical record of developmental screening using a standardized validated tool during the measurement year.Documentation must include a note indicating the standardized tool that was used, the date of screening and evidence that the tool was completed and scored.

Step 3: Calculate the age-specific indicators (1-3) by dividing the numerator by the denominator and multiplying by 100 to get a percentage.

Step 4. Create the measure of screening based on the age-specific measures.
Numerator: Numerator for Indicator 1 + Numerator for Indicator 2+ Numerator for Indicator3 (Divided by)
Denominator: Denominator for Indicator 1 + Denominator for Indicator 2+ Denominator for Indicator 3

Step 5: Multiply by 100 to get the percentage.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
Comparison of proportions and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
If administrative data are used, the entire population is used for the denominator. For hybrid measures (administrative plus chart review data sources),a random sample can be drawn. Preferred sample size would be 411.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Paper medical record/flow-sheet, Electronic administrative data/claims, Electronic Health/Medical Record

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
Claims data, Medical chart

2a.26-28 Data source/data collection instrument reference web page URL or attachment:

2a.29-31 Data dictionary/code table web page URL or attachment:

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Population: states, Program: QIO, Program: Other To evaluate the Assuring Better Child Development Efforts across the state and within specific communities of the state. These efforts were either within multiple practices or within specific geographic regions.

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)***:** No formal reliability testing has been conducted, however measures of screening have been collected with the ABCD community since 2003. The ABCD Screening Academy states built off work from the ABCD I and ABCD II efforts and the learnings gathered about medical chart abstraction instructions needed in order to ensure reliability (e.g. specific tools must be listed, scoring must clarified etc), Additionally, the ABCD Screening Academy conducted.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)***:**
See 2b.1

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
See 2b.1

2b
C☐
P☐
M☐
N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)***:** No formal validity testing has been conducted. Measures of screening have been collected within the ABCD community since 2003. The ABCD Screening Academy states built off work from the ABCD I and ABCD II efforts and the learnings gathered about medical chart abstraction instructions needed in order to ensure validity.

2c.2 Analytic Method *(type of validity & rationale, method for testing)***:**
No formal validity testing has been conducted. The measure presented is based on the shared learnings from NCQA's development work and CAHMI's technical assistance consulting to the ABCD Screening Academy. A detailed summary of the methodologies used by each state is attached and findings from the ABCD II can be found here (http://cahmi.org/ViewDocument.aspx?DocumentID=72). An executive summary can be found here: http://www.nashp.org/sites/default/files/screening_academy_results.pdf. Overall, 24 states Medicaid agencies (21 state/territories in the ABCD Screening Academy and then the states in ABCD II that were not in the Screening Academy) used claims or medical chart data using similar methods to those proposed here and found the data to be valid for assessing screening sensitive to the quality improvement efforts they were conducting.

It is important to note that some states have found that claims data can be inaccurate for screening that

2c
C☐
P☐
M☐
N☐

occurred in systems in which the payment is capitated (and therefore individual claims related to specific aspects of care provided are not submitted) or for health care providers for whom screening is not paid separately (e.g. Federally Qualified Health Centers). Thus, we recommend hybrid data collections for those settings. 2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*: The measure has been validity in being sensitive to quality improvement efforts. For those able to report baseline and follow-up data during the time-period of the ABCD Screening Academy, all reported an increase in the percent of children screened using a standardized tool (demonstrating validity and sensitivity). The average increase reported was 58 percentage points.	
2d. Exclusions Justified 2d.1 **Summary of Evidence supporting exclusion(s):** No Exclusions recommended at this time. 2d.2 **Citations for Evidence:** NA 2d.3 **Data/sample** *(description of data/sample and size)*: NA 2d.4 **Analytic Method** *(type analysis & rationale)*: NA 2d.5 **Testing Results** *(e.g., frequency, variability, sensitivity analyses)*: NA	**2d** C☐ P☐ M☐ N☐ NA☐
2e. Risk Adjustment for Outcomes/ Resource Use Measures 2e.1 **Data/sample** *(description of data/sample and size)*: No risk adjustment recommended at this time. 2e.2 **Analytic Method** *(type of risk adjustment, analysis, & rationale)*: NA 2e.3 **Testing Results** *(risk model performance metrics)*: NA 2e.4 **If outcome or resource use measure is not risk adjusted, provide rationale:** The measure assesses prevention and wellness in a general population; risk adjustment is not indicated.	**2e** C☐ P☐ M☐ N☐ NA☐
2f. Identification of Meaningful Differences in Performance 2f.1 **Data/sample from Testing or Current Use** *(description of data/sample and size)*: A detailed summary of the findings can be found here: http://www.nashp.org/sites/default/files/screening_academy_results.pdf. 2f.2 **Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*: Findings from the ABCD Screening Academy: Overall, 21 state/territories Medicaid agencies used claims or medical chart data using similar methods to those proposed here and found the data to be valid for assessing screening and sensitive to the quality improvement efforts they were conducting. For those able to report baseline and follow-up data during the time-period of the ABCD Screening Academy, all reported an increase in the percent of children screened using a standardized tool (demonstrating validity and sensitivity). The average increase reported was 58 percentage points. 2f.3 **Provide Measure Scores from Testing or Current Use** *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*: A detailed summary of the findings can be found here:	**2f** C☐ P☐ M☐ N☐

http://www.nashp.org/sites/default/files/screening_academy_results.pdf. Baselines findings amongst the screening academy states, prior to intervention, was between 0-20%. Follow-up results demonstrated sig. improvements, with an average increase of 58 percentage points.	
2g. Comparability of Multiple Data Sources/Methods 2g.1 **Data/sample** *(description of data/sample and size)*: 2g.2 **Analytic Method** *(type of analysis & rationale)*: 2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)*:	**2g** C☐ P☐ M☐ N☐ NA☐
2h. Disparities in Care 2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: To assess screening in each of the 1st three years of life, the measure should be stratified by age of child: • Indicator 1: Members who turn 12 months of age between January 1 of the measurement year and December 31 of the measurement year • Indicator 2: Members who turn 24 months of age of age between January 1 of the measurement year and December 31 of the measurement year • Indicator 3: Members who turn 36 months years of age between January 1 of the measurement year and December 31 of the measurement year A review of data provided to the CAHMI by the ABCD states stratified by age showed differences in screening rates, with rates for indicator 2 (screened by 24 months of age) being higher than the other age-specific indicators. 2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** NA	**2h** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* C☐ P☐ M☐ N☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	Eval Rating
3a. Meaningful, Understandable, and Useful Information 3a.1 **Current Use:** In use 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*): All 24 states involved in the ABCD efforts implemented measures of standardized screening, a majority of which used medical chart and claims data. A majority continue to track screening using similar methodologies that are based on claims/medical chart data and stratified by age of child. States such as Illinois are using specifications that are nearly identical to the measure described in this submission. 3a.3 **If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years***):**	**3a** C☐ P☐ M☐ N☐

This is a state-level measure designated as a measure in the CHIPRA Core Set. Many of the ABCD states are using the measure to drive quality improvement efforts at the state, community, program and practice-level. **Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) **3a.4 Data/sample** *(description of data/sample and size)*: Expert panel, other stakeholders. In addition, NCQA and CAHMI convened the ABCD Community (N=43 people from states) in a focus-group-like telephone interview that also included staff from the National Academy of State Health Policy (NASHP). Participants reviewed the specifications and provided comments on the specifications presented. In addition, participants were able to submit written comments to the CAHMI. **3a.5 Methods** *(e.g., focus group, survey, QI project)*: CAHMI vetted the measures with the ABCD community. NCQA vetted the measure concepts and specifications with other stakeholder groups, including the National Association of State Medicaid Directors and the American Academy of Pediatrician's Quality Improvement Innovation Network. **3a.6 Results** *(qualitative and/or quantitative results and conclusions)*: Stakeholders and potential measure users found the measure to be understandable and actionable for quality improvement and to inform policy-level improvements. The ABCD Community call participants indicated that the measure, as anchored to the goals for the CHIPRA measure on screening and for state-level assessment, was specified correctly and gave suggestions for further clarifying the measure. NCQA and CAHMI modified the measure specifications to ensure the lessons from all feedback were incorporated into the measure. The primary issue raised on the call for which there are different approaches taken by states in the number and types of measures on screening collected is whether additional measures should be collected that assess complimentary, but separate, screening for specific domains of development (e.g. social-emotional screening, autism screening). It was clarified that this measure is focused on general, standardized screening for children at risk for developmental, behavioral and social delays. Some states collect additional measures that capture screening for risk of social-emotional delays or for autism and they felt that a future priority should be placed on measures that assess this type of screening.	
3b/3c. Relation to other NQF-endorsed measures **3b.1 NQF # and Title of similar or related measures:** The National Quality Form has endorsed the Promoting Healthy Development Survey (PHDS) [NQF # 0011}, which includes a measure of screening for risk of developmental, behavioral and social delays based on family surveys. In additon, NCQA will be submitting in tandem a similar developmental screening measure specified for a physician population and the CAHMI submitted a measure based on the National Survey of Children´s Health. The information provided below is specific to comparisons of this measure to the NQF Measure #0011.	
(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): **3b.2 Are the measure specifications harmonized? If not, why?** The measure of screening based on the PHDS is complementary, but different from this measure in the following ways: • Data source: The screening measure in the PHDS is based on parental report (which the CAHMI validated). • Denominator: The PHDS sampling is anchored to children who have had 1 or more well-child visit. This measure is harmonized with the NCQA physician-level developmental screening measure.	**3b** C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** This measure is derived from a different data source (claims, medical chart) and is solely focused on a state-level unit of analysis. This measure complements the physician-level measure of screening submitted	**3c** C☐ P☐ M☐ N☐

by NCQA for the call for quality measures through which this measure is being submitted. 5.1 **If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:** NA	**NA☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** **C☐** **P☐** **M☐** **N☐**
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes 4a.1-2 **How are the data elements that are needed to compute measure scores generated?** Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition), Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** **C☐** **P☐** **M☐** **N☐**
4b. Electronic Sources 4b.1 **Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) No 4b.2 **If not, specify the near-term path to achieve electronic capture by most providers.** The CAHMI has worked with two private practice settings and Kaiser Permanente Northwest to develop standardized processes to enter in screening results to allow for tracking and monitoring of a child's development and also to allow for measurement of the quality of care provided. These templates could be disseminated for use by others.	**4b** **C☐** **P☐** **M☐** **N☐**
4c. Exclusions 4c.1 **Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No 4c.2 **If yes, provide justification.**	**4c** **C☐** **P☐** **M☐** **N☐** **NA☐**
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences 4d.1 **Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.**	**4d** **C☐** **P☐** **M☐** **N☐**
4e. Data Collection Strategy/Implementation 4e.1 **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** The measure presented is based on the shared learnings from NCQA's development work and CAHMI's	**4e** **C☐** **P☐** **M☐** **N☐**

technical assistance consulting to the ABCD Screening Academy. A detailed summary of the methodologies used by each state is attached and findings from the ABCD II can be found here (http://cahmi.org/ViewDocument.aspx?DocumentID=72). An executive summary can be found here: http://www.nashp.org/sites/default/files/screening_academy_results.pdf. Overall, 24 states Medicaid agencies (21 state/territories in the ABCD Screening Academy and then the states in ABCD II that were not in the Screening Academy) used claims or medical chart data using similar methods to those proposed here and found the data to be valid for assessing screening sensitive to the quality improvement efforts they were conducting. Additionally, some states have found that claims data can be inaccurate for screening that occurred in systems in which the payment is capitated (and therefore individual claims related to specific aspects of care provided are not submitted) or for health care providers for whom screening is not paid separately (e.g. Federally Qualified Health Centers). Thus, we recommend hybrid data collections for those settings. 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** Collecting measures from medical charts is time-consuming and can be burdensome. Adapting this measure in electronic health records may relieve some of this burden 4e.3 **Evidence for costs:** 4e.4 **Business case documentation:**	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility***, met?** **Rationale:**	**4** **C** ☐ **P** ☐ **M** ☐ **N** ☐
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited** ☐
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y** ☐ **N** ☐ **A** ☐
CONTACT INFORMATION	
Co.1 **Measure Steward (Intellectual Property Owner)** Co.1 **Organization** Child and Adolescent Health Measurement Initiative, 707 SW Gaines Drive, Mail Code CDRC-P, Portland, Oregon, 97239 Co.2 **Point of Contact** Colleen, Reuland, MS, reulandc@ohsu.edu, 503-494-0456-	
Measure Developer If different from Measure Steward Co.3 **Organization** Child and Adolescent Health Measurement Initiative, 707 SW Gaines Drive, Mail Code CDRC-P, Portland, Oregon, 97239 Co.4 **Point of Contact** Colleen, Reuland, MS, reulandc@ohsu.edu, 503-494-0456-	
Co.5 **Submitter If different from Measure Steward POC** Sepheen, Byron, MHS, Byron@ncqa.org, 202-955-3573-, National Committee for Quality Assurance	

Co.6 Additional organizations that sponsored/participated in measure development
To ensure clarity, the measure is being co-submitted by CAHMI and NCQA.

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The NCQA Child Health MAP advised NCQA during measure development of the Physician-Level measure for which this measure is harmonized. They evaluated the way staff specified measures, assessed the content validity of measures, and reviewed field test results. As you can see from the list, the MAP consisted of a balanced group of experts, including representatives from pediatricians, family physicians, researchers, Medicaid CHIP offices and health plans.

NCQA Child Health MAP:
Jeanne Alicandro
Barbara Dailey
Denise Dougherty, PhD
Ted Ganiats, MD
Foster Gesten, MD
Nikki Highsmith, MPA
Charlie Homer, MD, MPH
Jeff Kamil, MD
Elizabeth Siteman
Mary McIntyre, MD, MPH
Virginia Moyer, MD, MPH, FAAP
Lee Partridge
Xavier Sevilla, MD, FAAP
Michael Siegal
Jessie Sullivan

Secondly, states/consultants from the ABCD community participated in a conference call review of the measure, which included staff from the National Academy of State Health Policy. Below is a list of persons that attended the call and/or gave written comments to the CAHMI and what state they were from:

Mary Alice Lee, CT
Chris Kus, NY
Linda Dann, MI
Jenny Salesa, MI
Sonni Vierling, IA
Mary Noel, MT
Maude Holt, Washington DC
Molly Carpenter, VI
Julie Doetsch, IL
Laura McGuinn, OK
Trish Blake, CO
Viki Brant, AL
Carole Lannon, OH
Kevin Stanford, OH
Harvey Doremus, OH
Kim Elliot, AZ
Kathy Mayfield-Smith, SC
William Golden, AK
Molly Emmons, OR
Patrician Mack, IL
Kristi Plotner, MS
Russell Frank, VT
Eileen Bennet, CO
Mary Lundtke, MI

Margaret Bennett, NJ Lillian Garcia, AZ Suzanne Yockelson (Consultant- UCI) Amy Fine (Health Policy/Program Consultant- Washington DC) Anita Berry, IL Mary Timmerman, AL Juanona Brewster, IL Michelle Urban, WI Patrician Hagan, CT Vicky Hosey, IL Sheena Olson, AK Gina Robinson, CO Kim Davis Allen, AL Theresa Thomas, AL Sandra Watson, IL Susan Castellano, MN Charles Gallia, OR Norma Everret, Nemours Thus, our measures are the result of consensus from a broad and diverse group of stakeholders.
Ad.2 If adapted, provide name of original measure: **Ad.3-5 If adapted, provide original specifications URL or attachment**
Measure Developer/Steward Updates and Ongoing Maintenance **Ad.6 Year the measure was first released:** 2010 **Ad.7 Month and Year of most recent revision:** **Ad.8 What is your frequency for review/update of this measure?** **Ad.9 When is the next scheduled review/update for this measure?**
Ad.10 Copyright statement/disclaimers:
Ad.11 -13 Additional Information web page URL or attachment:
Date of Submission (*MM/DD/YY*)**:** 09/23/2010

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1382 **NQF Project:** Child Health Quality Measures 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 Measure Title: Percentage of low birthweight births
De.2 Brief description of measure*:* The percentage of births with birthweight <2,500 grams
1.1-2 Type of Measure: Outcome **De.3 If included in a composite or paired with another measure, please identify composite or paired measure** N/A
De.4 National Priority Partners Priority Area: Population health **De.5 IOM Quality Domain**: Safety **De.6 Consumer Care Need:** Getting better, Staying healthy, Living with illness

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* **A.1 Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes **A.2 Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** **A.3 Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary **A.4 Measure Steward Agreement attached:**	**A** **Y**☐ **N**☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** **Y**☐ **N**☐

C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Other Improving infant health and reducing infant mortality	**C** Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☐ N☐
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers, Leading cause of morbidity/mortality, High resource use, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Here is a quotation from reference 1 below: "Infants born at low birth weight (LBW) – conventionally defined as a birth weight less than 2,500 grams – experience severe health and developmental difficulties that can impose substantial costs on society. For example, the expected costs of delivery and initial care of a baby weighing 1000 grams at birth can exceed $100,000 (in year 2000 dollars), and the risk of death within one year of birth is over one-in-five. Even among babies weighing 2000-2100 grams, who have comparatively low mortality rates, an additional pound (454 grams) of weight is still associated with a $10,000 difference in hospital charges for inpatient services." 1a.4 **Citations for Evidence of High Impact:** 1. Almond D, Chay KY, Lee DS. The costs of low birthweight. National Bureau of Economic Research, Working Paper 10552, June 2004. Available at: http://www.nber.org/papers/w10552. 2. Petrou S, Eddama O, Mangham L. Arch Dis Child Fetal Neonatal Ed. 2010 May 20. [Epub ahead of print] A structured review of the recent literature on the economic consequences of preterm birth. 3. Dorling J, D'Amore A, Salt A, et al. Data collection from very low birthweight infants in a geographical region: methods, costs, and trends in mortality, admission rates, and resource utilisation over a fi ve-year period. Early Hum Dev 2006;82:117–24.	**1a** C☐ P☐ M☐ N☐

4. Tommiska V, Tuominen R, Fellman V. Economic costs of care in extremely low birthweight infants during the fi rst 2 years of life. Pediatr Crit Care Med 2003;4:157–63.

5. Russell RB, Green NS, Steiner CA, et al. Cost of hospitalization for preterm and low birth weight infants in the United States. Pediatrics 2007;120:e1–9.

6. Mistry H, Dowie R, Franklin RC, Jani BR. Acta Paediatr. 2009 Jul;98(7):1123-9. Epub 2009 Apr 30.Costs of neonatal care for low-birthweight babies in English hospitals.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: The percentage of low birthweight infants has increased by 22% from 6.7% of births in 1984 to 8.2% in 2007. Since a substantially lower percentage of low birthweight births has already been achieved in the United States in the past, there appears to be no reason why a substantially lower level could not be achieved again.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
The US percentage of low birthweight births has increased by 22% since 1984. The US percentage of low birthweight births is substantially higher than in most other developed countries.

1b.3 Citations for data on performance gap:
1. Martin JA, Hamilton BE, Sutton PD et al. Births: Final data for 2007. National vital statistics reports, vol 58 no 24, August 2010.

2. Organization for Economic Cooperation and Development Health Data 2010. Available at: http://www.ecosante.org/index2.php?base=OCDE&langh=ENG&langs=ENG&sessionid=

1b.4 Summary of Data on disparities by population group:
In 2007, the percentage of low birthweight births was 13.9% for non-Hispanic black women, 1.9 times the 7.3% for non-Hispanic white women. The higher percentage of low birthweight infants for non-Hispanic black women accounts for much of their elevated infant mortality risk, when compared to non-Hispanic white women.

1b.5 Citations for data on Disparities:
Martin JA, Hamilton BE, Sutton PD et al. Births: Final data for 2007. National vital statistics reports, vol 58 no 24, August 2010.

1b
C ☐
P ☐
M ☐
N ☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** In 2006, the infant mortality rate for low birthweight infants was 55.38 infant deaths per 1,000 live births, 25 times the rate of 2.24 for infants born weighing 2,500 grams or more. For very low birthweight infants (<1,500 grams), the infant mortality rate was 240.44 infant deaths per 1,000 live births, 107 times the rate for normal birthweight infants. Source: Mathews T.J., MacDorman MF. Infant mortality statistics from the 2006 period linked birth/infant death data set. National vital statistics reports vol 58 no 17. Hyattsville, MD: April 2010.

1c.2-3. Type of Evidence: Other Linked birth and infant death certificate data for the entire US population

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Prenatal care can assist women in eliminating or successfully managing pregnancy risk factors such as smoking during pregnancy, inadequate weight gain, pregnancy-associated diabetes, and others. Women who resolve pregnancy risks can substantially lower their chance of having a low birthweight infant. Source: Ricketts SA, Murray EK, Schwalberg R. Reducing low birthweight by resolving risks: results from Colorado´s prenatal plus program. Am J Public Health. 2005 Nov;95(11):1952-7. Epub 2005 Sep 29.

1c
C ☐
P ☐
M ☐
N ☐

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):** Women who smoke during pregnancy and who have late or no prenatal care have a higher percentage of low birthweight births, and higher infant mortality rates. This is from national birth certificate data and these relationships have been stable in the data each year since we began measuring these variables. **1c.6 Method for rating evidence:** To my knowledge, the evidence has not been formally rated, but since it is based on an accurate population data source, and these relationships have been found each year for the past 30 years of data collection, I believe that they constitute high quality evidence. **1c.7 Summary of Controversy/Contradictory Evidence:** None **1c.8 Citations for Evidence (***other than guidelines***):** Martin JA, Hamilton BE, Sutton PD et al. Births: Final data for 2007. National vital statistics reports, vol 58 no 24, August 2010. **1c.9 Quote the Specific guideline recommendation (***including guideline number and/or page number***):** Healthy People 2010 Objective 16-10: Reduce low birthweight and very low birthweight. **1c.10 Clinical Practice Guideline Citation:** http://www.healthypeople.gov/hpscripts/KeywordResult.asp?n269=269&n362=362&Submit=Submit **1c.11 National Guideline Clearinghouse or other URL:** http://www.healthypeople.gov/hpscripts/KeywordResult.asp?n269=269&n362=362&Submit=Submit **1c.12 Rating of strength of recommendation (***also provide narrative description of the rating and by whom***):** N/A **1c.13 Method for rating strength of recommendation** (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:** **1c.14 Rationale for using this guideline over others:** Scientific acceptability. Widely-used measure. Easy to measure, use and understand.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** The number of babies born weighing <2,500 grams at birth in the United States **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** A calendar year (for example, 2010)	**2a-specs** **C☐** **P☐** **M☐** **N☐**

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions***):**
Data are directly available from public-use data files of national birth certificate data produced by the National Center for Health Statistics.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
All births in the United States

2a.5 Target population gender: Female, Male
2a.6 Target population age range: Under 1 year (365 days) of age

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
A calendar year (for example, 2010)

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Data are directly available from public-use data files of national birth certificate data produced by the National Center for Health Statistics.

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** None

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
None

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
No stratification of this variable is required. However, the variable can be stratified by all variables available on the birth certificate, including maternal race/ethnicity, age, education, for example.

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
N/A

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Other Percentage
2a.20 Interpretation of Score: Better quality = Lower score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
The number of births weighing <2,500 grams/Total births at any birthweight * 100

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
percentage of low birthweight births significantly higher or lower than the national average.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
This measure is based on the complete population of 4.2 million births in the United States each year. As such, it is not a sample and is not subject to sampling limitations.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Public health data/vital statistics

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
National Center for Health Statistics, Natality Detail file. These publicly available data files contain individual record data for the 4.2 million births in the United States each year. Data are from birth certificates.

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL not needed http://www.cdc.gov/nchs/data_access/VitalStatsOnline.htm

2a.29-31 Data dictionary/code table web page URL or attachment: URL not needed ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/DVS/natality/UserGuide2007.pdf

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Population: national, Population: regional/network, Population: states, Population: counties or cities

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*
Other United States, states, counties

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)***:** Many studies have found a high degree of reliability in the percent low birthweight measure from the birth certificate. I describe two examples below.

Study 1. 110 birth certificates were randomly sampled from each of 4 different counties in New York State. Total sample size = 440. Birth certificates were traced back to their hospital of origin and birth certificate data were directly compared to hospital medical record data.

Study 2. A random sample of birth certificates from 20 hospitals in the Cleveland metropolitan area. Total sample size =33,616

2b.2 Analytic Method *(type of reliability & rationale, method for testing)***:**
Study 1 - Direct comparison of birth certificate data to medical records. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were computed.

Study 2 – Direct comparison of birth certificate data to data from medical records collected by the Cleveland Health Quality Choice Initiative, a voluntary regional initiative to compare hospital performance. Concordance, sensitivity, specificity, PPV, and NPV were computed.

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
Study 1 – Low birthweight (<2,500 g) - Sensitivity, specificity, PPV, NPV all 100%.

Study 2 – Birthweight <3000 g or > 3000 g. Concordance 99%, sensitivity 99% specificity 99% PPV 100% NPV 98%.

Source: Study 1 – Roohan PJ, Josberger RE, Acar J et al. Validation of birth certificate data in New York State. Journal of Community Health 2003;28:335-46.

Study 2 – DiGuiseppe DL, Aron DC, Ranbom L et al. Reliability of birth certificate data: A multi-hospital comparison to medical records information. Maternal and Child Health Journal 2002;6:169-179.

2b
C☐
P☐
M☐
N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)***:** Many studies have found a high degree of validity in the percent low birthweight measure from the birth certificate. I describe two examples below.

Study 1. 110 birth certificates were randomly sampled from each of 4 different counties in New York State.

2c
C☐
P☐
M☐
N☐

Total sample size = 440. Birth certificates were traced back to their hospital of origin and birth certificate data were directly compared to hospital medical record data.

Study 2. A random sample of birth certificates from 20 hospitals in the Cleveland metropolitan area. Total sample size =33,616

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
Study 1 - Direct comparison of birth certificate data to medical records. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were computed.

Study 2 – Direct comparison of birth certificate data to data from medical records collected by the Cleveland Health Quality Choice Initiative, a voluntary regional initiative to compare hospital performance. Concordance, sensitivity, specificity, PPV, and NPV were computed.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
Study 1 – Low birthweight (<2,500 g) - Sensitivity, specificity, PPV, NPV all 100%.

Study 2 – Birthweight <3000 g or > 3000 g. Concordance 99%, sensitivity 99% specificity 99% PPV 100% NPV 98%.

Source: Study 1 – Roohan PJ, Josberger RE, Acar J et al. Validation of birth certificate data in New York State. Journal of Community Health 2003;28:335-46.

Study 2 – DiGuiseppe DL, Aron DC, Ranbom L et al. Reliability of birth certificate data: A multi-hospital comparison to medical records information. Maternal and Child Health Journal 2002;6:169-179.

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
No exclusions needed.

2d.2 Citations for Evidence:

2d.3 Data/sample *(description of data/sample and size)*:

2d.4 Analytic Method *(type analysis & rationale)*:

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:

2d
C☐
P☐
M☐
N☐
NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: No risk adjustment needed.

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:

2e.3 Testing Results *(risk model performance metrics)*:

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:

2e
C☐
P☐
M☐
N☐
NA☐

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: Data are based on the complete population of 4.2 million birth certificates filed in the United States each year.

2f
C☐
P☐
M☐
N☐

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*: Any statisically significant increase or decrease,using standard methods for significance testing. **2f.3 Provide Measure Scores from Testing or Current Use** *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*: N/A - no scores needed	
2g. Comparability of Multiple Data Sources/Methods **2g.1 Data/sample** *(description of data/sample and size)*: Multiple data sources and/or methods are not needed as birth certificate data provide the gold standard for any measurement of this variable. **2g.2 Analytic Method** *(type of analysis & rationale)*: N/A **2g.3 Testing Results** *(e.g., correlation statistics, comparison of rankings)*: N/A	**2g** C☐ P☐ M☐ N☐ NA☐
2h. Disparities in Care **2h.1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: Measure can detect diferences by all birth certificate variables, for example maternal race/ethnicity, maternal age, maternal education, etc. **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** N/A	**2h** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties*?	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* C☐ P☐ M☐ N☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	Eval Rating
3a. Meaningful, Understandable, and Useful Information **3a.1 Current Use:** In use **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*): Publicly reported in many NCHS publications, such as: Martin JA, Hamilton BE, Sutton PD et al. Births: Final data for 2007. National vital statistics reports, vol 58 no 24, August 2010. http://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58_24.pdf. Widely used by the research and policy community. A Medline search of the term "low birthweight" yields 32,070 articles. **3a.3 If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years***):** Monitoring the percentage of low birthweight births is widely used in quality improvement programs in	**3a** C☐ P☐ M☐ N☐

maternity hospitals, health care systems, by the US Government's Healthy Start Program, and others too numerous to mention. **Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) 3a.4 **Data/sample** *(description of data/sample and size)*: The percentage of low birthweight births is widely reported in the US media, in hospitals throughout the country, and by the research and quality improvement communities. These are commonly understood constructs. I don't know what "testing of interpretability" could be done or would be needed. 3a.5 **Methods** *(e.g., focus group, survey, QI project)*: N/A 3a.6 **Results** *(qualitative and/or quantitative results and conclusions)*: N/A	
3b/3c. Relation to other NQF-endorsed measures 3b.1 **NQF # and Title of similar or related measures:**	
(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): 3b.2 **Are the measure specifications harmonized? If not, why?**	**3b** C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value 3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** 5.1 **If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:**	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes 4a.1-2 **How are the data elements that are needed to compute measure scores generated?** Other Data are from birth certificates filed for each US birth.	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources 4b.1 **Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*)	**4b** C☐ P☐ M☐

Yes 4b.2 **If not, specify the near-term path to achieve electronic capture by most providers.**	N☐
4c. Exclusions 4c.1 **Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No 4c.2 **If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences 4d.1 **Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** N/A - no data problems have been identified or are expected.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation 4e.1 **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** Data are of high quality and no modifications are needed. 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** 4e.3 **Evidence for costs:** 4e.4 **Business case documentation:**	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** C☐ P☐ M☐ N☐
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited** ☐
Steering Committee: Do you recommend for endorsement? **Comments:**	Y☐ N☐ A☐
CONTACT INFORMATION	
Co.1 **Measure Steward (Intellectual Property Owner)** Co.1 **Organization** Division of Vital Statistics, National Center for Health Statistics, CDC, 3311 Toledo Road, Room 7318, Hyattsville, Maryland, 20782	

Co.2 **Point of Contact** Marian, MacDorman, Ph.D., M.A., mfm1@cdc.gov, 301-458-4356-
Measure Developer If different from Measure Steward Co.3 **Organization** Division of Vital Statistics, National Center for Health Statistics, CDC, 3311 Toledo Road, Room 7318, Hyattsville, Maryland, 20782 Co.4 **Point of Contact** Marian, MacDorman, Ph.D., M.A., mfm1@cdc.gov, 301-458-4356-
Co.5 **Submitter If different from Measure Steward POC** Marian, MacDorman, Ph.D., M.A., mfm1@cdc.gov, 301-458-4356-, Division of Vital Statistics
Co.6 **Additional organizations that sponsored/participated in measure development**
ADDITIONAL INFORMATION
Workgroup/Expert Panel involved in measure development Ad.1 **Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.**
Ad.2 **If adapted, provide name of original measure:** Ad.3-5 **If adapted, provide original specifications URL or attachment**
Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 **Year the measure was first released:** Ad.7 **Month and Year of most recent revision:** Ad.8 **What is your frequency for review/update of this measure?** This measure has been used in vital statistics data since the 1930´s. Data quality reviewed annually Ad.9 **When is the next scheduled review/update for this measure?**
Ad.10 **Copyright statement/disclaimers:**
Ad.11 -13 **Additional Information web page URL or attachment:** not needed http://www.cdc.gov/nchs/data_access/VitalStatsOnline.htm
Date of Submission (*MM/DD/YY*)**:** 08/26/2010