NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 2528

Measure Title: Prevention: Topical Fluoride for Children at Elevated Caries Risk, Dental Services

Measure Steward: American Dental Association on behalf of the Dental Quality Alliance

Brief Description of Measure: Percentage of enrolled children aged 1-21 years who are at "elevated" risk (i.e., "moderate" or "high") who received at least 2 topical fluoride applications within the reporting year.

Developer Rationale: Inequalities in oral health status and inadequate use of oral health care services are well documented. Dental caries is the most common chronic disease in children in the United States (NCHS 2012). In 2009–2010, 14% of children aged 3 –5 years had untreated dental caries. Among children aged 6–9 years, 17% had untreated dental caries, and among adolescents aged 13–15, 11% had untreated dental caries (Dye, Li, and Thorton-Evans 2012). Dental decay among children has significant short- and long-term adverse consequences (Tinanoff and Reisine 2009). Childhood caries is associated with increased risk of future caries (Gray, Marchment, and Anderson 1991; O´Sullivan and Tinanoff 1996; Reisine, Litt, and Tinanoff 1994), missed school days (Gift, Reisine, and Larach 1992; Hollister and Weintraub 1993), hospitalization and emergency room visits (Griffin et al. 2000; Sheller, Williams, and Lombardi 1997) and, in rare cases, death (Casamassimo et al. 2009).

Identifying caries early is important to reverse the disease process, prevent progression of caries, and reduce incidence of future lesions. Evidence suggests that topical fluoride applied to children starting as early as six months of age is beneficial in preventing dental caries (Weyant et al. 2013). However, approximately three quarters of children younger than age 6 years did not have at least one visit to a dentist in the previous year (Edelstein & Chinn 2009). Evidence-based clinical recommendations suggest that topical fluoride should be applied at least every three to six months in children at elevated risk for caries (Weyant et al. 2013).

The proposed measure, Topical Fluoride for Children at Elevated Caries Risk – Dental Services, captures whether children at moderate or high caries risk received at least two topical fluoride applications as dental services. Because topical fluoride is indicated at 3-6 month intervals (2-4 times per year) for children at elevated caries risk, at least two applications are indicated during the reporting year. This measure directly reflects evidence-based guidelines regarding an effective caries prevention measure (professionally applied topical fluoride), including the frequency required for clinical effectiveness (at least every three-six months). Topical Fluoride allows plans and programs to assess whether children at risk for caries are receiving evidence-based preventive services and target performance improvement initiatives accordingly.

Note: Procedure codes contained within claims data are the most feasible and reliable data elements for quality metrics in dentistry, particularly for developing programmatic process measures to assess the quality of care provided by programs (e.g., Medicaid, CHIP) and health/dental plans. In dentistry, diagnostic codes are not commonly reported and

collected, precluding direct outcomes assessments. Although some programs are starting to implement policies to capture diagnostic information, evidence-based process measures are the most feasible and reliable quality measures at programmatic and plan levels at this point in time.

[Complete citations provided in 1c4 and in Evidence Submission Form.]

Numerator Statement: Unduplicated number of enrolled children aged 1-21 years who are at "elevated" risk (i.e., "moderate" or "high") who received at least 2 topical fluoride applications as a dental service

Denominator Statement: Unduplicated number of enrolled children aged 1-21 years who are at "elevated" risk (i.e., "moderate" or "high")

Denominator Exclusions: Medicaid/CHIP programs should exclude those individuals who do not qualify for dental benefits. The exclusion criteria should be reported along with the number and percentage of members excluded.

Measure Type: Process

Data Source: Claims

Level of Analysis: Health Plan, Integrated Delivery System

Original Endorsement Date: Sep 18, 2014 **Most Recent Endorsement Date:** Sep 18, 2014

Staff Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

- This measure assesses the percentage of children at moderate to high risk for caries who received at least two topical fluoride applications as dental services during the reporting year. Evidenced-based clinical guidelines recommends the specific topical fluoride agents for people who are at elevated risk of developing dental caries. (Weyant et al. 2013, full report, 0. 10)
- For children at elevated risk of developing caries specifically, the guidelines recommend applying 2.26 percent fluoride varnish at least every three to six months for children younger than 6 years old. For children 6-18 years

old, the guidelines recommend 2.26 percent fluoride varnish at least every three to six months or 1.23 percent fluoride (APF).

- 71 studies were included in evidence reviews, representing 82 citations. All studies included were controlled clinical trials.
- This evidence received a grade of moderate by an expert panel, which is second on a three-point scale and denotes that evidence statements "are based on preliminary determination from the current best available evidence, but confidence in the estimate is constrained by one or more factors, such as: the number, size, or risk of bias of individual studies; inconsistency of findings across individual studies; limited applicability due to the populations of interest; or lack of coherence in the chain of evidence. As more information becomes available, the magnitude or direction of the observed effect could change, and this change could be large enough to alter the conclusion."
- The clinical recommendations for fluoride among children and adolescents received an evidence grade of "in favor", which is the second highest out of six grading categories.

Citation:

Weyant RJ, Tracy SL, Anselmo TT, Beltrán-Aguilar ED, et al; American Dental Association Council on Scientific Affairs Expert Panel on Topical Fluoride Caries Preventive Agents. Topical fluoride for caries prevention: full report of the updated clinical recommendations and supporting systematic review. Available at: http://ebd.ada.org/contentdocs/Topical_fluoride_for_caries_prevention_2013_update_-_full_manuscript.pdf

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3)→ QQC presented (Box 4)→ Contains Quantity: High (71 studies, 82 citations) Quality: Moderate, Consistency: High (Box 5b) → Rate as MODERATE

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer used data from four sources and referred to "program" level information and "plan" level information (Texas Medicaid, Florida CHIP, and Florida Medicaid programs, as well as national commercial data from Dental Service of Massachusetts, Inc.). The developer presented the total number of children enrolled in each program/plan. In the data summaries, "Programs" refer to population data from (1) Texas Medicaid, (2) Florida CHIP, (3) Commercial Data, and (4) Florida Medicaid. "Plans" refer to data from the two dental plans that served Florida CHIP members in both 2010 and 2011.
- The measure testing findings are consistent with other data indicating that children have sub-optimal utilization of dental services in general and preventive dental services in particular.
- The data source and sample size are sufficient to assess gaps in performance. The performance range of **18% to 35% in CY 2010** (year in which data were available for all five programs) indicates variation in topical fluoride application across programs.

- The developer did not provide more recent performance data, stating that due to the start-up phase for integration of the measures into contracts and for programs and plans to prepare for reporting, in combination with a lag period for reporting measures calculated using administrative claims data, most of the entities that have adopted the measures are just getting underway and there is limited data reporting.

Disparities

- Disparities were detected for age, geographic location, and race/ethnicity for all programs. The developer also evaluated whether the measure could detect disparities by income (within program), children's health status (based on their medical diagnoses), Medicaid program type, CHIP dental plan, commercial product line, and preferred language for program communications. The developer detected disparities by income, health status, CHIP plan, and Medicaid program type, but data on all of these characteristics were not consistently available for all programs and present disparities data on those characteristics that were most consistently available and had the greatest standardization (i.e. race/ethnicity and geographic location).
- The developer provided an overview of the literature documenting the disparities in dental service use among children by age, race/ethnicity, and geographic region, including within vulnerable populations, much of which is summarized in three major national reports on oral health: the Surgeon General's report on Oral Health in America in 2000, the IOM report, *Improving Access to Oral Health Care for Vulnerable and Underserved Populations,* and the IOM report, *Advancing Oral Health in America*.

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments: Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing
2b. Validity: Testing; Exclusions**;** Risk-Adjustment**;** Meaningful Differences**;** Comparability Missing Data

2c. For composite measures: empirical analysis support composite approach

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if 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 enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**

Evaluation of Reliability and Validity (and composite construction, if applicable):

Staff evaluation of Scientific Acceptability

Questions for the Committee regarding reliability:

- *The staff is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?*

Questions for the Committee regarding validity:

- *The staff is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?*

Preliminary rating for reliability:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
Preliminary rating for validity:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**

Scientific Acceptability

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. **Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

RELIABILITY

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?
   NOTE: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.

   TIPS: Consider the following: Are all the data elements clearly defined? Are all appropriate codes included? Is the logic or calculation algorithm clear? Is it likely this measure can be consistently implemented?

   ☒Yes (go to Question #2)

   ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

2. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

   TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)

   ☒Yes (go to Question #4)

☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

3. Was **empirical VALIDITY testing** of patient-level data conducted?

   ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

   ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)

4. Was reliability testing conducted with computed performance measure scores for each measured entity?

   TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data

   ☐Yes (go to Question #5)

   ☒No (go to Question #8)

5. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? *NOTE: If multiple methods used, at least one must be appropriate.*

   TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

   ☐Yes (go to Question #6)

   ☐No (please explain below then go to Question #8)

6. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

   TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation?

   Do the results demonstrate sufficient reliability so that differences in performance can be identified?

   ☐High (go to Question #8)

   ☐Moderate (go to Question #8)

   ☐Low (please explain below then go to Question #7)

7. Was other reliability testing reported?

   ☐Yes (go to Question #8)

   ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

8. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

   TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)

   ☒Yes (go to Question #9)

   ☐No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

   Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

   ☒Yes (go to Question #10)

☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)

10. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?

    ☒Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)

    ☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)

    ☐Insufficient (go to Question #11)

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

☐High (NOTE: Can be HIGH only if score-level testing has been conducted)

☒Moderate (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

1. Were all potential threats to validity that are relevant to the measure empirically assessed?

   TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

   ☒Yes (go to Question #2)

   ☐No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]

2. Analysis of potential threats to validity: Any concerns with measure exclusions?

   TIPS: Consider the following: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

   ☐Yes (please explain below then go to Question #3)

   ☐No (go to Question #3)

   ☒Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

3. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

   ☒Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

   a. Is a conceptual rationale for social risk factors included? ☐Yes ☐No

   b. Are social risk factors included in risk model? ☐Yes ☐No

c. Any concerns regarding the risk-adjustment approach?

*TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? Are all of the risk adjustment variables present at the start of care (if not, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

☐Yes (please explain below then go to Question #4)

☐No (go to Question #4)

4. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

   ☐Yes (please explain below then go to Question #5)

   ☒No (go to Question #5)

5. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

   ☐Yes (please explain below then go to Question #6)

   ☐No (go to Question #6)

   ☒Not applicable (go to Question #6)

6. Analysis of potential threats to validity: Any concerns regarding missing data?

   ☐Yes (please explain below then go to Question #7)

   ☒No (go to Question #7)

ASSESSMENT OF MEASURE TESTING

7. Was <u>empirical</u> validity testing conducted using the measure as specified and appropriate statistical test?

   Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

   ☒Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

   ☐No (please explain below then go to Question #8)

8. Was <u>face validity</u> systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

   TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

   ☐Yes (go to Question #9)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

9. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

   ☐Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

10. Was validity testing conducted with computed performance measure scores for each measured entity?

    TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

    ☐Yes (go to Question #11)

    ☒No (please explain below and go to Question #13)

11. Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?

    TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

    ☐Yes (go to Question #12)

    ☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

12. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☐High (go to Question #14)

    ☐Moderate (go to Question #14)

    ☐Low (please explain below then go to Question #13)

    ☐Insufficient

13. Was other validity testing reported?

    ☒Yes (go to Question #14)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

14. Was validity testing conducted with patient-level data elements?

    TIPS: Prior validity studies of the same data elements may be submitted

    ☒Yes (go to Question #15)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

15. Was the method described and appropriate for assessing the accuracy of ALL critical data elements? *NOTE that data element validation from the literature is acceptable.*

    TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

    Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

☒Yes (go to Question #16)

☐No (please explain below and rate Question #16 as INSUFFICIENT)

16. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

    ☒Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

    ☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

    ☐Insufficient (go to Question #17)

17. OVERALL VALIDITY RATING

OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.

☐High (NOTE: Can be HIGH only if score-level testing has been conducted)

☒Moderate (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

Committee Pre-evaluation Comments: Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent
3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- This measure relies on **standard data elements in administrative claims data** (e.g., patient ID, patient birthdate, enrollment information, CDT codes, date of service, and provider taxonomy). These data are readily available and can be easily retrieved because they are routinely used for billing and reporting purposes.

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments: Criteria 3: Feasibility

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No** ☐ **UNCLEAR**

Accountability program details

- Texas Health and Human Services Commission: Texas Medicaid and CHIP: https://hhs.texas.gov/sites/default/files//documents/laws-regulations/handbooks/umcm/6-2-15.pdf

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The developer reports there has been no feedback indicating any significant issues related to the clarity or feasibility of implementing the measure specifications.

Additional Feedback:

- This measure was one of 10 performance measures approved by the Dental Quality Alliance (DQA) that focused on Dental Caries Prevention and Disease Management among children. The DQA was formed at the request of the CMS specifically for the purpose of bringing together recognized expertise in oral health to develop quality measures through consensus processes. As noted in the letter from the Director of the Center for Medicaid & CHIP Services within CMS: "The dearth of tested quality measures in oral health has been a concern to CMS and other payers of oral health services for quite some time."

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- The developer provides initial reporting data available from the Texas Medicaid/CHIP programs (https://thlcportal.com/qoc/dental), which started implementing this measure after approval by the DQA, but before NQF endorsement, as follows:

Texas Medicaid
Year, Program Denominator, Program Overall Score, DentaQuest(Plan) Score, MCNA(Plan) Score
2014, 1090952, 39.97, 41.57, 37.62
2015, 1334887, 41.75, 44.70, 38.15
Texas CHIP
Year, Program Overall, DentaQuest(Plan), MCNA(Plan)

2014, 108704, 33.01, 35.45, 32.99
2015, 79693, 37.50, 41.44, 37.71

- The developer noted that these data suggest a trend in improvement over time. These are initial performance data for one program, however, since most measure users are just now getting their quality measurement programs underway.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer reports no unintended or negative consequences have been identified.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

Preliminary rating for Usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments: Criteria 4: Usability and Use

Criterion 5: Related and Competing Measures

Related or competing measures

- N/A

Harmonization

- N/A

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: Month/Day/Year

- **Of the XXX NQF members who have submitted a support/non-support choice:**
  - XX support the measure
  - YY do not support the measure

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

4_Evidence_fluoride.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?
Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

No

1a Evidence (subcriterion 1a)

Measure Title: Prevention: Topical Fluoride for Children at Elevated Caries Risk, Dental Services

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 2/11/2014

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *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 10 pages (*incudes 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.

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

Subcriterion 1a. Evidence to Support the Measure Focus

The measure focus is a health outcome or is evidence-based, demonstrated as follows:

- Health outcome:[3] a rationale supports the relationship of the health outcome to processes or structures of care.
- Intermediate clinical outcome, Process,[4] or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a desired health outcome.
- Patient experience with care: evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is correlated with desired outcomes.

- Efficiency:[6] evidence for the quality component as noted above.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement.

5. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

6. Measures of efficiency combine the concepts of resource use and quality (NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of:

Outcome

☐ Health outcome:

Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)

☐ Intermediate clinical outcome:

☒ Process: Receipt of evidence-based preventive service - topical fluoride application - during the reporting period

☐ Structure:

☐ Other:

HEALTH OUTCOME PERFORMANCE MEASURE *If not a health outcome, skip to 1a.3*

1a.2. Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Not applicable.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

Not applicable.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Topical Fluoride for Children at Elevated Caries Risk indicates the percentage of children at moderate to high risk for caries who received at least two topical fluoride applications as dental services during the reporting year. Evidence suggests that topical fluoride applied to children starting as early as six months of age is beneficial in preventing dental caries (Weyant et al. 2013). Evidence-based clinical recommendations also suggest that topical fluoride should be applied at least every three to six months in children at elevated risk for caries (Weyant et al. 2013). This measure directly reflects evidence-based guidelines regarding an effective caries prevention measure (professionally applied topical fluoride), including the frequency required for clinical effectiveness (at least every three-six months). As described in 1b1 (Importance), dental caries is the most common chronic disease in children in the U.S. and a significant percentage of children have untreated dental caries. Dental decay causes significant short- and long-term adverse

consequences for children's health and functioning. As detailed below, professionally applied topical fluoride has demonstrated effectiveness in reducing caries among children at elevated caries risk, thereby improving oral health, overall health, and overall well-being.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections*** 1a.4***, and*** 1a.7

☐ US Preventive Services Task Force Recommendation – ***complete sections*** 1a.5 ***and*** 1a.7

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections*** 1a.6 ***and*** 1a.7

☐ Other – ***complete section*** 1a.8

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Full Report: Weyant RJ, Tracy SL, Anselmo TT, Beltrán-Aguilar ED, et al; American Dental Association Council on Scientific Affairs Expert Panel on Topical Fluoride Caries Preventive Agents. Topical fluoride for caries prevention: full report of the updated clinical recommendations and supporting systematic review. Available at: http://ebd.ada.org/contentdocs/Topical_fluoride_for_caries_prevention_2013_update_-_full_manuscript.pdf.

Condensed version: Weyant RJ, Tracy SL, Anselmo TT, Beltrán-Aguilar ED, et al; American Dental Association Council on Scientific Affairs Expert Panel on Topical Fluoride Caries Preventive Agents. J Am Dent Assoc. 2013 Nov;144(11):1279-91. Topical fluoride for caries prevention: executive summary of the updated clinical recommendations and supporting systematic review. Available at: http://ebd.ada.org/contentdocs/JADA_updated_executive_summary_Nov_2013.pdf.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Summary of Clinical Recommendations: "For people who are at an elevated risk of developing dental caries, the panel makes clinical recommendations for the use of specific topical fluoride agents; these recommendations are based on the evidence statements and the balance of benefits with potential harm. The panel recommends topical fluoride agents only for people who are at elevated risk of developing dental caries." (Weyant et al. 2013, full report, 0. 10)

Age –Specific Recommendations: "The panel recommends the following for people at risk of developing dental caries: 2.26% fluoride varnish or 1.23% fluoride (APF) gel, or a prescription-strength, home-use 0.5% fluoride gel or paste or 0.09% fluoride mouthrinse for 6 years or older. Only 2.26% fluoride varnish is recommended for children younger than 6 years. The strengths of the recommendations for the recommended products varied from "in favor" to "expert opinion for." As part of the evidence-based approach to care, these clinical recommendations should be integrated with the practitioner's professional judgment and the patient's needs and preferences." (Weyant et al. 2013, full report, p. 10)

For children at elevated risk of developing caries specifically, the recommendations are "in favor" for:

- "2.26 percent fluoride varnish at least every three to six months" for children younger than 6 years
- "2.26 percent fluoride varnish at least every three to six months OR 1.23 percent fluoride (APF) gel for four minutes at least every three to six months" for children 6-18 years

(Weyant et al., 2013, p. 11, Table 1)

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Grade: The grade for both bulleted items is **"in favor"** which is defined as: "Evidence favors providing this intervention." This is the second highest recommendation out of a six-point scale. The grading system was adapted from that used by the U.S. Preventive Services Task Force. (Weyant et al. 2013, full report, p. 11, Table 1; p. 20, Table 6)

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Strong: Evidence strongly supports providing this intervention.

In Favor: Evidence favors providing this intervention.

Weak: Evidence suggests implementing this intervention after alternatives have been considered.

Expert Opinion For:† Evidence is lacking; the level of certainty is low. Expert opinion guides this recommendation

Expert Opinion Against:† Evidence is lacking; the level of certainty is low. Expert opinion suggests not implementing this intervention.

Against: Evidence suggests not implementing this intervention or discontinuing ineffective procedures.

† The USPSTF system defines this category of evidence as "insufficient"; "grade I indicates that the evidence is insufficient to determine the relationship between benefits and harms (i.e., net benefit)." The corresponding recommendation grade "I" is defined as follows: "The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined."

Grades definitions can be found at Weyant et al. 2013, full report, p. 20, Table 6. The grading system was adapted from that used by the U.S. Preventive Services Task Force. (U.S. Preventive Services Task Force. Methods and processes. Available at: www.uspreventiveservicestaskforce.org/methods.htm.)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Same as that provided for the guidelines provided in 1a.4.1.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section*** 1a.7

☐ No → ***report on another systematic review of the evidence in sections*** 1a.6 ***and*** 1a.7***; if another review does not exist, provide what is known from the guideline review of evidence in*** 1a.7

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable.

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade: Not applicable.

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable.

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

The following three clinical questions were addressed:

- "In primary and permanent teeth, does the use of a topical fluoride compared to no topical fluoride reduce the incidence of new lesions, or arrest or reverse existing coronal and/or root caries?"
- "For primary and permanent teeth, is one topical fluoride agent more effective than another in reducing the incidence of, or arresting or reversing coronal and/or root caries?"
- "Does the use of prophylaxis before application of topical fluoride reduce the incidence of caries to a greater extent than the application of topical fluoride without prophylaxis?"
  (Weyant et al., 2013, full report, pp. 7-8)

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

For children at elevated risk of developing caries specifically, the recommendations are "in favor" for:

- "2.26 percent fluoride varnish at least every three to six months" for children younger than 6 years
- "2.26 percent fluoride varnish at least every three to six months OR 1.23 percent fluoride (APF) gel for four minutes at least every three to six months" for children 6-18 years

(Weyant et al., 2013, p. 11, Table 1)

Grade: The evidence grade for both bulleted items is **"moderate"** which is defined as: "This statement is based on preliminary determination from the current best available evidence, but confidence in the estimate is constrained by one or more factors, such as: the number, size, or risk of bias of individual studies; inconsistency of findings across individual studies; limited applicability due to the populations of interest; or lack of coherence in the chain of evidence. As more information becomes available, the magnitude or direction of the observed effect could change, and this change could be large enough to alter the conclusion." (Weyant et al., 2013, full report, pp. 18-19, Table 4)

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

High: This statement is strongly established by the best available evidence; the conclusion is unlikely to be affected strongly by the results of future studies. The body of evidence usually includes consistent results from well-designed, well-conducted studies in representative populations. This conclusion is unlikely to be strongly affected by the results of future studies.

Moderate: This statement is based on preliminary determination from the current best available

evidence, but confidence in the estimate is constrained by one or more factors, such as: the number, size, or risk of bias of individual studies; inconsistency of findings across individual studies; limited applicability due to the populations of interest; or lack of coherence in the chain of evidence. As more information becomes available, the magnitude or direction of the observed effect could change, and this change could be large enough to alter the conclusion."

Low: The available evidence is insuffi cient to support the statement, or the statement is based on

extrapolation from the best available evidence. Evidence is insufficient or the reliability of estimated effects is limited by factors such as: the limited number or size of studies; important flaws in study design or methods leading to high risk of bias; inconsistency of findings across individual studies; gaps in the chain of evidence; findings not applicable to the populations of interest; or a lack of information on important health outcomes. More information could allow a reliable estimation of effects on health outcomes.

(Weyant 2013, full report, pp. 18-19)

The grading system was adapted from that used by the U.S. Preventive Services Task Force (U.S. Preventive Services Task Force. Available at: www.uspreventiveservicestaskforce.org/methods.htm.)

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: 1965-2012

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

71 studies included in evidence reviews, representing 82 citations. All studies included were controlled clinical trials.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

The quality of the evidence was rated by the expert panel as "moderate" - i.e., the evidence statements "are based on preliminary determination from the current best available evidence, but confidence in the estimate is constrained by one or more factors, such as: the number, size, or risk of bias of individual studies; inconsistency of findings across individual studies; limited applicability due to the populations of interest; or lack of coherence in the chain of evidence. As more information becomes available, the magnitude or direction of the observed effect could change, and this change could be large enough to alter the conclusion."

However, the clinical recommendations for fluoride among children and adolescents received an evidence grade of "in favor", which is the second highest out of six grading categories. <u>The expert panel not only made recommendations based on the study designs, but also on an evaluation on the *net benefit* of the interventions, explicitly balancing benefits to potential harms in conjunction with the level of the certainty of the evidence</u>. The full methodology is provided in Weyant et al., full report, 2013.

The evidence directly pertains to both the measure focus and the measure target population.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Recommendations:

- "2.26 percent fluoride varnish at least every three to six months" for children younger than 6 years
- "2.26 percent fluoride varnish at least every three to six months OR 1.23 percent fluoride (APF) gel for four minutes at least every three to six months" for children 6-18 years

Estimates of Benefit in Support of Recommendations:

(1) 2.26% Fluoride Varnish

"The results of meta-analyses for primary teeth indicate tha the application of 2.26% fluoride varnish has a statistically significant effect (SMD -0.19 [95% CI: -0.31, -0.08)on caries prevention as measured by increment or incidence using surface-level data." Weyant et al., full report, 2013, p. 25

"The results of meta-analyses for permanent teeth indicate that 2.26% fluoride varnish has a statistically significant effect (SMD= -0.38 [95% CI: -0.53, -0.24])on caries prevention as measured by increment or incidence using surface-level data." Weyant et al., full report, 2013, p. 25

<u>Evidence Profile (Weyant et al., full report, 2013, pp. 26-27):</u>

(a) Primary teeth (children under age 6):

- Level of certainty: Moderate
- Benefit: Yes (smaller caries increment or incidence with topical fluoride use).
    - Standardized mean difference=-0.19 [-0.31, -0.08]
    - Prevented fraction=0.27
    - Number needed to treat for control rate of 1 DMFS per year = 4
- Adverse events or harms: Little potential for harms if swallowed
- Benefit-harm assessment (Net benefit rating): Benefits outweigh potential harms
- Strength of clinical recommendation: In favor

(b) Permanent teeth (children):

- Level of certainty: Moderate
- Benefit: Yes (smaller caries increment or incidence with topical fluoride use).
    - Standardized mean difference=-0.38 [-0.53, -0.24]
    - Prevented fraction=0.36

- Number needed to treat for control rate of 1 DMFS per year = 3
- Adverse events or harms: None if used as manufacturers recommend
- Benefit-harm assessment (Net benefit rating): Benefits outweigh potential harms
- Strength of clinical recommendation: In favor

The table below (Table 8 from the report) summarizes the findings.

Table 8. Summary of standardized mean differences from meta-analysis and individual studies for 2.26% fluoride varnish studies

Outcome Measures	Number and type* of studies	Number of participants**	Standardized Mean Difference [95% Confidence Interval] (negative favors intervention, positive favors control)
Meta-analysis results: Primary teeth			
d(e/m)fs, increment or incidence†	6 RCT[31-33, 35-39] and 2 CCT[40, 41]	3,409**	**-0.19 [-0.31,-0.08]**
Meta-analysis results: Permanent teeth			
D(M)FS, increment or incidence†	8 RCT[31-33, 42-44, 46-49, 52] and 1 CCT[54]	2,574	**-0.38 [-0.53, -0.24]**
Root caries, meta-analysis results			
Root caries increment	2 RCT[50, 51]	132	**-0.67 [-1.14, -0.20]**
Individual study results			
Combined dentition	1 CCT[55]	390	**DMFS + dmfs: -1.47 [-1.70, -1.25]** **DMFT + dmft: -1.15 [-1.37, -0.94]**
DMFT	1 CCT[53]	77	**-0.13 [-0.58, 0.32]**
DS occlusal surfaces	1 RCT[45]	79	**-0.54 [-1.06, -0.03]**

Notes: * RCT = randomized controlled trial; CCT = controlled clinical trial (non-randomized); **Including all participants (not using cluster-adjusted number of participants or number of clusters); †all stages used if cavitated data not available; parentheses indicate that component was included in some of the combined results and not others

(2) 1.23% fluoride (APF) gel

"The panel concluded with moderate certainty that there is a benefit of APF gel (1.23% fluoride) application up to every three months for 4G minutes for caries prevention in the permanent teeth of 6-14 year olds. This statement is based on meta-analysis of 12 studies with moderate to high bias scores and including over 4,000 participants; although there was some inconsistency, there was low statistical heterogeneity (I2=43) between the studies." (Weyant, full report, 2013, p. 33)

Evidence Profile (Weyant et al., full report, 2013, p. 34):

Permanent teeth (children):

- Level of certainty: Moderate
- Benefit: Yes (smaller caries increment or incidence with topical fluoride use).
  - Standardized mean difference=-0.25 [-0.33, -0.16]
  - Prevented fraction=0.27
  - Number needed to treat for control rate of 1 DMFS per year = 4
- Adverse events or harms: None if used as manufacturers recommend
- Benefit-harm assessment (Net benefit rating): Benefits outweigh potential harms
- Strength of clinical recommendation: In favor

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Potential harms evaluated included: (1) nausea and vomiting associated with the ingestion of topical fluorides and (2) dental fluorosis while tooth enamel is developing until approximately age 6, due to daily ingestion of topical fluoride, such as from toothpaste or from prescription home gels.

"There is less of a concern with professionally-applied topical fluorides that have much longer intervals between applications [citing Wong et al. 2010]. Additionally, fluoride varnish has less potential for harms than other forms of high concentration topical fluoride because the amount of fluoride that is placed in the mouth with fluoride varnish is approximately one-tenth that of other professionally-applied products [citing Beltran-Aguilar et al. 2000]. The panel judged that the benefits outweighed the potential for harms for all professionally-applied or prescription-strength

topical fluorides and age groups except for children under age 6, where the risk of swallowing and associated events (particularly nausea and vomiting) outweighed the potential benefits for all professionally-applied or prescription-strength topical fluorides except 2.26% fluoride varnish." (Weyant et al., 20130, p. 10)

Citations

Beltran-Aguilar ED, Goldstein JW, Lockwood SA. Fluoride varnishes - A review of their clinical use, cariostatic mechanism, efficacy and safety. JADA 2000;131(May):589-96.

Weyant RJ, Tracy SL, Anselmo TT, Beltrán-Aguilar ED, et al; American Dental Association Council on Scientific Affairs Expert Panel on Topical Fluoride Caries Preventive Agents. Topical fluoride for caries prevention: full report of the updated clinical recommendations and supporting systematic review. Available at: http://ebd.ada.org/contentdocs/Topical_fluoride_for_caries_prevention_2013_update_-_full_manuscript.pdf

Wong MC, Glenny AM, Tsang BW, et al. Topical fluoride as a cause of dental fluorosis in children. Cochrane Database of Systematic Reviews 2010;Jan 20(1).

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

This review was published on November 2013 and reflects the latest evidence.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

Not applicable.

1a.8.2. Provide the citation and summary for each piece of evidence.

Not applicable.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

<u>If a COMPOSITE</u> (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

Inequalities in oral health status and inadequate use of oral health care services are well documented. Dental caries is the most common chronic disease in children in the United States (NCHS 2012). In 2009–2010, 14% of children aged 3 –5 years had untreated dental caries. Among children aged 6–9 years, 17% had untreated dental caries, and among adolescents aged 13–15, 11% had untreated dental caries (Dye, Li, and Thorton-Evans 2012). Dental decay among children has significant short- and long-term adverse consequences (Tinanoff and Reisine 2009). Childhood caries is associated with increased risk of future caries (Gray, Marchment, and Anderson 1991; O´Sullivan and Tinanoff 1996; Reisine, Litt, and Tinanoff 1994), missed school days (Gift, Reisine, and Larach 1992; Hollister and Weintraub 1993), hospitalization and emergency room visits (Griffin et al. 2000; Sheller, Williams, and Lombardi 1997) and, in rare cases, death (Casamassimo et al. 2009).

Identifying caries early is important to reverse the disease process, prevent progression of caries, and reduce incidence of future lesions. Evidence suggests that topical fluoride applied to children starting as early as six months of age is beneficial in preventing dental caries (Weyant et al. 2013). However, approximately three quarters of children younger

than age 6 years did not have at least one visit to a dentist in the previous year (Edelstein & Chinn 2009). Evidence-based clinical recommendations suggest that topical fluoride should be applied at least every three to six months in children at elevated risk for caries (Weyant et al. 2013).

The proposed measure, Topical Fluoride for Children at Elevated Caries Risk – Dental Services, captures whether children at moderate or high caries risk received at least two topical fluoride applications as dental services. Because topical fluoride is indicated at 3-6 month intervals (2-4 times per year) for children at elevated caries risk, at least two applications are indicated during the reporting year. This measure directly reflects evidence-based guidelines regarding an effective caries prevention measure (professionally applied topical fluoride), including the frequency required for clinical effectiveness (at least every three-six months). Topical Fluoride allows plans and programs to assess whether children at risk for caries are receiving evidence-based preventive services and target performance improvement initiatives accordingly.

Note: Procedure codes contained within claims data are the most feasible and reliable data elements for quality metrics in dentistry, particularly for developing programmatic process measures to assess the quality of care provided by programs (e.g., Medicaid, CHIP) and health/dental plans. In dentistry, diagnostic codes are not commonly reported and collected, precluding direct outcomes assessments. Although some programs are starting to implement policies to capture diagnostic information, evidence-based process measures are the most feasible and reliable quality measures at programmatic and plan levels at this point in time.

[Complete citations provided in 1c4 and in Evidence Submission Form.]

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Below are the testing data and results that met scientific acceptability criteria for endorsement. Because there were no changes in the data source, level of analysis or setting, additional testing has not been conducted.

Data Sources:

We used data from four sources and refer to "program" level information and "plan" level information. We included data for publicly insured children in the Texas Medicaid, Florida CHIP, and Florida Medicaid programs as well as national commercial data from Dental Service of Massachusetts, Inc. Florida and Texas represent two of the largest and most diverse states. The two states also represent the upper and lower bounds of dental utilization based on dental utilization data available from the Centers for Medicare and Medicaid Services. The four programs collectively represent different delivery system models. The Texas Medicaid data represented dental fee-for-service. The Florida CHIP data included da ta from two dental MCOs. The Florida Medicaid data include dental fee-for-service and prepaid dental data. The commercial data included members in indemnity and preferred provider organization (PPO) product lines. Data from calendar years 2010 and 2011 were used for all programs except Florida Medicaid. Full-year data for CY 2011 were not available for Florida Medicaid. Therefore, we report only CY 2010 data for Florida Medicaid.

In the data summaries, "Programs" refer to population data from (1) Texas Medicaid, (2) Florida CHIP, (3) Commercial Data, and (4) Florida Medicaid. "Plans" refer to data from the two dental plans that served Florida CHIP members in both 2010 and 2011.

Below we provide summary data for each of the four programs and two plans individually.

Programs

Our source data for the testing included children 0-20 years in each program. The numbers of children ages 0-20 years enrolled at least one month in each program were as follows:

Texas Medicaid, 2011: 3,544,247

Texas Medicaid, 2010: 3,393,963

Florida CHIP, 2011: 317,146

Florida CHIP, 2010: 315,975

Commercial, 2011: 184,152

Commercial, 2010: 189,968

Florida Medicaid, 2010: 2,068,670

Within these programs, we had claims data available in both years for two dental managed care plans in Florida CHIP. We also report rates for those two plans separately.

Plan 1, 2010: 77,255

Plan 2, 2010: 116,388

Plan 1, 2011: 140,986

Plan 2, 2011: 168,191

Data 1b.2. Performance Scores for Topical Fluoride, Dental Services

Program, Year, Measure Score as % (Measure Score, SD, Lower 95% CI, Upper 95% CI)

Program 1, CY 2011:	37.13%	(	0.3713 ,	0.0004 ,	0.3704 ,	0.3722)
Program 2, CY 2011:	27.15%	(	0.2715 ,	0.0020 ,	0.2676 ,	0.2754)
Program 3, CY 2011:	22.04%	(	0.2204 ,	0.0020 ,	0.2165 ,	0.2243)
Program 1, CY 2010:	34.96%	(	0.3496 ,	0.0005 ,	0.3487 ,	0.3505)
Program 2, CY 2010:	22.63%	(	0.2263 ,	0.0019 ,	0.2225 ,	0.2301)
Program 3, CY 2010:	35.04%	(	0.3504 ,	0.0023 ,	0.3458 ,	0.3550)
Program 4, CY 2010:	18.16%	(	0.1816 ,	0.0009 ,	0.1799 ,	0.1833)
Plan 1, CY 2011:	25.50%	(	0.2550 ,	0.0030 ,	0.2491 ,	0.2609)
Plan 2, CY 2011:	28.69%	(	0.2869 ,	0.0027 ,	0.2815 ,	0.2923)
Plan 1, CY 2010:	23.24%	(	0.2324 ,	0.0048 ,	0.2230 ,	0.2418)
Plan 2, CY 2010 :	23.76%	(	0.2376 ,	0.0034 ,	0.2309 ,	0.2443)

The measure score range of 18% to 35% in CY 2010 (year in which data were available for all four programs) indicates a significant performance gap overall. Two-thirds or more of children identified as being at elevated risk for caries do not receive the evidence-based recommendations of at least two topical fluoride applications during the reporting year. In addition, these results demonstrate the ability of the measure to identify variations in performance between programs.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

The measure testing findings are consistent with other data indicating that children have sub-optimal utilization of dental services in general and preventive dental services in particular. Although comprehensive dental benefits are covered under Medicaid and the Children's Health Insurance Program (CHIP), there are significant variations in use of dental services overall across states, ranging from approximately 25% to 69% (CMS EPSDT Data, FY 2011). Similar variation between states is observed among children 0-20 years of age enrolled in commercial dental plans (ADA 2013). With respect to preventive dental services more specifically, 14% to 58% of children enrolled in Medicaid/CHIP for at least 90 continuous days receive any preventive dental services (CMS EPSDT Data, FY 2011). Even among the highest performing states, 42% of publicly-insured children do not receive any type of preventive dental service during the year.

[Complete citations provided in 1c4 and in Evidence Submission Form Template.]

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out",*

disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.

The same data sources were used as described in 1b.2. The data below summarizes performance data by age, geographic location, and race/ethnicity for CY 2011 (CY 2010 for one program) with the p-values from chi-square tests used to detect whether there were statistically significant differences in performance between groups. The results demonstrate that there are disparities by age, geographic location and race/ethnicity. In addition, we also evaluated whether the measure could detect disparities by income (within program), children's health status (based on their medical diagnoses), Medicaid program type, CHIP dental plan, commercial product line, and preferred language for program communications. We additionally detected disparities by income, health status, CHIP plan, and Medicaid program type, but data on all of these characteristics were not consistently available for all programs so we are presenting disparities data on those characteristics that were most consistently available and had the greatest standardization

Data1b.4. Disparities in Performance by Child Age, Geographic Location and Race/Ethnicity

PROGRAM 1

Overall performance score:	**37.13%**
Scores by Age	
Age 1-2 years:	6.21%
Age 3-5 years:	43.07%
Age 6-7 years:	43.64%
Age 8-9 years:	42.03%
Age 10-11 years:	40.50%
Age 12-14 years:	34.83%
Age 15-18 years:	24.93%
Age 19-20 years:	11.75%
p-value from Chi-square test:	<0.0001
Scores by Geographic Location	
Urban:	37.87%
Rural:	32.50%
p-value from Chi-square test:	<0.0001
Scores by Race	
Non-Hispanic White:	30.37%
Non-Hispanic Black:	29.68%
Hispanic:	40.84%
p-value from Chi-square test:	<0.0001

PROGRAM 2

Overall performance score:	**27.15%**
Scores by Age	
Age 1-2 years:	n/a
Age 3-5 years:	30.00%
Age 6-7 years:	37.81%
Age 8-9 years:	34.88%
Age 10-11 years:	31.60%
Age 12-14 years:	27.14%
Age 15-18 years:	18.60%
Age 19-20 years:	n/a
p-value from Chi-square test:	<0.0001
Scores by Geographic Location	
Urban:	26.96%
Rural:	30.64%
p-value from Chi-square test:	<0.0001
Scores by Race	
Non-Hispanic White:	n/a
Non-Hispanic Black:	n/a
Hispanic:	n/a
p-value from Chi-square test	n/a

PROGRAM 3

Overall performance score:	**22.04%**
Scores by Age	
Age 1-2 years:	25.93%
Age 3-5 years:	34.24%
Age 6-7 years:	34.11%
Age 8-9 years:	33.97%
Age 10-11 years:	32.26%
Age 12-14 years:	28.78%
Age 15-18 years:	15.08%
Age 19-20 years:	2.22%
p-value from Chi-square test:	<0.0001
Scores by Geographic Location	
Urban:	22.13%
Rural:	19.71%
p-value from Chi-square test:	0.025
Scores by Race	
Non-Hispanic White:	n/a
Non-Hispanic Black:	n/a
Hispanic:	n/a
p-value from Chi-square test	n/a

PROGRAM 4

Overall performance score:	**18.16%**
Scores by Age	
Age 1-2 years:	17.17%
Age 3-5 years:	21.43%
Age 6-7 years:	21.19%
Age 8-9 years:	21.44%
Age 10-11 years:	19.47%
Age 12-14 years:	16.86%
Age 15-18 years:	12.53%
Age 19-20 years:	7.45%
p-value from Chi-square test:	<0.0001
Scores by Geographic Location	
Urban:	18.16%
Rural:	17.32%
p-value from Chi-square test:	0.025
Scores by Race	
Non-Hispanic White:	21.64%
Non-Hispanic Black:	15.02%
Hispanic:	17.74%
p-value from Chi-square test:	<0.0001

Note: N/A for age indicates that those ages are not within the program's age eligibility. N/A for race/ethnicity indicates that those programs did not collect race/ethnicity data or had high rates of missing data

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

There is extensive literature documenting disparities in dental service use among children by age, race/ethnicity, and geographic region, including within vulnerable populations, much of which is summarized in three major national reports on oral health: the Surgeon General's report on Oral Health in America in 2000, the IOM report, Improving Access to Oral Health Care for Vulnerable and Underserved Populations, and the IOM report, Advancing Oral Health in America.

With respect to preventive dental services, there are documented disparities. Using data from the National Survey of Children's Health, Edelstein and Chinn (2009) noted disparities in access to preventive dental services by race and income: "Stepwise disparities in access to preventive dental services are evident by race and income in ways that parallel Medical Expenditure Panel Survey findings. White parents report higher use of preventive dental services than do black or Hispanic parents (77%, 66%, and 61%, respectively). Poor parents report less use of services than do low income, middle class, and higher-income parents (58%, 66%, 77%, and 82%, respectively)" (Edelstein & Chinn, 2009, p.418). A recent analysis by Bouchery (2013) of the Medicaid Analytic eXtract files for nine states found variations in the

percentage of children receiving a preventive dental visit by age, race and ethnicity, and geographic area. Specifically, relative to the reference group of 9 year olds, the percentage point change in the probability of having a dental preventive services was -27.6 for 3 years old; -8.6 for 6 years, -2.2 for 12 years and -15.4 for 15 years (all significant at p<0.0001); relative to the reference group of white, non-Hispanic, the percentage point change was -1.8 for black non-Hispanic and 7.8 for Hispanic (p<0.0001 for both); relative to the reference group of small metro area, the percentage point change was 5.9 for large metro area (p<0.0001).

Sources

Bouchery, E. 2013. "Utilization of Dental Services among Medicaid-Enrolled Children." Medicare & Medicaid Research Review. 3(3) E1-16. Available at: https://www.cms.gov/mmrr/Downloads/MMRR2013_003_03_b04.pdf.

Dietrich, T., C. Culler, R. Garcia, and M. M. Henshaw. 2008. Racial and ethnic disparities in children's oral health: The National Survey of Children's Health. Journal of the American Dental Association 139(11):1507-1517.

Dye BA, Li X, Thorton-Evans G. Oral health disparities as determined by selected healthy people 2020 oral health objectives for the United States, 2009-2010. NCHS Data Brief 2012(104):1-8.U.S. Dept. of Health and Human Services, National Institute of Dental and Craniofacial Research.

Edelstein, B. L. and C. H. Chinn. 2009. "Update on Disparities in Oral Health and Access to Dental Care for America´s Children." Acad Pediatr 9(6): 415-9.

Institute of Medicine (U.S.). Committee on an Oral Health Initiative. Advancing oral health in America. Washington, D.C.: National Academies Press; 2011.

Institute of Medicine and National Research Council. Improving access to oral health care for vulnerable and underserved populations. Washington, D.C.: National Academies Press; 2011.

Kenney, G. M., J. R. McFeeters, and J. Y. Yee. 2005. Preventive dental care and unmet dental needs among low-income children. American Journal of Public Health 95(8):1360-1366.

Lewis, C., W. Mouradian, R. Slayton, and A. Williams. 2007. Dental insurance and its impact on preventative dental care visits for U.S. children. Journal of the American Dental Association 138(3):369-380.

U.S. Dept. of Health and Human Services, National Institute of Dental and Craniofacial Research. Oral health in America : a report of the Surgeon General. Rockville, Md.: U.S. Public Health Service, Dept. of Health and Human Services; 2000.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

Ears, Nose, Throat (ENT)

De.6. Non-Condition Specific*(check all the areas that apply):*

Access to Care, Disparities Sensitive, Health and Functional Status : Change, Health and Functional Status : Total Health, Primary Prevention

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

Children, Populations at Risk

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

http://www.ada.org/~/media/ADA/Science%20and%20Research/Files/DQA_2018_Dental_Services_Topical_Fluoride.pdf?la=en

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

No data dictionary **Attachment:**

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

1. No changes to the measure specifications

2. Measure specification website updated to be more user friendly

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Unduplicated number of enrolled children aged 1-21 years who are at "elevated" risk (i.e., "moderate" or "high") who received at least 2 topical fluoride applications as a dental service

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Please see section S14.

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

Unduplicated number of enrolled children aged 1-21 years who are at "elevated" risk (i.e., "moderate" or "high")

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Please see Section S14.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

Medicaid/CHIP programs should exclude those individuals who do not qualify for dental benefits. The exclusion criteria should be reported along with the number and percentage of members excluded.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

There are no other exclusions than those described above

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

This measure is stratified by age using the following categories:

1-2; 3-5; 6-7; 8-9; 10-11; 12-14; 15-18; 19-20

No new data are needed for this stratification. Please see attached specifications for complete measure details.

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

Topical Fluoride Intensity Calculation for Children at Elevated Caries Risk

1. Use administrative enrollment and claims data for a single year. When using claims data to determine service receipt, include both paid and unpaid claims (including pending, suspended, and denied claims).

2. Check if the enrollee meets age criteria at the last day of the reporting year:

a. If child is >=1 and < 21, then proceed to next step.

b. If age criteria are not met or there are missing or invalid field codes (e.g., date of birth), then STOP processing. This enrollee does not get counted.

3. Check if subject is continuously enrolled for the reporting year (12 months) with a gap of no more than 31 days (one month gap for programs that determine eligibility on a monthly basis):

a. If subject meets continuous enrollment criterion, then proceed to next step.

b. If subject does not meet enrollment criterion, then STOP processing. This enrollee does not get counted.

YOU NOW HAVE THE COUNT OF THOSE WHO MEET THE AGE AND ENROLLMENT CRITERIA

4. Check if subject is at "elevated risk":

a. If subject meets ANY of the following criteria, then include in denominator:

i. the subject has a CDT Code among those in Table 1 in the reporting year,

OR

ii. the subject has a CDT Code among those in Table 1 in any of the three years prior to the reporting year, (NOTE: The subject does not need to be enrolled in any of the prior three years for the denominator enrollment criteria; this is a "look back" for enrollees who do have claims experience in any of the prior three years.)

OR

iii. the subject has a visit with a CDT code = (D0602 or D0603) in the reporting year.

b. If the subject does not meet any of the above criteria for elevated risk, then STOP processing. This enrollee will not be included in the measure denominator.

YOU NOW HAVE THE DENOMINATOR (DEN): Enrollees who are at "elevated risk"

5. Check if subject received at least two fluoride applications as dental service during the reporting year – at least two unique dates of service when topical fluoride was provided. Service provided on each date of service should satisfy the following criteria:

a. If [CDT CODE] = D1206 or D1208 , and

b. If [RENDERING PROVIDER TAXONOMY] code = any of the NUCC maintained Provider Taxonomy Codes in Table 1 below, then include in numerator; proceed to next step.

c. If both a AND b are not met, then the service was not a "dental service"; STOP processing. This enrollee is already included in the denominator but will not be included in the numerator.

Note 1: No more than one fluoride application can be counted for the same member on the same date of service.

Note 2: All claims with missing or invalid CDT CODE, missing or invalid NUCC maintained Provider Taxonomy Codes, or NUCC maintained Provider Taxonomy Codes that do not appear in Table 2 should not be included in the numerator.

YOU NOW HAVE NUMERATOR (NUM) COUNT: Enrollees at "elevated risk" who received fluoride as a dental service

6. Report

a. Unduplicated number of enrollees in numerator

b. Unduplicated number of enrollees in denominator

c. Measure Rate (NUM/DEN)

d. Rate stratified by age

Table 1: 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 D2941

D2392 D2610 D2710 D2783 D2950 D1354

D2393 D2620 D2712 D2790 D3110

Table 2: NUCC maintained Provider Taxonomy Codes classified as "Dental Service"*

122300000X	1223P0106X	1223X0008X	261QF0400X
1223D0001X	1223P0221X	1223X0400X	261QR1300X
1223D0004X	1223P0300X	124Q00000X+	125Q00000X
1223E0200X	1223P0700X	125J00000X	
1223G0001X	1223S0112X	125K00000X	

*Services provided by County Health Department dental clinics may also be included as "dental" services.

+Only dental hygienists who provide services under the supervision of a dentist should be classified as "dental" services. Services provided by independently practicing dental hygienists should be classified as "oral health" services and are not applicable for this measure.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

Not applicable.

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

Not applicable.

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Claims

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

Not applicable.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Health Plan, Integrated Delivery System

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Outpatient Services

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

Not applicable.

2. Validity – See attached Measure Testing Submission Form

5_Testing_top_flouride.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

No

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

No

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

No - This measure is not risk-adjusted

Measure Testing (subcriteria 2a2, 2b2-2b6)

Measure Title: Prevention: Topical Fluoride for Children at Elevated Caries Risk, Dental Services
Date of Submission: 2/12/2014

Type of Measure:

☐ **Composite –** *STOP – use composite testing form*	☐ **Outcome (*including PRO-PM*)**
☐ Cost/resource	☒ **Process**
☐ Efficiency	☐ Structure

Instructions

- 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.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (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 (2b2-2b6) 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 20 pages (*incuding 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.

Note: The information provided in this form is intended to aid the Steering 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.

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.

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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) 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.

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

16. 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 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:
☐ abstracted from paper record	☐ abstracted from paper record
☒ **administrative claims**	☒ **administrative claims**
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

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*).

The testing datasets were consistent with the measure specifications for the target populations and reporting entities. This measure was specified for administrative enrollment and claims data for children with private or public insurance coverage. We used data from four sources and refer to "program" level information and "plan" level information. We included data for publicly insured children in the Texas Medicaid, Florida CHIP, and Florida Medicaid programs as well as national commercial data from Dental Service of Massachusetts, Inc. Florida and Texas represent two of the largest and most diverse states. The two states also represent the upper and lower bounds of dental utilization based on dental utilization data available from the Centers for Medicare and Medicaid Services. The four programs collectively represent different delivery system models. The Texas Medicaid data represented dental fee-for-service. The Florida CHIP data included data from two dental MCOs. The Florida Medicaid data include dental fee-for-service and prepaid dental data. The commercial data included members in indemnity and preferred provider organization (PPO) product lines.

1.3. What are the dates of the data used in testing We used data from calendar years 2010 and 2011 for all programs except Florida Medicaid. Full-year data for 2011 were not available for Florida Medicaid.

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:
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ **health plan**	☒ **health plan**
☒ **other: Program (e.g., Medicaid, CHIP)**	☒ **other: Program (e.g., Medicaid, CHIP)**

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*)

Level of Analysis: Program, 4 Measured Entities

1. Texas Medicaid
    A. Size: # Members 0-20 years, CY 2011: 3,554,247; # Members 0-20 years, CY 2010: 3,393,963
    B. Location: Texas – Statewide
    C. Delivery Type – FFS
2. Florida CHIP
    A. Size: # Members 0-20 years, CY 2011: 317,146; # Members 0-20 years, CY 2010: 315,975
    B. Location: Florida – Statewide
    C. Delivery Type – Dental MCO (2 plans)
3. Commercial
    A. Size: # Members 0-20 years, CY 2011: 184,152; # Members 0-20 years, CY 2010: 189,968
    B. Location: National
    C. Delivery Type – Indemnity/FFS & PPO product lines
4. Florida Medicaid
    A. Size: # Members 0-20 years, CY 2010: 2,068,670
    B. Location: Florida – Statewide
    C. Delivery Type – FFS and Prepaid Dental

Note: At the time of testing, complete data were not available for Florida Medicaid for CY 2011.

Level of Analysis: Plan, 2 Measured Entities

The FL CHIP program had two separate dental plans that participate in the program in 2010 and 2011.

1) FL CHIP – Plan 1
    1) Size: # Members 0-20 years, CY 2011: 140,986; # Members 0-20 years, CY 2010: 77,255
    B. Location: Florida – Statewide
    C. Delivery Type – Dental MCO

2) FL CHIP – Plan 2

A. Size: # Members 0-20 years, CY 2011: 168,191; # Members 0-20 years, CY 2010: 116,388

B. Location: Florida – Statewide

C. Delivery Type – Dental MCO

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*)

Note that there were only three programs in CY 2011 because Florida Medicaid did not have complete claims data available for CY 2011 at the time testing was conducted.

Table 1.6A, Patient Characteristics, 0-20 Years Old, 2011

	Descriptive Characteristics of Individuals 0-20 Years Enrolled at Least One Month, CY 2011				
	Program 1	Program 2	Program 3	Plan 1	Plan 2
Total Number Patients	3,544,247	317,146	184,152	140,986	168,191
Age Group Distribution					
Age <1 years	7.05%	N/A	1.54%	N/A	N/A
Age 1-2 years	14.32%	N/A	5.75%	N/A	N/A
Age 3-5 years	19.46%	3.81%	12.68%	4.12%	3.60%
Age 6-7 years	11.21%	13.05%	9.57%	13.71%	12.55%
Age 8-9 years	9.85%	15.00%	10.18%	15.76%	14.41%
Age 10-11 years	9.03%	15.71%	10.55%	16.27%	15.25%
Age 12-14 years	11.63%	23.73%	16.09%	23.06%	24.31%
Age 15-18 years	13.19%	28.70%	22.13%	27.08%	29.88%
Age 19-20 years	4.27%	N/A	11.50%	N/A	N/A
Geographic Location					
Urban	83.63%	92.94%	95.95%	93.01%	92.91%
Rural	15.15%	5.02%	3.86%	4.83%	5.15%
Missing	1.22%	2.04%	0.19%	2.16%	1.94%
Race and Ethnicity					
Non-Hispanic White	17.36%	N/A	N/A	N/A	N/A
Non-Hispanic Black	15.08%	N/A	N/A	N/A	N/A
Hispanic	58.07%	N/A	N/A	N/A	N/A
Other & Unknown	9.49%	N/A	N/A	N/A	N/A

Table 1.6B, Patient Characteristics, 0-20 Years Old, 2010

	Descriptive Characteristics of Individuals 0-20 Years Enrolled at Least One Month, CY 2010					
	Program 1	Program 2	Program 3	Program 4	Plan 1	Plan 2
Total Number Patients	3,393,963	315,975	189,968	2,068,670	77,255	116,388
Age Group Distribution						
Age <1 years	7.35%	N/A	1.45%	6.05%	N/A	N/A
Age 1-2 years	15.16%	N/A	5.67%	14.23%	N/A	N/A
Age 3-5 years	19.48%	3.64%	12.73%	19.26%	5.72%	4.22%
Age 6-7 years	11.12%	13.32%	9.69%	10.47%	15.68%	12.54%
Age 8-9 years	9.70%	15.14%	10.24%	9.19%	16.99%	14.21%
Age 10-11 years	8.75%	15.84%	10.60%	8.74%	16.41%	15.18%
Age 12-14 years	11.23%	23.70%	16.20%	11.87%	21.40%	24.05%
Age 15-18 years	12.99%	28.37%	22.12%	14.73%	23.79%	29.81%
Age 19-20 years	4.22%	N/A	11.31%	5.47%	N/A	N/A
Geographic Location						
Urban	83.20%	92.08%	96.70%	91.47%	92.10%	92.11%
Rural	15.56%	5.07%	3.17%	7.30%	5.00%	5.19%
Missing	1.24%	2.85%	0.13%	1.23%	2.89%	2.70%
Race and Ethnicity						
Non-Hispanic White	18.21%	N/A	N/A	29.89%	N/A	N/A
Non-Hispanic Black	15.45%	N/A	N/A	29.39%	N/A	N/A
Hispanic	59.42%	N/A	N/A	29.65%	N/A	N/A
Other & Unknown	6.92%	N/A	N/A	11.06%	N/A	N/A

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.

These data were used for all testing aspects except two:

A. Part of the face validity assessments involved expert consensus processes, including conducting an environmental scan of measure concepts and using the RAND-UCLA modified Delphi process to rate the importance, feasibility and validity. Please see section 2b2.2 for a complete description.

B. Data element validation using medical chart reviews did not include all programs. Due to the cost of these activities, chart reviews were conducted only for the Texas Medicaid program. Texas has the third largest Medicaid program in the U.S. with significant diversity represented. In addition, the research team conducting the testing is the External Quality Review Organization for Texas and has years of experience conducting medical chart audits for the Texas Medicaid program for ongoing quality assurance purposes. Thus, an established infrastructure and expertise was in place to conduct chart reviews for these programs.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

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

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

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

2a2.2. For each level checked above, 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*)

Data Elements:

- See section 2b2 for validity testing of data elements.
- Note: Unlike measures that rely on medical record data for which issues such as inter-rater reliability are likely to introduce measurement concerns or measures that rely on survey data for which issues such as internal consistency may be a concern, this measure relies on standard data fields commonly used in administrative data for a wide range of billing and reporting purposes.

Measure Score – Threats to Measure Reliability

An important component of assessing reliability is assessing, testing, and addressing threats to measure reliability.

1. Evaluation of Clarity and Completeness of Measure Specifications

For a measure to be reliable – to allow for meaningful comparisons across entities – the measure specifications must be unambiguous: the denominator criteria, numerator criteria, exclusions, and scoring need to be clearly specified. The initial measure specifications were developed by the Dental Quality Alliance (DQA). The Dental Quality Alliance includes 30 members, representing a broad range of stakeholders, including federal agencies involved with oral health services, dental professional associations, medical professional associations, dental and medical health insurance commercial plans, state Medicaid and CHIP programs, quality accrediting bodies, and the general public. The initial specifications were developed based on (1) the evidence regarding the effectiveness of professionally applied topical fluoride in caries prevention, (2) an environmental scan, and (3) face validity assessments of the measure concept. These specifications were contained in the competitive Request for Proposals to conduct measure testing; a research team from the University of Florida was selected to conduct testing. The research team independently carefully evaluated whether the measure specifications identified all necessary data elements to calculate the numerators and denominators for each measure. In addition, the research team carefully reviewed the logic flow and made revision recommendations to improve the reliability of the resulting calculations. The DQA also solicited public comment on an Interim Report and posted the measurement specifications online for public comment. The research 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 research team and the DQA to ensure clear and detailed measure specifications.

2. Other Threats to Reliability - Sample Size

Our measured entities include very large numbers of patients; small sample size is not a concern.

2a2.3. For each level checked above, 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*)

See section 2b2 for validity testing of data elements.

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?*)

See section 2b2 for validity testing of data elements.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

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

☐ **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*)

2b2.2. For each level 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)*

We assessed (1) critical data element validity, (2) measure score validity, and (3) potential threats to validity.

1. CRITICAL DATA ELEMENT VALIDITY

Topical Fluoride measures the percentage of children aged 1-21 years at moderate to high risk for dental caries who had at least 2 topical fluoride applications during the reporting year. The critical data elements for this measure include: (1) member ID (to link between claims and enrollment data), (2) date of birth, (3) monthly enrollment indicator, (4) date of service, and (5) CDT codes. The first four items are core fields used in virtually all measures relying on administrative data and essential for any reporting or billing purposes. As such, it was determined that these fields have established reliability and validity. Thus, critical data element validity testing focused on assessing the accuracy of the dental procedure codes reported in the claims data as the data elements that contribute most to the measure score. To evaluate data element validity, we conducted reviews of dental records for the Texas Medicaid program. Validation of clinical codes in administrative claims data are most often conducted using manual abstraction from the patient's full chart as the authoritative source. As described in detail below, we evaluated agreement between the claims data and dental charts by calculating the sensitivity, specificity, positive predictive value, and negative predictive value as well as the kappa statistic.

A. Data Sources & Methodology

A. Data Sources

A random sample of encounters for members ages 3-18 years with at least one outpatient dental visit was selected for dental record reviews. The targeted number of records was 400. The expected response rate for returning records was 65%. Therefore, 600 records were requested. All outpatient dental records for members during an eight-month period were requested. Table 2b2.2-1 below summarizes the number of records requested and received. The number of eligible records received (414) exceeded the total targeted number of 400 records.

Table 2b2.2-1 Dental Records Requested and Received

# Requested	# Received	%Received
600	414	69%

B. Record Review Methodology

There were two components to the record reviews used to evaluate data element validity:

1. Encounter data validation (EDV) that provided an overall assessment of the accuracy of dental procedure codes found in the administrative claims data compared to dental records for the same dates of service.
2. Validation of topical fluoride application procedure codes specifically.

The record reviews were conducted by two coders certified as registered health information technicians (RHITs). At weekly intervals during the record review process, the two RHITs randomly selected a sample of records to evaluate inter-rater reliability. A total of 100 records and 1,830 fields were reviewed by both individuals with 100% agreement.

C. Encounter Data Validation – Overall Assessment

For the first component of validation, encounter data validation, the research team followed standard Encounter Data Validation processes following External Quality Review protocols from CMS that it has used in ongoing quality assurance activities for the Texas Health and Human Services Commission. [Centers for Medicare and Medicaid Services, External Quality Review Encounter Data Validation Protocol (http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Quality-of-Care/Quality-of-Care-External-Quality-Review.html)]. The first three procedure codes were reviewed for each claim. A total of 1,135 procedure codes were reviewed. The RHITs were provided with a pre-populated data

entry form with the codes from the claims data for the patient with the specified provider on a particular date of service. They evaluated whether the code in the claims data was supported by the dental record.

D. Critical Data Element Validation – Topical Fluoride Application Procedure Codes

Data Extraction. For the second component of validation, assessing whether the specific preventive service of topical fluoride application is accurately captured by claims data, chart abstraction forms were developed by the research team. The chart abstraction forms and process were reviewed and approved by the DQA R&D Committee. Claims data were validated against dental records by comparing the dental records to the codes in the claims data for a randomly selected date of service. Prior to conducting the reviews, a sample of 30 records from prior encounter data validation activities was used to test the data abstraction tool and refinements were made accordingly. During the chart abstraction testing process, the RHITs met with the research team, which included two dentists (including a pediatric dentist), to review questions about interpreting the records. They then evaluated the 414 dental records using the data abstraction form. The results were recorded in an Access database. Specifically, the chart abstracting process involved identifying and recording whether there was any evidence of fluoride application during the visit. The programming team extracted data from the administrative claims data for the same members and dates of service, recording the presence or absence of topical fluoride procedure codes. The data files from the record review team and the programming team were merged into a single data file.

Statistical Analysis. To assess validity, we calculated sensitivity (accuracy of administrative data indicating a service was received when it is present in the chart), specificity (accuracy of administrative data indicating a service was not received when it is absent in the chart), positive predictive value (extent to which a procedure that is present in the administrative data is also present in the charts), and negative predictive value (extent to which a procedure that is absent from the administrative data is also absent in the chart). Positive and negative predictive values are influenced by sensitivity and specificity as well as the prevalence of the procedure. Thus, interpretation of "high" and "low" values is not straightforward. In addition, although charts are typically used as the authoritative source for validating claims data, some question whether charts always represent an "authoritative" source versus being better characterized as a "reference" standard. The kappa statistic has been recommended as "a more 'neutral' description of agreement between the 2 data sources . . . ." (Quan H, Parsons GA, Ghali WA, Validity of procedure codes in International Classification of Diseases, 9th revision, clinical modification administrative data, Med Care, 2004;42(8):801-809.) Thus, the kappa statistic also was used to compare the degree of agreement between the two data sources. A kappa statistic value of 0 reflects the amount of agreement that would be expected to be observed by chance. A kappa statistic value of 1 indicates perfect agreement. Guidance on interpreting the kappa statistic is: <0 (poor/less chance of agreement; 0.00-0.20 (slight agreement); 0.21-0.40 (fair agreement); 0.41-0.60 (moderate agreement); 0.61-0.80 (substantial agreement); 0.81-0.99 (almost perfect agreement). (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.)

2. MEASURE SCORE - FACE VALIDITY

Face validity of this measure was assessed at several stages during the measure development and testing processes.

A. Face Validity Assessment – Measure Development

Face validity was systematically assessed by recognized experts. The Dental Quality Alliance (DQA) was formed at the request of the Centers of Medicare and Medicaid Services (CMS) specifically for the purpose of bringing together recognized expertise in oral health to develop quality measures through consensus processes. As noted in the letter from Cindy Mann, JD, Director of the Center for Medicaid & CHIP Services within CMS: "The dearth of tested quality measures in oral health has been a concern to CMS and other payers of oral health services for quite some time." (See Appendix)

During the measurement development process, the DQA Research and Development Committee, purposely comprised of individuals with recognized and appropriate expertise in oral health to lead quality measure development, undertook an environmental scan of existing pediatric oral health performance measures, which involved the following: (1) Literature Search, (2) Measure Solicitation, (3) Review of Measure Concepts, (4)Delphi Ratings of Measure Concepts, (5) Scan Results Analysis, (6) Gap Analysis, (7) Identification of Measures. A more detailed description of this process, the findings and the resulting measure concepts that were pursued is provided in reports published by the DQA. (Dental

Quality Alliance. Pediatric Oral Health Quality and Performance Measures: Environmental Scan. 2012; Dental Quality Alliance. Pediatric Oral Health Quality & Performance Measure Concept Set: Achieving Standardization & Alignment. 2012. Both reports available at: http://ada.org/7503.aspx.)

(1) Literature Search. The Committee began its work by identifying existing performance and quality measure concepts (description, numerator, and denominator) on pediatric populations defined as children younger than 21 years. Staff conducted a comprehensive online search for publicly available measure concepts. This search was conducted initially in August – September 2011 and then updated on February 8, 2012. The following searches were conducted: (1) PubMed Search. Staff used two specific search strategies to search Medline. Search 1: (performance OR process OR outcome OR quality) AND measure AND (oral or dental) AND (children OR child OR pediatric OR paediatric) – 1121 citations. Search 2 - "Quality Indicators, Health Care"[Mesh] AND (dental OR oral) - 150 citations. Staff included five articles based on title and abstract review of these citations. Measure concepts presented within these articles were included in the list of concepts for R&D Committee review. (2) Web Search. Staff then performed an internet search with keywords similar to the ones used for the PubMed search. (3) Search of relevant organization websites. Staff began this search through the links provided within the National Library of Medicine database of relevant organizations (http://www.nlm.nih.gov/hsrinfo/quality.html#760). Example of organizations involved in quality measurement include the National Quality Measures Clearinghouse (NQMC), National Quality Forum (NQF), and Maternal and Child Health Bureau (MCHB).

(2) Solicitation of Measures. In addition, the R&D Committee contacted staff at the Agency for Healthcare Research and Quality (AHRQ) in August 2011 to obtain the measures collected by the Subcommittee on Children's Healthcare Quality for Medicaid and CHIP programs (SNAC). The Committee solicited measures from other entities, such as the DentaQuest Institute, involved in measure development activities.

(3) Review of Measure Concepts. Using inclusion/exclusion criteria, the R&D Committee reviewed the measure concepts and identified the measures that would be reviewed and rated in greater depth.

(4) Delphi Ratings. The RAND-UCLA modified Delphi approach was used to rate the remaining measure concepts, applying the criteria and scoring system for importance, validity, and feasibility consistent with the process that was used by the SNAC. There were two rounds of Delphi ratings to identify a starter set of pediatric oral health performance measures. [Brook RH. The RAND/UCLA appropriateness method. In: McCormick KA, Moore SR, Siegel R, United States. Agency for Health Care Policy and Research. Office of the Forum for Quality and Effectiveness in Health Care., editors. Clinical practice guideline development : methodology perspectives.]

(5) Scan Results. There were a total of 112 measure concepts identified through the environmental scan: 59 met the inclusion criteria for being processed through the Delphi rating process and 53 did not. Among the 59 measures that were evaluated through the Delphi rating process, 38 were deemed "low-scoring measure concepts" and 21 were deemed "high-scoring measure concepts."

(6) Gap Analysis. The R&D Committee then identified the gaps in existing measures, including both gaps in terms of the care domains addressed (e.g., use of services, prevention, care continuity) as well as gaps based on good measurement practices (e.g., standardized measurement methodology, evidence-based, etc.). Although the Committee did identify content areas that were not addressed, a key finding was the lack of standardized, clearly-specified, validated measures.

(7) Identification of Measures. The findings were used to identify a starter set of measures that would achieve the following objectives: (a) uniformly assess the quality of care for comparison of results across private/public sectors and across state/community and national levels; (b) inform performance improvement projects longitudinally and monitor improvements in care; (c) identify variations in care, and (d) develop benchmarks for comparison.

B. Face Validity Assessment – Measure Testing

The research team and the DQA R&D Committee continued to assess face validity throughout the testing process. Face validity also was gauged through feedback solicited through public comment periods. In March 2013, an Interim Report describing the measures, testing process, and preliminary results was sent to a broad range of stakeholders, including representatives of federal agencies, dental professionals/professional associations, state Medicaid and CHIP programs, community health centers, and pediatric medical professionals/professional associations. Each comment received was

carefully reviewed and addressed by the research team and DQA, which entailed additional sensitivity testing and refinement of the measure specifications. Draft measure specifications were subsequently posted on the DQA's website in a public area and public comment was invited. National presentations, including presentations at the National Oral Health Conference, were made by the research team and DQA in the spring and summer of 2013, which included reference to the website containing the measure specifications and invitations to provide feedback. All comments received were reviewed and addressed by the research team and DQA, including additional sensitivity testing and refinement of the measure specifications.

The final face validity assessment was conducted at the July 2013 Dental Alliance Quality meeting at which the full membership, representing a broad range of stakeholders. A detailed presentation of the testing results was provided. The membership then participated in an open consensus process with observed unanimous agreement that the calculated measure scores can be used to evaluate quality of care.

Sample Presentations

Aravamudhan K. Dental Quality Alliance Measures. Presentation at 2013 National Oral Health Conference Pre-Conference Workshop on Objectives, Indicators, Measures and Metrics. 2013.

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.

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.

3. ADDITIONAL VALIDITY TESTING - IDENTIFYING ELEVATED RISK WITH CLAIMS DATA

Evidence based guideline indicate that fluoride is most effective for children at higher risk for caries. Thus, inclusion in the denominator is limited to children identified as being at moderate to high risk for caries. Administrative claims data for dental claims typically do not include diagnostic codes. Procedure codes for risk assessment that identify moderate and high risk were included in the measure logic. However, because these are newer codes, additional logic was included to identify children with recent history of restorations, which are indicative of caries. A systematic review found that prior caries experience to be an important predictor of future risk (Zero D, Fontana M, Lennon AM. 2001. Clinical applications and outcomes of using indicators of risk in caries management. J Dent Educ. 2001 Oct;65(10):1126-32.) Expert consensus and validation through chart reviews was done to finalize the procedure codes (indicated in the measure specifications) used to identify elevated risk. The test data results reported in this application demonstrate that it is feasible to use these validated codes to identify children at elevated risk who should receive preventive services.

4. ADDITIONAL VALIDITY EVALUATION – ASSESSMENT OF THREATS TO VALIDITY

A. Exclusions

As described in 2b3. of this form, there are no exclusions for this measure.

B. Risk Adjustment

Risk adjustment is not applicable for this process measure.

C. Missing Data

As described in measure evaluation criteria 3c1, this measure relies on standard data elements in claims data that are already collected and widely used for a range of reporting and billing purposes with very low rates of missing or invalid data (which we empirically assessed and reported in 3c1).

D. Multiple Sets of Specifications

This does not apply to the proposed measure.

E. Ability to Identify Statistically Significant and Meaningful Differences in Performance

As described in 2b5 of this form, this measure is able to identify statistically significant and meaningful differences in performance. We also demonstrate with empirical data and statistical testing the ability of this measure to detect disparities in 1b4 (Importance).

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

1. CRITICAL DATA ELEMENT VALIDITY

A. Encounter Data Validation – Overall Assessment

Encounter data validation of 1,135 procedure codes in the claims data against dental charts found agreement for 94% of the procedure codes (Table 2b2.3-1). Only 4.2% of procedure codes reported in the administrative data were not supported by evidence in the dental record. For 1.8% of the records reviewed, the documentation was insufficient to determine whether the service indicated by the procedure code had been rendered or not.

Table 2b2.3-1 Agreement between Records and Administrative Data for Procedures

Number of Procedure Codes	Record and Procedure Code on Claim Correlate	Record Did Not Correlate with Procedure Code on Claim	Unable to Determine Correlation
1,135	94.04%	4.22%	1.75%

B. Critical Data Element Validation – Topical Fluoride Application Procedure Codes

To assess whether the specific preventive service of topical fluoride application is accurately captured by claims data, the 414 records, representing 631 dates of service, were reviewed. Table 2b2.3-2 below summarizes the agreement between the dental records and administrative data for topical fluoride applications. Agreement (concordance) for topical fluoride application was 89.9%. Sensitivity was 90.7% and specificity was 88.4%. The positive predictive value was 93.5% and negative predictive value was 83.9%. As noted above, the kappa statistic provides a more neutral description of agreement and extends a comparison of simple agreement by taking into account agreement occurring by chance, thereby providing a more rigorous and conservative measure of agreement between the two data sources. The kappa statistic value was 0.782, which is at the high end of the "substantial agreement" category.

Table 2b2.3-2 Agreement between Record and Administrative Data for Specific Services

	Concordance	Prevalence	Sensitivity	Specificity	PPV	NPV	Kappa
Fluoride	89.91%	0.647	0.907	0.884	0.935	0.839	0.782
Dates of service: 317			(0.857-0.942)	(0.806-0.934)	(0.888-0.963)	(0.757-0.898)	(0.710-0.853)
# indeterminate: 0							

95% confidence intervals indicated in parentheses

Our findings are similar to those in the peer-reviewed literature. A study was conducted in 2004 that used data from 3,751 patient visits in 120 dental practices participating in the Ohio Practice-Based Research Network to examine the concordance of chart and billing data with direct observation of dental procedures. For fluoride, they found lower sensitivity (80%), higher specificity (98%) and similar kappa value (0.81) of billing data compared to direct observation. (Demko CA, Victoroff KZ, Wotman S. 2008. "Concordance of chart and billing data with direct observation in dental practice" Community Dent Oral Epidemiol. 36(5):466-74.)

2. FACE VALIDITY

The measures concept of preventive dental services identified using CDT codes (within which topical fluoride falls) was identified through the Delphi rating process as a high-scoring measure concept with a mean importance score of 7, mean feasibility score of 8, and mean validity score of 7 for specific evidence-based preventive services, all out of a 9-point scale. [Rating of 1-3: not scientifically sound and invalid; 4-6 – uncertain scientific soundness and uncertain validity; 7-9 – scientifically sound and valid.] Thus, the measure has face validity. However, gaps were identified with existing preventive services measures, including defining "preventive services" too broadly (encompassing services without sound evidence of their effectiveness in caries prevention), lack of clear specifications and lack of standardization. Although the scan included two measure concepts that were specific to fluoride, they were deemed to be low scoring because they pertained to "fluoride supplements" or "fluoride exposure assessment." Scientific soundness was limited due to lack of clarity in measure description.

Content Validity. In addition, the measure also demonstrates **content validity** – the extent to which the measure specifications reflect the intended domain of care. This measure directly reflects evidence-based guidelines regarding an effective caries prevention measure (professionally applied topical fluoride), including the frequency required for clinical effectiveness (at least every three-six months). Please see the Measure Evidence Form for more details.

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?*)

As noted above, the overall agreement between the administrative claims data and dental record data was high based on both simple agreement and using the more conservative Kappa statistic. Overall, we interpret these findings as evidence that validates the accuracy of administrative claims data for performance measurement purposes. These empirical findings, combined with our face validity and content validity assessments of the measure score, lead us to conclude that both the data elements and the measure score represent valid measures of the evidence-based preventive service topical fluoride application.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section* 2b4

The only exclusions were those that are standard exclusions in any measure reporting: children who do not qualify for dental benefits under their coverage were not included because this measure is intended only for children with dental coverage. For example, individuals 0-20 years with Medicaid coverage for emergency services only or for pregnancy-related services that do not provide dental coverage were not included.

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

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

Not applicable.

2b4.1. What method of controlling for differences in case mix is used?

☒ No risk adjustment or stratification

☐ Statistical risk model with _risk factors

☐ Stratification by _risk categories

☐ Other,

2b4.2. If an outcome or resource use 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 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 and not related to disparities*)

Not applicable

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

Not applicable.

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*)

Not applicable.

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:

Not applicable.

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*)

Not applicable.

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

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)*

This is a new measure. As noted in 1b, there were variations in the measure scores across the four programs included in the testing. For convenience we have included the performance score data from 1b below. In addition to providing the 95% confidence intervals for each score, we used chi-square tests to analyze whether there were statistically significant differences between (1) the 3 programs with performance data for 2011, (2) the 4 programs with performance data for 2010, (3) the two dental MCOs in FL CHIP in CY 2010 and (4) the two dental MCOs in FL CHIP in CY 2011. Because the measure score is the proportion of children who received two topical fluoride applications, the dichotomous outcome of had/did not have two topical fluoride applications can be used to conduct chi-square significance testing in order to evaluate whether there are statistically significant differences in the measure scores between programs and between plans.

Table 1b.2. Performance Scores

Program/Plan, Year, Measure Score as % (Measure Score, SD, Lower 95% CI, Upper 95% CI)

Program 1, CY 2011:	37.13%	(	0.3713 ,	0.0004 ,	0.3704 ,	0.3722)
Program 2, CY 2011:	27.15%	(	0.2715 ,	0.0020 ,	0.2676 ,	0.2754)
Program 3, CY 2011:	22.04%	(	0.2204 ,	0.0020 ,	0.2165 ,	0.2243)
Program 1, CY 2010:	34.96%	(	0.3496 ,	0.0005 ,	0.3487 ,	0.3505)
Program 2, CY 2010:	22.63%	(	0.2263 ,	0.0019 ,	0.2225 ,	0.2301)
Program 3, CY 2010:	35.04%	(	0.3504 ,	0.0023 ,	0.3458 ,	0.3550)
Program 4, CY 2010:	18.16%	(	0.1816 ,	0.0009 ,	0.1799 ,	0.1833)
Plan 1, CY 2011:	25.50%	(	0.2550 ,	0.0030 ,	0.2491 ,	0.2609)
Plan 2, CY 2011:	28.69%	(	0.2869 ,	0.0027 ,	0.2815 ,	0.2923)
Plan 1, CY 2010:	23.24%	(	0.2324 ,	0.0048 ,	0.2230 ,	0.2418)
Plan 2, CY 2010 :	23.76%	(	0.2376 ,	0.0034 ,	0.2309 ,	0.2443)

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*)

For both years, statistically significant differences were detected in the measure scores between programs in both years and between plans in one of the two years (Table 2b5.2).

Table 2b5.2. Chi-Square Test of Differences in Measure Scores

	Chi-Square Value	*p* - value
Program Results, 2011	5887.1	<0.0001
Program Results, 2010	23554.5	<0.0001
Plan Results, 2011	61.2	<0.0001
Plan Results, 2010	0.8	0.3711

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?*)

Statistically significant differences between measured entities were detected at both the program and plan reporting levels, with program-level performance scores ranging by approximately 17 percentage points. At the plan level, statistically significant differences were detected in 2011, but not in 2010. This is consistent with a greater difference in performance between the two plans in 2011 (25.50% and 28.69%) than in 2010 when the rates were almost equal (23.24% and 23.76%). This is precisely the purpose of performance measurement - to detect when there are differences in performance. In 2010, there was no appreciable difference in performance between the two plans. Collectively, however, it is clear that this measure detects differences in performance on the measure scores when they do exist. Our findings are consistent with evidence reported elsewhere in this application documenting a performance gap and disparities in performance. Thus, Topical Fluoride informs performance improvement efforts by allowing plans and programs to identify and monitor performance gaps and disparities both at any given point in time and over time.

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

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

Note: *This criterion is directed 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). ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different datasources/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 demonstrating comparability of 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.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic claims

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

This measure is specified for reporting at the program and plan level and there are currently no efforts to develop an eMeasure (eCQM) of this measure at these levels.

Our understanding is that the Feasibility Score Card is only for eMeasures; consequently, we have not submitted this. Feasibility criteria were met during the initial endorsement review.

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. <u>Required for maintenance of endorsement.</u> Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

<u>IF instrument-based,</u> consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

This measure relies on standard data elements in administrative claims data (e.g., patient ID, patient birthdate, enrollment information, CDT codes, date of service, and provider taxonomy). These data are readily available and can be easily retrieved because they are routinely used for billing and reporting purposes. A key advantage of using administrative claims data is that the time and cost of data collection for performance measurement purposes are relatively low because these data are already collected for other purposes.

Initial feasibility assessments were conducted using the RAND-UCLA modified Delphi process to rate the measure concepts with feasibility as one component of the assessment. On a 1-9 point scale, the measure concept of preventive dental services identified using CDT codes (within which topical fluoride falls) was rated as an 8 or "definitely feasible" by the expert panel. During the empirical testing phase, our testing found that the critical data elements had missing/invalid data of <1% (Data 3c.1.), meeting or exceeding the guidance from the Centers for Medicare and Medicaid Services regarding acceptable error rates. During measure development and testing, the measure specifications were made available through a publicly accessible website for public comment with additional broad email dissemination to a wide range of stakeholders. No concerns regarding feasibility were raised during this process.

Citation: 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.

Data 3c.1 Percentage of Missing and Invalid Values for Critical Data Elements

PROGRAM 1

Member ID: 0.00%

Date of Birth: 0.00%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.00%

Date of Service: 0.01%

Rendering Provider ID: 0.28%

PROGRAM 2

Member ID: 0.27%

Date of Birth: 0.00%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.28%

Date of Service: 0.00%

Rendering Provider ID: 0.18%

PROGRAM 3

Member ID: 0.00%

Date of Birth: 0.00%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.01%

Date of Service: 0.00%

Rendering Provider ID: 0.61%

PROGRAM 4

Member ID: 0.43%

Date of Birth: 0.02%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.00%

Date of Service: 0.00%

Rendering Provider ID: 0.67%

Endorsement Maintenance Update: There have been no reports of feasibility issues with implementing this measure. Please see Use and Usability section.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

This measure is intended to be transparent and available for widespread adoption. As such, it was purposefully designed to avoid using software or other proprietary materials that would require licensing fees. The measure specifications, including a companion User Guide, is accessible through a website and available free of charge for non-commercial purposes. The main requirements of users is to ensure the quality of their source data and expertise to program the measures within their information systems, following the clear and detailed specifications. Technical assistance is available to users.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting Texas Health and Human Services Commission: Texas Medicaid and CHIP Texas Health and Human Services Commission: Texas Medicaid and CHIP Payment Program https://hhs.texas.gov/sites/default/files//documents/laws-regulations/handbooks/umcm/6-2-15.pdf Texas Health and Human Services Commission: Texas Medicaid and CHIP Quality Improvement (external benchmarking to organizations) Covered California http://hbex.coveredca.com/insurance-companies/PDFs/2017-2019-Individual-Model-Contract.pdf Quality Improvement (Internal to the specific organization) State Medicaid Agencies http://www.msdanationalprofile.com/2015-profile/management-reporting-and-quality-measurement/quality-measurement/? Michigan Healthy Kids Dental RFP https://www.buy4michigan.com/bso/external/bidDetail.sdo?bidId=007117B0011386&parentUrl=activeBids

4a1.1 For each CURRENT use, checked above (update for <u>maintenance of endorsement</u>), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

1. Program and Sponsor: Texas Health and Human Services Commission - Texas Medicaid and CHIP
https://hhs.texas.gov/sites/default/files//documents/laws-regulations/handbooks/umcm/6-2-15.pdf
Purpose: Payment Program and Public Reporting
This measure has been adopted by the Texas Health and Human Services Commission as part of the Texas CHIP and Medicaid Dental Services Pay-for-Quality (P4Q) program. [Texas HHSC Uniform Managed Care Manual, Chapters 6.2.15. Effective Date 09/01/2017, Version 2.0].
This measure was also present in earlier iterations of the Texas Medicaid and CHIP quality programs since initial endorsement. We are referencing current use for this update.
Geographic Area and Number/Percentage of Accountable Entities and Patients:
This applies to the state of Texas CHIP and Medicaid programs (statewide application). There are two dental plans (i.e., the accountable entities) that serve Texas CHIP and Medicaid. In June 2017, there were 3,359,770 children enrolled in Texas Medicaid and CHIP (https://hhs.texas.gov/about-hhs/records-statistics/data-statistics/healthcare-statistics).
Level of Measurement and Setting: The measure is implemented at the plan and program level within the Texas Medicaid and CHIP programs.
2. Covered California, the California Health Benefit Exchange
http://hbex.coveredca.com/insurance-companies/PDFs/2017-2019-Individual-Model-Contract.pdf
http://hbex.coveredca.com/insurance-companies/PDFs/2017-2019-QDP-Issuer-Contract-and-Attachments.pdf
Purpose: Quality Improvement
This measure is included in the Covered California Qualified Health Plan Issuer Contract for 2017-019 For the Individual Market and the Covered California Qualified Dental Plan Issuer Contract for 2017-2019. The measure is to be reported annually.
Geographic Area and Number/Percentage of Accountable Entities and Patients:

This applies statewide. In March 2017 there were 85,000 enrollees 0-18 years old in CC health plans (which may offer dental benefits and would therefore report on the dental quality measures). There were 5,100 children enrolled specifically in Qualified Dental Plans. (http://hbex.coveredca.com/data-research/)
Level of Measurement and Setting. The measure is implemented at the plan level with the Covered California program.
3. State Medicaid Agencies
http://www.msdanationalprofile.com/2015-profile/management-reporting-and-quality-measurement/quality-measurement/?
(Note: To access the data, a public user account must be created. We can help facilitate access to the data if needed.)
Purpose: Quality Improvement
The Medicaid|Medicare|CHIP Services Dental Association conducts an annual survey of state Medicaid programs and collects data specifically on which programs report Dental Quality Alliance measures.
In its 2015 profile (the most recent available), 9 states reported that they currently use this measure in the Medicaid and/or CHIP programs.
Geographic Area and Number/Percentage of Accountable Entities and Patients:
The 9 states are: Alabama, Connecticut, Florida, Idaho, Illinois, Nevada, Oklahoma, Rhode Island, and West Virginia. Data are not provided on the number of accountable entities included.
4. Michigan Healthy Kids Dental Program
https://www.buy4michigan.com/bso/external/bidDetail.sdo?bidId=007117B0011386&parentUrl=activeBids
Note: Select Schedule A Work Statement link under File Attachments
Purpose: Quality Improvement
The Michigan Healthy Kids Dental Program has included this measure in the set of measures included in its Performance Monitoring Standards, which is currently included in the Request for Proposals and will be included in the contracts between the contracted dental plans and the State of Michigan.
Geographic Area and Number/Percentage of Accountable Entities and Patients:
The Healthy Kids dental program covers children enrolled in Michigan's Medicaid program statewide. The state intends to award two contracts. There are approximately 955,000 enrollees served by the Healthy Kids Dental Program.
Additional Information:
This measure was one of ten performance measures that focused on Dental Caries Prevention and Disease Management among children and that was approved by the DQA. The Dental Quality Alliance (DQA) was formed at the request of the Centers of Medicare and Medicaid Services (CMS) specifically for the purpose of bringing together recognized expertise in oral health to develop quality measures through consensus processes. As noted in the letter from Cindy Mann, JD, Director of the Center for Medicaid & CHIP Services within CMS: "The dearth of tested quality measures in oral health has been a concern to CMS and other payers of oral health services for quite some time." (See Appendix)

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Not applicable.

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Not applicable.

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

Per the annual survey conducted by the Medicaid|Medicare|CHIP Services Dental Association (MSDA), 9 Medicaid/CHIP agencies are implementing this measure. The measure is part of measure set included in the Request for Proposal (RFP)

released by the Michigan Healthy Kids Dental Program. This measure is included in the Pay-For-Quality program and publicly reported in the Texas Medicaid and CHIP programs. Additionally, this measure is a requirement for the Qualified Dental Plans to report to the Covered California, the state-based marketplace in California.

The DQA provides technical assistance to these and other users of DQA measures through webinars, resource document development, and one-on-one staff support. The DQA has an Implementation Committee dedicated to developing implementation and improvement resources.

In order to ensure transparency, incorporate learnings from implementation, establish proper protocols for timely assessment of the evidence and measure properties, and to comply with the NQF's endorsement agreement, the DQA has established an annual measure review and maintenance process. This measure review process is overseen by the DQA's Measures Development and Maintenance Committee (MDMC) which is comprised of subject matter experts. This annual review process includes: (1) call for public comments, (2) evaluation of the comments, (3) user group feedback, and (4) code set reviews.

In 2016, the DQA expanded its scope of review of its measures by convening conference calls for two user groups – one comprised of representatives from 6 state Medicaid programs (Alabama, Florida, Kentucky, Oregon, Nevada, and Pennsylvania) and the other comprised of representatives from 8 dental plans. Participants shared their experiences implementing DQA measures in their respective programs, including any challenges related to the DQA measures specifications and use of these measures in their quality improvement programs. Participants did not have any significant issues related to the clarity or feasibility of implementing the measure specifications.

This is the first 3-year maintenance endorsement review for this measure. As indicated above, the measure is being implemented in multiple programs. Because measure implementation requires a start-up phase for integration of the measures into contracts and for programs and plans to prepare for reporting, in combination with a lag period for reporting measures calculated using administrative claims data, most of the entities that have adopted the measures are just getting underway and there is limited data reporting. Implementation has mostly focused on addressing questions related to how to use the measures in the context of broader quality improvement and clarifying questions related to the specifications.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

In an effort to facilitate implementation of the DQA measures, the DQA provides technical assistance on an ongoing basis to users of DQA measures through webinars, resource document development and one-on-one staff support.

In 2016, the DQA expanded its scope of review of its measures by convening conference calls for two user groups – one comprised of representatives from 6 state Medicaid programs (Alabama, Florida, Kentucky, Oregon, Nevada, and Pennsylvania) and the other comprised of representatives from 8 dental plans. Participants shared their experiences implementing DQA measures in their respective programs, including any challenges related to the DQA measures specifications and use of these measures in their quality improvement programs. Participants did not have any significant issues related to the clarity or feasibility of implementing the measure specifications.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

In order to ensure transparency, establish proper protocols for timely assessment of the evidence and measure properties, and to comply with the NQF's endorsement agreement, the DQA has established an annual measure review and maintenance process. This measure review process is overseen by the DQA's Measures Development and Maintenance Committee (MDMC) which is comprised of subject matter experts. This annual review process includes: (1) call for public comments, (2) evaluation of the comments, (3) user group feedback, and (4) code set reviews.

DQA provides technical assistance on an ongoing basis to users of DQA measures through webinars,

resource document development and one-on-one staff support.

In 2016, the DQA expanded its scope of review of its measures by convening conference calls for two user groups – one comprised of representatives from 6 state Medicaid programs (Alabama, Florida, Kentucky, Oregon, Nevada, and Pennsylvania) and the other comprised of representatives from 8 dental plans. Participants shared their experiences implementing DQA measures in their respective programs, including any challenges related to the DQA measures specifications and use of these measures in their quality improvement programs. Participants did not have any significant issues related to the clarity or feasibility of implementing the measure specifications.

4a2.2.2. Summarize the feedback obtained from those being measured.

There has been no feedback indicating any significant issues related to the clarity or feasibility of implementing the measure specifications.

4a2.2.3. Summarize the feedback obtained from other users

There have been no significant issues related to the clarity or feasibility of implementing the measure specifications.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

There have been no significant issues related to the clarity or feasibility of implementing the measure specifications.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

This is the first 3-year maintenance endorsement review for this measure. As indicated above, the measure is being implemented in multiple programs. Because measure implementation requires a start-up phase for integration of the measures into contracts and for programs and plans to prepare for reporting, in combination with a lag period for reporting measures calculated using administrative claims data, most of the entities that have adopted the measures either have only limited baseline scores or will start reporting measures within the next year.

We are only aware of repeat measurements within the Texas Medicaid/CHIP programs (https://thlcportal.com/qoc/dental), which started implementing this measure after it was approved by the Dental Quality Alliance and before NQF endorsement, as follows:

Texas Medicaid

Year, Program Denominator, Program Overall Score, DentaQuest(Plan) Score, MCNA(Plan) Score

2014, 1090952, 39.97, 41.57, 37.62

2015, 1334887, 41.75, 44.70, 38.15

Texas CHIP

Year, Program Overall, DentaQuest(Plan), MCNA(Plan)

2014, 108704, 33.01, 35.45, 32.99

2015, 79693, 37.50, 41.44, 37.71

These data suggest a trend in improvement over time. However, as noted above, these are initial performance data for one program. Most measure users are just now getting their quality measurement programs underway.

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

No unintended or negative consequences have been identified.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

Not applicable.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Not applicable.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment Attachment: Appendix_Fluoride.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American Dental Association on behalf of the Dental Quality Alliance

Co.2 Point of Contact: Krishna, Aravamudhan, aravamudhank@ada.org, 312-440-2772-

Co.3 Measure Developer if different from Measure Steward: American Dental Association on behalf of the Dental Quality Alliance

Co.4 Point of Contact: Krishna, Aravamudhan, aravamudhank@ada.org, 312-440-2772-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

This project is headed by the DQA through its Measure Development and Maintenance Committee (formerly Research and Development Committee). The following individuals were responsible for executing and overseeing all scientific aspects of this project.

- Craig W. Amundson, DDS, General Dentist, HealthPartners, National Association of Dental Plans. Dr. Amundson serves as chair for the Committee.
- Mark Casey, DDS, MPH, Dental Director, North Carolina Department of Health and Human Services Division of Medical Assistance
- Natalia Chalmers, DDS, PhD, Diplomate, American Board of Pediatric Dentistry, Director, Analytics and Publication, DentaQuest Institute
- Frederick Eichmiller, DDS, Vice President & Science Officer, Delta Dental of Wisconsin
- Chris Farrell, RDH, BSDH, MPA, Oral Health Program Director, Michigan Department of Health and Human Services

This group oversees the maintenance. All work of this Committee was distributed for review and formal vote and approval by the entire Dental Quality Alliance. (http://ada.org/dqa) The DQA is made up of representatives from 38 stakeholder organizations.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2013

Ad.3 Month and Year of most recent revision: 01, 2017

Ad.4 What is your frequency for review/update of this measure? Annual

Ad.5 When is the next scheduled review/update for this measure? 01, 2018

Ad.7 Disclaimers: Dental Quality Alliance measures and related data specifications, developed by the Dental Quality Alliance (DQA), are intended to facilitate quality improvement activities. These Measures are intended to assist stakeholders in enhancing quality of care. These performance Measures are not clinical guidelines and do not establish a standard of care. The DQA has not tested its Measures for all potential applications.

Measures are subject to review and may be revised or rescinded at any time by the DQA. The Measures may not be altered without the prior written approval of the DQA. The DQA shall be acknowledged as the measure steward in any and all references to the measure.

Measures developed by the DQA, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Measures require a license agreement between the user and DQA. Neither the DQA nor its members shall be responsible for any use of these Measures.

THE MEASURES ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND

Limited proprietary coding is contained in the Measure specifications for convenience.

For Proprietary Codes:

The code on Dental Procedures and Nomenclature is published in Current Dental Terminology (CDT), Copyright © 2017 American Dental

Association (ADA). All rights reserved.

This material contains National Uniform Claim Committee (NUCC) Health Care Provider Taxonomy codes

(http://www.nucc.org/index.php?option=com_content&view=article&id=14&Itemid=125). Copyright © 2017 American Medical

Association. All rights reserved.

Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. The DQA, American Dental Association (ADA), and its members disclaim all liability for use or accuracy of any terminologies or other coding contained in the specifications.

THE SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

Ad.8 Additional Information/Comments: In 2008, the Centers for Medicare and Medicaid Services (CMS) asked the ADA to lead the development of a broad coalition of organizations that would lead dentistry to improve the oral health of Americans through quality measurement and quality improvement. The ADA subsequently established the DQA. The DQA is a multi-stakeholder alliance comprised of approximately 38 stakeholders (with organizations as members) from across the oral health community, including federal agencies, third-party payers, professional associations, and an individual member from the general public. The DQA’s mission is to advance the field of performance measurement to improve oral health, patient care, and safety through a consensus building process.