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 #: 2508

Measure Title: Prevention: Dental Sealants for 6-9 Year-Old 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 in the age category of 6-9 years at "elevated" risk (i.e., "moderate" or "high") who received a sealant on a permanent first molar tooth 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, L i, 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-based clinical recommendations recommend that sealants be placed on pits and fissures of children's primary and permanent teeth when it is determined that the tooth, or the patient, is at risk of experiencing caries (Beauchamp et al. 2008). The evidence for sealant effectiveness in permanent molars is stronger than evidence for primary molars (Beauchamp et al. 2008). Sealants benefit children across a wide age range; however, for greatest effectiveness in caries prevention, it is recommended that sealants be placed on teeth soon after they erupt (US DHHS 2010; CDC 2013).

The proposed measure, Prevention: Sealants for 6-9 Year-Old Children at Elevated Caries Risk, captures whether children at moderate or high caries risk received a sealant on a permanent first molar tooth. Permanent first molars usually erupt between ages 6 and 7 years. Thus, this measure addresses both the tooth type on which sealants are placed and the timeliness of care provision. The measure Sealants for 6-9 Year-Old Children allows plans and programs to assess whether children at risk for caries are receiving evidence-based prevention and target performance improvement initiatives accordingly.

This measure is a program/plan specific measure that contributes to the Healthy People 2020 Objective OH 12.2 that calls for increasing the percent children aged 6 to 9 years who received dental sealants on one or more of their first permanent molars.

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 age 6-9 years at "elevated" risk (i.e., "moderate" or "high") who received a sealant on a permanent first molar tooth as a dental service.

Denominator Statement: Unduplicated number of enrolled children age 6-9 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.

There are no other exclusions.

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

Maintenance of Endorsement – Preliminary Analysis

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

- Sealants for 6-9 Year-Old Children at Elevated Caries Risk indicates the percentage of children at moderate to high risk for caries who received a sealant on a first permanent molar. Evidence-based clinical

recommendations recommend that sealants be placed on pits and fissures of children's primary and permanent teeth when it is determined that the tooth, or the patient, is at risk of experiencing caries, with greater evidence of effectiveness in permanent molars compared to primary molars (Beauchamp et al. 2008).

- Grade/Strength of Recommendation: B which is defined as: "Directly based on category II evidence or extrapolated recommendation for category I evidence."

Citation:
Beauchamp J, Caufield PW, Crall JJ, Donly K, Feigal R, Gooch B, et al. Evidence-based clinical recommendations for the use of pit-and-fissure sealants: a report of the American Dental Association Council on Scientific Affairs. J Am Dent Assoc 2008;139(3):257-268. Available at: http://jada.ada.org/content/139/3/257.full.

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:**

Updates:

- A recent Cochrane Review on the effectiveness of sealants brings together all the evidence on this topic. The conclusions of this new review continue to support the recommendations of the ADA Sealant Guideline (Note: the ADA is currently updating this guideline).
    - Ahovuo-Saloranta A, Forss H, Walsh T, Hiiri A, Nordblad A, Mäkelä M, Worthington HV. Sealants for preventing dental decay in the permanent teeth. Cochrane Database Syst Rev. 2013 Mar 28;3:CD001830. doi: 10.1002/14651858.CD001830.pub4.

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review, but does note a recent Cochrane review collated all evidence and reached the same conclusions that supported the original guideline. 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 (7 systematic reviews and 14 individual clinical studies) Quality: High, Consistency: Moderate → Rate as High

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 five sources and refer to "program" level information and "plan" level information (Texas Medicaid, Texas CHIP, 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) Texas CHIP, (3) Florida CHIP, (4) Commercial Data, and (5) Florida Medicaid. "Plans" refer to data from the two dental plans that served Florida CHIP members in both 2010 and 2011.
- The data source and sample size are sufficient to assess gaps in performance. The performance range of **20% to 30% in CY 2010** (year in which data were available for all five programs) indicate variation in sealant replacement across programs. Data from the Centers for Medicare and Medicaid Services (CMS) indicate significant variation among state Medicaid programs, ranging from 6% to 31% of children 6-9 years old, who received a sealant on a permanent molar tooth (CMS-416 data, FY 2011).

- 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 by geographic location were detected for two programs. Statistically significant difference in performance by race and ethnicity also were detected in the two programs for which there were race/ethnicity data. In addition, the developers 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 developers detected disparities based on each of these various factors, but data on all of these characteristics were not consistently available for all programs so they presented disparities data on those characteristics that were most consistently available and had the greatest standardization (i.e. race/ethnicity and geographic location).

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.

Staff Scientific Acceptability Rating Logic

*The original testing was submit as permitted by NQF.

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

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

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

Staff Scientific Acceptability

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 empirical 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 face validity 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/CHIP https://hhs.texas.gov/sites/default/files//documents/lawsregulations/ 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

- This measure is included in the CHIPRA Core Measures Program. Some Medicaid programs noted that they do not receive complete data on tooth number from their contracted plans, which is a required data element for this measure. As a result, the affected programs must manually get these data from their contracted plans. Because tooth number is required for reimbursement, these data are readily accessible for plan level reporting. Despite initial concerns about this data element, 25 states reported this measure in FY 2015, and 34 reported in FY 2016.
- A dental benefits administrator (DBA) suggested that the DQA consider adding patient exclusions to the measure. The DQA considered exclusions previously during initial measure development and during annual reviews. Exclusions were not incorporated due to concerns about the introduction of biased measurement, increasing measurement complexity, and adversely affecting implementation feasibility. However, the DQA continues to monitor this issue and will revisit it during the 2018 annual review. The DQA has invited the DBA to

present its suggestion with supporting data to the DQA. The DQA also has invited other DBAs and Medicaid program administrators to provide input. All of this stakeholder feedback will be incorporated into the next annual review.

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 Dental Quality Alliance (DQA) was formed at the request of 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 and before NQF endorsement, as follows:

Texas Medicaid

Year, Program Denominator, Program Overall Score, DentaQuest(Plan) Score, MCNA(Plan) Score
2014, 461207, 25.41, 25.59, 25.53
2015, 503515, 24.99, 25.18, 24.91

Texas CHIP

Year, Program Overall, DentaQuest(Plan), MCNA(Plan)
2014, 76415, 20.17, 22.30, 21.69
2015, 58833, 20.20, 23.14, 22.43

The developer notes that these data also suggest fairly stable rates over the two-year period (i.e. improvement is not noted). However, as noted above, these are initial performance data; additional time may be needed to see improvement within this program because 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 [unexpected findings]

No unintended or negative consequences were identified by the developer.

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_NQF_Evidence_6-9.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: Dental Sealants for 6-9 Year-Old Children at Elevated Caries Risk

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

Date of Submission: 2/10/2014

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 dental service – sealants on permanent molars - 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.

Sealants for 6-9 Year-Old Children at Elevated Caries Risk indicates the percentage of children at moderate to high risk for caries who received a sealant on a first permanent molar. Evidence-based clinical recommendations recommend that sealants be placed on pits and fissures of children's primary and permanent teeth when it is determined that the tooth, or the patient, is at risk of experiencing caries, with greater evidence of effectiveness in permanent molars compared to primary molars (Beauchamp et al. 2008). Sealants benefit children across a wide age range; however, for

greatest effectiveness in caries prevention, it is recommended that sealants be placed on teeth soon after they erupt (US DHHS 2010; CDC 2013). This measure directly reflects evidence-based guidelines regarding an effective caries prevention measure (sealants) as well as the specific tooth type for which the evidence is the strongest (permanent molar) and the timing of sealant placement to maximize effectiveness (shortly after eruption – 6-9 years of age for permanent first molars). 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, timely placement of sealants on permanent first molars have demonstrated effectiveness in reducing caries among children, 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*):

Beauchamp J, Caufield PW, Crall JJ, Donly K, Feigal R, Gooch B, et al. Evidence-based clinical recommendations for the use of pit-and-fissure sealants: a report of the American Dental Association Council on Scientific Affairs. J Am Dent Assoc 2008;139(3):257-268. Available at: http://jada.ada.org/content/139/3/257.full.

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

"Caries Prevention: Sealants should be placed on pits and fissures of **children's** and **adolescents'** permanent teeth when it is determined that the tooth, or the patient, is at risk of developing caries." (Beauchamp et al. 2008, p. 263, Table 3)

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

Grade/Strength of Recommendation: B which is defined as: "Directly based on category II evidence or extrapolated recommendation for category I evidence." (Beauchamp 2008, pp. 261, 263, Tables 1, 2, 3)

[See grades for strength of evidence in section 1a7.]

Grading system adapted from: Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines: developing guidelines. BMJ 1999;318(7183):593-596.

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.*)

A: Directly based on category I evidence

B: Directly based on category II evidence or extrapolated recommendation from category I evidence

C: Directly based on category III evidence or extrapolated recommendation from category I or II evidence

D: Directly based on category IV evidence or extrapolated recommendation from category I, II or III evidence

Grading system adapted from: Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines: developing guidelines. BMJ 1999;318(7183):593-596.

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*):

Beauchamp J, Caufield PW, Crall JJ, Donly K, Feigal R, Gooch B, et al. Evidence-based clinical recommendations for the use of pit-and-fissure sealants: a report of the American Dental Association Council on Scientific Affairs. J Am Dent Assoc 2008;139(3):257-268. Available at http://jada.ada.org/content/139/3/257.full.

Ahovuo-Saloranta A, Forss H, Walsh T, Hiiri A, Nordblad A, Mäkelä M, Worthington HV. Sealants for preventing dental decay in the permanent teeth. Cochrane Database Syst Rev. 2013 Mar 28;3:CD001830. doi: 10.1002/14651858.CD001830.pub4.

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 four clinical questions were addressed:

- "Under what circumstances should sealants be placed to prevent caries?"
- "Does placing sealants over early (noncavitated) lesions prevent progression of the lesions?"
- "Are there conditions that favor the placement of resin-based versus glass ionomer cement sealants in terms of retention or caries prevention?"
- "Are there any techniques that could improve sealants' retention and effectiveness in caries prevention?"

(Beauchamp et al. 2008, pp. 259-260)

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

"Caries Prevention: Sealants should be placed on pits and fissures of **children's** and **adolescents'** permanent teeth when it is determined that the tooth, or the patient, is at risk of developing caries." (Beauchamp et al. 2008, p. 263, Table 3)

Grade: The evidence grade is **IA** which is defined as: "Evidence from systematic reviews of randomized controlled trials" (Beauchamp 2008, pp. 261, 263, Tables 1, 3). Grading system adapted from: Shekelle et al. (1999) cited in 1a.4.

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

Ia: Evidence from systematic reviews of randomized controlled trials

Ib: Evidence from at least one randomized controlled trial

IIa: Evidence from at least one controlled study without randomization

IIb: Evidence from at least one other type of quasiexperimental study, such as time series analysis or studies in which the unit of analysis is not the individual

III: Evidence from nonexperimental descriptive studies, such as comparative studies, correlation studies, cohort studies and case-control studies

IV: Evidence from expert committee reports or opinions or clinical experience of respected authorities

(Beauchamp et al. 2008, p. 261) Grading system adapted from: Shekelle et al. (1999).

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Literature studies for sealants were conducted to identify all systematic reviews through Oct. 4, 2006. To ensure new clinical studies published since the search within each review were included within the guideline development effort, additional searches were conducted for clinical trials until September 2006.

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*)

7 systematic reviews and 14 individual clinical studies were reviewed with respect to the clinical questions identified. The evidence guidelines do not provide summary data regarding the number of studies by type of study. (Beauchamp 2008, p. 260)

However, the guidelines provide the following details regarding the strength and quality of the evidence related to sealants for caries prevention:

Evidence Grade Ia (systematic reviews of randomized controlled trials)

Supports the following evidence statements based on the evidence review by the expert panel:

- "Placement of resin-based sealants on the permanent molars of children and adolescents is effective for caries reduction." (Beauchamp 2008, p. 260)
- "Reduction of caries incidence in children and adolescents after placement of resin-based sealants ranges from 86 percent at one year to 78.6 percent at two years and 58.6 percent at four years." (Beauchamp 2008, p. 260)

Studies with evidence grade of Ia cited:

Ahovuo-Saloranta A, Hiiri A, Nordblad A, Worthington H, Mäkelä M. Pit and fissure sealants for preventing dental decay in the permanent teeth of children and adolescents. Cochrane Database Syst Rev 2004(3):CD001830.

Llodra JC, Bravo M, Delgado-Rodriguez M, Baca P, Galvez R. Factors influencing the effectiveness of sealants: a meta-analysis. Community Dent Oral Epidemiol 1993;21(5):261-268.

Evidence Grade Ib (evidence from at least one randomized controlled trial)

Supports the following evidence statements based on the evidence review by the expert panel:

- "Sealants are effective in reducing occlusal caries incidence in permanent first molars of children, with caries reductions of 76.3 percent at four years, when sealants were reapplied as needed. Caries reduction was 65 percent at nine years from initial treatment, with no reapplication during the last five years." (Beauchamp 2008, p. 261)

Studies with evidence grade of Ib cited:

Bravo M, Montero J, Bravo JJ, Baca P, Llodra JC. Sealant and fluoride varnish in caries: a randomized trial. J Dent Res 2005;84(12):1138-1143.

Evidence Grade III (evidence from nonexperimental descriptive studies, such as comparative studies, correlation studies, cohort studies and case control studies)

Supports the following evidence statements based on the evidence review by the expert panel:

- "There is consistent evidence from private dental insurance and Medicaid databases that placement of sealants on first and second permanent molars in children and adolescents is associated with reductions in the subsequent provision of restorative service." (Beauchamp 2008, p. 261)
- "Evidence from Medicaid claims data for children who were continuously enrolled for four years indicates that sealed permanent molars are less likely to receive restorative treatment, that the time between receiving sealants and receiving restorative treatment is greater, and that the restorations were less extensive than those in permanent molars that were unsealed." (Beauchamp 2008, p. 261)

Studies with evidence grade of III cited:

Bhuridej P, Damiano PC, Kuthy RA, et al. Natural history of treatment outcomes of permanent first molars: a study of sealant effectiveness. JADA 2005;136(9):1265-1272.

Dennison JB, Straffon LH, Smith RC. Effectiveness of sealant treatment over five years in an insured population. JADA 2000;131(5):597-605.

Hotuman E, Rølling I, Poulsen S. Fissure sealants in a group of 3-4-year-old children. Int J Paediatr Dent 1998;8(2):159-160.

Weintraub JA, Stearns SC, Rozier RG, Huang CC. Treatment outcomes and costs of dental sealants among children enrolled in Medicaid. Am J Public Health 2001;91(11):1877-1881.

1a.7.6. What is the overall quality of evidence across studies 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 is high, grades of Ia (systematic reviews of randomized controlled trials), for sealants placed on permanent molars of children and adolescents.

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) across studies 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*)

Meta-analyses were not conducted as part of the evidence review. Please see the response in 1a.7.5. regarding the identified benefits and associated strength of evidence. However, a more recent Cochrane Review published in 2013 by Ahovuo-Saloranta et al. brings together all the evidence in a quantitative manner. More information from this review is provided below in Section 1.a.7.9

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

Harms were not evaluated as part of this systematic review. However this question was addressed in a recent Cochrane Review on the effectiveness of sealants (Ahovuo-Saloranta et al. 2013), and it was noted: "Only two studies (Bravo 2005; Liu 2012) assessed side effects of the sealants. No adverse effects were detected or reported by patients included in the studies."

Citations:

Ahovuo-Saloranta A, Forss H, Walsh T, Hiiri A, Nordblad A, Mäkelä M, Worthington HV. Sealants for preventing dental decay in the permanent teeth. Cochrane Database Syst Rev. 2013 Mar 28;3:CD001830. doi: 10.1002/14651858.CD001830.pub4.

Bravo M, Montero J, Bravo JJ, Baca P, Llodra JC. Sealant and fluoride varnish in caries: a randomized trial. Journal of Dental Research 2005;84(12):1138-43.

Liu BY, Lo ECM, Chu CH, Lin HC. Randomized trial on fluorides and sealants for fissure caries prevention. Journal of Dental Research 2012;91(8):753-8.

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.

A recent Cochrane Review on the effectiveness of sealants brings together all the evidence on this topic. The conclusions of this new review continue to support the recommendations of the ADA Sealant Guideline (Note: the ADA is currently updating this guideline). The summary of findings from the Cochrane review appears below:

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON *[Explanation]*

Resin-based sealant compared to control without sealant for preventing dental caries

Patient or population: Children and adolescents
Settings: Sealant applications for school children in USA, Canada, China & Colombia
Intervention: Resin-based sealant applications on occlusal tooth surfaces of permanent molars
Comparison: No sealant application

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control teeth	Sealed teeth				
Dentine caries in permanent molars Follow-up: 2 years	Incidence of carious first molars (40%) 400 per 1000[1]	Incidence of carious first molars (6.3%) 63 per 1000 (38 to 96)	OR 0.12 (0.07 to 0.19) [2]	1259 children randomised & 1066 evaluated after 2 years (6 studies [3,4,5])	⊕⊕⊕◯ **moderate**	Benefits of resin-sealant maintained up to at least 48 months of follow-up[6]
	Incidence of carious first molars (70%) 700 per 1000[1]	Incidence of carious first molars (19%) 190 per 1000 (122 to 272)	OR 0.12 (0.07 to 0.19) [2]	1259 children randomised & 1066 evaluated after 2 years (6 studies[3,4,5])	⊕⊕⊕◯ **moderate**	Benefits of resin-based sealant maintained up to at least 48 months of follow-up[6]

CI: confidence interval; **OR:** odds ratio

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] The incidence of carious control teeth in the five split-mouth trials included in this comparison ranged from 37% to 69% (studies published between 1976 and 1979). We have shown the effect of sealants at each end of this range. These studies did not give information on the baseline caries prevalence of the children.
The sixth study included in this meta-analysis (parallel group study published in 2012) reported clearly lower incidence of carious first molars than the five split-mouth studies. In sealant group, carious first molars were detected in 9 out of 121 children (7.4%) (11 carious teeth out of 367 sealed teeth) and in placebo group in 21 out of 124 children (17%) (28 carious teeth out of 379 placebo teeth). Caries prevalence: mean baseline dmft level of 3.4.
[2] There was considerable heterogeneity in this estimate (I^2 = 77% P = 0.0007) but all of the trials showed a statistically significant effect favouring sealants.
[3] Six studies at low risk of bias for the four key domains of allocation concealment, incomplete outcome data, selective reporting and baseline comparability of the groups.
[4]All studies recruited children aged 5-10 years. Three studies conducted in areas with fluoridated water, two studies stated water was not fluoridated and the remaining one study did not report whether water supplies were fluoridated.
[5] Five trials were published between 1976 and 1979 and one in 2012. One further parallel group trial from Thailand at unclear risk of bias reporting DFS increment published in 1995 also found a benefit in favour of resin-based sealant (mean difference in DFS increment -0.65, 95% CI -0.83 to -0.47, 276 children evaluated).
[6] The benefit associated with sealant use is maintained at all of the follow-up estimates (up to 9 years) though the number of studies and the number of children available for evaluation reduced markedly over this period (e.g. at 48 to 54 months of follow-up odds ratio 0.21, 95% CI 0.16 to 0.28, two studies at low risk of bias and two studies at high risk of bias, 482 children evaluated; risk ratio 0.24, 95% CI 0.12 to 0.45, one study at unclear risk of bias, 203 children evaluated).

Citations

Ahovuo-Saloranta A, Forss H, Walsh T, Hiiri A, Nordblad A, Mäkelä M, Worthington HV. Sealants for preventing dental decay in the permanent teeth. Cochrane Database Syst Rev. 2013 Mar 28;3:CD001830. doi: 10.1002/14651858.CD001830.pub4.

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)*

If a COMPOSITE (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, L i, 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-based clinical recommendations recommend that sealants be placed on pits and fissures of children's primary and permanent teeth when it is determined that the tooth, or the patient, is at risk of experiencing caries (Beauchamp et al. 2008). The evidence for sealant effectiveness in permanent molars is stronger than evidence for primary molars (Beauchamp et al. 2008). Sealants benefit children across a wide age range; however, for greatest effectiveness in caries prevention, it is recommended that sealants be placed on teeth soon after they erupt (US DHHS 2010; CDC 2013).

The proposed measure, Prevention: Sealants for 6-9 Year-Old Children at Elevated Caries Risk, captures whether children at moderate or high caries risk received a sealant on a permanent first molar tooth. Permanent first molars usually erupt between ages 6 and 7 years. Thus, this measure addresses both the tooth type on which sealants are placed and the timeliness of care provision. The measure Sealants for 6-9 Year-Old Children allows plans and programs to assess whether children at risk for caries are receiving evidence-based prevention and target performance improvement initiatives accordingly.

This measure is a program/plan specific measure that contributes to the Healthy People 2020 Objective OH 12.2 that calls for increasing the percent children aged 6 to 9 years who received dental sealants on one or more of their first permanent molars.

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 five sources and refer to "program" level information and "plan" level information. We included data for publicly insured children in the Texas Medicaid, Texas CHIP, 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 five programs collectively represent different delivery system models. The Texas Medicaid data represented dental fee-for-service, and Texas CHIP data reflected a single dental managed care organization (MCO). 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. 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) Texas CHIP, (3) Florida CHIP, (4) Commercial Data, and (5) Florida Medicaid. "Plans" refer to data from the two dental plans that served Florida CHIP members in both 2010 and 2011. [Technically, there were three plans represented in the data because Texas CHIP was served by a single dental plan. Since the program=plan in that case, we included it in the "program" level data.]

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

Programs

Our source data for the testing prior to applying the denominator age criteria of 6-9 years old included children 0-20 years in each program. The number 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

Texas CHIP, 2011: 842,454

Texas CHIP, 2010: 786,070

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

The number of children in the age range of 6-9 years specifically were:

Texas Medicaid, 2011: 746,535

Texas Medicaid, 2010: 706,596

Texas CHIP, 2011: 224,908

Texas CHIP, 2010: 210,624

Florida CHIP, 2011: 88,943

Florida CHIP, 2010: 89,897

Commercial, 2011: 36,905

Commercial, 2010: 38,390

Florida Medicaid, 2010: 406,698

Plan 1, 2010: 25,240

Plan 2, 2010: 31,126

Plan 1, 2011: 41,537

Plan 2, 2011: 45,348

Data 1b.2. Performance Scores for Dental Sealants for 6-9 Year-Olds at Elevated Risk

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

Program 1, CY 2011:	23.69%	(	0.2369 ,	0.0006 ,	0.2357 ,	0.2381)
Program 2, CY 2011:	23.01%	(	0.2301 ,	0.0017 ,	0.2267 ,	0.2335)
Program 3, CY 2011:	31.33%	(	0.3133 ,	0.0036 ,	0.3062 ,	0.3204)
Program 4, CY 2011:	22.59%	(	0.2259 ,	0.0042 ,	0.2176 ,	0.2342)
Program 1, CY 2010:	23.38%	(	0.2338 ,	0.0007 ,	0.2325 ,	0.2351)
Program 2, CY 2010:	19.82%	(	0.1982 ,	0.0017 ,	0.1949 ,	0.2015)
Program 3, CY 2010:	30.04%	(	0.3004 ,	0.0036 ,	0.2933 ,	0.3075)
Program 4, CY 2010:	26.68%	(	0.2668 ,	0.0043 ,	0.2583 ,	0.2753)
Program 5, CY 2010:	21.04%	(	0.2104 ,	0.0015 ,	0.2074 ,	0.2134)
Plan 1, CY 2011:	31.43%	(	0.3143 ,	0.0054 ,	0.3037 ,	0.3249)
Plan 2, CY 2011:	30.91%	(	0.3091 ,	0.0050 ,	0.2993 ,	0.3189)
Plan 1, CY 2010:	31.38%	(	0.3138 ,	0.0078 ,	0.2985 ,	0.3291)
Plan 2, CY 2010 :	29.97%	(	0.2997 ,	0.0067 ,	0.2866 ,	0.3128)

The measure rate range of 20% to 30% in CY 2010 (year in which data were available for all five programs) indicates variations in sealant prevalence across 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 there are significant variations in the percentage of children who received sealants. Data from the Centers for Medicare and Medicaid Services indicate significant variation among state Medicaid programs, ranging from 6% to 31% of children 6-9 years old, who received a sealant on a permanent molar tooth (Norris 2013; CMS-416 data, FY 2011).

[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 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. Disparities by geographic location were detected for two programs. Statistically significant difference in performance by race and ethnicity also were detected in the two programs for which there were race/ethnicity data. 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 health status, dental 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 Geographic Location and Race/Ethnicity

PROGRAM 1

Overall performance score: 23.69%

Scores by Geographic Location

Urban: 23.95%

Rural: 21.89%

p-value from Chi-square test: <.0001

Scores by Race

Non-Hispanic White: 22.07%

Non-Hispanic Black: 23.08%

Hispanic: 24.31%

p-value from Chi-square test <.0001

PROGRAM 2

Overall performance score: 23.01%

Scores by Geographic Location

Urban: 23.00%

Rural: 23.23%

p-value from Chi-square test: 0.6649

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: 31.33%

Scores by Geographic Location

Urban: 31.29%

Rural: 31.82%

p-value from Chi-square test: 0.7252

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: 22.59%

Scores by Geographic Location

Urban: 22.70%

Rural: 20.60%

p-value from Chi-square test: 0.3436

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 5

Overall performance score: 21.04%

Scores by Geographic Location

Urban: 21.07%

Rural: 19.33%

p-value from Chi-square test: 0.0087

Scores by Race

Non-Hispanic White: 21.24%

Non-Hispanic Black: 19.63%

Hispanic: 21.87%

p-value from Chi-square test <.0001

Note: 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 in general, 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 ($p<0.0001$).

In addition, there are documented disparities in dental sealant receipt specifically. For example, using data from the National Health and Nutrition Examination Survey, researchers at the National Center for Health Statistics identified variations in dental sealant prevalence among children by age, race, ethnicity, and poverty level (Dye, Li, and Thorton-Evans 2012). Specifically: "Dental sealant prevalence was lower among children [6-9 years] living at or below 100% of the federal poverty level (26%) compared with children living above the poverty level (34%). A similar pattern was found among adolescents aged 13–15, but the difference was not statistically significant. Dental sealant prevalence was significantly lower for non-Hispanic black adolescents (32%) compared with non-Hispanic white adolescents (56%), among those aged 13–15" (Dye, Li, and Thorton-Evans 2012, p. 2).

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.

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):*

Dental

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_Sealants_6-9_Years.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 age 6-9 years at "elevated" risk (i.e., "moderate" or "high") who received a sealant on a permanent first molar tooth 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 age 6-9 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.

There are no other exclusions.

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.)*

There are no stratifications for this measure.

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.)*

Sealants for 6 – 9 year olds - 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 >= 6 and <= 9, then proceed to next step.

b. If age criterion is 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 at least 180 days during the reporting year:

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":

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 a sealant as a dental service:

a. If [CDT CODE] = D1351 and;

b. If [RENDERING PROVIDER TAXONOMY] code = any of the NUCC maintained Provider Taxonomy Codes in Table 2 below, then 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: In this step, 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.

6. Check if sealant was placed on a permanent first molar:

a. If [TOOTH-NUMBER] = 3, 14, 19 or 30 then include in numerator; STOP processing.

b. If not, then service was not provided for the first permanent molar; STOP processing. This enrollee is already included in the denominator but will not be included in the numerator.

YOU NOW HAVE NUMERATOR (NUM) COUNT: Enrollees at "elevated risk" who received sealants on a permanent first molar as a dental service

7. Report

a. Unduplicated number of enrollees in numerator

b. Unduplicated number of enrollees in denominator

c. Measure rate (NUM/DEN)

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_6-9.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: Dental Sealants for 6-9 Year-Old Children at Elevated Caries Risk
Date of Submission: Click here to enter a date2/10/2014

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 five sources and refer to "program" level information and "plan" level information. We included data for publicly insured children in the Texas Medicaid, Texas CHIP, 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 five programs collectively represent different delivery system models. The Texas Medicaid data represented dental fee-for-service, and Texas CHIP data reflected a single dental managed care organization (MCO). 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 <u>all</u> 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 <u>measured entities</u> 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*)

<u>Level of Analysis: Program, 5 Measured Entities</u>

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. Texas CHIP
    A. Size: # Members 0-20 years, CY 2011: 842,454; # Members 0-20 years, CY 2010: 786,070
    B. Location: Texas – Statewide
    C. Delivery Type – Dental MCO (1 plan)
3. 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)

4. 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
5. 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. Technically, we had three plans represented because the Texas CHIP program was served by a single dental plan so the program=plan in that case. For the purposes of testing plan comparisons within a program, we focus on the two plans in FL CHIP.

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 four 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	Program 4	Plan 1	Plan 2
Total Number Patients	3,544,247	842,454	317,146	184,152	140,986	168,191
Age Group Distribution						
Age <1 years	7.05%	0.11%	N/A	1.54%	N/A	N/A
Age 1-2 years	14.32%	5.34%	N/A	5.75%	N/A	N/A
Age 3-5 years	19.46%	11.70%	3.81%	12.68%	4.12%	3.60%
Age 6-7 years	11.21%	12.30%	13.05%	9.57%	13.71%	12.55%
Age 8-9 years	9.85%	14.40%	15.00%	10.18%	15.76%	14.41%
Age 10-11 years	9.03%	14.03%	15.71%	10.55%	16.27%	15.25%
Age 12-14 years	11.63%	19.57%	23.73%	16.09%	23.06%	24.31%
Age 15-18 years	13.19%	22.54%	28.70%	22.13%	27.08%	29.88%
Age 19-20 years	4.27%	N/A	N/A	11.50%	N/A	N/A
Geographic Location						
Urban	83.63%	84.33%	92.94%	95.95%	93.01%	92.91%
Rural	15.15%	14.61%	5.02%	3.86%	4.83%	5.15%
Missing	1.22%	1.06%	2.04%	0.19%	2.16%	1.94%
Race and Ethnicity						
Non-Hispanic White	17.36%	N/A	N/A	N/A	N/A	N/A
Non-Hispanic Black	15.08%	N/A	N/A	N/A	N/A	N/A
Hispanic	58.07%	N/A	N/A	N/A	N/A	N/A
Other & Unknown	9.49%	N/A	N/A	N/A	N/A	N/A

Table 1.6B, Patient Characteristics, 6-9 Years Old (Age Range Targeted by Measure), 2011

	Descriptive Characteristics of Individuals 6-9 Years Enrolled at Least One Month, CY 2011					
	Program 1	Program 2	Program 3	Program 4	Plan 1	Plan 2
Total Number Patients	746,535	224,908	88,943	36,905	41,537	45,348
Age Group Distribution						
Age 6-7 years	53.23%	46.05%	46.53%	48.49%	46.52%	46.55%
Age 8-9 years	46.77%	53.95%	53.47%	51.51%	53.48%	53.45%
Geographic Location						
Urban	84.16%	84.46%	93.32%	96.19%	93.41%	93.29%
Rural	15.00%	14.54%	5.00%	3.58%	4.84%	5.11%
Missing	0.84%	1.00%	1.68%	0.23%	1.75%	1.61%
Race and Ethnicity						
Non-Hispanic White	16.77%	N/A	N/A	N/A	N/A	N/A
Non-Hispanic Black	14.90%	N/A	N/A	N/A	N/A	N/A
Hispanic	62.04%	N/A	N/A	N/A	N/A	N/A
Other & Unknown	6.29%	N/A	N/A	N/A	N/A	N/A

Table 1.6C, 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	Program 5	Plan 1	Plan 2
Total Number Patients	3,393,963	786,070	315,975	189,968	2,068,670	77,255	116,388
Age Group Distribution							
Age <1 years	7.35%	0.15%	N/A	1.45%	6.05%	N/A	N/A
Age 1-2 years	15.16%	5.37%	N/A	5.67%	14.23%	N/A	N/A
Age 3-5 years	19.48%	11.69%	3.64%	12.73%	19.26%	5.72%	4.22%
Age 6-7 years	11.12%	12.19%	13.32%	9.69%	10.47%	15.68%	12.54%
Age 8-9 years	9.70%	14.61%	15.14%	10.24%	9.19%	16.99%	14.21%
Age 10-11 years	8.75%	14.04%	15.84%	10.60%	8.74%	16.41%	15.18%
Age 12-14 years	11.23%	19.49%	23.70%	16.20%	11.87%	21.40%	24.05%
Age 15-18 years	12.99%	22.47%	28.37%	22.12%	14.73%	23.79%	29.81%
Age 19-20 years	4.22%	N/A	N/A	11.31%	5.47%	N/A	N/A
Geographic Location							
Urban	83.20%	84.46%	92.08%	96.70%	91.47%	92.10%	92.11%
Rural	15.56%	14.45%	5.07%	3.17%	7.30%	5.00%	5.19%
Missing	1.24%	1.08%	2.85%	0.13%	1.23%	2.89%	2.70%
Race and Ethnicity							
Non-Hispanic White	18.21%	N/A	N/A	N/A	29.89%	N/A	N/A
Non-Hispanic Black	15.45%	N/A	N/A	N/A	29.39%	N/A	N/A
Hispanic	59.42%	N/A	N/A	N/A	29.65%	N/A	N/A
Other & Unknown	6.92%	N/A	N/A	N/A	11.06%	N/A	N/A

Table 1.6D, Patient Characteristics, 6-9 Years Old (Age Range Targeted by Measure), 2010

	Descriptive Characteristics of Individuals 6-9 Years Enrolled at Least One Month, CY 2010						
	Program 1	Program 2	Program 3	Program 4	Program 5	Plan 1	Plan 2
Total Number Patients	706,596	210,624	89,897	36,905	406,698	25,240	31,126
Age Group Distribution							
Age 6-7 years	53.39%	45.48%	46.80%	48.49%	53.23%	47.98%	46.88%
Age 8-9 years	46.61%	54.52%	53.20%	51.51%	46.77%	52.02%	53.12%
Geographic Location							
Urban	83.80%	84.66%	92.92%	96.19%	91.39%	93.10%	92.97%
Rural	15.35%	14.28%	5.08%	3.58%	7.35%	4.86%	5.12%
Missing	0.85%	1.06%	1.99%	0.23%	1.26%	2.04%	1.91%
Race and Ethnicity							
Non-Hispanic White	17.00%	N/A	N/A	N/A	29.57%	N/A	N/A
Non-Hispanic Black	14.85%	N/A	N/A	N/A	29.19%	N/A	N/A
Hispanic	62.25%	N/A	N/A	N/A	30.95%	N/A	N/A
Other & Unknown	5.89%	N/A	N/A	N/A	10.30%	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 and CHIP programs. Texas has the third largest Medicaid program and second largest CHIP in the U.S., both 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 and CHIP programs 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 sealants 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; therefore, 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

Dental Sealants for 6-9 Year-Old Children at Elevated Caries Risk measures the percentage of children ages 6-9 years at moderate to high risk for dental caries who received a sealant on a permanent first molar tooth 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 and CHIP programs. 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

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 sealant procedure and tooth number 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 - Sealant and Tooth Number Codes

Data Extraction. For the second component of validation, assessing whether the specific preventive service of sealant placement and associated tooth type coding are 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 sealants applied to the teeth during the visit. If there was evidence of sealant placement, the RHITs then recorded whether sealants were applied to the child's permanent first molar, permanent second molar, and/or "other" tooth type. If there was no indication of the tooth to which the sealant was applied, the tooth number field was coded as "indeterminate." The programming team extracted data from the administrative claims data for the same members and dates of service, recording the presence or absence of CDT code D1351 (sealants); and, when D1351 was present, recording the associated tooth number (or noted as missing). Permanent first molars were identified in the claims data as tooth numbers 3, 14, 19, and 30; permanent second molars were identified as tooth numbers 2, 15, 18, and 31. 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 - RELEVANCE OF TOOTH TYPE

Evidence-based recommendations advise that sealants be placed on pits and fissures of children's primary and permanent teeth when the tooth, or patient, is at caries risk, with stronger evidence for effectiveness in permanent molars (Beauchamp et al. 2008). Sealants benefit children across a wide age range; however, for greatest effectiveness in caries prevention, it is recommended that sealants be placed on teeth soon after they erupt (US DHHS 2010; CDC 2013). Thus, we also sought to evaluate how well the specifications addressed both the tooth type on which sealants are placed and the timeliness of care provision. The research team ran frequency distributions of sealant placement by tooth number and age range for three programs. Specifically, the percentage of children with (1) any sealants (regardless of tooth type), (2) sealants on permanent first molars, and (3) sealants on permanent second molars was assessed by age for children enrolled at least one month in the program.

Citations

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

Centers for Disease Control and Prevention. 2013. Dental Sealants. Available at: http://www.cdc.gov/OralHealth/publications/faqs/sealants.htm. Accessed January 20, 2014.

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.

4. ADDITIONAL VALIDITY TESTING - DENOMINATOR ENROLLMENT CRITERIA

To finalize the denominator definition, several different enrollment criteria were tested: (1) enrolled at least one month, (2) enrolled at least three months, (3) enrolled at least 6 months, (4) enrolled the entire year (12 months), allowing a single one-month gap, and (5) average period of enrollment/person-time equivalent (weighting members in denominator by enrollment length). These were evaluated through the face validity consensus processes.

The first definition was ruled out because of concern that one month is an insufficient period of time to expect children to seek, schedule, and obtain a preventive care dental visit. The last definition was ruled out on the basis of usability as it was considered to be less readily interpretable by a wide range of stakeholders. Table 2a2.2-2 summarizes the percentage of members enrolled in the program during the reporting year who were eligible under each of the different enrollment intervals. Based on these data, a consensus was reached to adopt a six-month continuous enrollment requirement to balance sufficient enrollment duration that allows children adequate time to access care (seek, schedule and obtain a preventive care dental visit) with the number of children who drop out of the denominator due to stricter enrollment requirements.

Table 2b2.2-2. Percentage of All Enrolled Members Included in Different Denominator Definition

	Percentage of All Enrolled Members Included in Different Denominator Definitions				
	Program 1	Program 2	Program 3	Program 4	Program 5
At least 1 month	100%	100%	100%	100%	100%
At least 3 months	95%	85%	84%	93%	94%
At least 6 months	83%	63%	65%	81%	81%
11-12 months	64%	33%	42%	63%	59%

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

Evidence-based guidelines indicate that sealants are most effective for children at higher risk for caries (see Measure Evidence Form). 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.

6. 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 - Sealant and Tooth Number Codes

To assess whether the specific preventive service of dental sealants and associated tooth type are 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 sealants and tooth number. Agreement (concordance) for sealant placement was 95%. Sensitivity of sealant placements was moderately high (77.8%) and specificity was very high (98.8%). Similar findings were obtained for first molars. The positive predictive and negative predictive values were both high (>93%) for sealant placement with a lower negative predictive value for the specific tooth type. 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 for sealants was also very high at 0.8205 indicating “almost perfect” agreement. For dates of service in which there was agreement with the administrative data that sealants had been applied (n=84), we then assessed whether there was agreement on tooth type using the following categories: permanent first molar, permanent second molar, and other teeth. We report here on the findings for permanent first molar which is the focus of the proposed measure (we had similar findings for second molars). Overall, the simple agreement percentage was 84% for permanent first molars. The corresponding kappa statistic value was 0.691, indicating “substantial” agreement.

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

	Concordance	Prevalence	Sensitivity	Specificity	PPV	NPV	Kappa
Sealants Applied	95.22%	0.172	0.778	0.988	0.933	0.955	0.820
Dates of service: 613			(0.686-0.850)	(0.974-0.995)	(0.855-0.973)	(0.933-0.971)	(0.758-0.882)
# indeterminate: 4							
First Molar (if sealant)	84.34%	0.627	0.750	1.000	1.000	0.705	0.691
Dates of service: 613			(0.608-0.855)	(0.863-1.000)	(0.888-1.000)	(0.546-0.828)	(0.545-0.838)
# indeterminate: 1							

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 sealants, they found lower sensitivity (73%), higher specificity (100%) and similar kappa value (0.84) 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

Sealants on a Permanent Molar Tooth 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, 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 measures, including not associating tooth type and age range, lack of clear specifications, and lack of standardization. The proposed measure overcomes these limitations.

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 (sealants) as well as the specific tooth type for which the evidence is the strongest (permanent molar) and the timing of sealant placement to maximize effectiveness (shortly after eruption – 6-9 years of age for permanent first molars). Please see the Measure Evidence Form for more details.

3. ADDITIONAL VALIDITY TESTING - RELEVANCE OF TOOTH NUMBER

Analysis of sealant placement by tooth type and age range validated the importance of including specific teeth numbers in the measure specifications to identify permanent first molars and permanent second molars and associating those tooth numbers with the corresponding appropriate age ranges (6-9 years and 10-14 years, respectively) in order to have reliable indicators of whether children are getting recommended and timely prevention. Table 2b2.3-3 indicates the percentage of children in each of three programs who had (1) a sealant placed on any tooth, (2) a sealant placed on a permanent first molar, and (3) a sealant placed on a permanent second molar; the same child could be included in more than one category. In programs 3 and 4, the percentage of children ages 6-9 years with sealants on permanent first molars is very close to the percentage of children with sealants on any tooth, suggesting that most children ages 6-9 years in these two programs who received sealants received them for permanent first molars. However, in Program 1 there were substantial differences between the percentage of children with a sealant on any tooth compared to the percentage of children with a sealant on a permanent first molar. For example, 25% of children received a sealant, but only 14% received a sealant specifically on a permanent first molar. The differences reflect differences in benefit coverage between the programs; Program 1did not condition reimbursement for sealants on tooth type. These results indicate that children ages 6-9 years may have teeth other than permanent first molars (e.g., premolars or primary teeth) sealed that would get captured in the numerator and inflate the measure score if teeth numbers are not included, resulting in misleading comparisons of performance between programs. Thus, the research team concluded that the incorporation of teeth numbers in the DQA specifications is a significant and important improvement over existing sealant measures that have lacked this specificity.

Table 2b2.3-3 Sealant Placement by Age and Tooth Type

	Program 1			Program 3			Program 4		
Age (years)	% with Any Sealants (Any Tooth)	% with Sealant on Permanent 1st Molars	% with Sealant on Permanent 2nd Molars	% with Any Sealants (Any Tooth)	% with Sealant on Permanent 1st Molars	% with Sealant on Permanent 2nd Molars	% with Any Sealants (Any Tooth)	% with Sealant on Permanent 1st Molars	% with Sealant on Permanent 2nd Molars
6	25.02%	13.73%	0.04%	6.42%	6.32%	0.04%	8.21%	7.58%	0.01%
7	34.44%	26.20%	0.06%	15.03%	14.95%	0.09%	21.21%	20.92%	0.09%
8	31.02%	21.56%	0.08%	15.52%	15.49%	0.15%	18.85%	18.70%	0.12%
9	29.80%	14.00%	0.28%	12.45%	12.34%	0.18%	11.35%	11.06%	0.19%
10	35.36%	9.91%	1.87%	10.36%	9.90%	0.88%	7.63%	6.77%	0.74%
11	40.45%	7.42%	6.92%	10.18%	8.78%	3.07%	7.70%	4.92%	3.18%
12	40.96%	5.36%	12.76%	10.46%	7.67%	6.29%	11.99%	4.57%	9.05%
13	36.20%	3.73%	14.40%	10.40%	6.89%	8.27%	14.94%	4.04%	13.34%
14	29.85%	2.82%	11.64%	9.07%	5.93%	8.08%	12.44%	3.32%	11.51%

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. Although the agreement for the specific tooth type was not as strong as for sealant application in general, it was still "substantial," and we believe that data concordance will improve with increasing accountability as is often the case when new performance measures are implemented. 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 assessments of the measure score, lead us to conclude that both the data elements and the measure score represent valid measures of sealant placement prevalence among 6-9 year olds. In addition, our testing indicated that the incorporation of tooth number as part of the measure specifications was important for ensuring that the measure captures sealant placement on the tooth type (permanent first molars) for which there is the strongest evidence of effectiveness among this age group.

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 five 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 4 programs with performance data for 2011, (2) the 5 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 had a sealant, the dichotomous outcome of had/did not have a sealant 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:	23.69%	(0.2369 ,	0.0006 ,	0.2357 ,	0.2381)
Program 2, CY 2011:	23.01%	(0.2301 ,	0.0017 ,	0.2267 ,	0.2335)
Program 3, CY 2011:	31.33%	(0.3133 ,	0.0036 ,	0.3062 ,	0.3204)
Program 4, CY 2011:	22.59%	(0.2259 ,	0.0042 ,	0.2176 ,	0.2342)
Program 1, CY 2010:	23.38%	(0.2338 ,	0.0007 ,	0.2325 ,	0.2351)
Program 2, CY 2010:	19.82%	(0.1982 ,	0.0017 ,	0.1949 ,	0.2015)
Program 3, CY 2010:	30.04%	(0.3004 ,	0.0036 ,	0.2933 ,	0.3075)
Program 4, CY 2010:	26.68%	(0.2668 ,	0.0043 ,	0.2583 ,	0.2753)
Program 5, CY 2010:	21.04%	(0.2104 ,	0.0015 ,	0.2074 ,	0.2134)
Plan 1, CY 2011:	31.43%	(0.3143 ,	0.0054 ,	0.3037 ,	0.3249)
Plan 2, CY 2011:	30.91%	(0.3091 ,	0.0050 ,	0.2993 ,	0.3189)
Plan 1, CY 2010:	31.38%	(0.3138 ,	0.0078 ,	0.2985 ,	0.3291)
Plan 2, CY 2010 :	29.97%	(0.2997 ,	0.0067 ,	0.2866 ,	0.3128)

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 (Table 2b5.2).

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

	Chi-Square Value	*p* - value
Program Results, 2011	548.60	<0.0001
Program Results, 2010	1049.18	<0.0001
Plan Results, 2011	0.50	0.4795
Plan Results, 2010	1.88	0.1703

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 the program level with performance scores ranging by approximately 10 percentage points. We did not detect statistically significant differences between the two plans within FL CHIP for this measure. Performance between the two plans were similar on this measure with a 1/2 of one percentage point difference in 2010 (31.43% versus 30.91%) and a 1.41 percentage point difference in 2011 (31.39% versus 29.97%). We do not believe that this signifies the inability of the measure to detect differences in performance between plans; rather, the two plans we tested performed similarly on the measure. Presumably, testing does not require that all comparisons evaluated demonstrate statistically significant differences; rather, testing should demonstrate that where meaningful differences exist, they can be detected. However, we can also look to Program 2 for further comparisons at the plan level because Program 2 was served by a single dental plan so the program measure score also represents a plan-level score. Differences between the Program 2 measure scores (which also represents a single dental plan) are significantly different from those for Plan 1 and Plan 2 as can be seen by comparing the confidence intervals in Table 1b.2. Collectively, these findings are consistent with evidence reported elsewhere in this application documenting disparities in sealant receipt among children. Thus, this measure informs performance improvement efforts by allowing plans and programs to identify and monitor performance gaps and disparities in performance 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) at the same reporting level.

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. Required for maintenance of endorsement. 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.

IF instrument-based, 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, this measure concept was rated as an 8 or "definitely feasible" by the expert panel. During the empirical testing phase, our testing found that all of the critical data elements except one 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. The exception was tooth number associated with sealant procedure codes. Missing/invalid data rates ranged from 0.15% to 15%, with most programs

having missing/invalid rates <5%. We do not view the higher rates among a subset of the programs as a threat to feasibility, however. The high compliance by the majority of programs indicates that it is feasible to obtain missing and invalid rates of <1%. The Centers for Medicare and Medicaid Services already requires state Medicaid programs to report sealants placed on permanent molars among enrolled children, which requires data on tooth number, and tooth number also is typically required for reimbursement. 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 of collecting any of the data elements 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%

Tooth number: 6.18%

Date of Service: 0.01%

Rendering Provider ID: 0.28%

PROGRAM 2

Member ID: 0.00%

Date of Birth: 0.00%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.00%

Tooth number: 15.31%

Date of Service: 0.00%

Rendering Provider ID: 0.00%

PROGRAM 3

Member ID: 0.27%

Date of Birth: 0.00%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.28%

Tooth number: 0.18%

Date of Service: 0.00%

Rendering Provider ID: 0.18%

PROGRAM 4

Member ID: 0.00%

Date of Birth: 0.00%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.01%

Tooth number: 2.47%

Date of Service: 0.00%

Rendering Provider ID: 0.61%

PROGRAM 5

Member ID: 0.43%

Date of Birth: 0.02%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.00%

Tooth number: 0.15%

Date of Service: 0.00%

Rendering Provider ID: 0.67%

Endorsement Maintenance Update:

This measure is included in the CHIPRA Core Measures Program. Some Medicaid programs noted that they do not receive complete data on tooth number from their contracted plans, which is a required data element for this measure. As a result, the affected programs must get these data from their contracted plans. Because tooth number is required for reimbursement, these data are readily accessible for plan level reporting. Despite initial concerns about this data element, 25 states reported this measure in FFY 2015, and 34 reported in FFY 2016.

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, are accessible through a website and can be used free of charge for non-commercial purposes. The main requirement 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 CMS CHIPRA core set https://www.medicaid.gov/medicaid/quality-of-care/downloads/2017-child-core-set.pdf Texas Health and Human Services Commission: Texas Medicaid/CHIP https://hhs.texas.gov/sites/default/files//documents/laws-regulations/handbooks/umcm/6-2-15.pdf Payment Program Texas Health and Human Services Commission: Texas Medicaid/CHIP https://hhs.texas.gov/sites/default/files//documents/laws-regulations/handbooks/umcm/6-2-15.pdf Quality Improvement (external benchmarking to organizations) Covered California http://hbex.coveredca.com/insurance-companies/PDFs/2017-2019-Individual-Model-Contract.pdf CMS CHIPRA core set https://www.medicaid.gov/medicaid/quality-of-care/downloads/2017-child-core-set.pdf Michigan Healthy Kids Dental https://www.buy4michigan.com/bso/external/bidDetail.sdo?bidId=007117B0011386&parentUrl=activeBids Quality Improvement (Internal to the specific organization) State Medicaid Agencies http://www.msdanationalprofile.com/2015-profile/management-reporting-and-quality-measurement/quality-measurement/?

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/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 levels 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. Centers for Medicare and Medicaid Services, Core Set of Children's Health Care Quality Measures for Medicaid and CHIP (CMS CHIPRA Core Set)
https://www.medicaid.gov/medicaid/quality-of-care/downloads/2017-child-core-set.pdf
Purpose: Quality Improvement/Public Reporting
This measure was included in the CHIPRA Core Set, with reporting starting in FFY 2015. In the first year of reporting, 25 states reported this measure (https://www.medicaid.gov/medicaid/quality-of-care/downloads/performance-measurement/2016-child-chart-pack.pdf). In the second year of reporting (FFY 2016), 34 states reported this measure (https://data.medicaid.gov/Quality/2016-Child-Health-Care-Quality-Measures/wnw8-atzy).
Geographic Area and Number/Percentage of Accountable Entities and Patients:
34 states are currently reporting this measure. Information is not provided on the number of accountable entities and patients.
4. 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), 13 states reported that they currently use this measure in their Medicaid and/or CHIP programs.
Geographic Area and Number/Percentage of Accountable Entities and Patients:
The 13 states are: Alabama, Colorado, Connecticut, Florida, Idaho, Illinois, Nevada, Oklahoma, Rhode Island, South Carolina, Tennessee, Virginia,and West Virginia. Data are not provided on the number of accountable entities included.
5. Michigan Healthy Kids Dental
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 approved by the Dental Quality Alliance (DQA) that focused on Dental Caries Prevention and Disease Management among children. 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.

This measure is part of the CMS CHIPRA core set for public reporting by all state CHIP programs. In FFY 2016, 34 states reported on this measure. States also report using this measure in the annual survey conducted by the Medicaid|Medicare|CHIP Services Dental Association. The measure is part of measure set included in the Request for Proposals (RFP) released by the Michigan Healthy Kids Dental Program. This measure is included in the Pay-For-Quality program and public reporting 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 assistance 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 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.

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.

A dental benefits administrator (DBA) has suggested that the DQA consider adding patient exclusions to the measure. The DQA considered exclusions previously during initial measure development and during annual reviews. Exclusions were not incorporated due to concerns about the introduction of biased measurement, increasing measurement complexity, and adversely affecting implementation feasibility. However, the DQA continues to monitor this issue and will revisit it during the 2018 annual review. The DQA has invited the DBA to present its suggestion with supporting data to the DQA. The DQA has also invited other DBAs and Medicaid program administrators to provide input. All of this stakeholder feedback will be incorporated into the next annual review.

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

No other significant issues have been raised by other users.

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.

The DQA considered exclusions during initial measure development and during annual reviews. Exclusions were not incorporated due to concerns about the introduction of biased measurement, increasing measurement complexity, and adversely affecting implementation feasibility. However, the DQA continues to monitor this issue and will revisit it during the 2018 annual review. The DQA has invited the DBA to present its suggestion with supporting data to the DQA. The DQA has also invited other DBAs and Medicaid program administrators to provide input. All of this stakeholder feedback will be incorporated into the next annual review.

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.

Repeat measurements for two years are available from the CMS CHIPRA Child Health Care Quality Measures reporting. CMS has not released its formal report evaluating trends and changes. However, the data released indicate that in both FFY 2015 and FFY 2016 the median performance was 23.4% in both years across all states reporting the measure. As noted above, 9 additional states reported the measure in FFY 2016 (34 in 2016 versus 25 in 2015). CMS has not reported on improvement among the states who reported the measure in both years.

There also are initial reporting data available from the Texas Medicaid/CHIP programs (https://thlcportal.com/qoc/dental), which started implementing this measure after approval 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, 461207, 25.41, 25.59, 25.53

2015, 503515, 24.99, 25.18, 24.91

Texas CHIP

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

2014, 76415, 20.17, 22.30, 21.69

2015, 58833, 20.20, 23.14, 22.43

These data also suggest fairly stable rates over the two-year period. However, as noted above, these are initial performance data; additional time may be needed to see improvement within this 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_Sealants69.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 process of the measures. 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 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.