Testing Pediatric Oral Health Performance Measures in the Florida and Texas Medicaid and CHIP Programs

FINAL REPORT

Prepared for the
Dental Quality Alliance

Prepared by
Jill Boylston Herndon, PhD
University of Florida

August 2013

Contents

Purpose

The purpose of this report is to summarize the goals, methodology, high-level results, and key outcomes of the validation testing conducted by the University of Florida for the *Dental Caries in Children: Prevention and Disease Management DQA Measure Set.*

Background

In 2012, the Dental Quality Alliance (DQA) proposed a Starter Set of Pediatric Oral Health Performance Measures that could be calculated using administrative data (hereafter referred to as Pediatric Starter Set).[1] The DQA prepared draft specifications for eleven proposed measures. Measure specifications included exclusion criteria, denominator criteria, and numerator criteria as well as the process flows for calculating the denominators, numerators and resulting rates. The measures were proposed for calculation at the plan and program levels (e.g., Medicaid and CHIP). A multidisciplinary research team at the University of Florida (UF Team) was selected to conduct feasibility, reliability and validity testing of the measures through a competitive request for proposal (RFP) process.[2] The UF Team is described in Appendix A. Table 1 summarizes the eleven measures that were tested.

The DQA secured financial support for this project through a grant from the ADA Foundation. The University of Florida secured support from the Florida Agency for Health Care Administration, the Florida Healthy Kids Corporation, and the Texas Health and Human Services Commission, which collectively provided Florida and Texas Medicaid and CHIP data for measure testing. The DQA additionally secured support from DentaQuest to provide commercial data to the University of Florida for measure testing.

The RFP requested that testing be conducted in two phases. Phase 1 focused on testing that would allow the DQA R&D Committee to finalize the denominator criteria for each measure. Within this phase, there were two key components: (1) determining the enrollment requirements to be applied (for each of the eleven measures) and (2) evaluating methodologies for identifying children at elevated caries risk (for the four preventive services measures). Phase 2 focused on producing the measure rates for the measures as specified after Phase 1 determinations were made and completing feasibility, reliability, and validity testing.

The contract was awarded effective December 1, 2012, and testing was conducted during the period December 2012 through June 2013. Presentation of the results was made to the full DQA on July 19, 2013, which approved the first ten measures.[3] The finalized measurement set is posted on the DQA website.[4] This report summarizes the data, processes, methodologies, results, and outcomes of testing the Pediatric Starter Set. A comprehensive set of appendices reflecting the data generated during the project is on file with the DQA. These materials are intended for internal DQA use and not for public dissemination. The DQA and University of Florida are jointly preparing presentations and publications for the purposes of public dissemination.

All data sources and testing methodologies were approved by the University of Florida Institutional Review Board.

Table 1. Original DQA Pediatric Starter Set Measures

Utilization of Services	**Description:** Percentage of all enrolled children who received at least one dental (or oral health) service within the reporting year.
Preventive Services	**Description:** Percentage of a. all enrolled b. enrolled children who received at least one dental (or oral health) service who are at "elevated" caries risk (e.g. "moderate" or "high") who received topical fluoride application and/or sealants within the reporting year.
Treatment Services	**Description:** Percentage of a. enrolled children b. enrolled children who received at least one dental service who received treatment service within the reporting year.
Oral Evaluation	**Description**: Percentage of a. all enrolled children b. enrolled children who received at least one dental (or oral health) service who received a comprehensive or periodic oral evaluation within the reporting year.
Sealants in 6 – 9 years	**Description:** Percentage of a. enrolled children b. enrolled children who received at least one dental (or oral health) service in the age category of 6-9 years at "elevated" caries risk (e.g. "moderate" or "high") who received a sealant on a permanent first molar tooth within the reporting year.
Sealants in 10 – 14 years	**Description:** Percentage of a. enrolled children b. enrolled children who received at least one dental (or oral health) service in the age category of 10-14 years at "elevated" caries risk (e.g. "moderate" or "high") who received a sealant on a permanent second molar tooth in the reporting year.
Topical Fluoride Intensity	**Description:** Percentage of a. all enrolled b. enrolled children who received at least one dental service who are at "elevated" caries risk (e.g. "moderate" or "high") who received at least one topical fluoride application within the reporting year.
Usual Source of Services	**Description:** Percentage of a. all enrolled in two consecutive years b. enrolled children who received at least one dental (or oral health) service in both years who received care from the same practice or clinical entity in both years.
Care Continuity	**Description:** Percentage of a. all enrolled in two consecutive years b. enrolled children who received at least one dental (or oral health) service in both years who received a comprehensive or periodic oral evaluation in the reporting year and in the year prior to the reporting year.
Per Enrollee/User Cost of Clinical Services	**Description:** Total amount that is paid on direct provision of care per a. enrolled child b. enrolled child who accessed [dental/ oral health] services within the reporting year.
Percentage of Child Healthcare Expenditures	**Description:** Percentage of child health expenditures that is expended on [dental/oral health] care for the reporting year.

Data Sources

Administrative enrollment and claims data from the following programs were used:

- Florida Medicaid (only dental fee-for-service data were available for Phase 1 testing),
- Florida CHIP,
- Texas Medicaid,
- Texas CHIP, and
- DentaQuest commercial data (available for Phase 2 testing).

The Institute for Child Health Policy (ICHP) at the University of Florida has been the external evaluator for the Florida and Texas Medicaid and CHIP programs since 2000 and houses more than 10 years of administrative enrollment, claims, and encounter data (both medical and dental). These rich datasets provided the opportunity to test the proposed measure set for Medicaid and CHIP at the program and plan levels in two of the largest and most diverse states in the United States.[5,6] Florida and Texas account for 15% of all children enrolled in Medicaid nationally.[7] Moreover, these states have significant representation of African-American and Hispanic populations, which disproportionately experience low access to dental care.[8] These programs also represent different delivery system models and different forms of provider reimbursement.

Traditionally, the University of Florida has housed Florida Medicaid data mainly for its fee-for-service (FFS) program components. However, Florida Medicaid also had a prepaid dental program in Miami-Dade County during the study period. In addition, children enrolled in the medical managed care program may receive dental services as part of their medical MCO benefit package. (Medical MCO enrollees could receive dental services through their medical MCO if it is offered as a carve-out; otherwise, they received dental services through Medicaid dental FFS or through the prepaid dental program). Upon receiving notification of the award, the University of Florida entered into an additional data sharing agreement with the Florida Agency for Health Care Administration to obtain claims data for children in the prepaid dental program and medical MCOs. The agreement was executed in January 2013, and the datasets were transmitted to the University of Florida in March 2013. Given the time required to execute the data use agreement, develop dataset specifications, intake the data, and conduct basic quality assurance testing, only the FFS data were available for Phase 1 testing.

Because the Pediatric Starter Set was designed for use by all payer types, the DQA secured support from DentaQuest to provide commercial data for testing purposes and the University of Florida agreed to incorporate these data into the testing efforts. A data sharing agreement was executed between Dental Service of Massachusetts, Inc. and the University of Florida for the transmission of a Limited Data Set for research purposes as authorized under 45 CFR 164.514 Section (3) (1)-(e) (3). The agreement was executed in January 2013. The dataset was transmitted to the University of Florida in March 2013; thus, the commercial data were not available for Phase 1 testing.

Both of the data sharing agreements described above were reviewed by the University of Florida Privacy Office and use of the data for this project was approved by the University of Florida Institutional Review Board. Table 2 summarizes the main characteristics of each of the data sources used for measure testing.

Table 2. Summary of Data Sources, Delivery System Models, and Provider Reimbursement CY 2011

	Florida Medicaid*		Florida CHIP	Texas Medicaid	Texas CHIP	Commercial
Medical Delivery Models	Fee-For-Service (FFS), Primary Care Case Management (PCCM), Provider Service Network (PSN), Managed Care (MC)		MC	FFS, PCCM, MC	MC	N/A
Age Range	0-20 years		5-18 years	0-20 years	0-18 years	0-20 years
# Unique Enrollees, CY2011	2,079,616		331,285	3,556,915	889,501	187,065
Dental Delivery Models	FFS	Prepaid Dental Plan - single county pilot+	Dental MCOs	FFS+	Dental MCO - Single Dental Benefit Contractor	Includes members in indemnity and PPO product lines
Payment from Program (e.g., Medicaid/CHIP) to Dental Managed Care Organization (D-MCO)	N/A	Per Member Per Month (PMPM) capitation adjusted for eligibility category and age bands	PMPM Premium Rate – based on competitive bidding and legislated maximum	N/A	PMPM Premium based on historical claims experience and age bands	N/A
Payment from Program or D-MCO to Dental Provider	FFS based on fee schedule	Primary Care Dentists - capitation; Specialists – FFS	Negotiated FFS except for one plan that pays capitation to primary care dentists in two counties	FFS based on fee schedule	FFS	FFS or negotiated FFS based on fee schedule

*Florida Medicaid data reflect data from CY 2010.
+Some enrollees in Medicaid medical managed care may get dental benefits as a carve-out or value added service from the medical MCOs.

Time Frame

For all programs except Florida Medicaid, data from calendar years (CY) 2010 and 2011 were used. Full-year data for CY 2011 were not available for all Florida Medicaid program components when testing commenced; therefore, data from CY 2009 and CY 2010 were used.

Process

Throughout the testing, the UF Team engaged in an iterative and integrated process that involved providing regular and detailed feedback to the DQA R&D Committee during bi-weekly calls. For each bi-weekly call, the project PI (Herndon) prepared an agenda with focused questions, summary data reports, and proposed methodology for the next testing phase. We maintained detailed logs of all of the major issues discussed, decisions made, and action items (Appendix B). Throughout the iterative process, the UF Team assisted with refining the measure specifications, prepared additional data summaries requested by the DQA, and adapted the methodological approaches as needed.

Methodology

A. Finalize Denominator Definitions – Enrollment Interval

The RFP identified four approaches to defining enrollment for dental quality measures: (1) members enrolled for at least 90 continuous days (CMS-416 method for Early and Periodic Screening, Diagnosis, and Treatment – EPSDT- reporting), (2) members continuously enrolled during the measurement year, but having a single break in enrollment of no more than 45 days (HEDIS® method for the measure Annual Dental Visit), (3) members enrolled "any time" during the year, and (4) the "average period of enrollment/person-time" method. Although the RFP specified a pilot using only the utilization of services measure, the UF team expanded the scope to include all single-year measures. In addition, a fifth enrollment definition was added, which was 6-months continuous enrollment during the measurement year, as an intermediate option between 90-day and full-year enrollment.

Each measure specified two denominators: DEN1=all enrollees and DEN2=all enrollees who accessed a dental/oral health service. Both denominators were considered important in evaluating performance. The measure rate associated with the first denominator provides an indicator of access: the percentage of members meeting age and enrollment criteria who received the specified service(s). The measure rate associated with the second denominator indicates the percentage of children meeting age and enrollment criteria and who accessed any type of dental service who received the specified service. Thus, for many measures, DEN2 allows the measure to also serve as a process measure (i.e., are children who access any dental care getting recommended services). The UF Team calculated measure rates for the five different definitions for the eight single-year measures for both denominator definitions for the four public insurance programs. In addition, the UF Team calculated measure rates for oral health and dental/oral health measure variations for the two Medicaid programs. (Oral health services were not applicable to the CHIP programs.) Thus, more than 200 measure denominators, numerators, and rates were calculated and reported in this initial phase.

The UF Team presented data in the form of both tables and charts on (1) the number of children eligible for inclusion under each denominator definition and (2) the resulting measure rates for all single-year measures for each denominator definition (Appendix C). In addition, the UF Team provided feedback on the interpretability and usability of the different definitions. Based on the data presented, the DQA R&D Committee elected to use a six-month continuous enrollment requirement for all measures except two in order to balance sufficient enrollment duration to allow children adequate time to access care with the number of children who drop out of the denominator due to stricter enrollment requirements. The two measures with enrollment requirements different than six months are topical fluoride and per enrollee cost. Because topical fluoride is indicated for as many as four applications per year for children at elevated risk,[9] full-year enrollment was required combined with the number of applications per year in order to assess not only access but also intensity. The R&D Committee and UF Team also determined that specifying the per enrollee cost measure as a per-member-per-month (PMPM) measure would be consistent with existing cost measurement methodologies; therefore, only a single month of enrollment was required for this measure. In addition, the final measure specifications also include a 90-day continuous enrollment requirement for three measures (Utilization of Services, Oral Evaluation, and Treatment Services) to allow for historical comparisons to the CMS-416 measures.

B. Finalize Denominator Definitions – Identify Children at Elevated Risk

The RFP requested that the successful bidder identify children at elevated risk for the purposes of the measures related to caries prevention. Elevated risk was considered to be an important criterion because the evidence for topical fluoride and sealants is strongest for children at moderate to high risk. Based on existing evidence (as determined through a review of the literature), past caries experience was identified as the best predictor of future caries experience. Other predictors of caries risk, such as socio-demographic characteristics were considered, but the evidence base for these indicators is not as strong. Therefore, the focus was on how to identify caries experience using administrative claims data. The UF Team collaborated with the R&D Committee in identifying a set of CDT codes indicative of caries-related treatment as a proxy for caries diagnosis. The UF Team also provided enrollment data to help the R&D Committee determine whether to require enrollment in the year prior to the measurement year or to use prior claims experience if available without requiring prior enrollment. The resulting methodology was to identify children at elevated risk as those who had any of the caries-related treatment codes (Table 3) in the measurement year or in any of three prior years, where enrollment in prior years was not required. The purpose of this approach was clarified as not to identify all children at elevated risk, but to identify a subset of children for whom elevated risk could positively be confirmed.

Table 3: CDT Codes to Identify "Elevated Risk"

D2140	D2394	D2630	D2720	D2791	D3120
D2150	D2410	D2642	D2721	D2792	D3220
D2160	D2420	D2643	D2722	D2794	D3221
D2161	D2430	D2644	D2740	D2799	D3222
D2330	D2510	D2650	D2750	D2930	D3230
D2331	D2520	D2651	D2751	D2931	D3240
D2332	D2530	D2652	D2752	D2932	D3310
D2335	D2542	D2662	D2780	D2933	D3320
D2390	D2543	D2663	D2781	D2934	D3330
D2391	D2544	D2664	D2782	D2940	

C. Other Denominator Definition Considerations

The UF Team raised other clarification questions and recommendations related to the denominator definition for DQA R&D Committee consideration. These additional considerations are summarized below and also are reflected in Appendix B.

1. Reporting Year

There was discussion about whether the specifications should indicate how the reporting year is defined – e.g., calendar year (CY) or federal fiscal year (FFY). The UF Team calculated the Utilization of Services and Oral Evaluation measures, for both the Denominator 1 and Denominator 2 variations, using all 5 denominator definitions, for CY 2011 and FFY 2011, using data from Florida CHIP and Texas CHIP (Appendix D). The differences were less than one percentage point. Therefore, it was determined that the reporting year would be left unspecified, allowing program officials to determine the time frame to be used.

2. Date of Age Calculation

The original specifications provided for calculating age at the start of the reporting year The UF Team recommended changing this to the last day of the reporting year for consistency with existing dental measures (e.g., HEDIS Annual Dental Visit and CMS-Form 416 measures). This recommendation was adopted.

3. Anchor Date

The HEDIS Annual Dental Visit (ADV) measure includes an anchor date for the full-year continuous enrollment criteria, where the anchor date is the last day of the measurement year (i.e., the member must be enrolled on that date for inclusion). The UF Team tested the sensitivity of the rates with and without applying an anchor date criterion for the Utilization of Services measure using data from Florida CHIP and Texas CHIP (Appendix E). The rates with and without the anchor date were within 1 percentage point of each other. In addition, the UF Team sought guidance from National Committee for Quality Assurance (NCQA), which is the HEDIS ADV measure steward, about the rationale for using the anchor date. Based on the data provided and the response from NCQA, it was determined that there was little benefit from including this additional criterion; therefore, anchor dates were not applied to any of the measures.

4. Plan versus Program Enrollment

The measures are intended to be reported at the plan and program levels. Thus, the enrollment criteria must be clear at each reporting level. Specifically, there may be cases in which a member meets the denominator enrollment criteria at the program level, but not at the plan level. It was determined that the rates would be calculated at each level separately and that the program rate for a measure would not be a "roll up" of the plans' rates. Therefore, if a measure has a 6-month enrollment criterion and a Medicaid-enrolled child were enrolled in Plan A for 2 months and Plan B for 4 months, s/he would be included in the denominator for the measure rate calculated at the program level (Medicaid overall), but would not be included in the denominator for the measure rate for Plan A or in the denominator for the measure rate for Plan B.

5. Dental versus Oral Health Services

The Pediatric Starter Set distinguishes between "dental" services and "oral health" services. Based on guidance from the Code of Federal Regulations, "dental" services refers to services provided by or under the supervision of a dentist,[10] and "oral health" services refers to services not provided by or under the supervision of a dentist. For measurement implementation purposes, guidance on how to identify which services are "provided by or under the supervision of a dentist" was required.

Although all plans and programs capture provider type and specialty, many rely on their internally-developed classification systems rather than on standardized codes. The UF Team proposed using the Health Care Provider Taxonomy code set maintained by the National Uniform Claim Committee (NUCC codes), which are nationally standardized codes, to identify which provider categories should be classified as providing "dental" versus "oral health" services.[11] The UF Team and DQA jointly identified a set of NUCC codes that would be classified as being provided by "dental" providers. Services related to the oral cavity provided by providers with a taxonomy code outside of this set would be classified as "oral health" services. The specific codes and their application are described in the User Guide that accompanies the measure set.[12] The UF Team also recommended that guidance be provided regarding the use of "billing" or "rendering" provider for identifying provider type and assisted with developing that guidance, which also is contained in the User Guide. The UF Team provided frequency distributions of the billing and rendering provider types to help inform the guidance provided.

6. Denominators for Two-Year Measures

Consistent with other measures in the Pediatric Starter Set, a 6-month continuous enrollment requirement was specified for the two-year measures; the enrollment requirement applies in each of two consecutive years – 6 months continuous enrollment in the reporting year and 6 months continuous enrollment in the year prior to the reporting year. The UF Team conducted enrollment analyses and sensitivity analyses of the rates for the two-year measures to address two potential concerns described below

A. Proximity of Qualifying Visits

The first potential concern applied to both two-year measures, Usual Source of Services and Continuity of Care, and was whether children would be likely to have qualifying visits in each year that were in close proximity – for example, if a child had a qualifying visit in December of Year 1 and January of Year 2. The concern in this case was that the two visits are so close together that the measure would not capture the true intent of examining access to care over time.

The UF Team evaluated the gap between visits for Florida CHIP and Texas Medicaid for both of the two-year measures (Appendix F). For the Care Continuity measure, <1% of children had a gap of <6 months between the two visits (i.e., >99% had gaps between the two visits of $\geq$6 months). Requiring a minimum gap of six months affected the measure rate by less than one-half of one percentage point. For the Usual Source of Services measure, <1.5% of children meeting the age and enrollment criteria and <5% of children who visited the same practice in both years had a gap of <6 months between two qualifying visits. Requiring a minimum gap of six months affected the measure rate by 0.9 percentage point in one program and by 1.2 percentage point in the other.

B. Enrollment Gaps for Eligible Children

The second potential concern was with respect to the Care Continuity measure and was that children may qualify for the measure but have a significant enrollment gap – for example, if a child were continuously enrolled during the first six months of Year 1 and the second six months of Year 2 with a gap of one year in between the two enrollment spells. The concern in this case was whether a child with a significant enrollment gap could really be considered to have continuous care.

The UF Team evaluated the enrollment gap for all four public insurance programs (Appendix F). The percentage of children eligible for the measure with an enrollment gap >6 months ranged from 0.95%-2.8%; the percentage with an enrollment gap >9 months ranged from 0.23%–0.70%.

For both potential concerns, it was determined that the percentage of affected children was sufficiently small as to mitigate the concerns.

D. Feasibility Testing

A measure will be considered feasible if the data necessary to score the measure are available in administrative databases. –DQA RFP

The National Quality Forum (NQF) defines feasibility as the "[e]xtent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement."[13]

As part of its initial environmental scan, the DQA assessed measure feasibility through a systematic consensus process.[14] The UF Team conducted further feasibility testing through several approaches.

1. Evaluation of Availability of Critical Data Elements in Administrative Databases

The UF Team identified which data elements were "critical" for calculating each measure and which elements were needed for the proposed stratifications. Critical and stratification data elements were mapped to each measure to identify which critical data elements were needed most frequently (Appendix G). The UF Team calculated for each of the five data sources the percentage of missing and invalid data for each data element (Appendix G). Critical data elements typically had missing/invalid rates of <5% and frequently <1%. These rates are consistent with guidance from the Centers for Medicare and Medicaid services regarding acceptable error rates.[15] Stratification data elements were more variable in terms of data availability and completeness, which is consistent with the experience in health care quality measurement in general.[16]

2. Evaluation of Measurement Burden

Another consideration when assessing feasibility is the complexity of the measures. The UF Team took into account the personnel and system resources required to calculate the measures and provided feedback to the DQA. The UF Team also assessed and provided feedback on the measure data element requirements and specification logic flow with respect to complexity and reporting burden.

E. Reliability Testing

Reliability is the degree to which the measure is free from random error . . . and allows for meaningful comparisons across states, programs, individual providers or institutional providers. –DQA RFP

The NQF notes that the reliability of data elements refers to the "repeatability and reproducibility of the data elements for the same population in the same time period" and that the reliability of measure scores refers to the "proportion of variation in the performance scores due to systematic differences across the measured entities in relation to random error."[13] The NQF identifies the following as being important in demonstrating reliability: (1) unambiguous measure specifications, (2) data element reliability, which may be assumed for commonly used data elements (e.g., gender, age, date of service) and can be demonstrated through data element validity testing, and (3) measure score reliability as demonstrated by appropriate methods and resulting rates within acceptable norms. Accordingly, the UF Team evaluated reliability through evaluations of the measure specifications, data elements, and measure scores.

1. Evaluation of Clarity and Completeness of Measure Specifications

For a measure to be reliable – to allow for meaningful comparisons across entities – it is essential that the measure specifications are unambiguous: the denominator criteria, numerator criteria, exclusions, and scoring need to be clearly specified. The UF Team carefully evaluated whether the measure specifications identified all necessary data elements to calculate the numerators and denominators for each measure. In addition, we carefully reviewed the logic flow and made revision recommendations to improve the reliability of the resulting calculations. The DQA also solicited public comment on the Interim Report and posted the measurement specifications online for public comment. The UF Team worked with the DQA to evaluate and address all comments provided. Throughout the eight-month testing period, there were numerous reviews and revisions of the specifications conducted jointly by the UF Team and the DQA. The impact of these careful reviews and revisions on the measure specifications is illustrated in Appendix H for a sample measure, Oral Evaluation. The appendix provides the original specifications released in the RFP in July 2012 and the resulting specifications approved by the DQA in July 2013.

2. Evaluation of Data Element Reliability

Another consideration with respect to reliability is that the measures should be based on standardized data elements to the extent possible and missing/invalid data should be minimized. As described above, the UF Team evaluated missing and invalid rates for each critical and stratification data element. In addition, the UF Team provided feedback on whether data elements were commonly captured in administrative data using standardized formats or if they were subject to variation between plans and programs. In general, for the Pediatric Starter Set:

- most critical data elements are standardized and consistently available (e.g., date of birth, date of service, CDT codes);
- some critical data elements are consistently available but reported differently across entities (e.g., provider ID and type);
- some stratification data elements are standardized and consistently available data (e.g., date of birth); and
- some stratification data elements are either not currently available in administrative databases and/or are not standardized (e.g., race and ethnicity, language, and socioeconomic status).

The UF Team also made recommendations about how to move toward greater standardization, such as encouraging the use of NUCC-maintained Health Care Provider Taxonomy codes to identify dental versus oral health services. Data element reliability was further evaluated through the data element validation described below. (Data element reliability can be assumed if data element validity is established.)

3. Evaluation of Measure Score Reliability

The UF Team assessed measure score reliability – the ability to have meaningful comparisons between entities and over time - through several means:

- The UF Team calculated and compared measures across plans, programs, and states.
- The UF Team calculated each measure for each plan, program, and state for two time periods.
- The UF Team conducted sensitivity testing of rates to different measure specifications.
- The UF Team compared measure scores to local and national values where available to evaluate whether they were within expected ranges.

The UF Team calculated the ten measures ultimately approved by the DQA for all five programs, for two time periods, and for both denominator definitions. In addition, oral health and dental/oral health variations of the measure rates were calculated. Thus, more than 250 measure denominators, numerators, and rates were calculated and reported before stratification. The UF Team provided detailed tables that contained information on exclusions, denominators, numerators, and rates for each state-program-year-measure combination. In addition, the UF Team provided summary charts to facilitate review of the resulting rates. These detailed reports are on file with the Dental Quality Alliance.

F. Validity Testing

Validity demonstrates the extent to which a measure truly measures that which it is intended and designed to measure. –DQA RFP

The NQF notes that validity refers to the "correctness of measurement," including the "correctness of the data elements as compared to an authoritative source" and the "correctness of conclusions about the quality of measured entities that can be made on the measure scores."[13] Accordingly, the UF Team evaluated both data element validity and measure score validity using several approaches.

1. Evaluation of Data Element Validity

To evaluate data element validity, the UF Team conducted reviews of 414 dental records (representing 631 dates of service) collected as part of encounter data validation activities in Texas Medicaid and CHIP to validate critical data elements as well as broader care domains. Encounter data validation of 1,135 procedure codes in the claims data against dental charts found match rates greater than 93%. Additional validation of data elements and broader care domains in the DQA measures demonstrated simple agreement of 81%-96% with most kappa statistic values indicating "substantial" or "almost perfect" agreement.[17] The kappa statistic extends a comparison of simple agreement by taking into account agreement occurring by chance, thereby providing a more conservative measure of agreement between the two data sources. A more detailed report describing the record review processes and specific findings was provided to the DQA (Appendix I).

2. Evaluation of Measure Score Validity

A key aspect of validity is *face validity*, which refers to the evaluation by experts about whether the measure score accurately reflects quality. During initial measure development, the DQA assessed the face validity of each of the measures through a systematic consensus process.[14] The UF Team and the DQA R&D Committee continued to assess face validity throughout the testing process. Face validity also was gauged through feedback solicited during the public comment period on the Interim Report. In addition, measure score validity was assessed by comparison of measure rates to national values where available and by comparison to rates reported for similar measures in prior evaluation and research of the same plans and programs participating in the testing. These comparisons allowed the UF Team and R&D Committee to evaluate whether the measure scores were within acceptable norms. Stratification of the measure scores provides another form of validity testing and is described later in this report.

3. Additional Sealant Validity Testing

The UF Team proposed additional validity testing specifically for the sealant measures. Unlike measures such as preventive dental visits, which are recommended on an annual or more frequent basis, sealant receipt is "lumpy". For example, a child in the age range of 10-14 years may receive sealants in only one or two of those years. Thus, there is the potential that a child who is compliant with clinical guidelines may not be observed during the period in which s/he

received the sealants, resulting in a false-negative classification. To evaluate the implications of this for performance measurement, the UF Team identified a sample of children in who were enrolled throughout the ages of 6-9 years or 10-14 years in the Florida CHIP, Texas Medicaid, or commercial programs. Using these samples, we analyzed (1) the percentage who received sealants in any of those years; and (2) among those receiving sealants, (a) the frequency distribution by age (i.e., the percentage who received sealants at 10 years, 11 years, etc.) and (b) the percentage who received sealants in only one of the years, 2 of the years, and so forth while they were within the specified age range (Appendix J). The UF Team compared the percentage of children who received sealants during any time they were in the age range to the measure score with the final denominator definition(s). These findings reaffirmed that the appropriate age ranges were being captured. They also confirmed that children may receive sealants outside of the observation period. This finding regarding timing of sealant receipt relative to the reporting year led to the development of guidance in the measure specifications about how to appropriately interpret the sealant measure scores for performance measurement purposes.

Evidence-based recommendations advise that sealants be placed on pits and fissures of children's primary and permanent teeth when the tooth, or patient, is at caries risk, with stronger evidence for effectiveness in permanent molars.[18] Thus, we also sought to evaluate how well the specifications addressed both the tooth type on which sealants are placed and the timeliness of care provision. The UF Team ran frequency distributions of sealant placement by tooth number and age range using the Florida CHIP, Texas Medicaid, and commercial claims data (Appendix J). The findings validated the importance of including teeth numbers in the measure specifications to identify permanent first molars and permanent second molars with the corresponding appropriate age ranges (6-9 years and 10-14 years) in order to have reliable indicators of whether children are getting recommended and timely prevention. For example, testing revealed that children in the younger age group may have primary molars sealed that would get captured in the numerator if teeth numbers are not included. Similarly, children in the older age group may be receive sealants on premolars or replacement sealants on permanent first molars that would confound the findings about whether their permanent second molars are being sealed if specific teeth numbers were not identified in the measure specifications. Thus, the UF Team concluded that the incorporation of teeth numbers in the DQA specifications is a significant and important improvement over existing sealant measures that have lacked this specificity.

G. Stratifications

To stratify measure results, the denominator population is divided into different subsets based on different characteristics of interest, and the rates are reported for each sub-population. Stratifying measure scores by child and delivery system characteristics serves two important purposes. First, stratification of results is important for identifying disparities in care. Second, one key threat to validity for outcome and resource measures is lack of risk adjustment or risk stratification.[13] Stratification can also provide an additional form of measure validation by assessing whether measures scores are different for sub-populations know to have differences in access to and quality of care for the measured domain. Thus, the testing process included evaluating the feasibility of stratifying measures by different characteristics.

The RFP requested testing for the following stratification variables:

- age,
- geographic location,
- race/ethnicity,
- socioeconomic status,
- funding source (e.g., Medicaid, CHIP, private),
- provider payment mechanism (e.g., fee-for-service, PPO, capitation), and
- service location (e.g., private office, community clinic, school).

In addition, the UF Team proposed examining:

- language,
- health status,
- plan of enrollment, and
- plan or program types.

The data availability for the different stratification variables varied across programs. Table 4 summarizes which stratifications were tested in each program. Age and geographic location were the only stratification variables that could be tested in all five data sources because date of birth and member address are standard fields in health care claims data. Although there are significant national efforts to improve data collection related to race and ethnicity and language, these data currently are not consistently available.[16,19] The approach used for each category of stratification variable is described briefly below. The UF Team conducted chi-square tests to determine whether there were statistically significant differences in rates between sub-groups in each stratification category. Because these stratifications were tested for multiple measures and multiple programs, it is beyond the scope of this report to provide a summary of all of the results. A few key, high-level findings are provided. The detailed reports, including tables, charts and the results of the statistical tests, are on file with the DQA. Sample findings also are illustrated in presentation materials (Appendix K).[20-22]

Table 4. Stratifications Tested in Each Program

	FL Medicaid	TX Medicaid	FL CHIP	TX CHIP	Commercial
Age	X	X	X	X	X
Geographic Location (Rural/Urban)	X	X	X	X	X
Race/Ethnicity	X	X			
Language			X		
Socioeconomic Status					
Income Category			X	X	
CT-Poverty	X				
Health Status					
Clinical Risk Groups	X	X	X	X	
Title V Program	X				
Plan			X		
Product Line					X
Program Type	X				
Provider Payment			X		

1. Age

The following age groups identified by the DQA were used: <1; 1-2; 3-5; 6-7; 8-9; 10-11; 12-14; 15-18; and 19-20. Age was calculated as of the last day of the reporting year. These age groups are more refined than those used for existing dental measures, such as the HEDIS ADV and the CMS-Form 416 measures. Based on the results, which demonstrated variation between the different age groups, it was decided to retain the more refined breakouts since it is straightforward to collapse the results into broader categories if desired. Consistent with prior research and utilization reporting, children in the youngest and oldest age groups tended to have lower measure scores across the different measures.

2. Geographic Location

Geographic location was characterized as rural versus urban place of residence and was constructed from the Rural-Urban Commuting Areas (RUCA) codes, using categorization D.[23] Based on feedback from the DQA R&D Committee, we also created a 3-category variable with the classifications urban core, suburban, and rural. The dichotomous rural-urban variable was tested with all five data sources; the 3-category variable was tested in the two Medicaid programs. The patterns in measure scores by geographic region varied within and between both measures and programs.

3. Race and Ethnicity

Race and ethnicity data were well-filled only for the two Medicaid programs. Children were classified into the following race and ethnicity categories, using the programs' classifications: Hispanic, non-Hispanic white, non-Hispanic black, and other. The patterns in measure scores by race and ethnicity varied within and between both measures and programs.

4. Language

National data collection efforts related to identifying, measuring, and monitoring disparities include language as one of the priority areas for data collection.[16,24] Although we only had language data available in our Florida CHIP database, many programs and plans collect language data for the purposes of oral and written communication with their members. In Florida CHIP database, the language categories were English, Spanish, and Haitian Creole.

5. Socioeconomic Status

Despite the strong association between socioeconomic status (SES) and health disparities, there has not been as much progress (compared to race/ethnicity and language) in developing standardized data collection and measurement approaches to effectively incorporate SES into health performance measures.[19] The UF Team captured SES using two different approaches. In the two CHIP programs, the children's income category that is used to determine the family's monthly subsidized premium payment was used as an indicator of SES. The income categories were 100-150% of the federal poverty level (FPL) and 151-200% of the FPL. In Florida Medicaid, the enrollment database had been geocoded by the UF Team to include the children's census tract. Using these data, we created an area-based SES indicator ("CT-

poverty"). CT-poverty is a categorical variable that indicates the percentage of individuals within the member's census tract living below the poverty level, using the categories of 0-4.9%, 5-9.9%, 10-19.9%, 20-39.9%, and >=40% below poverty where >20% of people living below poverty constitutes the federal definition of a "poverty area" and >=40% represents "extreme poverty". CT-poverty has been identified as an accessible and standardized SES measure for monitoring health disparities that is as reliable an SES indicator as more complicated SES indices as well as individual SES measures that frequently are not available in administrative databases (e.g., education level).[25,26] The limitation of using income categories in the CHIP programs and CT-poverty in the Florida Medicaid program is that there is less SES variation within these populations than for the overall U.S. population. Measures of SES will be more likely to detect disparities in databases that contain more economically diverse populations, such as all-payer claims databases. Funding source (e.g., Medicaid, CHIP, private) may serve as reasonable proxies for SES in such databases. Although we did not have an integrated all-payer claims database, the range of our data sources allowed for comparisons of the measures scores between different payer types (Medicaid, CHIP, and commercial).

6. Health Status

The UF Team recommended examining health status for two reasons. Having a special health care need is a risk factor for developing dental caries,[27] and dental care is the most common unmet health service need reported by families who have children with special health care needs (CSHCN).[28] In addition, the NQF recommends risk adjustment for outcome and resource measures, and additionally recommends risk adjusting by clinical factors that influence the measure score and not by factors related to disparities in care (e.g., race and ethnicity, SES, and sex). Health status is commonly used for risk adjustment for medical quality of care measures.[13]

The UF Team measured health status for children enrolled in the four public insurance programs using the Clinical Risk Groups (CRGs) software, which uses ICD-9-CM diagnosis codes to classify children into hierarchically defined health status groups and has been validated for identifying CSHCN.[29] For classification, children <12 months old must be enrolled > 3 months, and children > 12 months old must be enrolled > 6 months. The CRGs has nine categories that were collapsed to the following five groups: (1) healthy, (2) significant acute conditions (e.g., meningitis and traumatic brain injury), (3) minor chronic conditions (e.g., attention deficit disorder), (4) moderate chronic conditions (e.g., asthma, diabetes and depression), and (5) major chronic conditions (e.g., cystic fibrosis, cancer, and schizophrenia). Statistically significant variation in measure scores based on health status was detected. Because this approach requires specialized software and medical claims data and, therefore, may be of limited utility for oral health performance measures, we additionally used a child's Title V classification status to identify Medicaid-enrolled children with special health care needs. The results were similar to those using the CRGs. An advantage of the CRGs relative to Title V classification is that the CRGs uses a standardized approach, whereas there is variation in how states determine Title V eligibility. In addition, Title V eligibility applies to a limited population. More significantly, however, both approaches are limited in that they rely on classification based on a child's medical diagnoses, which may be poor proxies for identifying children with special dental care needs. Thus, health status was not determined to be a high priority for stratification of the measures in the Pediatric Starter Set.

7. Plan, Product Line, and Program Type

Plan, product line, and program type are all different methods of stratifying measure scores by health care delivery system characteristics. How care is financed and organized may influence access, quality, and outcomes. Florida CHIP was the only program that had more than one dental plan participating in the program, and the measure scores were stratified by plan in order to detect any differences in performance between the two plans. Within public insurance programs, there may be different delivery system mechanism through dental services are provided. The Florida Medicaid system is an example. In Florida Medicaid, there is a dental fee-for-service program component, a prepaid dental plan program component, or children may get dental services as a carve-out benefit through their medical managed care plan. The commercial data reflected two different product lines: one product line was a standard fee-for-service product line and the other was a preferred provider organization with discounted fee-for-service reimbursement to providers.

8. Provider Payment

The DQA requested that testing include stratification of measure scores by provider payment. To stratify measure results, the denominator population is divided into different subsets based on different characteristics of interest, and the rates are reported for each sub-population. This requires that each member in the denominator be uniquely classified into a stratification category. Because a child may see different providers who are reimbursed under different mechanisms, it is not straightforward to stratify measures by provider payment. However, the type of delivery system (e.g., traditional FFS plan, PPO, prepaid plan) may serve as a rough proxy for provider payment.

9. Site of Service

The DQA also requested that testing explore reporting measure scores by where children received services. As with provider payment, this is not equivalent to stratifying measure scores by such characteristics as race or age. However, measure implementers can conduct queries on their claims data that would provide relevant contextual information about care provision. For example, among children who received sealants, one could examine what percentage of those children received sealants in a private dentist's office, in a school-based clinic, or in a federally qualified health center. This is a different type of question, though, than asking what percentage of children who are Hispanic, non-Hispanic white, and non-Hispanic black, respectively, received sealants during the reporting year. Because the nature of the questions is fundamentally different, as are the methodologies, site of service was recommended as a contextual variable but not as a stratification variable.

H. Cost Measures

There were two cost measures in the original Pediatric Starter Set: (1) User Cost of Clinical Services and (2) Percentage of Child Health Care Expenditures Expended on Dental Care.

1. User Cost of Clinical Services

As noted above, the UF Team and DQA R&D Committee specified this measure as a per member per month (PMPM) measure to be consistent with standard reporting for costs. Thus, children enrolled at least one month who meet the other denominator eligibility criteria are included in this measure. Performance measurement, particularly for resource use and outcome measures, frequently provides for risk adjustment of the results to account for variations in patient case mix. Case mix refers to the distribution of different patient types within a particular health care setting (e.g., provider, facility, plan, or program) based on characteristics associated with health status and health care resource requirements. The NQF recommends that risk adjustment (1) be based on factors that influence the outcome but do not influence disparities in care and (2) not obscure disparities in care.[13] Therefore, the NQF recommends against including such factors as race, SES, and sex in risk adjustment methodologies. Risk adjustment methodologies typically involve adjusting for individuals' health status. However, the lack of standardized and widely adopted dental diagnostic codes has posed a significant barrier to risk adjustment in dentistry both in terms of practical implementation as well as in having an evidence base for appropriate risk adjustment methodologies. The UF Team and DQA R&D Committee did not consider risk adjustment based on diagnoses in medical claims data to be an adequate substitute for dental diagnosis codes. In addition, appropriate risk adjustment methodologies in dentistry have not been established. Therefore, risk adjustment is currently not recommended for this measure. Instead, implementers are encouraged to stratify this measure.

2. Percentage of Child Health Care Expenditures Expended on Dental Care

This measure was specified as total dental expenditures, including both the amount paid for clinical services and administrative costs, divided by total child health expenditures (medical and dental) to include both the amount paid for health services and associated administrative costs. Administrative costs were an important component of this measure. However, administrative costs are not captured in administrative claims data. The UF Team contacted Medicaid and CHIP program administrators and dental plan administrators to explore whether there would be effective ways to capture these data in a consistent and standardized manner. The UF Team found that there are variable definitions of administrative costs among plans and programs and varying degrees regarding the extent to which administrative costs are tracked. The DQA R&D Committee determined it would be outside of the scope of the RFP to pursue this measure further as part of the formal testing process. However, this measure remains on the research agenda with the goal of developing recommendations and methodology for capturing this resource domain.

Outcomes

There have been several important outcomes of the testing process.

A. Successful Completion of Testing

The key outcome is that all of the original eleven measures in the Pediatric Starter Set underwent rigorous testing, and the testing was completed consistent with what was requested in the DQA RFP and with what was proposed by the UF Team. Taking into account the ten measures calculated for all five programs; for two reporting periods; for two denominator definitions; for dental, oral health and dental/oral health variations (select programs and measures); and for the different stratifications, more than 5,000 denominators, numerators, and measure scores were calculated and reported. Formatted tables and charts for these variations were provided in a series of packets to the DQA R&D Committee.

B. Public Dissemination, Awareness, and Education

The measures and testing processes have been widely disseminated not only for the purposes of soliciting stakeholder feedback but also to create awareness and provide education around oral health performance measurement. The UF Team has been highly engaged in these activities and has given national presentations.[20-22] In addition, the UF Team and the DQA R&D Committee plan to collaborate on manuscripts for submission to peer-reviewed journals.

C. Approval of Pediatric Starter Set

The results of the testing were presented to the full DQA at its July 19, 2013 meeting.[3] The DQA formally approved this initial set of performance measures, which are posted on the DQA's website.

Concluding Remarks

The UF Team greatly appreciates the opportunity to have been involved in this important initiative and looks forward to continued collaboration and engagement with the Dental Quality Alliance.

End Notes

1. Dental Quality Alliance. Dental Quality Alliance Measure Activities. 2013; http://www.ada.org/7503.aspx. Accessed July 28, 2013.
2. Dental Quality Alliance. Request for Proposals to Establish Feasibility, Reliability and Validity of Implementation of Claims Based Pediatric Oral Health Measures Developed by the Dental Quality Alliance. 2012.
3. Herndon JB. Results of Testing Dental Quality Alliance Pediatric Oral Health Performance Measures. Presentation at Dental Quality Alliance Meeting. 2013.
4. Dental Quality Alliance. Dental Quality Alliance Measure Sets. 2013; http://www.ada.org/8472.aspx. Accessed June 20, 2013.
5. Brewer CA, Suchan TA. *Mapping Census 2000 : the geography of U.S. diversity*. Washington, D.C: U.S. Government Printing Office; 2001.
6. U.S. Census Bureau. *Statistical abstract of the United States, 2012*. 131st ed. Washington, DC: U.S. Census Bureau; 2011.
7. Kaiser Family Foundation. Monthly Medicaid Enrollment for Children. 2013; http://statehealthfacts.org/comparemaptable.jsp?ind=612&cat=4. Accessed April 29, 2013.
8. Edelstein BL, Chinn CH. Update on Disparities in Oral Health and Access to Dental Care for America's Children. *Acad Pediatr.* Nov-Dec 2009;9(6):415-419.
9. American Dental Association. Report of the Council on Scientific Affairs. Evidence-Based Clinical Recommendations: Professionally Applied Topical Fluoride. 2006.
10. Code of Federal Regulations. 42 C.F.R. Section 400.100 Dental services. (2006). http://www.gpo.gov/fdsys/pkg/CFR-2006-title42-vol4/xml/CFR-2006-title42-vol4-sec440-100.xml. Accessed July 30, 2013.
11. National Uniform Claim Committee. Health Care Provider Taxonomy. 2013; http://nucc.org/index.php?option=com_content&view=article&id=14&Itemid=125. Accessed July 13, 2013.
12. Dental Quality Alliance. Dental Quality Alliance Measures User Guide. 2013; http://www.ada.org/8472.aspx. Accessed July 25, 2013.
13. National Quality Forum. Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties.January 2011.

14. Dental Quality Alliance. Pediatric Oral Health Quality and Performance Measures: Environmental Scan. 2012; http://ada.org/sections/professionalResources/pdfs/dqa_environmental_scan.pdf. Accessed June 18, 2013.

15. Centers for Medicare & Medicaid Services. Medicaid and CHIP Statistical Information System (MSIS) File Specifications and Data Dictionary. 2010; http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MSIS/downloads/msisdd2010.pdf. Accessed August 10, 2013.

16. Institute of Medicine (U.S.). Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement., Ulmer C, McFadden B, Nerenz DR. *Race, ethnicity, and language data : standardization for health care quality improvement*. Washington, D.C.: National Academies Press; 2009.

17. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics.* Jun 1977;33(2):363-374.

18. Beauchamp J, Caufield PW, Crall JJ, et al. Evidence-based clinical recommendations for the use of pit-and-fissure sealants: a report of the American Dental Association Council on Scientific Affairs. *J Am Dent Assoc.* Mar 2008;139(3):257-268.

19. Dorsey R, Graham G. New HHS data standards for race, ethnicity, sex, primary language, and disability status. *JAMA.* Dec 7 2011;306(21):2378-2379.

20. Herndon JB. DQA Pediatric Oral Health Performance Measure Set: Overview of Measures and Validation Process. Presentation at 2013 National Oral Health Conference Pre-Conference Workshop on Objectives, Indicators, Measures and Metrics. 2013; http://www.nationaloralhealthconference.com/docs/presentations/2013/04-21/Jill%20Boylston%20Herndon.pdf. Accessed August 15, 2013.

21. Herndon JB. DQA Pediatric Oral Health Performance Measure Set: Overview of Measures and Validation Process. Presentation at 2013 Texas Medicaid and CHIP Managed Care Quality Forum. 2013.

22. Herndon JB. Disparities and Performance Measurement. Presentation at Dental Quality Alliance Conference: Improving Dental Quality through Measurement. 2013;

23. Economic Research Service USDA. Rural-Urban Commuting Area Codes. 2000; http://www.ers.usda.gov/Data/RuralUrbanCommutingAreaCodes/. Accessed May 27, 2013.

24. U.S. Department of Health and Human Services. Final Data Collection Standards for Race, Ethnicity, Primary Language, Sex, and Disability Status Required by Section 4302 of the Affordable care Act. 2013.; http://minorityhealth.hhs.gov/templates/browse.aspx?lvl=2&lvlID=208. Accessed June 18, 2013.

25. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am J Public Health.* Feb 2005;95(2):312-323.

26. Krieger N, Waterman PD, Chen JT, Soobader MJ, Subramanian SV. Monitoring socioeconomic inequalities in sexually transmitted infections, tuberculosis, and violence: geocoding and choice of area-based socioeconomic measures--the public health disparities geocoding project (US). *Public Health Rep.* May-Jun 2003;118(3):240-260.

27. American Academy of Pediatric Dentistry. Guideline on Caries-risk Assessment and Management for Infants, Children, and Adolescents. . *2012-2013 Definitions, Oral Health Policies, and Clinical Guidelines* 2011; http://www.aapd.org/policies/. Accessed May 27, 2013

28. Lewis CW. Dental care and children with special health care needs: a population-based perspective. *Acad Pediatr.* Nov-Dec 2009;9(6):420-426.

29. Neff JM, Sharp VL, Muldoon J, Graham J, Popalisky J, Gay JC. Identifying and classifying children with chronic conditions using administrative data with the clinical risk group classification system. *Ambul Pediatr.* Jan-Feb 2002;2(1):71-79.