

NATIONAL QUALITY FORUM—Composite Measure Testing (subcriteria 2a2, 2b2-2b7, 2c)

Measure Number (if previously endorsed): 0011

Composite Measure Title: [Promoting Health Development Survey \(PHDS\)](#)

Date of Submission: [12/20/2016](#)

Composite Construction:

- ☒ Two or more individual performance measure scores combined into one score
- ☐ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)

Instructions: Please contact NQF staff before you begin.

- If a component measure is submitted as an individual performance measure, the non-composite measure testing form must also be completed and attached to the individual measure submission.
- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **Sections 1, 2a2, 2b2, 2b3, 2b5, 2b7, and 2c must be completed.**
- **For composites with outcome and resource use measures**, section **2b4** also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b7) and composites (2c) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 25 pages (*including questions/instructions*; minimum font size 11 pt; do not change margins). **Contact NQF staff if more pages are needed.**
- Contact NQF staff regarding questions. Check for resources at [Submitting Standards webpage](#).
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 7.0 of the Measure Testing Attachment and the 2016 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing ¹⁰ demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing ¹¹ demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12](#)

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, 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). [13](#)

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15](#) and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16](#) **differences in performance;**

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements 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 of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by

identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

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

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. 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 percent v. 75 percent) 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 less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing (e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation. **If different data sources are used for different components in the composite, indicate the component after the checkbox. If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.**)

Measure Specified to Use Data From: (must be consistent with data sources entered in S.23)	Measure Tested with Data From:
<input type="checkbox"/> abstracted from paper record	<input type="checkbox"/> abstracted from paper record
<input type="checkbox"/> administrative claims	<input type="checkbox"/> administrative claims
<input type="checkbox"/> clinical database/registry	<input type="checkbox"/> clinical database/registry
<input type="checkbox"/> abstracted from electronic health record	<input type="checkbox"/> abstracted from electronic health record
<input type="checkbox"/> eMeasure (HQMF) implemented in EHRs	<input type="checkbox"/> eMeasure (HQMF) implemented in EHRs
<input checked="" type="checkbox"/> other: Patient reported data	<input checked="" type="checkbox"/> other: Patient reported data

1.2. If an existing dataset was used, identify the specific dataset (the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

For the current submission, we are using 2 data sources: (1) Online PHDS dataset and (2) A study funded by Health Resources and Services and Administration's (HRSA) Maternal and Child Health Bureau. (Patient Centered Quality Improvement of Well-Child Care, Final Report, Supported by a grant from the Maternal and Child Health Bureau Research Grants Program, Health Resources and Services Administration, R40 MC08959 03-00.) The previous submission was based on data obtained from clinics and health plans.

1.3. What are the dates of the data used in testing? The current submission is based on two data sources: HRSA Evaluation Study: 2010-2012 and the Online PHDS dataset: 2008-2016

1.4. What levels of analysis were tested? (testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan)

Measure Specified to Measure Performance of: (must be consistent with levels entered in item S.26)	Measure Tested at Level of:
<input checked="" type="checkbox"/> individual clinician	<input checked="" type="checkbox"/> individual clinician
<input checked="" type="checkbox"/> group/practice	<input checked="" type="checkbox"/> group/practice
<input type="checkbox"/> hospital/facility/agency	<input type="checkbox"/> hospital/facility/agency
<input type="checkbox"/> health plan	<input type="checkbox"/> health plan
<input type="checkbox"/> other: Click here to describe	<input type="checkbox"/> other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample)

The Online PHDS contains n=5,670 surveys reporting on quality of care provided by individual pediatricians and primary care providers from 88 clinics in 36 states. Over one fourth of completed surveys (20.6%, n=1736) are linked to individual providers. All surveys are linked to clinics.

Participation: Voluntary self-selection process based on knowledge and interest in PHDS and quality improvement in their practice or health system.

Table 1a and 1b show prevalence of quality measures for top 5 providers (1a) and 5 clinics (1b).

Table 1a. Quality measures by providers (Online PHDS) – Top 5 providers out of 372 providers

Provider ID	Anticipatory guidance and parental education		Parenting information and resources in community		Ask about concerns and addressing concerns		Family centered care		Assessment of smoking, drug and alcohol use and safety		Assessment of family psychosocial well-being	
	Total n	% met criteria (n)	Total n	% met criteria (n)	Total n	% met criteria (n)	Total n	% met criteria (n)	Total n	% met criteria (n)	Total n	% met criteria (n)
948	50	62.0% (31)	50	70.0% (35)	50	86.0% (43)	50	92.0% (46)	50	6% (3)	50	6% (3)
802	39	74.4% (29)	42	81.0% (34)	42	90.5% (38)	-	-	42	21.4% (9)	42	16.7% (7)
747	32	81.3% (26)	32	87.5% (28)	32	65.6% (21)	-	-	33	42.4% (14)	33	39.4% (13)
756	32	71.9% (32)	35	85.7% (30)	35	74.3% (26)	-	-	35	11.4% (4)	35	34.3% (12)

927	29	86.2% (25)	28	92.9% (26)	29	96.6% (28)	29	100% (29)	29	34.5% (10)	29	44.8% (13)
978	-	-	-	-	-	-	28	78.6% (22)	-	-	-	-
954	-	-	-	-	-	-	23	82.6% (19)	-	-	-	-
942	-	-	-	-	-	-	20	90.9% (20)	-	-	-	-

Table 1b. Quality measures by clinic (Online PHDS) - Top 5 clinics out of 88 clinics

Clinic ID (description)	Anticipatory guidance and parental education		Parenting information and resources in community		Ask about concerns and addressing concerns		Family centered care		Assessment of smoking, drug and alcohol use and safety		Assessment of family psychosocial well-being	
	Total n	% met criteria (n)	Total n	% met criteria (n)	Total n	% met criteria (n)	Total n	% met criteria (n)	Total n	% met criteria (n)	Total n	% met criteria (n)
146 (urban pediatric PC clinic)	411	59.6% (245)	393	78.4% (208)	400	70.0% (280)	414	85.7% (355)	404	29.5% (119)	405	28.6% (116)
39 (urban children's hospital)	287	67.2% (193)	285	73.7% (210)	286	78.0% (223)	-	-	284	20.8% (59)	286	17.5% (50)
126 (rural pediatric clinic)	287	61.7% (177)	276	76.1% (210)	285	74.4% (212)	287	88.2% (253)	287	12.9% (37)	285	14.4% (41)
152 (rural pediatric and women clinic)	196	60.7% (119)	185	77.3% (143)	196	58.7% (115)	201	76.1% (153)	188	30.9% (58)	188	39.9% (75)
64 (urban outpatient clinic)	163	60.1% (98)	165	83.0% (137)	170	70.0% (119)	-	-	169	20.7% (35)	171	12.3% (21)
132 (urban pediatric outpatient clinic)	-	-	-	-	-	-	146	88.4% (129)	-	-	-	-
172 (rural community health center)	-	-	-	-	-	-	138	81.9% (113)	-	-	-	-

HRSA Evaluation Study:

A. *Population studied.* 3 pediatric clinics in Oregon: 1) a rural site, Pediatricians of North Bend Medical Center (4 pediatricians), 2) an urban site, Broadway Medical Clinic (8 pediatricians), and 3) an urban site, The Children's Clinic (12 pediatricians).

B. Sample selection. All pediatricians and select clinic and office staff participated in relevant baseline and follow up data collection. The study inclusion criteria were used to determine which parents/guardians of children were invited to participate in the interventions and/or evaluation from each participating study site:

- Parent has a well-child visit scheduled at this intervention site for one or more of their children.
- The child is scheduled for their 4-month to 3-year-old well-child visit and, therefore, is between the ages of 4 and 40 months (e.g. 40 month old children could be there for their 3 year well-child visit)
- The parent can read and understand English and is able to complete the intervention and evaluation tools.
- For intervention, the parent was able to access the online version of the Plan My Child's Well-Visit tool and the online evaluation survey.

551 surveys were completed at baseline 551 and 275 surveys were completed at follow-up.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample)

Online PHDS: Children age 3-48 months of age whose parents completed the online publicly available PHDS survey were included in the testing. Between 2008-2016, we received 5670 completed patient surveys. Table 1 shows demographic and health characteristics of the children for selected top 5 providers and 5 clinics by each quality measure.

Table 2a. Children's characteristics for selected providers (out of 372 providers)

Characteristics	All children (372 providers)		Selected provider IDs							
	n	%	948	802	747	756	927	978	954	942
All children	5670	100	50	42	33	35	29	28	23	22
Age groups										
3-8 months	2301	40.6	52.0 (26)	57.1 (24)	39.4 (13)	54.3 (19)	62.1 (18)	14.3 (4)	39.1 (9)	54.5 (12)
9-18 months	1936	34.1	38.0 (19)	38.1 (16)	36.4 (12)	37.1 (13)	24.1 (7)	53.6 (15)	43.5 (10)	45.5 (10)
19-48 months	1433	25.3	10.0 (5)	4.8 (2)	24.2 (8)	8.6 (3)	13.8 (4)	32.1 (9)	17.4 (4)	0
Race/ethnicity										
Hispanic	1984	35.0	10.2 (5)	2.4 (1)	6.3 (2)	2.9 (1)	28.6 (8)	25.9 (7)	13.0 (3)	10.5 (2)
White non-Hispanic	2857	50.4	75.5 (37)	97.6 (41)	90.6 (29)	91.2 (31)	50.0 (14)	70.4 (19)	82.6 (19)	73.7 (14)
Non-White non-Hispanic	565	10.0	14.3% (7)	0	3.1 (1)	5.9 (2)	21.4 (6)	3.7 (1)	4.3 (1)	15.8 (3)
Adult survey responds education level										
Did not complete high school	627	11.1	0	0	0	0	3.6 (1)	0	0	0

Characteristics	All children (372 providers)		Selected provider IDs							
			948	802	747	756	927	978	954	942
	n	%	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Completed high school or higher education	4800	84.7	100 (49)	100 (42)	100 (32)	100 (35)	96.4 (27)	100 (28)	100 (23)	100 (21)
CSHCN status										
Non-CSHCN	5093	89.9	90.0 (45)	81.0 (34)	90.9 (30)	94.3 (33)	82.8 (24)	92.9 (26)	87.0 (20)	86.4 (9)
CSHCN	577	10.2	10.0 (5)	19.0 (8)	9.1 (3)	5.7 (2)	17.2 (5)	7.1 (2)	13.0 (3)	13.6 (3)

Table 2b. Children's characteristics for selected clinics (out of 88 clinics)

Characteristics	All children (88 clinics)		Selected clinic IDs						
			146	39	126	152	64	132	172
	n	%	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
All children	5670	100	419	287	287	202	171	146	140
Age groups									
3-8 months	2301	40.6	27.5 (115)	78.7 (226)	41.1 (118)	37.6 (76)	46.2 (79)	36.3 (53)	29.3 (41)
9-18 months	1936	34.1	39.5 (165)	21.3 (61)	38.3 (110)	33.7 (68)	34.5 (59)	43.2 (63)	42.9 (60)
19-48 months	1433	25.3	33 (138)	0	20.6 (59)	28.7 (58)	19.3 (33)	20.5 (30)	27.9 (39)
Race/ethnicity									
Hispanic	1984	35.0	66.4 (267)	12.9 (36)	9.7 (25)	71.8 (135)	2.4 (4)	60.0 (84)	94.7 (124)
White non-Hispanic	2857	50.4	24.9 (100)	76.6 (213)	78.7 (203)	19.1 (36)	92.1 (152)	32.1 (45)	4.6 (6)
Non-White non-Hispanic	565	10.0	8.7 (35)	10.4 (29)	11.6 (30)	9.0 (17)	5.5 (9)	7.9 (11)	0.8 (1)
Adult survey responds education level									
Did not complete high school	627	11.1	16.3 (65)	2.2 (6)	100 (281)	36.5 (66)	1.2 (2)	6.4 (9)	22.4 (30)
Completed high school or higher education	4800	84.7	83.7 (333)	97.8 (272)	0	63.5 (115)	98.8 (166)	93.6 (132)	77.6 (104)
CSHCN status									
Non-CSHCN	5093	89.9	90.0 (376)	91.6 (263)	89.9 (258)	91.1 (184)	92.4 (158)	95.9 (140)	92.1 (129)
CSHCN	577	10.2	10.0 (42)	8.4 (24)	10.1 (29)	8.9 (18)	7.6 (13)	4.1 (6)	7.9 (11)

HRSA Evaluation Study: 551 surveys were completed at baseline 551 and 275 surveys were completed at follow-up.

Table 2c. Children's Characteristics, by study site

Characteristics	All Study Sites	The Children's Clinic	North Bend Medical Center	Broadway Medical Clinic
	% (n)	% (n)	% (n)	% (n)
All children				
Age				
3-9 months	40.0 (988)	37.8 (414)	43.5 (192)	40.9 (382)
10-18 months	33.6 (831)	35.9% (394)	30.6 (135)	32.3 (302)
19-48 months	26.4 (652)	26.3 (288)	25.9 (114)	26.8 (250)
Race				
Hispanic	7.4 (146)	8.6 (61)	8.8 (36)	5.7 (49)
White	88.2 (1748)	87.4 (623)	87.2 (355)	89.4 (770)
Asian	2.3 (45)	3.1 (22)	0.7 (3)	2.3 (20)
Multiple or other	2.1 (42)	1.0 (7)	3.2 (13)	2.6 (22)
Insurance type				
Private or private and public	82.2 (1764)	89.4 (703)	44.7 (193)	93.5 (868)
Public only (includes Medicaid, Medicare, CHIP, and Military)	16.4 (353)	9.0 (71)	53.0 (229)	5.7 (53)
Other insurance type	0.5 (10)	0.6 (5)	0.5 (2)	0.3 (3)
Uninsured	0.9 (19)	0.9 (7)	1.9 (8)	0.4 (4)
At risk of developmental delay				
Low/no risk	77.7 (1547)	74.5 (489)	74.3 (326)	81.5 (759)
High/moderate risk	22.3 (452)	25.5 (656)	25.7 (113)	18.5 (172)

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Online PHDS data was used for reliability testing and stratification analysis.

Validity findings were taken from a peer-reviewed publication on PHDS measurement (Bethell C, Reuland CH, Halfon N, Schor EL. Measuring the quality of preventive and developmental services for young children: national estimates and patterns of clinicians' performance. Pediatrics. 2004 Jun;113(6 Suppl):1973-83.)

Performance analysis was conducted using both the Online PHDS and HRSA Evaluation Study data.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Online PHDS: Children's age, sex, race/ethnicity, and respondent (parent) age, race/ethnicity, and education level

HRSA Study: Children's age, race-ethnicity, and insurance type

2a2. RELIABILITY TESTING

2a2.1. What level of reliability testing was conducted? *(may be one or both levels)*

Note: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.

☒ **Performance measure score** (e.g., signal-to-noise analysis)

2a2.2. Describe the method of reliability testing and what it tests *(describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used)*

Internal consistency reliability: Cronbach's alpha for each quality measure and PHDS composite measure was calculated using the online PHDS data collected from 88 clinics. Cronbach's alpha is the most widely used in health care research when multiple-item measures of a concept or construct are employed. The acceptable values of alpha ranges from 0.70 to 0.95.

Internal consistency was tested at clinic level data only due to sample size limitation for provider level data. The quality of care measures of interest for this analysis included 1) whether parents' needs were met with regard to age-specific areas of anticipatory guidance and parent education; 2) whether families received anticipatory guidance information from their doctor and health care providers about parenting information and resources in community; 3) if parents received family-centered care; 4) if providers talk with parents about safety, alcohol and drug use in the family; and 5) if providers talk with parents about their own well-being (psychosocial screening). Table 3 presents Cronbach's alpha for the quality measures by selected clinics.

Survey questions on anticipatory guidance and parental education are age-specific. Therefore, testing was done by children's age group (3-9, 10-18 and 19-48 months). The anticipatory guidance and parent education questions can be also grouped by topical area.

2a2.3. What were the statistical results from reliability testing? *(e.g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis)*

Table 3. PHDS Quality Measures: Content, Scoring and Internal Consistency by selected clinics (Online PHDS based on 88 clinics data)-2008-2016

PHDS Quality Measure	What is measured	Scoring	Internal Consistency (Cronbach's Alpha) for each selected clinic (by ID)							
			All Clinics	146	39	126	152	64	132	172
1. Anticipatory Guidance and Parent Education	Four multi-part items assess whether general and age specific anticipatory guidance topics are addressed. Includes feeding and nutrition, sleeping and physically caring for child, safety and injury prevention, child growth, development, communication and behavior	Mean score on a multi-item scale								
	By age group:									
	- 3-9 months		.90*	.95*	.96*	.88*	.91*	.78*	.89*	.90*
	- 10-18 months		.89*	.91*	.93*	.87*	.93*	.70	.82*	.79*
	- 19-48 months		.90*	.91*	.94*	.79*	.92*	.74*	.88*	.92*
	By topical area:									
	- Physical		.77*	-	-	-	-			-
	- Behavioral or developmental		.86*	-	-	-	-			-
	- Injury prevention		.73*	-	-	-	-			-
2. Parenting Information and Resources in Community	One multi-part item assesses whether families receive anticipatory guidance information from their doctor and health care providers about parenting information and resources in community.	Mean score on a multi-item scale	.76*	.81*	.73*	.57	.81*	.59	.77*	.61

PHDS Quality Measure	What is measured	Scoring	Internal Consistency (Cronbach's Alpha) for each selected clinic (by ID)							
			All Clinics	146	39	126	152	64	132	172
3. Family Centered Care	Two multi-part items assess the degree to which care is provided in a family centered manner. Includes respect, understanding specific needs of child and concerns of parent, asking how feeling as a parent, understand family "culture" and talking about resources and issues in the community	Mean score on a multi-item scale	.88*	.88*	-	.80*	.90*	.90*	.75*	.78*
4. Assessment of smoking, drug and alcohol use and safety in the family	Two items within the family assessment scale assess whether providers talk with parents about smoking, alcohol and drug use in the family	Average proportion answering "yes" to each assessment topic.	.81*	.80*	.81*	.79*	.79*	.86*	.81*	.60
5. Assessment of family psychosocial well-being	Two multi-part items assess whether providers talk with parents about their own well-being and safety within the family. Topics include depression, history of abuse, stressors and emotional support, firearms, feeling safe in home.	Average proportion answering "yes" to each assessment topic.	.88*	.91*	.84*	.81	.92*	.87*	.88*	.91*
6. Ask About and Address Parental Concern†	One multi-part item assesses whether doctors and health care provider ask parental concerns and addressed the parental concerns	Proportion of parents responses - Not appropriate	N/A							

PHDS Quality Measure	What is measured	Scoring	Internal Consistency (Cronbach's Alpha) for each selected clinic (by ID)							
			All Clinics	146	39	126	152	64	132	172
		for psychometric analysis								

*Met criteria for reliability and internal consistency.

†Variable used for reporting purposes.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.e., what do the results mean and what are the norms for the test conducted?)

Psychometric analyses demonstrated that the five of the six individual PHDS quality measure scales have strong internal consistency (mean Cronbach's alpha ranges 0.73-0.90 across the measures (overall)). The sixth measure -- *Asking and Addressing Parental Concerns* -- is not appropriate for psychometric analysis. The PHDS provides psychometrically reliable assessment of the provision of nationally recommended well-child care with strong internal consistency.

2b2. VALIDITY TESTING

Note: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.

2b2.1. What level of validity testing was conducted?

Validity testing was completed for original endorsement and then re-evaluated for each improvement project in which it has been employed. A standard, multistage process was used to ensure validity of the survey items/measures:

- Focus groups and in-depth cognitive interviews were conducted throughout the survey development process;
- A review of literature identified through Medline or during key informant interviews; and,
- Three Advisory Groups comprised of pediatricians, family practitioners, consumer representatives, public health experts, and researchers, regularly reviewed and provided input on the identification of quality measurement topics and the development of the PHDS.

More detailed information on psychometric testing of the quality measures including factor analysis, testing of concurrent validity and others can be found in the peer-reviewed publication: Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. *Pediatrics*. 2001 May;107(5):1084-94.

☒ **Composite performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

☐ **Systematic assessment of content validity**

☒ **Validity testing for component measures** (check all that apply)

Note: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.

☐ **Endorsed (or submitted) as individual performance measures**

☒ **Critical data elements** (data element validity must address ALL critical data elements)

☒ **Empirical validity testing of the component measure score(s)**

- ☐ **Systematic assessment of face validity of component measure score(s) as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used*)

A “gold standard” does not exist for determining the criterion validity of patient-reported measures of quality.¹ However, to ensure the validity of the PHDS survey results, we followed rigorous procedures representing best practices within the field to develop the survey. To ensure the content validity of measures of parent experiences, we used qualitative methods, including both focus groups and cognitive interviews, to inform development and evaluation of the survey questions. Validity testing was conducted for the initial NQF submission in 2007 and the tool has not been changed since then.

Focus groups with families aimed to identify the aspects of health care quality that are important to parents in the area of preventive care for their children. In-depth cognitive testing of the draft PHDS instrument was conducted with 15 families representing a range of racial, income, and education groups, as well as different types of health insurance coverage, age of child, age and sex of parent, and number of children in family. Readability assessments indicated that the PHDS is written at the 8th- to 9th- grade reading level. Focus groups and cognitive interviews with 35 health care providers in Vermont and Washington and 20 parents of young children in Vermont were conducted to inform item-reduction, administration specifications, and reporting templates. Survey modifications were made based on findings in order to improve the reliability, validity and cognitive ease of the PHDS items.

We used quantitative methods, including measure-to-measure correlations, to evaluate the validity the survey survey. We also conducted a literature review and environmental scan of best practices.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Using behavior coding methods, for each item in the PHDS, instances where the respondent required clarification or did not appropriately answer an item were noted. Also, items where the interviewer had difficulty asking the question without edits to the wording were noted. Data analysis of PHDS data was used to inform item-reduction. Content was revised and refined iteratively with each set of interviews.

Cognitive testing confirmed the readability of the PHDS for people across a range of educational levels. Parents reporting positive parenting behaviors had significantly higher scores on the anticipatory guidance quality measure compared with parents not reporting positive behaviors. Parents who reported that their questions on specific anticipatory guidance topics were answered were more likely to report higher confidence in related parenting activities (odds ratio [OR]: 5.9, 95% confidence interval [CI]: 3.4-10.2; OR: 8.3, 95% CI: 5-13.8) and were less likely to report concerns about their child's development in related areas compared with parents who reported they wished they had talked more with their child's doctor about these topics (OR: 0.46, 95% CI: 0.29-0.72; OR: 0.58, 95% CI: 0.37-0.89).

Factor analysis demonstrated a strong factor structure within the PHDS. Each of the survey items used to construct the 6 PHDS scale-based quality measures were used in the factor analysis. The results of the validity testing has been published in peer reviewed journals:

1. Bethell C1, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. *Pediatrics*. 2001 May;107(5):1084-94.
2. Christina Bethell, PhD, MPH, MBA*; Colleen H. Peck Reuland, MS*; Neal Halfon, MD, MPH‡; Edward L. Schor, Measuring the Quality of Preventive and Developmental Services for Young Children: National Estimates and Patterns of Clinicians' Performance. *PEDIATRICS* Vol. 113 No. 6 June 2004.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.e., what do the results mean and what are the norms for the test conducted?)

The PHDS provide psychometrically valid and reliable assessments of the provision of nationally recommended preventive care services, with strong construct validity (mean factor loading: 0.69).

2b3. EXCLUSIONS ANALYSIS

Note: Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.

☒ no exclusions — skip to section [2b4](#)

2b3.1. Describe the method of testing exclusions and what it tests (describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used)

Not applicable

2b3.2. What were the statistical results from testing exclusions? (include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores)

Not applicable

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (i.e., the value outweighs the burden of increased data collection and analysis. **Note:** If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion)

Not applicable

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

Note: Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section [2b5](#).

2b4.1. What method of controlling for differences in case mix is used? (check all that apply)

☐ Endorsed (or submitted) as individual performance measures

☐ No risk adjustment or stratification

☐ Statistical risk model

☒ Stratification by risk categories

☐ Other, [Click here to enter description](#)

2b4.1.1 If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

Not applicable.

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of $p < 0.10$; correlation of x or higher; patient factors should be present at the start of care)

Child and parent demographics such as age, sex, race-ethnicity, parent behavior, CSHCN screener and follow-up for children at risk can be used for stratification. Several studies have documented differences in access and quality of care provided to children, as well as in parent-reported satisfaction with care.¹⁻² One study found: "Overall, 94.0% of parents reported 1 or more unmet needs for parenting guidance, education, and screening by pediatric clinician(s) in 1 or more of the content of care areas evaluated. Uninsured children and children aged 18 to 35 months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need in each of the 4 areas of care. ... There are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors."³

References:

1. Halfon N, Regalado M, Sareen H, Inkelas M, Reuland CH, Glascoe FP, Olson LM. Assessing development in the pediatric office. *Pediatrics*. 2004 Jun;113(6 Suppl):1926-33.
2. Weech-Maldonado R, Morales LS, Spritzer K, Elliott M, Hays RD. Racial and ethnic differences in parents' assessments of pediatric care in Medicaid managed care. *Health Serv Res*. 2001 Jul;36(3):575-94.
3. Bethell C, Reuland CH, Halfon N, Schor EL. Measuring the quality of preventive and developmental services for young children: national estimates and patterns of clinicians' performance. *Pediatrics*. 2004 Jun;113(6 Suppl):1973-83.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Variation in quality measure scores according to a child's child demographic and health-related characteristics such as age; gender; race/ethnicity; level of risk for developmental, behavioral, or social delays; and children with special health care needs status; as well as mother's educational level based on online PHDS data are shown below.

Table 4: Quality measures by demographics and other characteristics for all clinics.

Table 4a. Anticipatory guidance and parental education by child demographics and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	1347	60.4%
9-18 months	1104	57.4%
19-48 months	889	63.1%
p values (Pearson chi-square)	-	0.003
Gender		
Male	372	59.7%
Female	364	57.7%
p values (Pearson chi-square)	-	0.48
Race/ethnicity		
Hispanic	1150	58.9%
White non-Hispanic	1735	61.9%
Black non-Hispanic	58	56.9%
Asian non-Hispanic	47	42.7%
Other/Multi race, non-Hispanic	52	60.5%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	341	55.3%
Completed high school or higher education	2879	60.9%
p values (Pearson chi-square)	-	0.01
CSHCN status		
Non-CSHCN	3002	60.1%
CSHCN	338	59.3%
p values (Pearson chi-square)	-	0.70
At risk for developmental delay (online only)		
Low/No risk	1394	63.0%
High/Moderate risk	423	56.2%
p values (Pearson chi-square)	-	0.001

Table 4b. Parenting information and resources in community by child demographics and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	1656	74.9%
9-18 months	1432	76.3%
19-48 months	1074	78.7%
p values (Pearson chi-square)	-	0.04
Gender		
Male	472	77.3%
Female	466	75.9%
p values (Pearson chi-square)	-	0.59
Race/ethnicity		
Hispanic	1383	72.8%
White non-Hispanic	2218	79.2%
Black non-Hispanic	72	73.5%
Asian non-Hispanic	71	66.4%
Other/Multi race, non-Hispanic	65	76.5%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	421	71.4%
Completed high school or higher education	3597	77.0%
p values (Pearson chi-square)	-	0.002
CSHCN status		
Non-CSHCN	3745	76.5%
CSHCN	417	74.9%
p values (Pearson chi-square)	-	0.39
At risk for developmental delay (online only)		
Low/No risk	1763	79.7%
High/Moderate risk	516	70.3%
p values (Pearson chi-square)	-	<0.0001

Table4c. Family centered care by child demographics and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	1309	78.0%
9-18 months	1393	81.4%
19-48 months	1138	80.1%
p values (Pearson chi-square)	-	0.04
Gender		
Male	505	81.2%
Female	497	78.6%
p values (Pearson chi-square)	-	0.26
Race/ethnicity		
Hispanic	1530	78.7%
White non-Hispanic	1779	82.9%
Black non-Hispanic	71	74.7%
Asian non-Hispanic	70	66.0%
Other/Multi race, non-Hispanic	60	76.9%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	431	70.2%
Completed high school or higher education	3264	81.8%
p values (Pearson chi-square)		<0.0001
CSHCN status		
Non-CSHCN	3461	80.3%
CSHCN	379	75.8%
p values (Pearson chi-square)	-	0.02
At risk for developmental delay (online only)		
Low/No risk	1346	84.1%
High/Moderate risk	437	74.3%
p values (Pearson chi-square)	-	<0.0001

Table 4d. Assessment of smoking, drug and alcohol use and safety (asked about all items) in the family by child demographics and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	705	31.3%
9-18 months	531	28.1%
19-48 months	403	29.1%
p values (Pearson chi-square)	-	0.06
Gender		
Male	215	34.7%
Female	210	33.6%
p values (Pearson chi-square)	-	0.67
Race/ethnicity		
Hispanic	723	37.1%
White non-Hispanic	680	23.9%
Black non-Hispanic	32	32.0%
Asian non-Hispanic	29	26.1%
Other/Multi race, non-Hispanic	29	34.5%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	260	42.6%
Completed high school or higher education	1319	27.8%
p values (Pearson chi-square)		<0.0001
CSHCN status		
Non-CSHCN	1464	29.5%
CSHCN	175	31.0%
p values (Pearson chi-square)	-	0.47
At risk for developmental delay (online only)		
Low/No risk	516	23.0%
High/Moderate risk	225	30.2%
p values (Pearson chi-square)	-	<0.0001

Table 4e. Assessment of family psychosocial well-being (asked about all items) by child demographic and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	779	34.7%
9-18 months	432	22.9%
19-48 months	279	20.1%
p values (Pearson chi-square)	-	<0.0001
Gender		
Male	188	30.3%
Female	191	30.5%
p values (Pearson chi-square)	-	0.96
Race/ethnicity		
Hispanic	610	31.3%
White non-Hispanic	664	23.3%
Black non-Hispanic	33	33.0%
Asian non-Hispanic	30	27.3%
Other/Multi race, non-Hispanic	24	28.2%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	227	37.1%
Completed high school or higher education	1213	25.5%
p values (Pearson chi-square)	-	<0.0001
CSHCN status		
Non-CSHCN	1333	26.9%
CSHCN	157	27.6%
p values (Pearson chi-square)	-	0.71
At risk for developmental delay (online only)		
Low/No risk	476	21.3%
High/Moderate risk	216	29.2%
p values (Pearson chi-square)	-	<0.0001

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Pearson's chi-square test was used to compare the prevalence of each quality measure across the stratification characteristics.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to [2b4.9](#)

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*):

Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable

2b4.9. Results of Risk Stratification Analysis:

Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays across all quality measures. For the Anticipatory Guidance and Parental Education (AGPE) measure, care for non-Hispanic white children is most likely to meet scoring criteria. In contrast, non-Hispanic white children are less likely to meet criteria on the Family Assessment (both) measures. Children of lower educated mothers are less likely than those with more educated mothers to have high AGPE or Family Centered Care scores but more likely to have high Family Assessment (both) scores. The similar evidence is found for children at risk.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (*i.e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

Note: *Applies to the composite performance measure.*

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified *(describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Peer-reviewed publication: Comprehensive analyses in performance measure scores are described in peer-reviewed publication: Bethell C, Reuland CH, Halfon N, Schor EL. Measuring the quality of preventive and developmental services for young children: national estimates and patterns of clinicians' performance. *Pediatrics*. 2004 Jun;113(6 Suppl):1973-83.

Online PHDS: Performance scores for the 4 composite quality measures were calculated using the scoring methods described in the PHDS_Final_Appendix, Item #4.

HRSA Study: Quantitative data results for the baseline and follow-up study of the intervention sites (using the HRSA Evaluation Study data) were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the PHDS measures for the baseline and follow-up samples.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? *(e.g., number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined)*

Peer-reviewed publication: Bethell C, Reuland CH, Halfon N, Schor EL. Measuring the quality of preventive and developmental services for young children: national estimates and patterns of clinicians' performance. *Pediatrics*. 2004 Jun;113(6 Suppl):1973-83.

Online PHDS data: Table 1a and 1b above present the proportion of children whose care met for each measure (for top 5 providers and 5 clinics). The proportion of parents who reported discussion of all anticipatory guidance and parental education topics or reported no need of discussion among unaddressed topics ranged 60-86%. Only 6%-45% of parents of young children reported that their child's pediatric clinician discussed psychosocial topics such as parent emotional well-being and partner support in parenting.

The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to report their needs met for anticipatory guidance at the follow-up assessment than at the baseline assessment; and parents were more likely to be asked about one or more psychosocial (family assessment) topics at follow-up. The tables below present comparison of percent of children who received care met the quality care criteria between baseline and follow-up survey data for each measure and overall composite comprehensive care measure.

Table 5a: Anticipatory Guidance & Parent Education Measure by Children's Characteristics
Parent had their needs met on all AGPE topics

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	38.9% (216)	45.2% (146)	0.08
10-18 months	48.5% (208)	45.7% (150)	0.46
19-48 months	55.0% (193)	65.9% (147)	0.01
Race			
Hispanic	46.0% (46)	47.8% (46)	0.86
White	46.2% (475)	51.9% (372)	0.02
Asian	35.7% (10)	52.9% (9)	0.35
Multiple or other	62.5% (15)	33.3% (6)	0.12
Insurance type			
Private or private and public	46.4% (502)	49.9% (339)	0.15
Public only (includes Medicaid, Medicare, CHIP, and Military)	44.7% (85)	54.9% (89)	0.07
Other insurance type	(3)	(1)	-
Uninsured	50.0% (6)	(4)	-
At risk of developmental delay			
Low/no risk	47.5% (487)	52.1% (285)	0.09
High/moderate risk	40.7% (114)	44.4% (76)	0.49

Table 5b. Asking about Parent's Concerns about Development Measure, by Children's Characteristics
Parent was asked if they had concerns about their child's development

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	64.6% (357)	73.7% (235)	0.01
10-18 months	78.6% (319)	76.6% (246)	0.59
19-48 months	80.3% (282)	82.3% (181)	0.59
Race			
Hispanic	72.0% (72)	84.1% (37)	0.14
White	72.5% (745)	76.4% (542)	0.07
Asian	75.0% (21)	70.6% (12)	0.74
Multiple or other	87.5% (21)	77.8% (14)	0.44
Insurance type			
Private or private and public	72.4% (784)	75.5% (509)	0.16
Public only (includes Medicaid, Medicare, CHIP, and Military)	78.4% (149)	82.2% (129)	0.42
Other insurance type	71.4% (5)	(3)	-
Uninsured	50.0% (6)	100% (7)	0.04
At risk of developmental delay			
Low/no risk	72.2% (741)	76.3% (411)	0.09
High/moderate risk	76.2% (214)	82.4% (140)	0.13

Table 5c. Family Assessment*, by Children's Characteristics: ---
Parent was asked about one or more family assessment topics

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	23.9% (132)	45.0% (145)	<0.0001
10-18 months	21.5% (87)	34.4% (111)	<0.0001
19-48 months	29.3% (103)	50.5% (112)	<0.0001
Race			
Hispanic	26.0% (26)	47.8% (22)	0.01
White	24.5% (252)	42.7% (305)	<0.0001
Asian	28.6% (8)	(4)	-
Multiple or other	50.0% (12)	50.0% (9)	1.00
Insurance type			
Private or private and public	22.9% (248)	37.9% (257)	<0.0001
Public only (includes Medicaid, Medicare, CHIP, and Military)	33.5% (64)	39.1% (63)	<0.0001
Other insurance type	(2)	(1)	N/A
Uninsured	(4)	(3)	N/A
At risk of developmental delay			
Low/no risk	24.3% (248)	41.6% (227)	<0.0001
High/moderate risk	26.0% (73)	44.7% (76)	<0.0001

Table 5d. Family Centered Care Measure Comparison by Children's Characteristics
Parent received family-centered care

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	61.9% (343)	65.1% (209)	0.38
10-18 months	64.9% (261)	69.3% (224)	0.23
19-48 months	65.2% (227)	68.0% (151)	0.53
Race			
Hispanic	62.0% (62)	71.1% (32)	0.35
White	64.7% (667)	67.7% (485)	0.20
Asian	39.3% (11)	70.6% (12)	0.07
Multiple or other	66.7% (16)	66.7% (12)	1.00
Insurance type			
Private or private and public	64.9% (704)	67.9% (461)	0.20
Public only (includes Medicaid, Medicare, CHIP, and Military)	58.1% (111)	66.9% (107)	0.10
Other insurance type	(4)	(2)	N/A
Uninsured	66.7% (8)	(4)	N/A
At risk of developmental delay			
Low/no risk	65.5% (669)	69.4% (379)	0.12
High/moderate risk	57.4% (159)	57.9% (99)	0.92

Table 5e. Comprehensive Care Measure, by Children's Characteristics

Met comprehensive care measure

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	6.6% (36)	16.2% (51)	<0.0001
10-18 months	9.5% (38)	10.7% (34)	0.62
19-48 months	13.5% (47)	25.5% (56)	<0.0001
Race			
Hispanic	6.0% (6)	23.3% (10)	0.01
White	9.4% (96)	16.7% (118)	<0.0001
Asian	- (2)	- (1)	N/A
Multiple or other	29.2% (7)	- (2)	N/A
Insurance type			
Private or private and public	8.8% (95)	14.3% (96)	0.001
Public only (includes Medicaid, Medicare, CHIP, and Military)	11.6% (22)	27.3% (42)	<0.0001
Other insurance type	(1)	(0)	N/A
Uninsured	(2)	(1)	N/A
At risk of developmental delay			
Low/no risk	9.3% (94)	17.5% (94)	<0.0001
High/moderate risk	9.4% (26)	11.8% (20)	0.43

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.e., what do the results mean in terms of statistical and meaningful differences?)

Findings show significant variation in performance across content of care areas. Parents generally reported receiving the highest quality of care in the areas of Family Centered Care and lowest quality of care in the areas of AGPE and Family Assessment (both family psychosocial screening and asking about alcohol, drug use and safety within the family).

The HRSA study findings based on PHDS survey responses (clinic level of analysis) indicate that the PHDS tool plays an important role in improvement in engagement, efficiency and team effectiveness in delivering well child care.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

Note: Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.

If only one set of specifications, this section can be skipped.

Note: This item is directed to measures that are risk-adjusted (with or without SDS factors) **OR** to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). **Comparability is not required when comparing performance scores with and**

without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

Not applicable.

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (*i.e., what do the results mean and what are the norms for the test conducted?*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

Note: *Applies to the overall composite measure.*

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The PHDS questions were developed based on several rounds of cognitive interviews with parents to ensure quality of responses appropriate to the questions and minimize missing responses.

Online PHDS: Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the PHDS quality measure scoring protocol, if a parent answered less than half of the items in the quality measure, their score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Online PHDS data shows that 2.6% of parents who started the survey did not complete the survey. Table 6 presents the frequency of missing values for each composite measure.

Table 6. The frequency of missing values for each composite measure, top 5 clinics

Quality measures	Overall	Clinic ID				
		146	39	126	152	64
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Anticipatory guidance and parental education	1.9 (107)	1.7 (7)	0	0	3.0 (6)	4.7 (8)
Parenting information and resources in community	3.8% (217)	6.0 (25)	3.8 (11)	0.7 (2)	8.4 (17)	3.5 (6)
Ask about concerns and addressing concerns	2.4% (138)	4.3 (18)	0.7 (3)	0.3 (1)	3.0 (6)	0.6 (1)
Family centered care	3.6% (205)	1.0 (4)	-	0	0.5 (1)	-
Assessment of smoking, drug and alcohol use and safety	2.5% (142)	3.1 (13)	0.3 (1)	0.7 (2)	6.9 (14)	0
Assessment of family psychosocial well-being	2.6% (149)	3.3 (14)	1.0 (3)	0	6.9 (14)	1.2 (2)
Missing to ANY of 5 quality measures	8.6% (503)	9.8% (41)	9.4% (27)	7.0% (6)	13.9% (28)	10.5% (18)
Missing to all 5 quality measures	0.2% (13)	0.1% (1)	0	0	1.5% (3)	0

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias? (i.e., *what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Information about non-respondent is not available to compare with who responded the survey since online PHDS is publicly available tool. Overall, all quality measures had less than 4% of missing cases, ranging 0-8.4% across top 5 clinics. Few missing cases to all quality measures suggest that the measure level results unlikely to be biased by non-response to the survey questions.

2c. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH

Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification*)

Pearson correlation coefficients were calculated across all measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality measures provide unique information.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., correlations, contribution of each component to the composite score, etc.; if no empirical analysis, identify the components that were considered and the pros and cons of each)

Table 6. PHDS Quality Measures: Intercorrelations among composites

Pearson Correlation Coefficients among PHDS Quality Measures (online PHDS)

Correlation among PHDS quality measures						
	Anticipatory Guidance and Parent Education	Parenting Information and Resources in Community	Family Centered Care	Ask About and Address Parental Concern	Assessment of smoking, drug and alcohol use and safety in the family	Assessment of family psychosocial well-being
Anticipatory Guidance and Parent Education						
Parenting Information and Resources in Community	.34					
Family Centered Care	.52	.33				
Ask About and Address Parental Concern	.16	.23	.14			
Assessment of smoking, drug and alcohol use and safety in the family	.16	.12	.13	.07		
Assessment of family psychosocial well-being	.19	.11	.16	.09	.54	

Average correlation: 0.34

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)

The strength of the observed correlations were not so high as to suggest redundancy across measures (average correlation: 0.34). The highest correlation observed was between the “Assessment of family psychosocial well-being” & “Assessment of smoking, drug and alcohol use and safety in the family” (0.54) and “anticipatory guidance from providers” & the “family-centered care” measures (0.52).

Conclusion: Each of the four individual quality measures provides unique information about performance. Regardless of the population group or the aspect of health care assessed, the quality of health care rarely can be represented accurately by either a single composite performance measure or by assessing whether a single recommended service is provided.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

Not applicable

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification*)

Not applicable

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (*e.g., results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

Not applicable

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (*i.e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

Not applicable