NATIONAL QUALITY FORUM
Driving measurable health improvements together

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: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3701

Measure Title: Prevention: Topical Fluoride for Children, Oral Health Services

Measure Steward: American Dental Association

Brief Description of Measure: Percentage of children aged 1 through 20 years who received at least 2 topical fluoride applications as oral health services within the reporting year.

The measure is specified for reporting at the program and plan levels for both public and private/commercial reporting.

Developer Rationale: Dental caries remains one of the most common, yet preventable, diseases of childhood. Dental decay in children has significant short- and long-term adverse consequences on children's health and overall well-being. Updated national surveillance data for the period 2011-2016 indicates that 23% of children aged 2-5 years experience dental caries related lesions, increasing to 52% among children aged 6- 8 years. Untreated decay was 10% among children aged 2-5 years and 16% among children aged 6-8 years. On permanent teeth, the prevalence of caries related tooth lesions was 17% for children 6-11 years and 57% among adolescents aged 12-19 years. Low-income children are twice as likely to have dental decay as higher-income children. Untreated decay was 5% among children aged 6-11 years and 17% among adolescents aged 12-19 years. (Centers for Disease Control and Prevention, 2019)

Oral health disparities are well documented and persist. Poor and near-poor children (children living in households with <=200% of the federal poverty level) are approximately twice as likely to have dental caries and untreated decay compared to higher income children (>200% FPL). National surveillance data indicate that Mexican American and non-Hispanic black children are more likely to have dental caries and untreated decay than non-Hispanic white children. For example, the prevalence of dental caries related lesions in primary teeth among Mexican American children aged 6-8 years was 73% compared to 44% for non-Hispanic white children. (Centers for Disease Control and Prevention, 2019)

Although dental caries can be managed and caries-related lesions can be treated and restored, it is important to prevent the disease process from developing in the first place. As noted in the evidence section, multiple systematic reviews with meta-analyses find evidence supporting professionally applied topical fluoride, starting as early as six months of age and applied at least twice per year, as beneficial in preventing dental caries and associated decay (USPSTF 2021, Marinho 2013, Weyant 2013).

The proposed measure, Topical Fluoride for Children, Oral Health Services, captures whether children received at least two topical fluoride applications as an oral health service, by a physician or health care provider other

than a dentist nor under supervision of a dentist. This measure directly reflects evidence-based guidelines regarding an effective caries prevention measure (professionally applied topical fluoride in medical care setting), including the frequency required for clinical effectiveness (at least every six months). This is an important process of care measure that enables identification of topical fluoride receipt among children without a dental home. Consequently this stand-alone measure also allows plans and programs to link children to a regular source of dental care as a quality improvement strategy.

References

Centers for Disease Control and Prevention. Oral Health Surveillance Report: Trends in Dental Caries and Sealants, Tooth Retention, and Edentulism, United States, 1999–2004 to 2011–2016. Atlanta, GA: Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2019. Available at: https://www.cdc.gov/oralhealth/publications/OHSR-2019-index.html

Marinho VCC, Worthington HV, Walsh T, Clarkson JE. 2013. Fluoride Varnishes for Preventing Dental Caries in Children and Adolescents (Review). Cochrane Database of Systematic Reviews 2013, Issue 7. Art. No.: CD002279. DOI: 10.1002/14651858.CD002279.pub2. PMID: 23846772. Available at: https://www.cochrane.org/CD002279/ORAL_fluoride-varnishes-for-preventing-dental-caries-in-children-and-adolescents.

US Preventive Services Task Force. Screening and Interventions to Prevent Dental Caries in Children Younger Than 5 Years, Final Recommendation Statement. December 7, 2021. *Available at:* https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/prevention-of-dental-caries-in-children-younger-than-age-5-years-screening-and-interventions1#bootstrap-panel--12

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

Numerator Statement: Unduplicated number of children who received at least 2 topical fluoride applications as oral health services

Denominator Statement: Unduplicated number of children aged 1 through 20 years

Denominator Exclusions: There are no measure-specific exclusions. There is a standard exclusion as part of determining denominator eligibility: Medicaid/CHIP programs should exclude those individuals who do not qualify for dental benefits.

Measure Type: Process

Data Source: Claims, Enrollment Data

Level of Analysis: Health Plan, Program

IF this measure is paired/grouped, NQF#/title: Prevention: Topical Fluoride for Children

#3701 - Prevention: Topical Fluoride for Children, Oral Health Services

#2528 - Prevention: Topical Fluoride for Children, Dental Services

#3700 - Prevention: Topical Fluoride for Children, Dental or Oral Health Services

IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:

Although this measure can be reported as a stand-alone measure, it is being grouped with two complementary measures (2528 and 3700) to enable more robust quality improvement efforts. The DQA considered submitting a single measure with three numerators (denominator population is the same). But NQF evaluation criteria state: “Measures with multiple measure components that are assessed for each patient, but that result in multiple scores for an accountable entity rather than a single score. These generally should be submitted as separate measures and indicated as paired/grouped measures.” (Measure Evaluation

Criteria and Guidance, September 2021, p. 52). Based on this and discussions with NQF staff, we are submitting as three distinct measures.

This measure – NQF 3701: Topical Fluoride for Children, Oral Health Services – focuses on topical fluoride delivered as an "oral health" service (by a physician or health care provider other than a dentist nor under supervision of a dentist).

Because many children, especially very young children, do not have a dental home, many state Medicaid programs and MCOs pay for the application of topical fluoride as an "oral health" service (by a physician or health care provider other than a dentist nor under supervision of a dentist). Consequently, state Medicaid programs, as well as commercial integrated medical-dental benefit MCOs or integrated medical-dental healthcare delivery sites, have a strong interest in tracking whether children receive **any** topical fluoride regardless of provider type ("dental" or "oral health" services). They also have a strong interest in understanding whether, and for whom, topical fluoride is being delivered through "dental" providers and "oral health" providers.

Measures of topical fluoride provision by provider type, in addition to a measure of overall provision, are important because multi-pronged quality improvement strategies may be used to improve rates of topical fluoride application among a population of children. Dental providers and/or medical providers may be the focus of these efforts. Without measures that track the effectiveness by provider type, it is more difficult for programs and plans to assess which efforts are most effective. In addition, the accountability and delivery systems are typically distinct. Improving fluoride application by dental providers is accomplished through the dental delivery system and related financing/reimbursement structures, whereas topical fluoride application by medical providers is accomplished through medical delivery system and related financing/reimbursement structures. Some measure users will benefit by implementing and using all three measures (such as Medicaid programs and private payers/delivery systems that include both medical and dental). Other users, focused specifically on either dental or medical care delivery, respectively, will be able to report using one of the measures: either the measure related to "dental" services or the measure related to "oral health" services.

The need for three grouped measures comes directly from user community requests. In considering a topical fluoride measure for inclusion in the Centers for Medicare and Medicaid Services' Core Set of Children's Health Care Quality Measures, a specific request was made for a single measure that included three numerators (dental or oral health services, dental services, and oral health services) because of the recognized need by Medicaid programs to track not only overall receipt of topical fluoride but also topical fluoride provided through the dental and medical delivery systems specifically. Review by the DQA's Measures Development and Maintenance Committee, which includes representation of providers, community health centers and payers, affirmed the value of reporting three numerators across public and private/commercial measure applications for the reasons described above.

Consequently, the DQA is submitting two complementary measures for endorsement to be "grouped" with this measure: (1) Topical Fluoride for Children, Dental Services (NQF#2528) and (2) Topical Fluoride for Children, Dental or Oral Health Services (NQF#3700). This grouping provides users with measurement options that appropriately support population-based assessments of quality. It enables measure users, including Medicaid programs and their contracted MCOs, integrated medical-dental MCOs, and integrated medical-dental delivery systems, to examine the overall provision of topical fluoride by provider type, which has been identified by stakeholders as integral for quality improvement and accountability purposes.

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a ***structure, process or intermediate outcome*** measure are 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 description for this measure:

- This is a new process measure at the health plan and program level that evaluates the percentage of children aged 1 through 20 years who received at least 2 topical fluoride applications as oral health services within the reporting year. This measure is paired with two measures that also focus on topical fluoride but are delivered through dental services and through either dental or oral health services.
- The developer provides a logic model that depicts topical fluoride applied to children starting as early as six months of age is beneficial in preventing dental caries.

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

Summary:

- The developers identified a Cochrane systematic review as evidence.
  - There were 22 studies included in this review. All were randomized or quasi-randomized controlled trials. The evidence found that fluoride varnish on permanent dentition is associated on average with a 43 percent reduction in decayed, missing, and filled tooth surfaces. For fluoride varnish on primary dentition, there was a 37 percent reduction. In general, they stated there was little information on possible adverse effects.
  - The review received a quality rating of moderate which was defined as further research will likely have an impact on the confidence in the estimate of the effect. This was the second highest option in the scale used, the higher option being high quality and the lower options being low quality and very low quality.
  - The study concluded that fluoride varnish in both permanent and primary teeth have a substantial caries-inhibiting effect.
  - Because this is not a clinical guideline, no formal recommendations were made. Rather, the author provided an implication for practice. The implication stated that the review found the application of fluoride varnishes two to four times a year was associated with a reduction in caries.
- The developer also cited the United States Preventive Services Task Force (USPSTF) systematic review and recommendation on the *Prevention of Dental Caries in Children Younger than 5 Years*.
  - This study addressed two key questions. The first is asking how effective preventive treatment is in preventing dental caries in children younger than 5 years old. The second is asking what

harms are associated with preventive oral health interventions in children younger than 5 years old.
    - There were 32 studies and one systematic review that included 19 studies in the USPSTF review. Twenty-two studies addressed the "how effective…" key question stated above. Fifteen of which specifically addressed topical fluoride application. These studies were all randomized controlled trials. Four studies reported on adverse events associated with topical fluoride. These studies were all randomized controlled trials. The trials found that there were decreased caries increments and decreased likelihood of incident caries. Further, the studies found that there were no reported differences between fluoride varnish and placebo in risk of fluorosis or adverse events. The only "adverse events" reported were that children did not like the smell of the fluoride varnish. Additionally, one study found that a few children vomited due to the smell, texture, or taste.
    - The studies within this review all supported that the recommendation that primary care clinicians should apply fluoride varnish to the primary teeth of all infants and children starting from primary tooth eruption. This recommendation received a B rating which suggests that the service should be offered and provided. This rating is the second highest option in a five-letter grade scale.
    - The authors gave the evidence a moderate rating which is the second highest option available in the scale. The authors stated that the evidence concluded there was a moderate certainty that a moderate benefit of preventing dental caries with fluoride varnish applications exists in all children younger than age 5.
- The developers then presented a systematic review done by the American Dental Association (ADA).
    - The evidence-based clinical guidelines recommend the specific topical fluoride agents for people who are at elevated risk of developing dental caries.
    - Additionally, the guidelines in the report recommend applying 2.26 percent fluoride varnish at least every three to six months for children younger than six years old and for children six-18, they recommend 2.26 percent fluoride varnish at least every three to six months or 1.23 percent acidulated phosphate fluoride (APF) gel for four minutes at least every three to six months.
    - Seventy-one studies were included in evidence reviews. All studies included were controlled clinical trials. Seventeen randomized and five non-randomized control trials evaluated 2.26 percent fluoride varnish while 11 randomized and four non-randomized control trials evaluated 1.23 percent APF gel.
    - The only potential harms identified in the study were nausea and vomiting when topical fluoride is ingested and dental fluorosis while tooth enamel is developing.
    - The evidence received a moderate grade by an expert panel which was the second highest rating it could receive.
    - The clinical recommendations for fluoride among children and adolescents received an evidence grade of '"in favor" which was the second highest recommendation out of a six-point scale.

Exception to evidence

- N/A

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance from the Evidence Algorithm

Process measure based on systematic review and grading (Box 3) → QQC provided in the submission (Box 4) → Quantity: High; Quality: Moderate; Consistency: Moderate/High (Box 5b) → Moderate

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

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

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

- Data on performance gap was derived from 14 state Medicaid programs including Alaska, Arizona, Delaware, Idaho, Michigan, Mississippi, Nevada, New Mexico, North Carolina, Oklahoma, Oregon, South Carolina, Washington, and Wyoming.
    - These states were specifically selected based on their quality of data and diverse geographic location, population size, demographic characteristics, and Medicaid dental delivery system. The data came from Medicaid enrollment and claims data contained within the Transformed Medicaid Statistical Information System (T-MIS) Analytic Files (TAFs).
    - There were over seven million enrollees aged zero to 20 years across the 14 programs for each of the calendar years (2016, 2017, and 2018) included in the analysis.
    - In the most recent year of data, 2018, measure scores ranged from 0.16 percent to 3.6 percent. A similar amount of variation exists in 2017 and 2016 as well. The developer states this variation indicates a gap in care.

Disparities

- The developer stratified measures scores from 2018 into four categories: 1) age; 2) rural/urban geographic location; 3) race and ethnicity; and 4) sex assigned at birth. The did not report stratification when more than 10 percent of the data for the variable was missing.
- There were differences between all four categories, but the developer noted that the most significant differences existed between age and race/ethnicity.
    - Children in the youngest age group had the highest performance scores while children in the oldest age group had the lowest performance scores.
    - Additionally, non-Hispanic White, non-Hispanic Black, and Hawaiian/Pacific Islander children had lower performance scores than non-Hispanic Asian, non-Hispanic American Indian/Alaska Native, and Hispanic children.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

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

Committee Pre-evaluation Comments:

1a. Evidence

- I noted my concerns in the other 2 related measures. This one in particular, whereby data is not reliant on Dental CPT codes, with expanded age and lack of risk factors is one that I find hard to justify as a necessary metric

- There is evidence to support this measure. It is a new process measure.
- Rating: low, numerator definition & evidence is problematic, low for children w/o permanent teeth, moderate for children w permanent teeth, insufficient for physically mature "older children" with permanent teeth, systematic review w QQC graded evidence for use on "permanent dentition"
- Strong evidence reflected in systematic reviews and practice guidelines. Adding a measure for oral health service practitioners to complement the previous measure (NQF #2528) focused only on dental providers makes this evidence more relevant to the measures.
- Strong evidence provided
- This is a new process measure at the health plan and program level that evaluates the percentage of children aged 1 through 20 years who received at least 2 topical fluoride applications as oral health services within the reporting year. This measure is paired with two measures that also focus on topical fluoride but are delivered through dental services and through either dental or oral health services. Evidence is systematic literature review, graded, with quality, quantify and consistency of evidence report.
- The evidence offered for this measure is moderately strong and supports improved dental health, particularly for vulnerable populations.

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

- Addressed in previous answer
- There is data regarding performance and disparity gaps. This measure is an attempt to collect additional data that could inform the performance gaps since children may not receive fluoride from a dentist.
- Rating: Low, no age appropriate evidence-based risk stratification to support performance gap
- Strong evidence of overall performance gaps and disparities, especially by age and race/ethnicity.
- Strong evidence of both gaps and inequities
- There were gaps of treatment within selected states, age groups, rural/urban and ethnicities.
- Disparities by age, geographic location, race/ethnicity and sex were reported, underscoring the need for improvement.

Criteria 2: Scientific Acceptability of Measure Properties

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

Evaluators: Staff

2a. Reliability: Specifications and Testing

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

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.

Specifications:

- This measure is the oral health services version of NQF #2528 Topical Fluoride for Children, Dental Services.
    - The developer states that since many children, especially very young children, do not have a dental home, many state Medicaid programs and managed care organizations (MCOs) pay for the application of topical fluoride as an “oral health” service (by a physician or health care

provider other than a dentist nor under supervision of a dentist). Consequently, state Medicaid programs, as well as commercial integrated medical-dental benefit MCOs or integrated medical-dental delivery sites, have a strong interest in tracking whether children receive any topical fluoride regardless of provider type ("dental or oral health" services). They also have a strong interest in understanding whether, and for whom, topical fluoride is being delivered through "dental" providers and "oral health" providers. Consequently, the developer is submitting three complementary measures for endorsement

Reliability Testing:

- The level of analysis states health plan and program; however, the testing is only conducted at the program level. The developer justified this by stating that the program data is transferrable to the plan level. They noted further that the only potential concern would be denominator size. They stated that the denominator requirement from the measure captures a broad population and to-date they have not encountered issues with small denominators.
- Reliability testing conducted at the accountable-entity level:
    - The developers used a random-split sample methodology and intraclass correlation coefficient (ICC). For each state Medicaid program, they randomly split the population and the denominator, numerator, and measure score were calculated for each sample. The ICC is used to calculate agreement between the samples.
        - The developer stated that the variation between split samples for each of the Medicaid programs is relatively small, signifying that the samples were similar. It is also noted that the measure scores have overlapping 95 percent confidence intervals (CI). Additionally, the ICC scores were reported by year.
            - 2018 (n=14): ICC = 0.999 with a 95 percent CI of (0.9972, 0.9997) and a p-value of less than 0.0001
            - 2017 (n=14): ICC = 0.999 with a 95 percent CI of (0.9961, 0.9996) and a p-value of less than 0.0001
            - 2016 (n=12): ICC = 0.999 with a 95 percent CI of (0.9990,0.9999) and a p-value of less than 0.0001
    - Additionally, an evaluation of relative rankings between split samples and year were done. For the relative ranking between split samples, they were evaluating whether measure scores remained stable. For the relative ranking between years, they were evaluating if any dramatic changes happened that would threaten reliability. To report the findings, they calculated Kendall's Tau-b which is a rank correlation coefficient that measures associations based on concordant and discordant pairs. They also reported the Spearman's rank correlation coefficient. They noted that while they did both, they felt Kendall's Tau-b was the more appropriate statistical test given the relatively small sample size (n=14).
        - The developer noted that the relative rankings based on measure score are stable across split samples.
        - The relative rankings between years were as follows:
            - For 2017 & 2018 (n=14): Kendall's Tau-b = 0.6484 (p-value = 0.0015) and Spearman's Rank = 0.7714 (p-value = 0.0012)
            - For 2016 & 2017 (n=12): Kendall's Tau-b = 0.5636 (p-value = 0.0195) and Spearman's Rank = 0.6636 (p-value = 0.026)
- Reliability testing conducted at the patient/encounter level:

- The developer presented patient/encounter level validity testing from NQF #2528 for their reliability testing at the patient/encounter level. This information is presented in more detail in the validity section.
    - The developer validated the encounter data by comparing claims data against dental charts. They noted that record and procedure codes on the claims had a 100 percent inter-rater agreement rate.
    - Additionally, the developer assessed whether the preventive service of topical fluoride application was accurately captured by claims data. The developer reported concordance (89.9 percent), sensitivity (90.7 percent), specificity (88.4 percent), positive predictive value (93.5 percent), and negative predictive value (83.9 percent). The developer also reported a kappa statistic which was 0.782.
    - The developer conducted an analysis on the CPT code 99188 which was not included in NQF #2528 to ensure that it was valid. They found that there was nothing in the data to suggest that the code was being used improperly.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are measure specifications adequate)?*
- *Are there concerns that the measure was not tested at the plan level and will program level testing translate to the plan level?*

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

2b. Validity: Validity testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

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.

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

Validity Testing

- The developer provided only patient/encounter level validity testing for this measure. The patient/encounter validity testing was similar to the testing done for NQF #2528; however, the developer did an additional analysis of CPT code 99188.
- The level of analysis states health plan and program; however, the testing is only conducted at the program level. The developer justified this by stating that the program data is transferrable to the plan level. They noted further that the only potential concern would be denominator size. They stated that the denominator requirement from the measure captures a broad population and to-date they have not encountered issues with small denominators.
- Validity testing at the patient/encounter level
    - Analysis of CPT code 99188 and CDT codes D1206/D1208
        - The developer noted that the main difference between this measure and NQF #2528 Topical Fluoride, Dental Services was the inclusion of CPT code 99188 in the numerator in addition to CDT codes D1206 and D1208, and the identification of oral health provider taxonomy codes. They stated that #2528 focused on validating the procedure codes used in the numerator, which did not include CPT code 99188 as it did not exist at the time the measure was created. The developer additionally noted

due to the very specific description of CPT code 99188 and the guidance from the American Academy of Pediatrics on coding there is no reason to believe the code is not valid. To verify this theory, they ran analyses on the claims data to figure out if there was any reason to suggest the code is not being used properly.

- The developer found that of the 14 state Medicaid programs, CPT code 99188 was not used in four states. However, those states used CDT codes D1206 and D1208. The developer also found that two states use both CPT code 99188 and CDT code D1206. The other eight states used CPT code 99188.
- The developer also analyzed CPT 99188 code usage by provider type. They found that in all states except for one the provider that accounted for the highest percentage of CPT 99188 usage was "Pediatrics Physicians".
- Additionally, the developer analyzed CPT 99188 codes usage in states that reimburse for using this code in relation to age distribution. The developer found that the patterns were as expected based on the reimbursable age ranges in each state.

- The developer concluded that the analyses of the CPT code 99188 were as expected with the expected provider types rendering CPT 99188 services and services being concentrated within the age ranges eligible for reimbursement.

- Critical Data Element Validation
  - The developer used the critical element validation data from the NQF #2528 submission. They noted that there were five critical data elements (below). The developer also noted that the first four elements are core field used in virtually all measures relying on administrative data. Therefore, they determined that these have established reliability and validity. So, they focused testing on assessing the accuracy of topical fluoride procedure codes.
    - Member ID
    - Date of birth
    - Monthly enrollment indicator
    - Date of service
    - CDT codes
  - The developer used a random sample of 414 encounters for members ages three through 18 with at least one outpatient dental visit. The records were reviewed by two coders certified as registered health information technicians.
    - The first part of the validation was "encounter data validation" where the first three procedure codes were reviewed for each claim and evaluated on if the claims data was supported by the dental record. A 100 percent inter-rater agreement rate was reported for this step.
    - The second part of the validation was assessing whether the preventive service of topical fluoride application is accurately captured by claims data. The claims data were validated against dental records by comparing the records to the codes in the claims data.
      - To assess the validity, they calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the kappa statistic.
        - Concordance: 89.91 percent

- Prevalence: 0.647
- Sensitivity: 0.907 with a 95 percent CI of (0.857-0.942)
- Specificity: 0.884 with a 95 percent CI of (0.888-0.963)
- PPV: 0.935 with a 95 percent CI of (0.888-0.963)
- NPV: 0.839 with a 95 percent CI of (0.757-0.898)
- Kappa: 0.782 with a 95 percent CI of (0.710-0.853)

Exclusions

- The measure does not use exclusions.

Risk-Adjustment

- The measure is not risk adjusted or stratified.

Meaningful Differences

- The developer provided performance score data for each program across three calendar years with 95 percent confidence intervals which were used to calculate the mean, median, standard deviation, and percentile distributions used for a chi-square test. The developers noted that the 95 percent confidence intervals did not overlap across any of the programs.
- The developer calculated the interquartile range for the measure scores and then conducted a chi-square test to evaluate the statistical significance of the differences in the measure scores between the lowest and highest quartiles. The following results were reported:
    - 2018 – Mean = 0.0079, interquartile range = 0.0037, p-value = less than 0.0001
    - 2017 – Mean = 0.0082, interquartile range = 0.0047, p-value = less than 0.0001
    - 2016 – Mean = 0.0073, interquartile range = 0.0028, p-value = less than 0.0001
- The developer notes that these results suggest that the measure can identify statistically and clinically meaningful differences in performance.

Missing Data

- The developers assessed missing data in two ways.
    - The first is the Medicaid and CHIP Business Information Solutions (MACBIS) conducted data quality assessments of the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAFs). Then a value (low concern, medium concern, high concern, unusable, and unclassified) was assigned to each of the states based on if the data is usable, reliable, and accurate. Then the developers reviewed the results for age, Medicaid enrollment, claims file completeness (claims volume), claims file completeness (service users), and service use (procedure codes). The second is the developer did an analysis of date of birth, beneficiary ID, and CDT codes and reported the same values and cut points that MACBIS uses.
        - There were two states in calendar year 2016 for which the level of concern for missing data was medium and/or high
            - In the state that had medium concern for missing data, the developer chose to include their data for measure reporting purposes as examination of measure scores found that performance was within expected ranges and similar between years.

        - In the state that had medium and high concerns, they excluded 2016 reporting from the testing as examination of measure scores found that the missing data significantly impacted performance.
    - For the fields assessed by the developer (date of birth, beneficiary ID, CDT codes, and rendering provider taxonomy for dental procedure codes), the rate of missing and invalid data was generally less than one percent.
- The developer noted that because there was generally low concern and low rates of missing data, no rules were developed for handling it. They also stated that because the data are standard fields contained within administrative claims, the measure users are encouraged to improve data collection and quality as part of the quality improvement efforts rather than using statistical methods to address missing data.

Comparability

- The measure only uses one set of specifications for this measure.

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*

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

Committee Pre-evaluation Comments:

2a1. Reliability – Specifications

- Addressed in previous answer
- No concerns about reliability testing.
- Rating: Low, specification for denominator & numerator who benefit from alignment with the evidence provided that is limited to "permanent teeth"
- Changing the denominator from “children at elevated caries risk” (as in NQF #2528) to all children probably makes specifications clearer, as I do not know how “elevated caries risk” was defined.
- No concerns
- The level of analysis states health plan and program; however, the testing is only conducted at the program level. The developer justified this by stating that the program data is transferrable to the plan level.
- Reliability is acceptable for accountable-entity level and patient/encounter level.

2a. Reliability – Testing

- Addressed in previous answerNo concerns
- See comments above
- Testing at the program vs. plan level is a minor concern. Split sample ICCs and relative rankings compared – appropriate and strong results. Previous submission (for NQF #2528) included testing at the patient/encounter level with strong results.
- No concerns
- No
- No concerns that the measure cannot be consistently implemented.

2b1. Validity – Testing

- Addressed in previous answer
- No
- CPT & CDT code irregularities, validation using measure 2528 data
- No concerns. Face validity from previous submission still relevant. Appropriate validity testing conducted at the patient/encounter level and critical data element validation.
- No concerns
- No concerns
- The validity rating for this measure is moderate.

2b2-2b3. Threats to validity (Exclusions, Risk Adjustment)

- Addressed in previous answer
- Risk adjustment is not used.
- See comments above
- No concerns.
- I do wonder about whether HPSAs/dental provider access should be taken into account (as a risk adjustor, to stratify)? Do "oral health" providers fill the gaps?
- No risk adjustment or risk stratification
- This measure does not have exclusions nor is it risk adjusted or stratified.

2b4-2b7. Threats to validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)

- Addressed in previous answer
- Missing data is addressed
- Rating: Low, risk adjustment is needed due to 1 year old to 20 year old age bracket use with the evidence provided for "permanent teeth"
- No concerns.
- Missing data do pose a small threat
- No threat to validityThe measure developer assesses aspects of validity listed above with special focus on the availability of administrative claims codes across the sites included in the assessment. No concerns with validity.

Criterion 3. Feasibility

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.

- The data elements are coded by someone other than the person obtaining the original information and all data elements are defined fields in electronic claims. The measure uses data elements that are found in standard fields in administrative claims data which are routinely collected. Therefore, the developer concluded that the time and cost of data collection is relatively low.
- Additionally, the measure was designed to avoid using software or other materials that require licensing fees. The measure specifications are accessible through a website and are free.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*

- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee Pre-evaluation Comments:

3. Feasibility

- Addressed in previous answer
- Data elements are collected. Since it is claims data, it may not represent uninsured individuals not receiving vaccination which is likely one of the most significant barriers.
- Rating: Low, measure logic weak based on age bracket for a developmental (time specific) aspect of care
- Removing elevated risk designation from the previous measure (NQF #2528) makes the new measure more feasible.
- No concerns
- This measure relies on standard fields in administrative claims data (e.g., patient ID, patient birthdate, enrollment information, date of service, procedure codes) that are routinely collected for billing and other purposes.
- There is high feasibility for the measure because it uses electronic claims data.

Criterion 4: Use and Usability

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

4a. Use evaluates 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
Planned use in an accountability program?	☒	Yes	☐	No	☐	NA

Accountability program details

- The developer provided the measure's planned use:
  - Adopted by CMS for Child Core Health Care Quality Measurement for fiscal year 2022 reporting by state Medicaid and CHIP.
  - Included in the Center for Oral Health Systems Integration and Improvement (COHSII) Oral Health Quality Indicators for the Maternal and Child Health Population, which is funded by the

Health Services and Resources Administration (HRSA) Maternal and Child Health Bureau for 2022 reporting.

- The developer anticipates widespread adoption of the measure within three years.

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

- During the development phase results and data were shared with stakeholders including the CMS Oral Health Technical Advisory Group, NCQA, and the Center for Oral Health Systems Integration and Improvement. The Dental Quality Alliance (DQA) also provides technical assistance for measures through webinars, resource documents, and staff support. The DQA has also released a State Oral Health Quality Dashboard that has information on key oral healthcare measures.
- DQA has a process in place for reviewing and updating all measures. The process is overseen by the DQA's Measures Development and Maintenance Committee (MDMC). The review process includes public commenting, evaluation of comments, user group feedback, and code set reviews. In general, during this process the stakeholders responded positively to the measure and its ability to increase quality improvement efforts.

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

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

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

4b. Usability evaluates 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 states that the initial testing data suggests a performance gap exists.
- The developer notes that performance data will be shared via DQA's State Oral Health Quality Dashboard once reporting in the CMS Child Core Health Care Quality Measurement set becomes mandatory in 2024 which will facilitate the ability to identify performance, establish improvement goals, evaluate any changes over time, and how improvement varies across entities.

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

- No negative unintended impacts have been identified.

Potential harms

- The developer states the potential for harm is minimal.
- The developer expects the benefits to be consistent with the evidence.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

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

Committee Pre-evaluation Comments:

4a. Use

- Addressed in previous answer
- This measure is a bit confusing with the aligned measures that incorporate who offered the fluoride treatment. Those being measured have been given performance results.
- Rating: low, see comments above
- Measure developed based on feedback on the previous measure (NQF #2528), with systems in place to use.
- No concerns
- This measure is currently under consideration for use by the National Committee for Quality Assurance for Healthcare Effectiveness Data and Information Set (HEDIS) plan-level reporting. This measure has been included in the Center for Oral Health Systems Integration and Improvement (COHSII) Oral Health Quality Indicators for the Maternal and Child Health Population, which is funded by Health Services and Resources Administration (HRSA) Maternal and Child Health Bureau, for 2022 reporting. The measure will be reported with the two related measures proposed to be "grouped" with this measure (Topical Fluoride for Children, Dental Services, and Topical Fluoride for Children, Dental or Oral Health Services).
- The measure has been adopted by CMS for Child Core Health Care Quality Measurement for 2022, included in the Center for Oral Health Systems Integration and Improvement Oral Health Quality Indicators, and widespread adoption is anticipated. Feedback on measure performance was performed. Use of the measure will likely result in health care quality and efficiency.

4b. Usability

- Addressed in previous answer
- No identified unintended consequences.
- Rating: Low, no evidence provided on potential harm of fluoride use on primary (deciduous) teeth in the youngest strata of children
- No comments.
- No concerns
- The developers anticipate widespread adoption of this measure within 3 years. As noted above, this measure has been included in the CMS Child Core Set for use by Medicaid programs and their contracted MCOs. This measure is also under consideration by NCQA for plan-level reporting. The

measure has also been included in the COHSII Oral Health Quality Indicators for the Maternal and Child Health Population.

- Usability benefits outweigh potential risks.

Criterion 5: Related and Competing Measures

Related measures

- NQF #2511 Utilization of Services, Dental Services
- NQF #2517 Oral Evaluation, Dental Services
- NQF #2689 Ambulatory Care Sensitive Emergency Department Visits for Dental Caries in Children
- NQF #2695 Follow-Up after Emergency Department Visits for Dental Caries in Children

Harmonization

- The developer states that the measures have harmonized to the extent possible. Additionally, the developer noted that the above stated measures all address the same population, however, the denominators are different. Therefore, the measures are complementary to each other but, in fact, distinct.

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- Addressed in previous answer
- There are related measures based on the type of provider offering fluoride treatment. They are harmonized but the value of each may not be great.
- Lack of precision, presence of ambiguous denominator related to evidence provided, see comments above
- Addition of 3700 and 3701 to the previous measure (NQF #2528) is an important improvement.
- No concerns
- The NQF-endorsed DQA measures (#2511, #2517, #2689 and #2695) all address broadly the same population - children enrolled in Medicaid and CHIP. But the denominators are specified differently. The measures are complementary to one another but distinct. As noted above, this measure is proposed to be "grouped" with NQF #2528 (existing endorsed measure) and NQF #3700 (new measure).
- There are four related measures that address the same population, but are distinct and complimentary.

Public and NQF Member Comments (Submitted as of June 15, 2022)

Member Expression of Support

- No member submitted an expression of support.

Comments

- No NQF member and public comments were received in advance of the Standing Committee evaluation.

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☐ No**
2. **Briefly summarize any changes to the measure specifications and/or concerns about the measure specifications.**
   - This is a new measure, so no previous review/changes to specifications.
   - The level of analysis states health plan and program; however, the testing is only conducted at the program level. The developer justified this by stating that the program data is transferrable to the plan level. They noted further that the only potential concern would be denominator size. They stated that the denominator requirement from the measure captures a broad population and to-date they have not encountered issues with small denominators.

RELIABILITY: TESTING

3. **Reliability testing level: ☒ Accountable-Entity Level ☒ Patient/Encounter Level ☐ Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure:**
   ☒ Yes ☐ No
5. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?
   ☐ Yes ☐ No
6. **Assess the method(s) used for reliability testing:**
   - The developers presented reliability testing at the accountable-entity level using a random split-sample methodology and reporting an intraclass correlation coefficient (ICC). Additionally, they evaluated relative rankings between split-samples and years reporting Kendall's Tau-b and Spearman's rank correlation coefficients.
   - The developers presented patient/encounter level validity testing that stood for patient/encounter reliability testing.
     - They presented data element validation from NQF #2528 where they used a random sample of 414 encounters for members ages three-18 with at least one outpatient dental visit. The records were then reviewed by two coders.
       - The first part of the validation was "encounter data validation" where the first three procedure codes were reviewed for each claim and evaluated on if the claims data was supported by the dental record.
       - The second part of the validation was assessing whether the preventive service of topical fluoride application is accurately captured by claims data. The claims data were validated against dental records by comparing the records to the codes in the claims data.
         - They then calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and a kappa statistic.
       - Additionally, the developers presented validation information for CPT code 99188 and CDT codes D1206/D1208 because at the time of data element validation for NQF #2528, CPT code 99188 did not exist. The developer noted that due to the very specific description of CPT code 99188 and the guidance from the American Academy of Pediatrics on coding there is no reason to believe the code is not valid. However, to

verify this, they analyzed the claims data to determine if there was any reason to believe that the code was not being used properly.

7. **Assess the results of reliability testing**
    - For reliability testing at the accountable entity level the developer reported the following results
        - ICC correlations per calendar year
            - The variation between split samples for each of the Medicaid programs is relatively small, signifying that the samples were similar. Additionally, the ICC scores were reported by year.
                - 2018 (n=14): ICC = 0.999 with a 95% CI of (0.9972, 0.9997) and a p-value of less than 0.0001
                - 2017 (n=14): ICC = 0.999 with a 95% CI of (0.9961, 0.9996) and a p-value of less than 0.0001
                - 2016 (n=12): ICC = 0.999 with a 95% CI of (0.9990,0.9999) and a p-value of less than 0.0001
        - Relative rankings between split-samples
            - The developer reported that the relative rankings between split-samples demonstrated high stability
        - Relative rankings between years
            - For 2017 & 2018 (n=14): Kendall's Tau-b = 0.6484 (p-value = 0.0015) and Spearman's Rank = 0.7714 (p-value = 0.0012)
            - For 2016 & 2017 (n=12): Kendall's Tau-b = 0.5636 (p-value = 0.0195) and Spearman's Rank = 0.6636 (p-value = 0.026)
    - For patient/encounter level validity testing that stood for patient/encounter level reliability testing the developer reported the following results. This information is stated in further detail in the validity testing section.
        - The developer presented patient/encounter level validity testing from NQF #2528 for their reliability testing at the patient/encounter level.
            - The developer validated the encounter data by comparing claims data against dental charts. They noted that record and procedure codes on the claims had a 100 percent inter-rater agreement rate.
            - Additionally, the developer assessed whether the preventive service of topical fluoride application was accurately captured by claims data. The developer reported concordance (89.9 percent), sensitivity (90.7 percent), specificity (88.4 percent), positive predictive value (93.5 percent), and negative predictive value (83.9 percent). The developer also reported a kappa statistic which was 0.782.
            - The developer conducted an analysis on the CPT code 99188 which was not included in NQF #2528 to ensure that it was valid. They found that there was nothing in the data to suggest that the code was being used improperly.
8. 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.

    ☒ **Yes** ☐ **No** ☐ **Not applicable**
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

    ☒ **Yes** ☐ **No** ☐ **Not applicable** (patient/encounter level testing was not performed)
10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

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

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

☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - While the ICC values are near perfect, the rank correlations suggest a moderate association. Additionally, though the developers provided a rationale for why program level data transfers to plan level data, that is ultimately up to the committee to determine appropriateness.

VALIDITY: TESTING

12. **Validity testing level (check all that apply):**

    ☐ **Accountable-Entity Level** ☒ **Patient or Encounter-Level** ☐ **Both**

13. **If patient/encounter level validity testing was provided, was the method described and appropriate for assessing the accuracy of ALL critical data elements? NOTE:** Data element validation from the literature is acceptable.

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (patient/encounter level testing was not performed)

14. **Method of establishing validity at the accountable-entity level:**

    ☐ **Face validity**

    ☐ **Empirical validity testing at the accountable-entity level**

    ☒ **N/A (accountable-entity level testing not conducted)**

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

    ☐ **Yes**

    ☐ **No**

    ☒ **Not applicable** (accountable-entity level testing was not performed)

16. **Assess the method(s) for establishing validity**
    - The developers presented the following patient/encounter level validity testing.
        - They presented data element validation from NQF #2528 where they used a random sample of 414 encounters for members ages three-18 with at least one outpatient dental visit. The records were then reviewed by two coders.
            - The first part of the validation was “encounter data validation” where the first three procedure codes were reviewed for each claim and evaluated on if the claims data was supported by the dental record.
            - The second part of the validation was assessing whether the preventive service of topical fluoride application is accurately captured by claims data. The claims data were validated against dental records by comparing the records to the codes in the claims data.

- They then calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and a kappa statistic.

- Additionally, the developers presented validation information for CPT code 99188 and CDT codes D1206/D1208 because at the time of data element validation for NQF #2528, CPT code 99188 did not exist. The developer noted that due to the very specific description of CPT code 99188 and the guidance from the American Academy of Pediatrics on coding there is no reason to believe the code is not valid. However, to verify this, they analyzed the claims data to determine if there was any reason to believe that the code was not being used properly.

17. **Assess the results(s) for establishing validity**

- For patient/encounter level validity testing the developer reported the following results
  - For the analysis of CPT code 99188 and CDT codes D1206/D1208
    - The developer found that of the 14 state Medicaid programs, CPT code 99188 was not used in four states. However, those states used CDT codes D1206 and D1208. The developer also found that two states use both CPT code 99188 and CDT code D1206. The other eight states used CPT code 99188.
    - The developer also analyzed CPT 99188 code usage by provider type. The developer found that in all states except for one the provider that accounted for the highest percentage of CPT 99188 usage was "Pediatrics Physicians".
    - Additionally, the developer analyzed CPT 99188 codes usage in states that reimburse for using this code in relation to age distribution. The developer found that the patterns were as expected based on the reimbursable age ranges in each state.
    - The developer stated that their analyses of CPT code 99188 were as expected with the expected provider types rendering CPT 99188 services and services being concentrated within the age ranges eligible for reimbursement.
  - For critical data element validation
    - Inter-rater agreement: 100 percent
    - Concordance: 89.91 percent
    - Prevalence: 0.647
    - Sensitivity: 0.907 with a 95 percent CI of (0.857-0.942)
    - Specificity: 0.884 with a 95 percent CI of (0.888-0.963)
    - PPV: 0.935 with a 95 percent CI of (0.888-0.963)
    - NPV: 0.839 with a 95 percent CI of (0.757-0.898)
    - Kappa: 0.782 with a 95 percent CI of (0.710-0.853)

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

18. **Please describe any concerns you have with measure exclusions.**

- This measure has no exclusions.

19. **Risk Adjustment**

19a. **Risk-adjustment method**

☒ None (only answer Question 20b and 20e) ☐ Statistical model ☐ Stratification

☐ Other method assessing risk factors (please specify)

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☐ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☐ Not applicable

19c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

19d.**Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

19d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

19d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

19e**. Assess the risk-adjustment approach**

- N/A

20. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    For cost/resource use measures, does this measure identify meaningful differences about cost and resource use between the measured entities?

    - The developer calculated the interquartile range for the measure rates and then conducted a chi-square test to evaluate the statistical significance of the differences in the measure scores between the lowest and highest quartiles. The following results were reported:
        - 2018 – Mean = 0.0079, interquartile range = 0.0037, p-value = less than 0.0001
        - 2017 – Mean = 0.0082, interquartile range = 0.0047, p-value = less than 0.0001
        - 2016 – Mean = 0.0073, interquartile range = 0.0028, p-value = less than 0.0001
    - The developer states that these results suggest that the measure can identify statistically and clinically meaningful differences in performance.

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

    - No concerns because only one data source is used.

22. **Please describe any concerns you have regarding missing data.**

    - No concerns with missing data

For cost/resource use measures ONLY:

If not cost/resource use measure, please skip to question 25.

23. **Are the specifications in alignment with the stated measure intent?**

    ☐ Yes ☐ Somewhat ☐ No (If "Somewhat" or "No", please explain)

24. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

25. **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 accountable-entity level testing has been conducted)

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

☐ **Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the accountable-entity level and the patient/encounter level is required; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**
    - The results suggest moderate validity.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**
    - N/A

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**
    - N/A

Criteria 1: Importance to Measure and Report

1a. Evidence

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

---

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

2021 Submission:

Updated evidence information here.

2018 Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

[Response Begins]

Topical Fluoride for Children, Oral Health Services, indicates the percentage of children who received at least two topical fluoride applications as oral health services during the reporting year. Evidence suggests that topical fluoride applied to children starting as early as six months of age is beneficial in preventing dental caries (USPTF 2021, Weyant et al. 2013; Marinho et al. 2013). This measure directly reflects the findings of systematic reviews regarding an effective caries prevention measure (professionally applied topical fluoride), including the frequency required for clinical effectiveness (at least every three-six months). As described in 1b.01 (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, including pain, tooth loss, impaired quality of life, and negative effects on school performance. As detailed below, professionally applied topical fluoride has demonstrated effectiveness in reducing caries among children, thereby improving oral health, overall health, and overall well-being.

[Response Ends]

1a.02. Select the type of source for the systematic review of the body of evidence that supports the performance measure.

A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data.

[Response Begins]

Clinical Practice Guideline recommendation (with evidence review)

US Preventive Services Task Force Recommendation

Other systematic review and grading of the body of evidence (e.g., Cochrane Collaboration, AHRQ Evidence Practice Center)

[Response Ends]

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking "Add" after the final question in the group.

Evidence - Systematic Reviews Table (Repeatable)

Group 1 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

Fluoride Varnishes for Preventing Dental Caries in Children and Adolescents (Review). Marinho VCC, Worthington HV, Walsh T, Clarkson JE. 2013. Cochrane Database of Systematic Reviews 2013, Issue 7. Art. No.: CD002279. DOI: 10.1002/14651858.CD002279.pub2. PMID: 23846772. Available at: https://www.cochrane.org/CD002279/ORAL_fluoride-varnishes-for-preventing-dental-caries-in-children-and-adolescents.

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

Conclusions (abstract):

"The conclusions of this updated review remain the same as those when it was first published. The review suggests a substantial caries- inhibiting effect of fluoride varnish in both permanent and primary teeth, however the quality of the evidence was assessed as moderate, as it included mainly high risk of bias studies, with considerable heterogeneity."

Implications for Practice (specific section in conclusion section of review):

"This review has found that the application of fluoride varnishes two to four times a year, either in the permanent or primary dentition, is associated with a substantial reduction in caries increment. We found that this relative effect applies in populations with different levels of caries risk and exposure to other sources of fluoride. We also found no evidence that this relative effect was dependent on frequency of varnish application, length of follow-up, whether prophylaxis was undertaken prior to application of the varnish, concentration of fluoride in the varnish and use of a placebo rather than a no treatment control, although these results should be interpreted with

caution. The review does not provide any information on the likelihood of side effects with this treatment and inconclusive information on acceptability."

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]

The authors rate the quality of the evidence as **moderate**, which is defined as: "Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate."

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

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

Description of Cochrane's evidence grading system

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

This is not a clinical guideline so there are no formal recommendations with assigned grades. As noted above, the authors offer their perspective on "implications for practice."

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

Not applicable.

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

Number of studies:

22 trials published between 1975-2012 in which a total of 12,455 children were randomized to treatment with either fluoride varnish or placebo/no treatment.

Type of studies:

- Included: Randomised or quasi-randomised controlled trials using or indicating blind outcome assessment, in which fluoride varnish is compared concurrently to a placebo or no treatment group during at least one year.
- Excluded: Randomised or quasi-randomised controlled trials using within-group paired comparison designs (e.g. split-mouth trials), or with open outcome assessment or no indication of blind outcome assessment, or lasting less than one year, or controlled trials where random or quasi-random allocation was not used or indicated.

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

Measure of treatment effect: Prevented fraction (PF) was the measure of treatment effect presented for caries increment. The prevented fraction is calculated as the mean increment in the control group minus the mean increment in the intervention group divided by the mean increment in the control group. For an outcome such as caries increment (where discrete counts are considered to approximate to a continuous scale and are treated as continuous outcome), this measure was considered more appropriate than the mean difference or standardized mean difference since it allowed combination of different ways of measuring caries increment and a meaningful investigation of heterogeneity between trials. It is also simple to interpret.

Overall findings:

- The evidence from meta-analysis of the 13 trials assessing the effect of fluoride varnish on the permanent dentition is that the use of fluoride varnish is associated on average with a 43% (95% CI 30% to 57%) reduction in decayed, missing and filled tooth surfaces.
- The meta-analysis of the 10 trials assessing the effect of fluoride varnish on the primary dentition suggests a 37% (95% CI 24% to 51%) reduction in decayed, missing and filled tooth surfaces.
- There was considerable statistical heterogeneity in both these estimates.

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

The authors collected adverse effects information from the included trials and noted the following:

- There was little information concerning possible adverse effects or acceptability of treatment.
- Three studies provided data, reporting no adverse effects.

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

The recently released updated USPSTF guidelines included in Group 2 Evidence include more recent studies.

[Response Ends]

Group 2 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

Full Report: Screening and Interventions to Prevent Dental Caries in Children Younger Than 5 Years, Final Recommendation Statement. US Preventive Services Task Force. December 7, 2021. *Available at:* https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/prevention-of-dental-caries-in-children-younger-than-age-5-years-screening-and-interventions1#bootstrap-panel--12

> Recommendation Statement: Screening and Interventions to Prevent Dental Caries in Children Younger Than 5 Years: US Preventive Services Task Force Recommendation Statement. U.S. Preventive Services Task Force. 2021. *JAMA.* 2021;326(21):2172–2178. doi:10.1001/jama.2021.20007. PMID: 34874412. Available at: https://jamanetwork.com/journals/jama/fullarticle/2786823
>
> Screening and Interventions to Prevent Dental Caries in Children Younger Than 5 Years: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. Chou R, Pappas M, Dana T, Selph S, Hart E, Fu RF, Schwarz E. 2021. JAMA. 2021 Dec 7;326(21):2179-2192. doi: 10.1001/jama.2021.15658. PMID: 34874413. Available at: https://jamanetwork.com/journals/jama/fullarticle/2786824

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

Recommendation:

> "The USPSTF recommends that primary care clinicians apply fluoride varnish to the primary teeth of all infants and children starting at the age of primary tooth eruption."

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]

The USPTF Recommendations specifically addressed the following questions:

1. How Effective Are Preventive Interventions (Dietary Fluoride Supplementation, Topical Fluoride Application, Silver Diamine Fluoride, or Xylitol) in Preventing Dental Caries in Children Younger Than Age 5 Years?
2. What Are the Harms of Specific Oral Health Interventions to Prevent Dental Caries in Children Younger Than Age 5 Years (Parental or Caregiver/Guardian Oral Health Education, Referral to a Dental Health Care Professional, and Preventive Interventions)?

   We focus our reporting from this review on the studies and findings related to topical fluoride application specifically. These findings were reported separately within the review.

   Evidence statement related to topical fluoride application: "The USPSTF concludes **with moderate certainty** that there is a **moderate net benefit** of preventing future dental caries with fluoride varnish application in all children younger than 5 years."

Moderate	The available evidence is sufficient to determine the effects of the preventive service on health outcomes, but confidence in the estimate is constrained by such factors as: • The number, size, or quality of individual studies. • Inconsistency of findings across individual studies. • Limited generalizability of findings to routine primary care practice. • Lack of coherence in the chain of evidence. As more information becomes available, the magnitude or direction of the observed effect could change, and this change may be large enough to alter the conclusion.

Table detailing the USPSTF definition of moderate

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

Level of Certainty	Description
High	The available evidence usually includes consistent results from well-designed, well-conducted studies in representative primary care populations. These studies assess the effects of the preventive service on health outcomes. This conclusion is therefore unlikely to be strongly affected by the results of future studies.
Moderate	The available evidence is sufficient to determine the effects of the preventive service on health outcomes, but confidence in the estimate is constrained by such factors as: • The number, size, or quality of individual studies. • Inconsistency of findings across individual studies. • Limited generalizability of findings to routine primary care practice. • Lack of coherence in the chain of evidence. As more information becomes available, the magnitude or direction of the observed effect could change, and this change may be large enough to alter the conclusion.

Level of Certainty	Description
Low	The available evidence is insufficient to assess effects on health outcomes. Evidence is insufficient because of: • The limited number or size of studies. • Important flaws in study design or methods. • Inconsistency of findings across individual studies. • Gaps in the chain of evidence. • Findings not generalizable to routine primary care practice. • Lack of information on important health outcomes. More information may allow estimation of effects on health outcomes.

Table describing the USPSTF evidence grades.

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

Population: Children younger than 5 years

Recommendation: The USPSTF recommends that primary care clinicians apply fluoride varnish to the primary teeth of all infants and children starting at the age of primary tooth eruption.

Grade: B

Grade Definition: The USPSTF recommends this service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. Offer or provide this service.

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

Grade	Definition	Suggestions for Practice
A	The USPSTF recommends the service. There is high certainty that the net benefit is substantial.	Offer or provide this service.
B	The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.	Offer or provide this service.

Grade	Definition	Suggestions for Practice
C	The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small.	Offer or provide this service for selected patients depending on individual circumstances.
D	The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits.	Discourage the use of this service.
I Statement	The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.	Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.

Table describing the USPSTF evidence grades.

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

Number of studies:

32 studies and 1 systematic review (19 studies) were included in the entire review (covering all key questions). Comparisons were against placebo or no intervention. Outcomes were dental caries (incidence or caries burden, measured based on the number of decayed, missing, or filled teeth [dmft] or decayed, missing, or filled surfaces), morbidity, quality of life, and harms (including fluorosis).

- 22 studies (representing 25 articles) addressed key question "How effective are preventive interventions (dietary fluoride supplementation, topical fluoride application, silver diamine fluoride, or xylitol) in preventing dental caries in children younger than 5 years?" **15 of these studies addressed topical fluoride application specifically.**
- 4 (representing 6 articles) studies reported on adverse events associated with topical fluoride.

Type of studies:

- 15 studies on topical fluoride: RCTs (n=9,541); inconsistent (high statistical heterogeneity), precise, moderate strength of evidence
- 4 studies reporting on adverse events: RCTs (n=4,141); consistency cannot be determined (single trials reported different adverse events), precise; low-moderate strength of evidence

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

Based on 15 trials (5 trials in the prior USPSTF review and 10 new trials) topical fluoride (administered as fluoride varnish in all trials except for one) versus placebo or no varnish was associated with:

- decreased caries increment (13 trials, N=5733, mean difference -0.94, 95% CI, -1.74 to -0.34, I2=86%)
- decreased likelihood of incident caries (12 trials, N=8177, RR 0.80, 95% CI, 0.66 to 0.95, I2=79%; absolute risk difference [ARD] -7%, 95% CI, -12% to - 2%)

  The authors identified high statistical heterogeneity across studies, but precise estimates.

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

- Four trials (n=4,141) reported no differences between fluoride varnish versus placebo or no varnish in risk of fluorosis or the likelihood of any adverse event.
- Two studies (n=2,864) reported children did not like the smell of the fluoride varnish and one study reported that a few children vomited due to the smell, texture, or taste.
- These findings led the USPSTF to conclude: "There is adequate evidence to bound the harms for dietary fluoride supplementation and topical fluoride application as no greater than small, based on limited evidence of harms."

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

This recommendation from the USPSTF was released in December 2021.

[Response Ends]

Group 3 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

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

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

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

"Clinical recommendations for use of professionally applied or prescription-strength, home-use topical fluorides for caries prevention in patients at elevated risk of developing caries." The recommendations are "in favor" for:

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

  (Weyant et al., 2013, Table 4)

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]

The recommendations are based on the following evidence statements that were graded as **moderate** level of certainty:

- < 6 years: There is a benefit of 2.26 percent fluoride varnish application at least twice per year for caries prevention.
- 6-18 years: There is a benefit of 2.26 percent fluoride varnish application at least twice per year for caries prevention.
- 6-18 years: There is a benefit of APF gel (1.23 percent fluoride) application up to every three months for four minutes for caries.

  † No studies were found regarding professionally applied fluoride APF gels with an application time of less than three minutes.

Moderate: "This statement is based on preliminary determination from the current best available evidence; as more information becomes available, the magnitude or direction of the observed effect could change, and this change could be large enough to alter the conclusion."

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

High: This statement is strongly established by the best available evidence; the conclusion is unlikely to be affected strongly by the results of future studies. The body of evidence usually includes consistent results from

well-designed, well-conducted studies in representative populations. This conclusion is unlikely to be strongly affected by the results of future studies.

> **Moderate:** This statement is based on preliminary determination from the current best available evidence, but confidence in the estimate is constrained by one or more factors, such as: the number, size, or risk of bias of individual studies; inconsistency of findings across individual studies; limited applicability due to the populations of interest; or lack of coherence in the chain of evidence. As more information becomes available, the magnitude or direction of the observed effect could change, and this change could be large enough to alter the conclusion."
>
> **Low**: The available evidence is insufficient to support the statement, or the statement is based on extrapolation from the best available evidence. Evidence is insufficient or the reliability of estimated effects is limited by factors such as: the limited number or size of studies; important flaws in study design or methods leading to high risk of bias; inconsistency of findings across individual studies; gaps in the chain of evidence; findings not applicable to the populations of interest; or a lack of information on important health outcomes. More information could allow a reliable estimation of effects on health outcomes. (Weyant 2013, full report, pp. 18-19)
>
> The grading system was adapted from that used by the U.S. Preventive Services Task Force

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

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

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

(Weyant et al., 2013, p. 1282, Table 4)

Grade: The grade for the recommendations is **"in favor"** which is **defined** as: "Evidence favors providing this intervention." This is the **second highest recommendation out of a six-point scale**. The grading system was adapted from that used by the U.S. Preventive Services Task Force. (Weyant et al. 2013, p. 1281, Table 3)

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

Strong: Evidence strongly supports providing this intervention.

> **In Favor:** Evidence favors providing this intervention.
>
> **Weak:** Evidence suggests implementing this intervention after alternatives have been considered.
>
> **Expert Opinion For:**† Evidence is lacking; the level of certainty is low. Expert opinion guides this recommendation

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

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

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

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

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

- 17 randomized and five non-randomized clinical trials evaluated 2.26% fluoride varnish with control groups of no treatment, oral health counseling or placebo varnish. (6 randomized, 2 non-randomized concerned primary dentition; 11 randomized, 2 non-randomized concerned permanent; 1 combined results) The varnish was professionally applied every 3 to 12 months, with the majority of studies applying varnish every 6 months.
- 11 randomized and 4 non-randomized clinical trials evaluated 1.23% APF gel with comparison groups of no treatment, placebo, prophylaxis or non-fluoride placebo gel. All studies except one were on permanent teeth. All studies applied fluoride gel for four minutes.

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

Recommendations:

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

Estimates of Benefit in Support of Recommendations:

(1) 2.26% Fluoride Varnish

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

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

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

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

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

(b) Permanent teeth (children):

- Level of certainty: Moderate
- Benefit: Yes (smaller caries increment or incidence with topical fluoride use).
    - Standardized mean difference=-0.38 [-0.53, -0.24]
    - Prevented fraction=0.36
    - Number needed to treat for control rate of 1 DMFS per year = 3
- Adverse events or harms: None if used as manufacturers recommend
- Benefit-harm assessment (Net benefit rating): Benefits outweigh potential harms
- Strength of clinical recommendation: In favor

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

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

Outcome Measures	**Number and type* of studies**	**Number of participants****	**Standardized Mean Difference [95% Confidence Interval]** (negative favors intervention, positive favors control)
Meta-analysis results: Primary teeth	*	*	*
d(e/m)fs, increment or incidence+	6 RCT and 2 CCT	3, 409**	**-0.19 [-0.31, -0.08]**
Meta-analysis results: Permanent teeth	*	*	*
D(M)FS, increment or incidence+	8 RCT and 1 CCT	2, 574	**-0.38 [-0.53, -0.24]**
Root caries, meta-analysis results	*	*	*

Outcome Measures	**Number and type* of studies**	**Number of participants****	**Standardized Mean Difference [95% Confidence Interval]** (negative favors intervention, positive favors control)
Root caries increment	2 RCT	132	**-0.67 [-1.14, -0.20]**
Individual study results	*	*	*
Combined dentition	1 CCT	390	**DMFS + dmfs: -1.47 [-1.70, -1.25]** **DMFT + dmft: -1.15 [-1.37, -0.94]**
DMFT	1 CCT	77	**-0.13 [-0.58, 0.32]**
DS occlusal surfaces	1 RCT	79	**-0.54 [-1.06, -0.03]**

Table summarizing the standard mean differences from meta-analyses and individual studies for 2.26% fluoride varnish studies.

*Cell left intentionally blank

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

(2) 1.23% fluoride (APF) gel

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

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

Permanent teeth (children):

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

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

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

> "There is less of a concern with professionally-applied topical fluorides that have much longer intervals between applications [citing Wong et al. 2010]. Additionally, fluoride varnish has less potential for harms than other forms of high concentration topical fluoride because the amount of fluoride that is placed in the mouth with fluoride varnish is approximately one-tenth that of other professionally-applied products [citing Beltran-Aguilar et al. 2000]. The panel judged that the benefits outweighed the potential for harms for all professionally-applied or prescription-strength topical fluorides and age groups except for children under age 6, where the risk of swallowing and associated events (particularly nausea and vomiting) outweighed the potential benefits for all professionally-applied or prescription-strength topical fluorides except 2.26% fluoride varnish." (Weyant et al., 2013, p. 10)

Citations

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

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

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

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

The recently released updated USPSTF guidelines included in Group 2 Evidence include more recent studies.

[Response Ends]

1a.13. If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

[Response Begins]

[Response Ends]

1a.14. Briefly synthesize the evidence that supports the measure.

[Response Begins]

[Response Ends]

1a.15. Detail the process used to identify the evidence.

[Response Begins]

[Response Ends]

1a.16. Provide the citation(s) for the evidence.

[Response Begins]

[Response Ends]

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

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

Dental caries remains one of the most common, yet preventable, diseases of childhood. Dental decay in children has significant short- and long-term adverse consequences on children's health and overall well-being. Updated national surveillance data for the period 2011-2016 indicates that 23% of children aged 2-5 years experience dental caries related lesions, increasing to 52% among children aged 6- 8 years. Untreated decay was 10% among children aged 2-5 years and 16% among children aged 6-8 years. On permanent teeth, the prevalence of caries related tooth lesions was 17% for children 6-11 years and 57% among adolescents aged 12-19 years. Low-income children are twice as likely to have dental decay as higher-income children. Untreated decay was 5% among children aged 6-11 years and 17% among adolescents aged 12-19 years. (Centers for Disease Control and Prevention, 2019)

Oral health disparities are well documented and persist. Poor and near-poor children (children living in households with <=200% of the federal poverty level) are approximately twice as likely to have dental caries and untreated decay compared to higher income children (>200% FPL). National surveillance data indicate that Mexican American and non-Hispanic black children are more likely to have dental caries and untreated decay than non-Hispanic white children. For example, the prevalence of dental caries related lesions in primary teeth among Mexican American children aged 6-8 years was 73% compared to 44% for non-Hispanic white children. (Centers for Disease Control and Prevention, 2019)

Although dental caries can be managed and caries-related lesions can be treated and restored, it is important to prevent the disease process from developing in the first place. As noted in the evidence section, multiple systematic reviews with meta-analyses find evidence supporting professionally applied topical fluoride, starting as early as six months of age and applied at least twice per year, as beneficial in preventing dental caries and associated decay (USPSTF 2021, Marinho 2013, Weyant 2013).

The proposed measure, Topical Fluoride for Children, Oral Health Services, captures whether children received at least two topical fluoride applications as an oral health service, by a physician or health care provider other than a dentist nor under supervision of a dentist.. This measure directly reflects evidence-based guidelines regarding an effective caries prevention measure (professionally applied topical fluoride in medical care setting), including the frequency required for clinical effectiveness (at least every six months). This is an

important process of care measure that enables identification of topical fluoride receipt among children without a dental home. Consequently this stand-alone measure also allows plans and programs to link children to a regular source of dental care as a quality improvement strategy.

References

Centers for Disease Control and Prevention. Oral Health Surveillance Report: Trends in Dental Caries and Sealants, Tooth Retention, and Edentulism, United States, 1999–2004 to 2011–2016. Atlanta, GA: Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2019. Available at: https://www.cdc.gov/oralhealth/publications/OHSR-2019-index.html

Marinho VCC, Worthington HV, Walsh T, Clarkson JE. 2013. Fluoride Varnishes for Preventing Dental Caries in Children and Adolescents (Review). Cochrane Database of Systematic Reviews 2013, Issue 7. Art. No.: CD002279. DOI: 10.1002/14651858.CD002279.pub2. PMID: 23846772. Available at: https://www.cochrane.org/CD002279/ORAL_fluoride-varnishes-for-preventing-dental-caries-in-children-and-adolescents.

US Preventive Services Task Force. Screening and Interventions to Prevent Dental Caries in Children Younger Than 5 Years, Final Recommendation Statement. December 7, 2021. *Available at:* https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/prevention-of-dental-caries-in-children-younger-than-age-5-years-screening-and-interventions1#bootstrap-panel--12

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

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and 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 (4b) under Usability and Use.

[Response Begins]

Data Source Description

We used Medicaid enrollment and claims data contained within the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAFs) available from the Centers for Medicare & Medicaid Services (https://www.medicaid.gov/medicaid/data-systems/macbis/transformed-medicaid-statistical-information-system-t-msis/index.html).

Dates: Calendar years 2016, 2017, 2018

Number of Measured Entities: Data from 14 state Medicaid programs were included for this submission: Alaska, Arizona, Delaware, Idaho, Michigan, Mississippi, Nevada, New Mexico, North Carolina, Oklahoma, Oregon, South Carolina, Washington, and Wyoming. These states were selected based both on the quality of their data submissions to CMS and because they represent diversity in geographic location, population size, population demographic characteristics, and Medicaid dental delivery system.

Number of Patients, all measured entities included in testing, by year:

Total enrollees 0-20 years across all 14 programs:

2018: 7,720,412

2017: 7,854,440

2016: 7,850,885

In all cases, statewide program data are used (i.e., location is statewide).

Number of Patients, 0-20 Years Enrolled at Least 1 Month, by State Medicaid Program and Year:

Medicaid Program	Dates	# Mem	Dental Delivery
Alaska	CY 2018	101,273	FFS
*	CY 2017	99,296	FFS
*	CY 2016	94,550	FFS
Arizona	CY 2018	974,161	Managed care carve in
*	CY 2017	994,391	Managed care carve in
*	CY 2016	981,695	Managed care carve in
Delaware	CY 2018	118,646	FFS
*	CY 2017	118,295	FFS
*	CY 2016	120,348	FFS
Idaho	CY 2018	214,879	Dental only PAHP
*	CY 2017	220,084	Dental only PAHP
*	CY 2016	201,253	Dental only PAHP
Michigan	CY 2018	1,163,658	Dental only PAHP
*	CY 2017	1,182,388	Dental only PAHP
*	CY 2016	1,182,388	Dental only PAHP
Mississippi	CY 2018	444,432	Managed care carve in
*	CY 2017	456,123	Managed care carve in
*	CY 2016	492,813	Managed care carve in
Nevada	CY 2018	379,289	Dental only PAHP & FFS
*	CY 2017	378,460	FFS
*	CY 2016	370,394	Managed care carve in & FFS
New Mexico	CY 2018	376,379	Managed care carve in
*	CY 2017	387,255	Managed care carve in
*	CY 2016	383,056	Managed care carve in
North Carolina	CY 2018	1,231,829	FFS
*	CY 2017	1,259,699	FFS
*	CY 2016	1,241,882	FFS
Oklahoma	CY 2018	531,222	PCCM/FFS

Medicaid Program	Dates	# Mem	Dental Delivery
*	CY 2017	553,905	PCCM/FFS
*	CY 2016	557,138	PCCM/FFS
Oregon	CY 2018	435,074	Dental only PAHP
*	CY 2017	463,301	Dental only PAHP
*	CY 2016	479,469	Dental only PAHP
South Carolina	CY 2018	767,719	FFS
*	CY 2017	762,747	FFS
*	CY 2016	752,206	FFS
Washington	CY 2018	932,270	FFS
*	CY 2017	945,583	FFS
*	CY 2016	939,142	FFS
Wyoming	CY 2018	49,581	FFS
*	CY 2017	52,127	FFS
*	CY 2016	54,551	FFS

Table showing program enrollment and dental delivery system type for 14 state Medicaid programs in each year 2016 through 2018.

Performance Scores

(1) Performance scores, overall summary for all included state Medicaid programs

*	2018 (n=14)	2017 (n=14)	2016 (n=12)
Mean	0.0079	0.0082	0.0073
SD	0.0086	0.0085	0.0092
Minimum	0.0016	0.0017	0.0023
10th Percentile	0.0029	0.0020	0.0023
25th Percentile	0.0037	0.0033	0.0038
Median	0.0057	0.0063	0.0046
75th Percentile	0.0074	0.0080	0.0066
90th Percentile	0.0138	0.0133	0.0068
Maximum	0.0360	0.0353	0.0348

Table summarizing descriptive statistics for performance scores of 14 state Medicaid programs.

*Cell left intentionally blank

(2) Performance Scores with 95% CIs by State Medicaid Program and Year:

CY 2018 [lowest to highest performing]

Program	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
DE, 2018	0.16%	(	0.0001	,	0.0013	,	0.0019	)
OK, 2018	0.29%	(	0.0001	,	0.0027	,	0.0031	)
WA, 2018	0.32%	(	0.0001	,	0.0031	,	0.0034	)
ID, 2018	0.37%	(	0.0002	,	0.0034	,	0.0041	)
NV, 2018	0.38%	(	0.0001	,	0.0035	,	0.0040	)
OR, 2018	0.53%	(	0.0001	,	0.0050	,	0.0055	)
WY, 2018	0.55%	(	0.0005	,	0.0046	,	0.0064	)
AZ, 2018	0.59%	(	0.0001	,	0.0057	,	0.0060	)
MI, 2018	0.61%	(	0.0001	,	0.0059	,	0.0062	)
MS, 2018	0.71%	(	0.0002	,	0.0068	,	0.0074	)
NM, 2018	0.74%	(	0.0002	,	0.0071	,	0.0078	)
AK, 2018	0.90%	(	0.0004	,	0.0083	,	0.0097	)
SC, 2018	1.38%	(	0.0002	,	0.0135	,	0.0141	)
NC, 2018	3.60%	(	0.0002	,	0.0356	,	0.0364	)

Table showing performance scores for each program in coverage year 2018, with standard deviations and 95% confidence intervals

*Cell left intentionally blank

CY 2017 [lowest to highest performing]

Program	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
DE, 2017	0.17%	(	0.0001	,	0.0014	,	0.0020	)
ID, 2017	0.20%	(	0.0001	,	0.0018	,	0.0022	)
OK, 2017	0.25%	(	0.0001	,	0.0023	,	0.0027	)
WA, 2017	0.33%	(	0.0001	,	0.0031	,	0.0034	)
OR, 2017	0.50%	(	0.0001	,	0.0047	,	0.0053	)
AZ, 2017	0.50%	(	0.0001	,	0.0048	,	0.0052	)
MS, 2017	0.63%	(	0.0001	,	0.0060	,	0.0065	)
NM, 2017	0.64%	(	0.0002	,	0.0061	,	0.0067	)
MI, 2017	0.71%	(	0.0001	,	0.0069	,	0.0072	)
WY, 2017	0.74%	(	0.0005	,	0.0064	,	0.0084	)

Program	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
AK, 2017	0.80%	(	0.0003	,	0.0073	,	0.0087	)
SC, 2017	1.17%	(	0.0001	,	0.0114	,	0.0120	)
NV, 2017	1.33%	(	0.0002	,	0.0128	,	0.0138	)
NC, 2017	3.53%	(	0.0002	,	0.0350	,	0.0357	)

Table showing performance scores for each program in coverage year 2017, with standard deviation and 95% confidence intervals

*Cell left intentionally blank

CY 2016 [lowest to highest performing]

Program	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
OK, 2016	0.23%	(	0.0001	,	0.0021	,	0.0025	)
AK, 2016	0.23%	(	0.0002	,	0.0020	,	0.0027	)
WA, 2016	0.38%	(	0.0001	,	0.0037	,	0.0040	)
DE, 2016	0.40%	(	0.0002	,	0.0036	,	0.0044	)
OR, 2016	0.40%	(	0.0001	,	0.0038	,	0.0043	)
SC, 2016	0.44%	(	0.0001		0.0042		0.0046	)
MI, 2016	0.46%	(	0.0001	,	0.0045	,	0.0048	)
AZ, 2016	0.48%	(	0.0001	,	0.0046	,	0.0049	)
NM, 2016	0.59%	(	0.0001	,	0.0056	,	0.0062	)
NV, 2016	0.66%	(	0.0002	,	0.0063	,	0.0070	)
WY, 2016	0.68%	(	0.0005	,	0.0058	,	0.0077	)
NC, 2016	3.48%	(	0.0002	,	0.0344	,	0.0351	)

Table showing performance scores for each program in coverage year 2016, with standard deviation and 95% confidence intervals

*Cell left intentionally left blank

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, 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. Include citations.

[Response Begins]

In 2018, the performance on this measure ranged from 0.16% of children receiving at least two topical fluoride applications during the calendar year from oral health providers in the lowest performing state to 3.60% in the highest performing state. Consequently, there was substantial variation between the highest and lowest performing states, indicating significant variation across state Medicaid programs.

[Response Ends]

1b.04. 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.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. 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 (4b) under Usability and Use.

[Response Begins]

The same data source, measured entities, and patients were used as described in 1b.03) above.

We report stratified measure scores for CY 2018.

We have reported the measure scores by the following four demographic variables:

- age (required stratification),
- rural/urban geographic location based on the patient's zip code,
- race and ethnicity, and
- sex assigned at birth.

We do not report stratifications where data are missing > 10% for the stratification variable.

Table 1b.04-A summarizes the extent of missing data across all 50 state Medicaid programs plus the District of Columbia (n=51) as well as among our sample of 14 state Medicaid programs specifically. We note the ongoing deficiencies in race and ethnicity in state Medicaid enrollment data.

Table 1b.04-A Percentage of States Plus District of Columbia Missing Data on Age, Geographic Location, Race/Ethnicity and Sex, CY 2018

*	**missing<=10%**	**10%<missing<=20%**	**20%<missing<=50%**	**missing>50%**
All 50 States + DC (n=51)	*	*	*	*
Age	100%	0%	0%	0%
Biological Sex	100%	0%	0%	0%
Geographic Location	96%	0%	2%	2%
Race/Ethnicity	31%	20%	37%	12%
Testing Sample (n=14)	*	*	*	*
Age	100%	0%	0%	0%
Biological Sex	100%	0%	0%	0%
Geographic Location	100%	0%	0%	0%
Race/Ethnicity	50%	21%	21%	7%

Table showing percentage of state Medicaid programs missing data on age, geographic location, race/ethnicity and sex for calendar year 2018.

*Cell left intentionally blank

Table 1b.04-B below provides descriptive statistics that summarize performance scores for all 14 state Medicaid programs stratified by age, biological sex, geographic location, and race/ethnicity. On average:

- By age, performance scores were highest for the youngest age group (e.g., 4.74% for 1-2 years) and lowest for the oldest age group (e.g., 0.04% for 19-20 years).
- By geographic location, performance was somewhat lower for children living in rural areas (0.75%) compared with those living in urban areas (0.84%).
- By race and ethnicity, performance was lower for non-Hispanic white, non-Hispanic black, and Hawaiian/Pacific Islander children compared with non-Hispanic Asian, non-Hispanic AIAN, and Hispanic children.
- By biological sex, performance was similar for male children (0.81%) compared with female children (0.78%).

Table 1b.04-B Performance Scores Stratified by Age, Geographic Location, Race/Ethnicity and Sex: Descriptive Statistics, CY 2018

*	**Mean**	**SD**	**Minimum**	**10th Percentile**	**25th Percentile**	**Median**	**75th Percentile**	**90th Percentile**	**Maximum**
Overall Performance Score	0.0079	0.0086	0.0016	0.0029	0.0037	0.0057	0.0074	0.0138	0.0360
Age Group (n=14)	*	*	*	*	*	*	*	*	*
1-2 yrs	0.0474	0.0656	0.0051	0.0094	0.0191	0.0302	0.0441	0.0666	0.2674
3-5 yrs	0.0048	0.0036	0.0007	0.0012	0.0022	0.0030	0.0074	0.0100	0.0117
6-7 yrs	0.0024	0.0032	0.0000	0.0000	0.0001	0.0004	0.0064	0.0073	0.0079
8-9 yrs	0.0025	0.0033	0.0000	0.0000	0.0001	0.0002	0.0066	0.0073	0.0083
10-11 yrs	0.0023	0.0032	0.0000	0.0000	0.0000	0.0001	0.0052	0.0074	0.0079
12-14 yrs	0.0020	0.0027	0.0000	0.0000	0.0000	0.0002	0.0045	0.0059	0.0069
15-18 yrs	0.0013	0.0018	0.0000	0.0000	0.0000	0.0002	0.0029	0.0040	0.0052
19-20 yrs	0.0004	0.0006	0.0000	0.0000	0.0000	0.0000	0.0006	0.0017	0.0019
Geographic Location (n=14)	*	*	*	*	*	*	*	*	*

*	Mean	SD	Minimum	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile	Maximum
Rural	0.0075	0.0086	0.0003	0.0015	0.0024	0.0060	0.0072	0.0148	0.0343
Urban	0.0084	0.0088	0.0019	0.0031	0.0036	0.0057	0.0087	0.0136	0.0366
Race/ Ethnicity	*	*	*	*	*	*	*	*	*
White, non-Hispanic (n=7)	0.0091	0.0109	0.0007	0.0007	0.0030	0.0058	0.0120	0.0324	0.0324
Black, non-Hispanic (n=6)	0.0090	0.0115	0.0023	0.0023	0.0026	0.0038	0.0101	0.0318	0.0318
Asian, non-Hispanic (n=6)	0.0114	0.0192	0.0014	0.0014	0.0025	0.0028	0.0084	0.0502	0.0502
AIAN, non-Hispanic (n=7)	0.0114	0.0130	0.0000	0.0000	0.0027	0.0074	0.0147	0.0384	0.0384
Hawaiian/Pacific Islander (n=5)	0.0085	0.0110	0.0000	0.0000	0.0007	0.0041	0.0109	0.0266	0.0266
Multiracial, non-Hispanic (n=3)	0.0173	0.0216	0.0000	0.0000	0.0000	0.0105	0.0415	0.0415	0.0415
Hispanic, all races (n=7)	0.0113	0.0177	0.0000	0.0000	0.0011	0.0033	0.0159	0.0495	0.0495
Sex (n=14)	*	*	*	*	*	*	*	*	*
Female	0.0078	0.0085	0.0015	0.0028	0.0033	0.0055	0.0076	0.0141	0.0351
Male	0.0081	0.0088	0.0016	0.0029	0.0036	0.0059	0.0073	0.0136	0.0368

Table showing mean, median, minimum, maximum and percentiles for scores stratified by geographic location, race/ethnicity and sex for calendar year 2018.

*Cell left intentionally blank

Table 1b.04-C provides the stratified performance scores for each state Medicaid program individually. We believe it is meaningful to examine disparities for each program rather than just examining an overall summary because the disparities patterns may vary across states, and each state Medicaid program is urged to examine disparities specific to the populations served.

By age, children in the youngest age cohorts were most likely to receive at least two topical fluoride applications as oral health services compared with those in the older cohorts. In most states, measure performance was lower for children living in rural areas compared with those living in urban areas although there were some state Medicaid programs for which performance was higher among children living in rural areas or for which performance was similar between rural and urban areas. There was quite a bit of variation by race and ethnicity. The greatest consistency among those states for which race and ethnicity stratifications

could be calculated was for non-Hispanic AIAN children who were more likely to receive at least two topical fluoride applications as oral health services. Generally, the measure scores for males and females were similar.

Table 1b.04) Performance Scores Stratified by Age, Geographic Location, Race/Ethnicity and Sex, CY 2018

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
Overall Performance Score	0.90%	0.59%	0.16%	0.37%	0.62%	0.71%	0.38%
Age Group	*	*	*	*	*	*	*
1-2 yrs	3.36%	0.94%	1.24%	2.69%	5.15%	4.41%	0.51%
3-5 yrs	1.00%	0.55%	0.07%	0.18%	0.22%	0.60%	0.32%
6-7 yrs	0.64%	0.67%	0.01%	0.01%	0.01%	0.03%	0.34%
8-9 yrs	0.66%	0.68%	0.01%	0.02%	0.01%	0.02%	0.38%
10-11 yrs	0.52%	0.65%	0.00%	0.01%	0.01%	0.01%	0.41%
12-14 yrs	0.43%	0.59%	0.00%	0.02%	0.01%	0.03%	0.45%
15-18 yrs	0.20%	0.40%	0.01%	0.02%	0.00%	0.02%	0.35%
19-20 yrs	0.02%	0.19%	0.00%	0.00%	0.00%	0.00%	0.06%
Missing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Geographic Location	*	*	*	*	*	*	*
Rural	0.51%	0.19%	0.03%	0.34%	0.59%	0.60%	0.99%
Urban	1.22%	0.66%	0.19%	0.39%	0.63%	0.87%	0.34%
Missing	0.16%	2.68%	0.00%	0.00%	0.00%	0.01%	0.05%
Race/ Ethnicity	*	*	*	*	*	*	*
White, non-Hispanic	1.20%	NR	0.07%	0.37%	NR	NR	0.58%
Black, non-Hispanic	1.01%	NR	0.26%	NR	NR	NR	0.26%
Asian, non-Hispanic	0.84%	NR	0.14%	NR	NR	NR	0.29%
AIAN, non-Hispanic	0.44%	NR	0.00%	0.74%	NR	NR	1.24%
Hawaiian/Pacific Islander	1.09%	NR	0.00%	NR	NR	NR	0.07%
Multiracial, non-Hispanic	1.05%	NR	NR	NR	NR	NR	0.00%
Hispanic, all races	1.59%	NR	0.11%	0.00%	NR	NR	0.32%
Missing	5.94%	34.18%	0.01%	0.00%	91.32%	12.53%	3.19%
Sex	*	*	*	*	*	*	*

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
Female	0.91%	0.57%	0.15%	0.39%	0.60%	0.73%	0.33%
Male	0.89%	0.60%	0.16%	0.36%	0.64%	0.70%	0.42%
Missing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
*	**New Mexico**	**North Carolina**	**Oklahoma**	**Oregon**	**South Carolina**	**Washington**	**Wyoming**
Overall Performance Score	0.74%	3.60%	0.29%	0.53%	1.38%	0.32%	0.55%
Age Group	*	*	*	*	*	*	*
1-2 yrs	2.16%	26.74%	1.91%	4.01%	6.66%	2.38%	4.14%
3-5 yrs	0.74%	0.93%	0.12%	0.28%	1.17%	0.28%	0.27%
6-7 yrs	0.73%	0.00%	0.00%	0.05%	0.79%	0.08%	0.03%
8-9 yrs	0.83%	0.00%	0.00%	0.08%	0.73%	0.03%	0.00%
10-11 yrs	0.74%	0.00%	0.00%	0.06%	0.79%	0.02%	0.00%
12-14 yrs	0.52%	0.00%	0.00%	0.09%	0.69%	0.02%	0.00%
15-18 yrs	0.29%	0.00%	0.00%	0.06%	0.52%	0.01%	0.00%
19-20 yrs	0.06%	0.00%	0.00%	0.04%	0.17%	0.00%	0.00%
Missing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Geographic Location	*	*	*	*	*	*	*
Rural	0.72%	3.43%	0.24%	0.62%	1.48%	0.15%	0.61%
Urban	0.75%	3.66%	0.31%	0.51%	1.36%	0.36%	0.45%
Missing	0.05%	0.03%	0.02%	2.35%	0.53%	0.00%	0.00%
Race/ Ethnicity	*	*	*	*	*	*	*
White, non-Hispanic	0.64%	3.24%	0.30%	NR	NR	NR	NR
Black, non-Hispanic	0.50%	3.18%	0.23%	NR	NR	NR	NR
Asian, non-Hispanic	0.27%	5.02%	0.25%	NR	NR	NR	NR
AIAN, non-Hispanic	1.47%	3.84%	0.27%	NR	NR	NR	NR
Hawaiian/Pacific Islander	NR	2.66%	0.41%	NR	NR	NR	NR
Multiracial, non-Hispanic	NR	4.15%	NR	NR	NR	NR	NR
Hispanic, all races	0.59%	4.95%	0.33%	NR	NR	NR	NR
Missing	1.13%	0.35%	5.27%	21.56%	37.04%	10.58%	17.77%
Sex	*	*	*	*	*	*	*

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
Female	0.76%	3.51%	0.28%	0.51%	1.41%	0.31%	0.52%
Male	0.73%	3.68%	0.29%	0.54%	1.36%	0.34%	0.59%
Missing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table showing performance scores stratified by geographic location, race/ethnicity and sex for each of 14 state Medicaid programs for calendar year 2018.

NR: Not reportable due to missing data>10% or specific category has 0 denominator or is not reported by the state.

*Cell left intentionally blank

We also tested for significantly significant differences in stratified measure scores within each state using bivariate logistic regression. Table 1b.04-D summarizes the number of states with odds ratios that are greater than the reference category ($p<0.05$), not statistically significant different from the reference category, less than the reference category ($p<0.05$), or not reportable due to missing data for that specific category. The greatest variation in measure scores was by age, with the youngest children most likely to receive topical fluoride applications as oral health services. In most states (8 of 14), children in urban areas were more likely to receive topical fluoride applications as oral health services compared with those living in rural areas. In most states (10 of 14), statistically significant differences in performance by sex were not detected. There was a fair amount of variability between states in disparities by race and ethnicity.

Table 1b.04-D Bivariate Logistic Regression, Reporting Number of States with Statistically Significant Differences from the Reference Category for Each Stratification Variable, CY 2018

*	**Topical Fluoride, Dental or Oral Health Services**	*	*	*
*	**OR< Ref (p<0.05)**	**Not signicant**	**OR> Ref (p<0.05)**	**Not reportable**
AGE	**n=14 states**	**n=14 states**	**n=14 states**	**n=14 states**
1-2 yrs	**0**	**1**	**13**	**0**
Odds ratio range	*	1.2	(1.41 to 29749.00	*
3-5 yrs	**3**	**0**	**8**	**3**
Odds ratio range	(0.62 to 0.79)	*	(1.29 to 674.00)	*
6-7 yrs	**1**	**9**	**0**	**4**
Odds ratio range	0.79	(0.61 to 3.05)	*	*
8-9 yrs	**Ref**	**Ref**	**Ref**	**Ref**
10-11 yrs	**0**	**10**	**0**	**4**
Odds ratio range	*	(0.32 to 1.09)	*	*
12-14 yrs	**3**	**6**	**0**	**5**

*	**Topical Fluoride, Dental or Oral Health Services**	*	*	*
Odds ratio range	(0.64 to 0.88)	0.53 to 1.59)	*	*
15-18 yrs	**5**	**5**	**0**	**4**
Odds ratio range	(0.16 to 0.78)	(0.31 to 1.24)	*	*
19-20 yrs	**5**	**1**	**0**	**8**
Odds ratio range	(0.03 to 0.28)	0.59	*	*
GEOGRAPHIC LOCATION	**n=14 states**	**n=14 states**	**n=14 states**	**n=14 states**
Rural	**Ref**	**Ref**	**Ref**	**Ref**
Urban	**3**	**3**	**8**	**0**
Odds ratio range	(0.32 to 0.80)	(0.75 to 1.20)	(1.07 to 7.52)	*
BIOLOGICAL SEX	**n=14 states**	**n=14 states**	**n=14 states**	**n=14 states**
Female	**Ref**	**Ref**	**Ref**	**Ref**
Male	**0**	**10**	**4**	**0**
Odds ratio range	*	(0.93 to 1.14)	(1.05 to 1.27)	*
RACE/ETHNICITY	**n=7 states**	**n=7 states**	**n=7 states**	**n=7 states**
White, non-Hispanic	**Ref**	**Ref**	**Ref**	**Ref**
Black, non-Hispanic	**2**	**3**	**1**	**1**
Odds ratio range	(0.44 to 0.75)	(0.68 to 0.98)	4.57	*
Asian, non-Hispanic	**2**	**3**	**1**	**1**
Odds ratio range	(0.52 to 0.63)	(p.25 to 2.61)	1.55	*
AIAN, non-Hispanic	**1**	**1**	**4**	**1**
Odds ratio range	0.36	0.89	(1.21 to 2.47)	*
Hawaiian/Pacific Islander	**1**	**3**	**0**	**3**
Odds ratio range	0.06	(0.83 to 1.38)	*	*
Multiracial, non-Hispanic	**0**	**1**	**1**	**5**
Odds ratio range	*	0.86	1.28	*

*	**Topical Fluoride, Dental or Oral Health Services**	*	*	*
Hispanic, all races	**1**	**3**	**2**	**1**
Odds ratio range	0.55	(0.90 to 1.31)	(1.57 to 2.02)	*
Ref=reference category; NS=not significant; NR=not reportable due to missing data; NA - not applicable - lower bound of age range for Topical Fluoride is 1 yr	*	*	*	*

Table shows the bivariate logistic regression, reporting on the number of states with statistically significant differences from the reference category for each stratification variable in calendar year 2018.

*Cell left intentionally blank

In summary, our results reflect variations in performance between states, disparities within each state, and variations in the nature and extent of disparities between states.

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, 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 above.

[Response Begins]

Performance data are provided above.

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

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

---

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]

Prevention: Topical Fluoride for Children, Oral Health Services

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

Percentage of children aged 1 through 20 years who received at least 2 topical fluoride applications as oral health services within the reporting year.

The measure is specified for reporting at the program and plan levels for both public and private/commercial reporting.

[Response Ends]

sp.03. Provide a rationale for why this measure must be reported with other measures to appropriately interpret results.

[Response Begins]

Although this measure can be reported as a stand-alone measure, it is being grouped with two complementary measures (2528 and 3700) to enable more robust quality improvement efforts. The DQA considered submitting a single measure with three numerators (denominator population is the same). But NQF evaluation criteria state: "Measures with multiple measure components that are assessed for each patient, but that result in multiple scores for an accountable entity rather than a single score. These generally should be submitted as separate measures and indicated as paired/grouped measures." (Measure Evaluation Criteria and Guidance, September 2021, p. 52). Based on this and discussions with NQF staff, we are submitting as three distinct measures.

This measure – NQF 3701: Topical Fluoride for Children, Oral Health Services – focuses on topical fluoride delivered as an "oral health" service (by a physician or health care provider other than a dentist nor under supervision of a dentist).

Because many children, especially very young children, do not have a dental home, many state Medicaid programs and MCOs pay for the application of topical fluoride as an "oral health" service (by a physician or health care provider other than a dentist nor under supervision of a dentist). Consequently, state Medicaid programs, as well as commercial integrated medical-dental benefit MCOs or integrated medical-dental healthcare delivery sites, have a strong interest in tracking whether children receive **any** topical fluoride regardless of provider type ("dental" or "oral health" services). They also have a strong interest in understanding whether, and for whom, topical fluoride is being delivered through "dental" providers and "oral health" providers.

Measures of topical fluoride provision by provider type, in addition to a measure of overall provision, are important because multi-pronged quality improvement strategies may be used to improve rates of topical fluoride application among a population of children. Dental providers and/or medical providers may be the focus of these efforts. Without measures that track the effectiveness by provider type, it is more difficult for programs and plans to assess which efforts are most effective. In addition, the accountability and delivery systems are typically distinct. Improving fluoride application by dental providers is accomplished through the dental delivery system and related financing/reimbursement structures, whereas topical fluoride application by medical providers is accomplished through medical delivery system and related financing/reimbursement structures. Some measure users will benefit by implementing and using all three measures (such as Medicaid programs and private payers/delivery systems that include both medical and dental). Other users, focused specifically on either dental or medical care delivery, respectively, will be able to report using one of the measures: either the measure related to "dental" services or the measure related to "oral health" services.

The need for three grouped measures comes directly from user community requests. In considering a topical fluoride measure for inclusion in the Centers for Medicare and Medicaid Services' Core Set of Children's Health Care Quality Measures, a specific request was made for a single measure that included three numerators (dental or oral health services, dental services, and oral health services) because of the recognized need by Medicaid programs to track not only overall receipt of topical fluoride but also topical fluoride provided through the dental and medical delivery systems specifically. Review by the DQA's Measures Development and Maintenance Committee, which includes representation of providers, community health centers and payers, affirmed the value of reporting three numerators across public and private/commercial measure applications for the reasons described above.

Consequently, the DQA is submitting two complementary measures for endorsement to be "grouped" with this measure: (1) Topical Fluoride for Children, Dental Services (NQF#2528) and (2) Topical Fluoride for Children, Dental or Oral Health Services (NQF#3700). This grouping provides users with measurement options that appropriately support population-based assessments of quality. It enables measure users, including Medicaid programs and their contracted MCOs, integrated medical-dental MCOs, and integrated medical-dental delivery systems, to examine the overall provision of topical fluoride by provider type, which has been identified by stakeholders as integral for quality improvement and accountability purposes.

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]

Dental

Dental: Caries

[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]

Disparities Sensitive

Primary Prevention

[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Children (Age < 18)

[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Health Plan

Other

[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

[Response Begins]

Outpatient Services

[Response Ends]

sp.09. 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. If no URL is available, indicate "none available".

[Response Begins]

None available.

[Response Ends]

sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

Attachment: 3701_NQF3701_sp11_NUCC provider taxonomy codes_2022Spring.xlsx

sp.12. State the numerator.

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.

[Response Begins]

Unduplicated number of children who received at least 2 topical fluoride applications as oral health services

[Response Ends]

sp.13. Provide details needed to calculate the numerator.

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 sp.11.

[Response Begins]

Please see section sp 22.

[Response Ends]

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

Unduplicated number of children aged 1 through 20 years

[Response Ends]

sp.15. Provide details needed to calculate the denominator.

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 sp.11.

[Response Begins]

Please see section sp 22.

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

There are no measure-specific exclusions. There is a standard exclusion as part of determining denominator eligibility: Medicaid/CHIP programs should exclude those individuals who do not qualify for dental benefits.

[Response Ends]

sp.17. Provide details needed to calculate the denominator exclusions.

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 sp.11.

[Response Begins]

There are no measure-specific exclusions.

[Response Ends]

sp.18. Provide all information required to stratify the measure results, if necessary.

Include 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 in the Data Dictionary field.

[Response Begins]

This measure is stratified by age (in years) using the following categories:

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

No new data are needed for this stratification. Please see sp. 22 and attached specifications for complete measure details.

[Response Ends]

sp.19. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

[Response Begins]

No risk adjustment or risk stratification

[Response Ends]

sp.20. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

sp.21. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score

[Response Begins]

Better quality = Higher score

[Response Ends]

sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]

Topical Fluoride for Children, Oral Health Services, Measure Score Calculation

(1) Check if the subject meets age criteria at the last day of the reporting year:[1]

(a) If child is >=1 and <21,[2] then proceed to next step.

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

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

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

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

YOU NOW HAVE THE DENOMINATOR (DEN): SUBJECTS WHO MEET THE AGE AND ENROLLMENT CRITERIA

(3) Check if subject received at least two fluoride applications as **oral health services** during the reporting year – at least two unique dates of service when topical fluoride was provided. Service provided on each date of service should satisfy the following criteria:

(a) If [SERVICE CODE] = CDT D1206 or D1208 or CPT99188,[4] AND

(b) If [RENDERING PROVIDER TAXONOMY] code is a valid NUCC maintained Provider Taxonomy code but NOT included in the NUCC maintained Provider Taxonomy Codes in Table 1 below, then include in numerator;[5] proceed to next step.

(c) If both a AND b are not met, then STOP processing. This subject is already included in the denominator but will not be included in the numerator.

Note 1: Some states may use additional codes to reimburse for fluoride provided by non-dental providers.[5] These codes should be included in the [SERVICE CODE] codes in addition to CDT D1206, CDT D1208 and CPT 99188.

Note 2: No more than one fluoride application can be counted for the same member on the same date of service.

Note 3: In this step, all claims with missing or invalid SERVICE CODE or with missing or invalid NUCC maintained Provider Taxonomy Codes should be excluded.

YOU NOW HAVE NUMERATOR (NUM) COUNT: Subjects who received at least two fluoride applications as oral health services

(6) Report

(a) Unduplicated number of subjects in denominator (DEN)

(b) Unduplicated number of subjects in numerator (NUM)

(c) Measure rate (NUM/DEN)

(d) Rate stratified by age

Table 1: NUCC maintained Provider Taxonomy Codes classified as "Dental Service"+ +

Note: See Excel file attached in sp.11) for code descriptions.

122300000X	1223P0106X	1223X0008X	125Q00000X	126800000X
1223D0001X	1223P0221X	1223X0400X	261QF0400X	261QD0000X
1223D0004X	1223P0300X	124Q00000X+	261QR1300X	204E00000X
1223E0200X	1223P0700X	125J00000X	1223X2210X	261QS0112X
1223G0001X	1223S0112X	125K00000X	122400000X	*

Table showing NUCC-maintained Provider Taxonomy Codes classified as "Dental Service"

*Cell left intentionally blank

++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.

[1] **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.

[2] **Age**: Medicaid/CHIP programs use under age 21(<21) as upper bound of age range; Exchange quality reporting use under age 19 (<19) as the upper bound of the age range; other programs check with program officials. The age criteria should be reported with the measure score.

[3] **Enrollment in "same" plan vs. "any" plan**: At the state program level (e.g., Medicaid/CHIP) a criterion of "any" plan applies versus at the health plan (e.g., MCO) level a criterion of "same" plan applies. The criterion used should be reported with the measure score. While this prevents direct aggregation of results from plan to program, each entity is given due credit for the population it serves. Thus, states with multiple MCOs should not merely "add up" the plan level scores but should calculate the state score from their database to allow inclusion of individuals who may be continuously enrolled but might have switched plans in the interim.

[4] **Topical Fluoride codes**: For reporting years prior to 2013, use CDT codes D1203 or D1204 or D1206.

[5] **Services provided by medical providers:** CPT 99188 is a dedicated code for "application of topical fluoride varnish by a physician or other qualified health care professional." In some instances, additional CPT or other

codes may be used for reimbursement of oral health services (e.g., medical primary care providers providing oral evaluation, risk assessment, anticipatory guidance or fluoride varnish). Details available at AAP Table. For such states these additional codes must be considered. The AAP also provides an Oral Health Coding Fact Sheet for Primary Care Physicians: https://www.aap.org/en-us/Documents/coding_factsheet_oral_health.pdf. Accessed May 25, 2021.

[Response Ends]

sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

[Response Begins]

Not applicable.

[Response Ends]

sp.28. Select only the data sources for which the measure is specified.

[Response Begins]

Claims

[Response Ends]

sp.29. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]

Not applicable.

[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

2a. Reliability

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

• 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.

• All required sections must be completed.

• For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.

• If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.

• An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.

• Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.

• For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

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

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

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) 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).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

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

OR

• rationale/data support no risk adjustment/ stratification.

2b4. 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.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

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

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

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

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

Definitions

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).

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. The degree of consensus and any areas of disagreement must be provided/discussed.

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.

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

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

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.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:

2021 Submission:

Updated testing information here.

2018 Submission:

Testing from the previous submission here.

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]

Claims

Other (specify)

[Other (specify) Please Explain]

Enrollment Data.

[Response Ends]

2a.02. 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).

[Response Begins]

This measure, Topical Fluoride for Children –Oral Health Services - is being submitted as part of a "grouped" set of measures. The other two measures are: Topical Fluoride, Dental Services (NQF #2528) and Topical Fluoride, Dental or Oral Health Services (NQF #3700). Topical Fluoride, Dental Services (NQF #2528) has been previously endorsed by NQF.

This measure was specified for reporting at the program and plan level using administrative enrollment and claims data for children enrolled in Medicaid programs or care delivery systems that have access to both medical and dental claims.

We used Medicaid enrollment and claims data contained within the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAFs) available from the Centers for Medicare & Medicaid Services (https://www.medicaid.gov/medicaid/data-systems/macbis/transformed-medicaid-statistical-information-system-t-msis/index.html).

We selected a sample of 14 states: Alaska, Arizona, Delaware, Idaho, Michigan, Mississippi, Nevada, New Mexico, North Carolina, Oklahoma, Oregon, South Carolina, Washington, and Wyoming. These states were selected based both on the quality of their data submissions to CMS and because they represent diversity in geographic location, population size, population demographic characteristics, and Medicaid dental delivery systems.

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

Administrative Claims data

01-01-2016 through 12-31-2018

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Health Plan

Other (specify)

[Other (specify) Please Explain]

Program (e.g., Medicaid, CHIP)

[Response Ends]

2a.05. List the measured entities 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.

[Response Begins]

Measured Entities used for Testing Measure Score Reliability

14 State Medicaid Programs: Alaska, Arizona, Delaware, Idaho, Michigan, Mississippi, Nevada, New Mexico, North Carolina, Oklahoma, Oregon, South Carolina, Washington, and Wyoming

Total enrollees 0-20 years across all 14 programs:

2018: 7,720,412

2017: 7,854,440

2016: 7,850,885

In all cases, statewide program data are used (i.e., location is statewide).

Number of Beneficiaries, 0-20 Years Enrolled at Least 1 Month, by State Medicaid Program and Year:

Medicaid Program	Dates	# Mem	Dental Delivery
Alaska	CY 2018	101,273	FFS
*	CY 2017	99,296	FFS
*	CY 2016	94,550	FFS
Arizona	CY 2018	974,161	Managed care carve in
*	CY 2017	994,391	Managed care carve in
*	CY 2016	981,695	Managed care carve in

Medicaid Program	Dates	# Mem	Dental Delivery
Delaware	CY 2018	118,646	FFS
*	CY 2017	118,295	FFS
*	CY 2016	120,348	FFS
Idaho	CY 2018	214,879	Dental only PAHP
*	CY 2017	220,084	Dental only PAHP
*	CY 2016	201,253	Dental only PAHP
Michigan	CY 2018	1,163,658	Dental only PAHP
*	CY 2017	1,182,388	Dental only PAHP
*	CY 2016	1,182,388	Dental only PAHP
Mississippi	CY 2018	444,432	Managed care carve in
*	CY 2017	456,123	Managed care carve in
*	CY 2016	492,813	Managed care carve in
Nevada	CY 2018	379,289	Dental only PAHP & FFS
*	CY 2017	378,460	FFS
*	CY 2016	370,394	Managed care carve in & FFS
New Mexico	CY 2018	376,379	Managed care carve in
*	CY 2017	387,255	Managed care carve in
*	CY 2016	383,056	Managed care carve in
North Carolina	CY 2018	1,231,829	FFS
*	CY 2017	1,259,699	FFS
*	CY 2016	1,241,882	FFS
Oklahoma	CY 2018	531,222	PCCM/FFS
*	CY 2017	553,905	PCCM/FFS
*	CY 2016	557,138	PCCM/FFS
Oregon	CY 2018	435,074	Dental only PAHP
*	CY 2017	463,301	Dental only PAHP
*	CY 2016	479,469	Dental only PAHP
South Carolina	CY 2018	767,719	FFS
*	CY 2017	762,747	FFS
*	CY 2016	752,206	FFS
Washington	CY 2018	932,270	FFS
*	CY 2017	945,583	FFS
*	CY 2016	939,142	FFS

Medicaid Program	Dates	# Mem	Dental Delivery
Wyoming	CY 2018	49,581	FFS
*	CY 2017	52,127	FFS
*	CY 2016	54,551	FFS

Table showing program enrollment and dental delivery system type for 14 state Medicaid programs in each year 2016 through 2018.

*Cell left intentionally blank

Plan-Level Data. This measure is also specified for both program and plan-level reporting. Our original testing of NQF #2528 Topical Fluoride for Children, Dental Services, included plan-level data. This measure is being grouped with NQF #2528 as well as with NQF #3700. All three measures share the same denominator. This measure focuses specifically on topical fluoride applications provided as "oral health" services. Consequently, the numerator captures topical fluoride applications by non-dental providers. DQA measures, including NQF #2528, have been implemented by programs and plans operating in commercial, Marketplace, and Medicaid markets. However, our testing for this submission does not include plan-specific data. The T-MSIS data used for testing currently do not enable reliable identification of which topical fluoride services are provided by which managed care organizations (MCOs). Based on prior testing of NQF #2528 and other administrative claims-based oral healthcare measures, including measures that require both medical and dental claims, we would not expect to see marked differences in the reliability or validity of plan-level reporting compared with program-level reporting given that the data sources (administrative claims) and measure specifications are the same for the two reporting levels. The only potential concern would be if the plan level denominators were too small to yield reliable results. The denominator requirements for this measure and other oral healthcare measures capture a broad population, and we have not encountered issues with small denominator sizes in our testing or in feedback from the user community. The DQA membership includes MCO representatives that operate in state Medicaid programs throughout the United States, and insufficient denominator sizes have never arisen as a concern. The DQA is also in frequent communication with the Centers for Medicare and Medicaid Services, as well as with Mathematica which serves as CMS's technical resource to state Medicaid programs for quality measure implementation; again, there have been no issues raised related to challenges with plan-level implementation.

The T-MSIS claims data are missing the managed care plan identifier for more than 90% of topical fluoride services in the states that we examined. This does **not** represent a feasibility issue for Medicaid programs and their participating plans to calculate the measures. We know from working with state Medicaid programs and state Marketplaces that it is highly feasible to have plan level reporting of oral healthcare quality measures. Rather, this reflects a limitation of the database that we used for testing. This is a relatively new database (released for public use in September 2020) for which data completeness and quality are continually being improved. In addition, because the measure specifications were updated during the DQA 2021 Annual Measure Review, there is no public reporting yet of the revised measure.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

Patient Characteristics: Measure Score Reliability Testing

Tables 2a.06)A-C below provides the patient characteristics for children enrolled in Medicaid (with comprehensive benefits) for at least one month included in the T-MSIS analytic files for the 14 Medicaid programs for each year CY 2016 through CY 2018. In CY 2018, program enrollment ranged from 49,581 in Wyoming Medicaid to 1,231,829 in North Carolina Medicaid. Age and biological sex distributions were similar across programs. There was substantial variation in the geographic location and race/ethnicity distributions between states. Two states were excluded from testing analysis in CY2016 due to data quality issues (see 2b.08 regarding analysis of missing data).

Table 2a.06-A, State Medicaid Program Patient Characteristics, <21 Years Old, CY2018 (T-MSIS Data)

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
Total # Patients	101,273	974,161	118,646	214,879	1,163,658	444,432	379,289
Age Group	*	*	*	*	*	*	*
<1 yr	5.35%	5.08%	5.29%	5.12%	5.48%	6.02%	5.69%
1-2 yrs	12.07%	10.75%	11.09%	11.49%	11.12%	11.80%	11.95%
3-5 yrs	16.76%	15.36%	15.49%	16.87%	15.51%	16.42%	16.63%
6-7 yrs	10.23%	9.70%	10.03%	10.33%	9.88%	9.98%	10.21%
8-9 yrs	9.90%	9.78%	10.20%	10.18%	9.81%	9.93%	9.97%
10-11 yrs	9.41%	10.32%	10.28%	10.48%	9.72%	10.42%	10.09%
12-14 yrs	12.94%	14.47%	13.87%	14.51%	13.72%	14.17%	13.62%
15-18 yrs	15.79%	17.06%	16.23%	16.63%	16.52%	16.55%	15.36%
19-20 yrs	7.53%	7.49%	7.51%	4.40%	8.23%	4.71%	6.48%
Missing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Geographic Location	*	*	*	*	*	*	*
Rural	44.49%	13.36%	18.44%	33.46%	19.27%	57.16%	5.86%
Urban	55.34%	84.19%	81.56%	66.20%	80.72%	42.84%	94.07%
Missing	<1%	2.45%	<1%	<1%	<1%	<1%	<1%
Race/ Ethnicity	*	*	*	*	*	*	*
White, non-Hispanic	32.50%	44.05%	34.56%	97.48%	0.00%	29.61%	24.94%
Black, non-Hispanic	3.62%	9.15%	41.42%	<1%	0.00%	53.53%	21.42%
Asian, non-Hispanic	5.37%	1.59%	1.97%	<1%	0.00%	<1%	2.86%
AIAN, non-Hispanic	36.17%	9.22%	<1%	2.15%	0.00%	<1%	1.31%
Hawaiian/Pacific Islander	4.86%	<1%	<1%	0.00%	0.00%	<1%	1.46%
Multiracial, non-Hispanic	8.11%	0.00%	0.00%	0.00%	0.00%	0.00%	<1%
Hispanic, all races	2.94%	<1%	21.73%	<1%	8.99%	<1%	43.90%

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
non-Hispanic, race unspecified	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Missing	6.43%	34.93%	<1%	<1%	91.01%	14.75%	4.09%
Sex	*	*	*	*	*	*	*
Female	48.77%	49.19%	49.41%	48.98%	49.20%	49.68%	49.27%
Male	51.23%	50.81%	50.59%	51.02%	50.80%	50.32%	50.73%
Missing	0.00%	0.00%	<1%	0.00%	0.00%	<1%	0.00%
*	**New Mexico**	**North Carolina**	**Oklahoma**	**Oregon**	**South Carolina**	**Washington**	**Wyoming**
Total # Patients	376,379	1,231,829	531,222	435,074	767,719	932,270	49,581
Age Group	*	*	*	*	*	*	*
<1 yr	4.68%	5.80%	6.42%	5.42%	4.90%	4.79%	5.80%
1-2 yrs	10.02%	11.68%	13.00%	11.19%	10.86%	10.49%	12.37%
3-5 yrs	15.55%	16.29%	17.85%	15.13%	16.42%	15.76%	16.54%
6-7 yrs	9.91%	10.01%	10.69%	9.64%	10.30%	10.21%	10.44%
8-9 yrs	9.75%	9.73%	10.31%	9.67%	10.40%	10.27%	10.27%
10-11 yrs	10.27%	9.96%	10.43%	9.91%	10.65%	10.36%	10.50%
12-14 yrs	14.33%	13.52%	13.99%	13.60%	14.36%	14.11%	13.92%
15-18 yrs	17.24%	16.11%	13.15%	16.81%	16.62%	16.83%	15.80%
19-20 yrs	8.24%	6.90%	4.17%	8.63%	5.48%	7.19%	4.34%
Missing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Geographic Location	*	*	*	*	*	*	*
Rural	33.89%	26.30%	38.40%	25.39%	22.78%	15.50%	65.44%
Urban	66.06%	72.91%	61.58%	72.78%	76.71%	84.49%	34.56%
Missing	<1%	<1%	<1%	1.84%	<1%	<1%	<1%
Race/ Ethnicity	*	*	*	*	*	*	*
White, non-Hispanic	20.83%	37.88%	39.65%	32.07%	24.71%	40.01%	54.43%
Black, non-Hispanic	1.93%	35.24%	13.79%	2.50%	27.71%	6.91%	2.20%
Asian, non-Hispanic	<1%	1.66%	1.70%	1.46%	<1%	3.43%	<1%
AIAN, non-Hispanic	15.25%	1.29%	17.07%	1.79%	<1%	3.06%	8.92%
Hawaiian/Pacific Islander	0.00%	<1%	<1%	<1%	<1%	3.01%	<1%
Multiracial, non-Hispanic	0.00%	3.98%	0.00%	<1%	0.00%	2.18%	0.00%

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
Hispanic, all races	60.05%	19.10%	21.79%	39.72%	5.07%	29.04%	11.46%
non-Hispanic, race unspecified	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Missing	1.32%	<1%	5.54%	21.79%	41.60%	12.36%	22.42%
Sex	*	*	*	*	*	*	*
Female	49.42%	49.39%	49.59%	49.23%	49.25%	49.21%	48.91%
Male	50.58%	50.61%	50.41%	50.77%	50.74%	50.79%	51.09%
Missing	0.00%	0.00%	0.00%	<1%	<1%	0.00%	0.00%

Table showing demographic characteristics for individuals<21 years old enrolled in each of 14 state Medicaid programs in CY 2018

*Cell left intentionally blank

Table 2a06-B, State Medicaid Program Patient Characteristics, <21 Years Old, CY2017 (T-MSIS Data)

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
Total # Patients	99,296	994,391	118,295	220,084	1,163,174	456,123	378,460
Age Group	*	*	*	*	*	*	*
<1 yr	5.72%	5.16%	5.49%	5.46%	5.34%	5.99%	5.88%
1-2 yrs	12.66%	11.04%	11.33%	11.92%	11.39%	11.77%	12.14%
3-5 yrs	16.46%	15.43%	15.29%	17.02%	15.45%	16.33%	16.57%
6-7 yrs	10.29%	9.68%	10.39%	10.34%	9.95%	10.22%	10.32%
8-9 yrs	9.86%	10.27%	10.53%	10.45%	9.95%	10.43%	10.30%
10-11 yrs	9.30%	10.32%	10.02%	10.34%	9.72%	10.62%	10.16%
12-14 yrs	12.54%	14.00%	13.56%	14.02%	13.44%	13.46%	13.17%
15-18 yrs	15.74%	16.63%	15.83%	16.07%	16.54%	16.42%	14.94%
19-20 yrs	7.45%	7.46%	7.56%	4.37%	8.22%	4.77%	6.51%
Missing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Geographic Location	*	*	*	*	*	*	*
Rural	44.74%	12.99%	18.45%	33.61%	19.31%	57.00%	6.02%
Urban	55.06%	81.41%	81.55%	66.00%	80.67%	42.99%	93.91%
Missing	<1%	5.61%	<1%	<1%	<1%	<1%	<1%
Race/ Ethnicity	*	*	*	*	*	*	*
White, non-Hispanic	32.60%	44.14%	35.10%	97.51%	0.00%	30.83%	25.33%
Black, non-Hispanic	3.66%	8.92%	41.31%	<1%	0.00%	55.30%	20.85%
Asian, non-Hispanic	5.47%	1.61%	1.93%	<1%	0.00%	<1%	2.82%

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
AIAN, non-Hispanic	36.14%	9.24%	<1%	2.06%	0.00%	<1%	1.35%
Hawaiian/Pacific Islander	4.83%	<1%	<1%	0.00%	0.00%	<1%	1.23%
Multiracial, non-Hispanic	8.21%	0.00%	0.00%	0.00%	0.00%	0.00%	<1%
Hispanic, all races	2.94%	2.03%	21.35%	<1%	8.95%	<1%	44.03%
non-Hispanic, race unspecified	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Missing	6.14%	33.72%	<1%	<1%	91.05%	11.78%	4.37%
Sex	*	*	*	*	*	*	*
Female	48.92%	49.17%	49.47%	49.01%	49.19%	49.75%	49.28%
Male	51.08%	50.83%	50.53%	50.99%	50.81%	50.25%	50.72%
Missing	0.00%	0.00%	<1%	0.00%	0.00%	<1%	0.00%
*	**New Mexico**	**North Carolina**	**Oklahoma**	**Oregon**	**South Carolina**	**Washington**	**Wyoming**
Total # Patients	387,255	1,259,699	553,905	463,301	762,747	945,583	52,127
Age Group	*	*	*	*	*	*	*
<1 yr	4.74%	5.82%	6.27%	5.41%	5.07%	4.94%	5.83%
1-2 yrs	10.37%	11.67%	13.08%	11.50%	11.32%	10.74%	12.68%
3-5 yrs	15.72%	16.30%	17.68%	15.08%	16.28%	15.78%	16.68%
6-7 yrs	9.98%	10.20%	10.76%	9.75%	10.41%	10.27%	10.66%
8-9 yrs	10.26%	10.15%	10.70%	10.07%	10.82%	10.58%	10.54%
10-11 yrs	10.22%	9.90%	10.39%	9.78%	10.64%	10.23%	10.35%
12-14 yrs	13.83%	13.07%	13.62%	13.24%	13.65%	13.66%	13.46%
15-18 yrs	16.87%	16.00%	13.25%	16.68%	16.60%	16.68%	15.30%
19-20 yrs	8.00%	6.88%	4.26%	8.49%	5.21%	7.12%	4.49%
Missing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Geographic Location	*	*	*	*	*	*	*
Rural	33.71%	26.47%	38.71%	24.60%	22.87%	15.45%	65.42%
Urban	66.25%	72.71%	61.25%	71.97%	76.62%	84.55%	34.58%
Missing	<1%	<1%	<1%	3.43%	<1%	<1%	0.00%
Race/ Ethnicity	*	*	*	*	*	*	*
White, non-Hispanic	20.82%	38.41%	40.18%	33.66%	26.23%	40.37%	56.33%
Black, non-Hispanic	1.94%	35.12%	13.51%	2.46%	29.24%	6.66%	2.12%

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
Asian, non-Hispanic	<1%	1.68%	1.67%	1.46%	<1%	3.46%	<1%
AIAN, non-Hispanic	15.04%	1.28%	17.07%	1.76%	<1%	2.88%	8.63%
Hawaiian/Pacific Islander	0.00%	<1%	<1%	<1%	<1%	2.89%	<1%
Multiracial, non-Hispanic	0.00%	3.91%	0.00%	<1%	0.00%	1.78%	0.00%
Hispanic, all races	60.22%	18.69%	21.45%	36.70%	5.42%	28.38%	11.87%
non-Hispanic, race unspecified	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Missing	1.34%	<1%	5.66%	23.21%	38.17%	13.59%	20.37%
Sex	*	*	*	*	*	*	*
Female	49.37%	49.43%	49.54%	49.19%	49.33%	49.18%	48.93%
Male	50.63%	50.57%	50.46%	50.81%	50.67%	50.82%	51.07%
Missing	0.00%	0.00%	0.00%	<1%	<1%	0.00%	0.00%

Table showing demographic characteristics for individuals<21 years old enrolled in each of 14 state Medicaid programs in CY 2017

*Cell left intentionally blank

Table 2a06-C, State Medicaid Program Patient Characteristics, <21 Years Old, CY2016 (T-MSIS Data)

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
Total # Patients	94,550	981,695	120,348	201,253	1,182,388	492,813	370,394
Age Group	*	*	*	*	*	*	*
<1 yr	6.43%	5.47%	5.32%	5.27%	5.38%	5.60%	5.97%
1-2 yrs	12.42%	11.59%	11.12%	12.65%	11.33%	11.20%	12.18%
3-5 yrs	16.26%	15.02%	15.68%	17.15%	15.33%	15.61%	16.60%
6-7 yrs	10.56%	10.06%	10.63%	10.98%	10.01%	10.64%	10.70%
8-9 yrs	9.87%	10.57%	10.60%	11.23%	9.94%	11.12%	10.63%
10-11 yrs	9.11%	10.03%	9.92%	10.42%	9.50%	10.35%	9.76%
12-14 yrs	12.50%	13.64%	13.34%	13.93%	13.28%	13.71%	12.90%
15-18 yrs	15.85%	16.34%	15.84%	14.44%	17.07%	17.08%	14.76%
19-20 yrs	7.00%	7.28%	7.56%	3.93%	8.16%	4.69%	6.51%
Missing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Geographic Location	*	*	*	*	*	*	*
Rural	45.02%	12.88%	18.31%	33.63%	18.03%	56.69%	<1%
Urban	54.77%	80.34%	81.69%	66.36%	75.62%	43.30%	7.87%
Missing	<1%	6.78%	<1%	<1%	6.34%	<1%	91.70%

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
Race/ Ethnicity	*	*	*	*	*	*	*
White, non-Hispanic	32.49%	43.12%	35.80%	98.02%	53.04%	5.63%	25.99%
Black, non-Hispanic	3.80%	8.69%	41.11%	<1%	27.90%	7.05%	20.23%
Asian, non-Hispanic	5.52%	1.61%	1.96%	<1%	<1%	<1%	2.78%
AIAN, non-Hispanic	36.49%	9.39%	<1%	1.95%	<1%	<1%	1.45%
Hawaiian/Pacific Islander	4.84%	<1%	<1%	0.00%	<1%	<1%	1.01%
Multiracial, non-Hispanic	8.16%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Hispanic, all races	3.07%	3.88%	20.83%	<1%	8.19%	<1%	43.93%
non-Hispanic, race unspecified	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Missing	5.64%	32.99%	<1%	<1%	9.43%	86.63%	4.61%
Sex	*	*	*	*	*	*	*
Female	48.95%	49.25%	49.54%	48.50%	49.22%	49.67%	49.34%
Male	51.05%	50.75%	50.45%	51.50%	50.78%	50.32%	50.66%
Missing	0.00%	0.00%	<1%	0.00%	0.00%	<1%	0.00%
*	**New Mexico**	**North Carolina**	**Oklahoma**	**Oregon**	**South Carolina**	**Washington**	**Wyoming**
Total # Patients	383,056	1,241,882	557,138	479,469	752,206	939,142	54,551
Age Group	*	*	*	*	*	*	*
<1 yr	4.94%	5.58%	6.52%	5.53%	5.12%	5.20%	6.11%
1-2 yrs	10.76%	11.91%	13.06%	11.70%	11.52%	10.84%	12.52%
3-5 yrs	15.78%	16.42%	17.58%	15.17%	16.15%	15.77%	16.93%
6-7 yrs	10.17%	10.65%	11.14%	10.02%	10.69%	10.55%	10.97%
8-9 yrs	10.55%	10.41%	10.85%	10.28%	10.98%	10.70%	11.04%
10-11 yrs	9.91%	9.62%	10.04%	9.59%	10.13%	9.90%	9.86%
12-14 yrs	13.65%	12.94%	13.41%	13.02%	13.35%	13.35%	13.03%
15-18 yrs	16.58%	15.91%	13.26%	16.61%	16.59%	16.61%	14.96%
19-20 yrs	7.67%	6.57%	4.14%	8.07%	5.45%	7.09%	4.58%
Missing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Geographic Location	*	*	*	*	*	*	*
Rural	33.89%	26.58%	38.85%	22.21%	23.14%	15.40%	0.00%
Urban	66.06%	72.95%	61.12%	67.14%	76.37%	84.60%	0.00%

*	Alaska	Arizona	Delaware	Idaho	Michigan	Mississippi	Nevada
Missing	<1%	<1%	<1%	10.65%	<1%	<1%	100.00%
Race/ Ethnicity	*	*	*	*	*	*	*
White, non-Hispanic	20.97%	38.76%	40.55%	36.46%	28.36%	41.07%	59.76%
Black, non-Hispanic	1.93%	35.36%	13.44%	2.55%	31.64%	6.54%	2.23%
Asian, non-Hispanic	<1%	1.68%	1.68%	1.35%	<1%	3.50%	<1%
AIAN, non-Hispanic	15.13%	1.28%	16.73%	1.76%	<1%	2.56%	8.23%
Hawaiian/Pacific Islander	0.00%	<1%	<1%	<1%	<1%	2.81%	<1%
Multiracial, non-Hispanic	0.00%	3.76%	0.00%	<1%	0.00%	1.49%	0.00%
Hispanic, all races	59.94%	18.34%	21.26%	33.92%	5.82%	27.92%	0.00%
non-Hispanic, race unspecified	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Missing	1.38%	<1%	5.91%	23.08%	33.22%	14.11%	29.08%
Sex	*	*	*	*	*	*	*
Female	49.40%	49.46%	49.55%	49.20%	49.39%	49.24%	48.79%
Male	50.60%	50.54%	50.45%	50.79%	50.61%	50.76%	51.21%
Missing	0.00%	0.00%	0.00%	<1%	<1%	0.00%	0.00%

Table showing demographic characteristics for individuals<21 years old enrolled in each of 14 state Medicaid programs in CY 2016

*Cell left intentionally blank

[Response Ends]

2a.07. 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.

[Response Begins]

Measure score reliability and validity tests were conducted using Medicaid claims and enrollment data contained within the T-MSIS analytic files from 14 state programs.

[Response Ends]

2a.08. List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

[Response Begins]

The measure scores were stratified by the following patient characteristics (when sufficient data were available): age, geographic location (rural or urban), race and ethnicity, and biological sex. These will be reported in section 1b: Importance to Measure and Report: Gap in Care/Disparities.

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.

[Response Begins]

Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)

Accountable Entity Level (e.g., signal-to-noise analysis)

[Response Ends]

2a.10. For each level of reliability testing 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.

[Response Begins]

(1) Accountable Entity Measure Score Reliability Testing using a random split-sample methodology:

Reliability indicates the extent to which repeated measurements yield consistent results. We conducted accountable entity measure score level reliability testing using a **random split-sample methodology and the intraclass correlation coefficient** (ICC) of the measure scores. For each of the 14 state Medicaid programs, we randomly split the population of children aged <21 years present in the T-MSIS demographic and eligibility file. The denominator, numerator, and measure score were calculated for each sample. Thus, the measure score is calculated twice for each state Medicaid program among two distinct and randomly selected sets of children contained within the analytic files. We used the ICC to calculate the agreement between the randomly selected samples (Koo & Li 2016; McGraw & Wong 1996; Shrout & Fleiss 1979). A higher ICC value indicates greater agreement and, therefore, greater reliability. We follow the guidance in Koo and Li (2016) regarding the interpretation of reliability using the 95% confidence interval of the ICC: <0.5 = poor; 0.5–0.75 = moderate; 0.75–0.9 = good; and > 0.9 = excellent.

(2) Evaluation of Relative Rankings: Between Split Samples

We compared the **relative rankings** for the split samples to evaluate whether the relative measure scores for the state Medicaid programs remained stable between the split samples.

(3) Evaluation of Relative Rankings: Between Years

We compared the **relative rankings** of the overall measure scores between 2017 and 2018 and between 2016 and 2017 to evaluate whether there were any dramatic changes that could suggest a threat to

reliability. Using the measure scores for the 14 state Medicaid programs in 2017 and 2018 and 10 state Medicaid programs in 2016 and 2017, we calculated Kendall's Tau-b, which is a rank correlation coefficient that measures association based on the number of concordant and discordant pairs. For reference, we also report the more commonly reported Spearman's rank correlation coefficients. Although the strength of these associations is stronger, we felt Kendall's tau was the more appropriate test to report given the relatively small sample size.

(4) Data Element Reliability/Validity

We include in this application testing data used to support critical data element reliability and validity for Topical Fluoride for Children, Dental Services (NQF #2528) given the significant overlap in critical data elements. Because data element reliability is established with demonstration of data element validity, we refer to Section 2b1 below which describes the chart audit process. **Note:** Unlike measures that rely on patient 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.

(5) Threats to Measure Reliability

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

(A) 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 measure specifications were developed by the Dental Quality Alliance (DQA) which represents and solicits input from 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 specifications are carefully evaluated to ensure that they identify all necessary data elements to calculate the numerators and denominators for each measure, and the specifications are refined during measure testing. The DQA also solicits public comment on measure specifications and works with its Measures Development and Maintenance Committee to address all comments to ensure clear and detailed measure specifications.

(B) Other Threats to Reliability - Sample Size

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

References

Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155-63. doi: 10.1016/j.jcm.2016.02.012. Epub 2016 Mar 31.

McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1:30–46.

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86:420–428.

Plan-Level Data. This measure is also specified for both program and plan-level reporting. Our original testing of NQF #2528 Topical Fluoride for Children, Dental Services, included plan-level data. This measure is being grouped with NQF #2528 as well as with NQF #3700. All three measures share the same denominator. This measure focuses specifically on topical fluoride applications provided as "oral health" services. Consequently, the numerator captures topical fluoride applications by non-dental providers. DQA measures, including NQF #2528, have been implemented by programs and plans operating in commercial, Marketplace, and Medicaid markets. However, our testing for this submission does not include plan-specific data. The T-MSIS data used for testing currently do not enable reliable identification of which topical fluoride services are provided by which managed care organizations (MCOs). Based on prior testing of NQF #2528 and other administrative

claims-based oral healthcare measures, including measures that require both medical and dental claims, we would not expect to see marked differences in the reliability or validity of plan-level reporting compared with program-level reporting given that the data sources (administrative claims) and measure specifications are the same for the two reporting levels. The only potential concern would be if the plan level denominators were too small to yield reliable results. The denominator requirements for this measure and other oral healthcare measures capture a broad population, and we have not encountered issues with small denominator sizes in our testing or in feedback from the user community. The DQA membership includes MCO representatives that operate in state Medicaid programs throughout the United States, and insufficient denominator sizes have never arisen as a concern. The DQA is also in frequent communication with the Centers for Medicare and Medicaid Services, as well as with Mathematica which serves as CMS's technical resource to state Medicaid programs for quality measure implementation; again, there have been no issues raised related to challenges with plan-level implementation.

The T-MSIS claims data are missing the managed care plan identifier for more than 90% of topical fluoride services in the states that we examined. This does **not** represent a feasibility issue for Medicaid programs and their participating plans to calculate the measures. We know from working with state Medicaid programs and state Marketplaces that it is highly feasible to have plan level reporting of oral healthcare quality measures. Rather, this reflects a limitation of the database that we used for testing. This is a relatively new database (released for public use in September 2020) for which data completeness and quality are continually being improved. In addition, because the measure specifications were updated during the DQA 2021 Annual Measure Review, there is no public reporting yet of the revised measure.

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

(1) Accountable Entity Measure Score Reliability Testing: Split-Half

(A) Performance Scores with 95% CI

Tables 2a.11)A-C below provides the performance scores for the split samples with their 95% confidence intervals.

Table 2a.11)A, Split Sample Topical Fluoride for Children, Oral Health Services,

Measure Scores and 95% CI, CY2018 (T-MSIS Data)

Program	Sample	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
AK, 2018	1	0.92%	(	0.0005	,	0.0082	,	0.0102	)
*	2	0.88%	(	0.0005	,	0.0078	,	0.0097	)
AZ, 2018	1	0.58%	(	0.0001	,	0.0055	,	0.0060	)

Program	Sample	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
*	2	0.59%	(	0.0001	,	0.0057	,	0.0062	)
DE, 2018	1	0.14%	(	0.0002	,	0.0010	,	0.0018	)
*	2	0.18%	(	0.0002	,	0.0013	,	0.0022	)
ID, 2018	1	0.35%	(	0.0002	,	0.0031	,	0.0039	)
*	2	0.41%	(	0.0002	,	0.0036	,	0.0045	)
MI, 2018	1	0.63%	(	0.0001	,	0.0060	,	0.0065	)
*	2	0.61%	(	0.0001	,	0.0059	,	0.0064	)
MS, 2018	1	0.69%	(	0.0002	,	0.0065	,	0.0074	)
*	2	0.74%	(	0.0002	,	0.0069	,	0.0078	)
NV, 2018	1	0.37%	(	0.0002	,	0.0034	,	0.0041	)
*	2	0.38%	(	0.0002	,	0.0034	,	0.0041	)
NM, 2018	1	0.73%	(	0.0002	,	0.0068	,	0.0078	)
*	2	0.75%	(	0.0002	,	0.0070	,	0.0080	)
NC, 2018	1	3.64%	(	0.0003	,	0.0358	,	0.0369	)
*	2	3.56%	(	0.0003	,	0.0350	,	0.0361	)
OK, 2018	1	0.29%	(	0.0001	,	0.0026	,	0.0031	)
*	2	0.29%	(	0.0001	,	0.0026	,	0.0032	)
OR, 2018	1	0.52%	(	0.0002	,	0.0048	,	0.0055	)
*	2	0.54%	(	0.0002	,	0.0050	,	0.0058	)
SC, 2018	1	1.36%	(	0.0002	,	0.0131	,	0.0140	)
*	2	1.41%	(	0.0002	,	0.0136	,	0.0145	)
WA, 2018	1	0.31%	(	0.0001	,	0.0029	,	0.0033	)
*	2	0.34%	(	0.0001	,	0.0032	,	0.0035	)
WY, 2018	1	0.54%	(	0.0006	,	0.0041	,	0.0066	)
*	2	0.56%	(	0.0007	,	0.0043	,	0.0069	)

Table showing split sample performance scores with standard deviation and 95% confidence intervals for 14 state Medicaid programs in 2018.

*Cell left intentionally blank

Table 2a.11)B, Split Sample Topical Fluoride for Children, Oral Health Services, Measure Scores and 95% CI, CY2017 (T-MSIS Data)

Program	Sample	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
AK, 2017	1	0.84%	(	0.0005	,	0.0074	,	0.0094	)
*	2	0.77%	(	0.0005	,	0.0068	,	0.0086	)
AZ, 2017	1	0.51%	(	0.0001	,	0.0049	,	0.0053	)
*	2	0.49%	(	0.0001	,	0.0047	,	0.0052	)
DE, 2017	1	0.20%	(	0.0002	,	0.0015	,	0.0024	)
*	2	0.14%	(	0.0002	,	0.0010	,	0.0018	)
ID, 2017	1	0.20%	(	0.0002	,	0.0017	,	0.0023	)
*	2	0.20%	(	0.0002	,	0.0017	,	0.0023	)
MI, 2017	1	0.72%	(	0.0001	,	0.0069	,	0.0074	)
*	2	0.72%	(	0.0001	,	0.0069	,	0.0074	)
MS, 2017	1	0.61%	(	0.0002	,	0.0058	,	0.0065	)
*	2	0.63%	(	0.0002	,	0.0060	,	0.0067	)
NV, 2017	1	1.31%	(	0.0003	,	0.0124	,	0.0138	)
*	2	1.35%	(	0.0004	,	0.0128	,	0.0142	)
NM, 2017	1	0.63%	(	0.0002	,	0.0059	,	0.0067	)
*	2	0.66%	(	0.0002	,	0.0062	,	0.0070	)
NC, 2017	1	3.52%	(	0.0003	,	0.0347	,	0.0358	)
*	2	3.54%	(	0.0003	,	0.0348	,	0.0360	)
OK, 2017	1	0.26%	(	0.0001	,	0.0023	,	0.0029	)
*	2	0.25%	(	0.0001	,	0.0022	,	0.0027	)
OR, 2017	1	0.50%	(	0.0002	,	0.0046	,	0.0053	)
*	2	0.49%	(	0.0002	,	0.0045	,	0.0053	)
SC, 2017	1	1.19%	(	0.0002	,	0.0115	,	0.0123	)
*	2	1.16%	(	0.0002	,	0.0112	,	0.0120	)
WA, 2017	1	0.32%	(	0.0001	,	0.0031	,	0.0034	)
*	2	0.33%	(	0.0001	,	0.0031	,	0.0035	)
WY, 2017	1	0.79%	(	0.0008	,	0.0064	,	0.0094	)
*	2	0.67%	(	0.0007	,	0.0053	,	0.0082	)

Table showing split sample performance scores with standard deviation and 95% confidence intervals for 14 state Medicaid programs in 2017.

*Cell left intentionally blank

Table 2a.11)C, Split Sample Topical Fluoride for Children, Oral Health Services, Measure Scores and 95% CI, CY2016 (T-MSIS Data)

Program	Sample	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
AK, 2016	1	0.24%	(	0.0003	,	0.0019	,	0.0029	)
*	2	0.23%	(	0.0003	,	0.0018	,	0.0028	)
AZ, 2016	1	0.48%	(	0.0001	,	0.0046	,	0.0050	)
*	2	0.47%	(	0.0001	,	0.0045	,	0.0049	)
DE, 2016	1	0.39%	(	0.0003	,	0.0033	,	0.0045	)
*	2	0.39%	(	0.0003	,	0.0033	,	0.0045	)
MI, 2016	1	0.48%	(	0.0001	,	0.0045	,	0.0050	)
*	2	0.47%	(	0.0001	,	0.0045	,	0.0049	)
NV, 2016	1	0.66%	(	0.0002	,	0.0061	,	0.0071	)
*	2	0.67%	(	0.0002	,	0.0062	,	0.0072	)
NM, 2016	1	0.58%	(	0.0002	,	0.0054	,	0.0062	)
*	2	0.60%	(	0.0002	,	0.0056	,	0.0064	)
NC, 2016	1	3.46%	(	0.0003	,	0.0341	,	0.0352	)
*	2	3.49%	(	0.0003	,	0.0343	,	0.0354	)
OK, 2016	1	0.21%	(	0.0001	,	0.0019	,	0.0024	)
*	2	0.25%	(	0.0001	,	0.0022	,	0.0027	)
OR, 2016	1	0.41%	(	0.0002	,	0.0038	,	0.0044	)
*	2	0.40%	(	0.0002	,	0.0037	,	0.0043	)
SC, 2016	1	0.44%	(	0.0001	,	0.0042	,	0.0047	)
*	2	0.44%	(	0.0001	,	0.0042	,	0.0046	)
WA, 2016	1	0.38%	(	0.0001	,	0.0036	,	0.0040	)
*	2	0.38%	(	0.0001	,	0.0036	,	0.0040	)
WY, 2016	1	0.65%	(	0.0007	,	0.0052	,	0.0078	)
*	2	0.69%	(	0.0007	,	0.0056	,	0.0082	)

Table showing split sample performance scores with standard deviation and 95% confidence intervals for 12 state Medicaid programs in 2016.

*Cell left intentionally blank

(B) ICC

The ICCs for each year 95% CI are shown in the table below:

Intraclass Correlation Coefficients	*	*	*	*	*	*	*
Year	ICC	*	95% CI, lower bound	*	95% CI, upper bound	*	p-value

Intraclass Correlation Coefficients	*	*	*	*	*	*	*
2018 (n=14)	0.999	(	0.9972	,	0.9997	)	<0.0001
2017 (n=14)	0.999	(	0.9961	,	0.9996	)	<0.0001
2016 (n=12)	0.999	(	0.9990	,	0.9999	)	<0.0001

Table showing intraclass correlation coefficient results for split sample measure scores in each year, 2016 through 2018

*Cell left intentionally blank

Following the guidance in Koo and Li (2016), this indicates "excellent" reliability given that even the lower bound of the 95%CI are greater than 0.90.

(2) Relative Rankings: Split Samples

The figures below show the relative rankings of the state Medicaid programs, overall and for the split samples, for each year 2016-2018. The only changes in relative rankings were not significant. In 2018, comparing split samples 1 and 2, Idaho and Nevada switched places (between ranks 10 and 11). In 2017, Michigan and Wyoming switched places (between ranks 5 and 6). In 2016, when comparing the relative rankings between the split samples, Nevada and Wyoming switched places (between ranks 2 and 3) and Alaska and Oklahoma switched places (between ranks 11 and 12). In all cases, each of these pairs of states have overlapping confidence intervals. Consequently, the relative rankings demonstrate high stability.

Figure 2a.11)-1, Relative Rankings, Overall Score and Split Samples for Topical Fluoride for Children, Oral Health Services, CY 2018

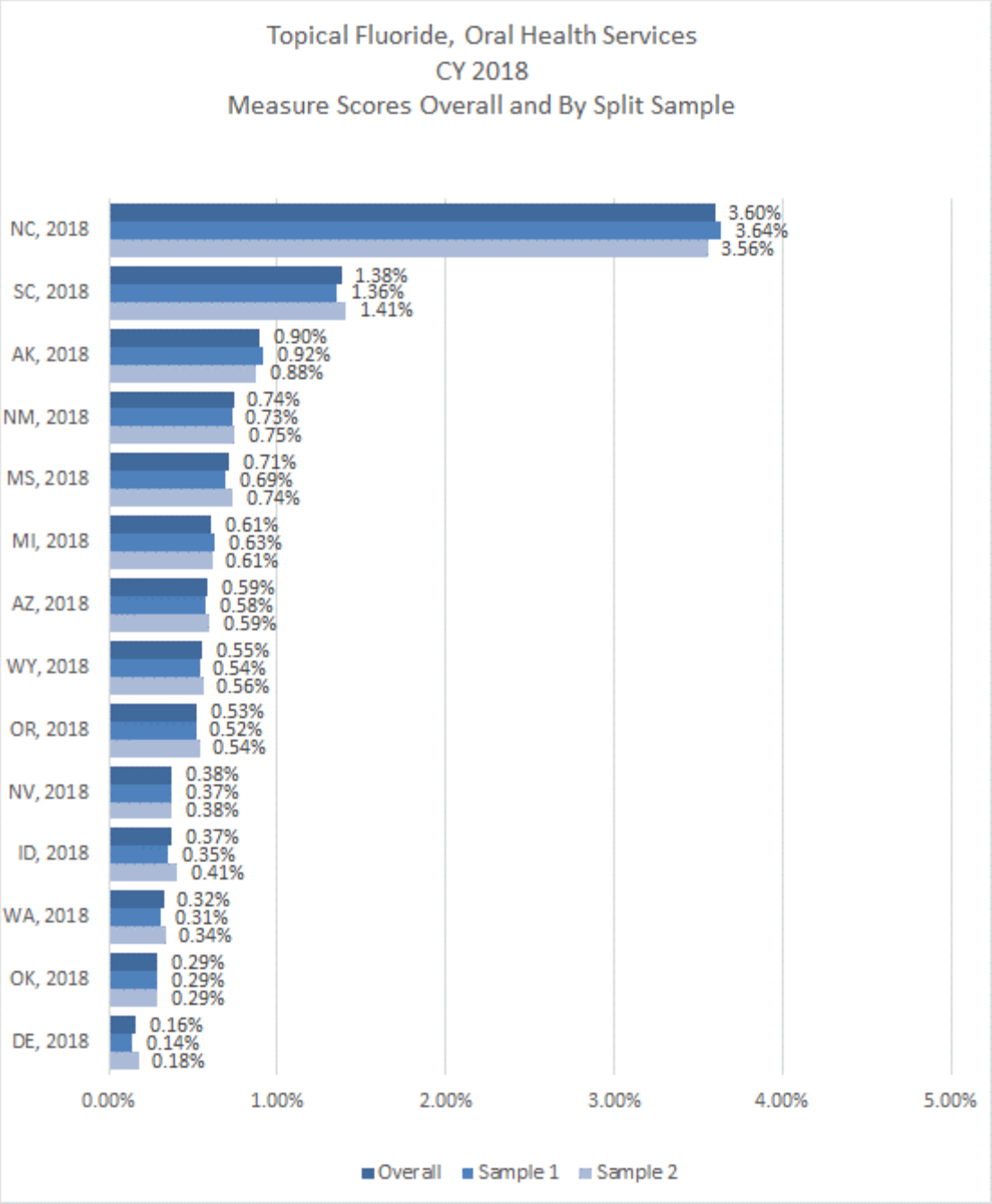


Figure 2a.11)-2, Relative Rankings, Overall Score and Split Samples for Topical Fluoride for Children, Oral Health Services, CY 2017

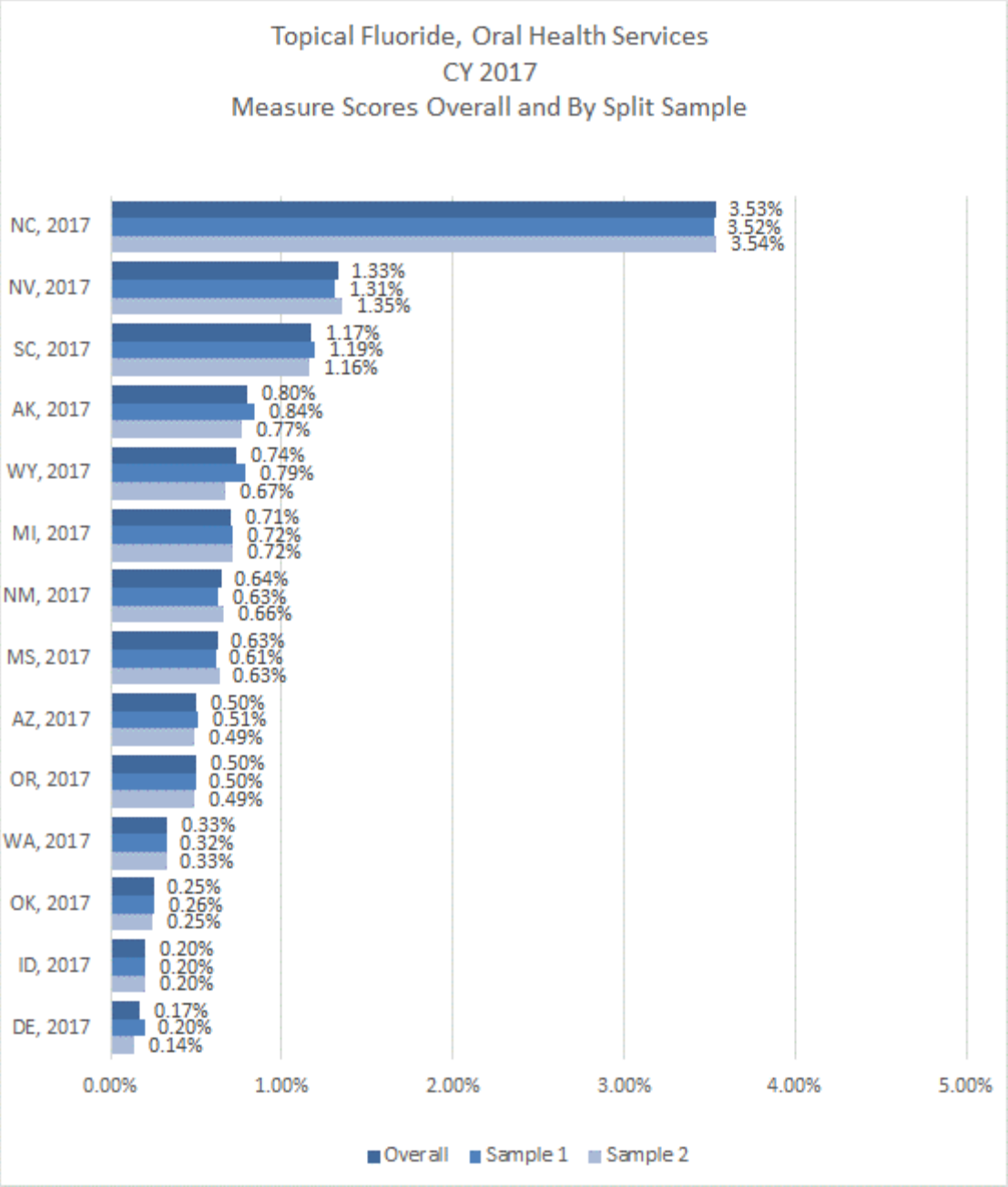
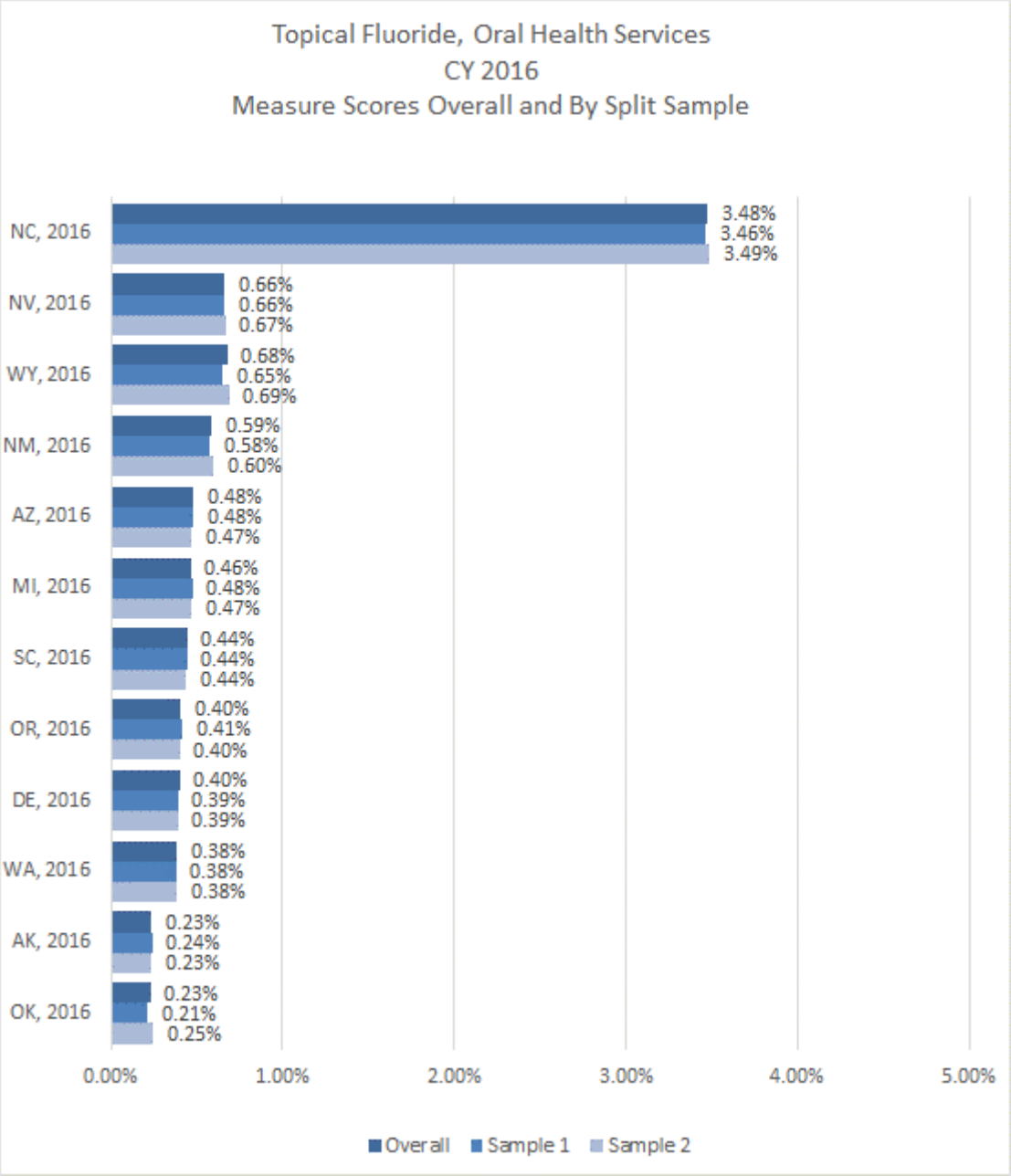


Figure 2a.11)-3, Relative Rankings, Overall Score and Split Samples for Topical Fluoride for Children, Oral Health Services, CY 2016



(3) Relative Rankings: Between Years

The correlation results are in the table below:

Year-to-Year Comparisons	Kendall's tau	p-value	Spearman's rank correlation coefficient	p-value
2017 & 2018 (n=14)	0.6484	0.0015	0.7714	0.0012
2016 & 2017 (n=12)	0.5636	0.0195	0.6636	0.026

Table showing the relative rankings between years using both Kendall's tau and Spearman's rank correlation coefficient

The Kendall tau-b results indicate a positive correlation with a "moderate" strength of association in the state measure scores between years.

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

The testing results indicate that the accountable entity measure scores are reliable:

1. The measure scores between the split samples for each state are similar and have overlapping 95% CIs.
2. The ICC of the split sample measure scores is >0.9, including the lower bound of the 95% CI for the ICC. Following the guidance of Koo and Li (2016), this indicates "excellent" reliability.
3. The relative rankings based on measure scores are stable across the split samples.
4. The relative rankings based on measure scores are fairly stable between years. The Kendall's tau value indicates a "moderate" degree of association. We have not located a definitive source regarding absolute cut points for what constitutes "weak", "moderate", or "strong" association. But based on what we have found in the literature collectively (e.g., Akoglu, 2018), we consider it is a fair characterization to classify the association as moderate.

Testing results demonstrate the reliability of the performance measure scores at the accountable entity level.

References:

Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med. 2018;18(3):91-93. Published 2018 Aug 7. doi:10.1016/j.tjem.2018.08.001

Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155-63. doi: 10.1016/j.jcm.2016.02.012. Epub 2016 Mar 31

[Response Ends]

2b. Validity

2b.01. Select the level of validity testing that was conducted.

[Response Begins]

Patient or Encounter-Level (data element validity must address ALL critical data elements)

Empirical validity testing

[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]

1. Patient / Encounter-Level (data element) Validity

I. Topical Fluoride Procedure Codes

The differences between the measure specifications for this measure and Topical Fluoride for Children, Dental Services (NQF #2528) is the inclusion of CPT code 99188 in the numerator, in addition to CDT codes D1206 and D1208, and the identification of "oral health" provider taxonomy codes that identify fluoride provided by "non-dental" providers, such as physicians and other primary care medical professionals. Because these measures rely on administrative claims data, our data element validation for the dental services version of the measure focused on validating the procedure codes used in the numerator (we have provided this testing below). Our original testing did not include validation of CPT code 99188, because it did not exist at the time. CPT 99188 is a dedicated code for "application of topical fluoride varnish by a physician or other qualified health care professional." This code was introduced in 2015. However, not all states have converted to using this code; some states still reimburse medical primary care providers for fluoride varnish using CDT codes D1206 and/or D1208.

To conduct chart reviews to validate the single procedure code, CPT 99188, is cost prohibitive. The American Academy of Pediatrics provides specific coding guidance for this code - and for coding oral health procedures by primary care physicians more generally (https://downloads.aap.org/AAP/PDF/coding_factsheet_oral_health.pdf). Given the extensive use and validation of dental and medical procedure codes generally and the very specific description of this particular code, combined with AAP coding guidance, we have no reason to expect that this would not be validly coded in the claims data. However, to verify this, we ran analyses of the claims data to identify whether there were any red flags that would suggest that CDT 99188 is not being used appropriately. Specifically, we:

1. Ran frequency distributions of the provider types that performed CPT 99188 (rendering provider type in the claims data).
2. Analyzed the distribution of CPT 99188 by age to see how well it matched with reimbursable age ranges in each state.

II. Data Validation for Topical Fluoride for Children, Dental Services

We also provide testing data from 2012 that was used to support critical data element validity (and reliability) for Topical Fluoride for Children, Dental Services (NQF #2528) given the significant overlap in critical data elements. This testing was conducted by a research team at the University of Florida under contract with the DQA.

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, (5) procedure 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 topical fluoride procedure codes reported in the claims data as the data elements that contribute most to the measure score.** To evaluate data element validity, we conducted reviews of dental records for the Texas Medicaid program. Validation of clinical codes in administrative claims data are most often conducted using manual abstraction from the patient's full chart as the authoritative source. As described in detail below, we evaluated agreement between the claims data and dental charts by calculating the sensitivity, specificity, positive predictive value, and negative predictive value as well as the kappa statistic.

A. Data Sources

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.

Records Requested: 600

#(%) Received: 414 (69%)

The response rate was somewhat higher than expected, and the number of eligible records received (414) exceeded the total targeted number of 400 records.

B. Record Review Methodology

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

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

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

C. Encounter Data Validation – Overall Assessment

For the first component of validation, encounter data validation, the research team followed standard Encounter Data Validation processes following External Quality Review protocols from CMS that it has used in ongoing quality assurance activities for the Texas Health and Human Services Commission. [Centers for Medicare and Medicaid Services, External Quality Review Encounter Data Validation Protocol]. The first three procedure codes were reviewed for each claim. A total of 1,135 procedure codes were reviewed. The RHITs were provided with a pre-populated data entry form with the codes from the claims data for the patient with the specified provider on a particular date of service. They evaluated whether the code in the claims data was supported by the dental record.

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

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

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

agreement; 0.00-0.20 (slight agreement); 0.21-0.40 (fair agreement); 0.41-0.60 (moderate agreement); 0.61-0.80 (substantial agreement); 0.81-0.99 (almost perfect agreement). (Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. Jun 1977;33(2):363-374.)

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

1. Patient / Encounter-Level (Data Element) Validity

A. Topical Fluoride Procedure Codes, Oral Health Services

The table below indicates the codes and age ranges reimbursed by each state in our testing sample:

State	Topical Fluoride Codes	Age Range Reimbursable
Alaska	D1206, D1208	<21 yrs
Arizona	99188	6 mos-2 yrs
Delaware	99188	6 mos-5 yrs
Idaho	99188	<21 yrs
Michigan	99188	0-5 yrs
Mississippi	D1206	0-3 yrs
Nevada	99188	<21 yrs
New Mexico	99188, 1206	6 mos-20 yrs
North Carolina	1206	<3.5 yrs
Oklahoma	99188	6 mos-5 yrs
Oregon	99188, D1206	<19 yrs
South Carolina	99188	<13 yrs
Washington	D1206, D1208	<7 yrs
Wyoming	99188	6 mos-3 yrs

Table showing the codes and age ranges reimbursed by each state in our testing sample.

(1) Rendering Provider Type for CPT 99188

- All CPT 99188 codes had rendering providers who are classified as an "oral health" provider. In all states except one (South Carolina), the rendering provider accounting for the highest percentage of CPT 99188 services was "Pediatrics Physician". For South Carolina it was "General Practice Physician." The other rendering provider types were as expected and included such provider types as family medicine physician, pediatric nurse practitioner, family nurse practitioner, and nurse practitioner.

(2) Age Distribution

The table below provides the age distribution of CPT 99188 in states that reimburse medical primary care providers using this code. In general, the patterns are as expected based on the reimbursable age ranges in each state. In states that reimburse a broader age range, such as Idaho, Nevada and New Mexico, it is expected that the services would still be concentrated among the younger children since those children are less likely to have established a dental home. But we can also see differences in reimbursement policies reflected in these distributions.

*	Arizona	Delaware	Idaho	Michigan	Nevada	New Mexico
*	6 mos-2 yrs	6 mos-5 yrs	<21 yrs	0-5 yrs	<21 yrs	6 mos-20 yrs
CPT 99188	*	*	*	*	*	*
<1 yr	5.14%	1.19%	2.32%	4.11%	1.11%	6.84%
1-2 yrs	93.19%	78.54%	70.02%	82.20%	34.88%	79.05%
3-5 yrs	1.62%	18.18%	22.71%	12.97%	26.82%	12.49%
6-7 yrs	0.34%	1.56%	2.39%	0.43%	9.50%	0.09%
8-9 yrs	0.01%	0.22%	1.05%	0.12%	7.05%	0.40%
10-11 yrs	0.00%	0.07%	0.53%	0.07%	6.40%	0.11%
12-14 yrs	0.00%	0.07%	0.42%	0.04%	8.18%	0.07%
15-18 yrs	0.00%	0.15%	0.24%	0.02%	5.87%	0.11%
19-20 yrs	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%

Table showing age distribution.

*Cell left intentionally blank

*	Oklahoma	Oregon	South Carolina	Wyoming
*	6 mos-5 yrs	<19 yrs	<13 yrs	6 mos-3 yrs
CPT 99188	*	*	*	*
<1 yr	1.19%	1.88%	1.69%	2.74%
1-2 yrs	72.27%	58.42%	56.54%	70.57%
3-5 yrs	25.96%	23.49%	24.95%	20.25%
6-7 yrs	0.56%	4.24%	6.14%	3.80%
8-9 yrs	0.00%	2.06%	3.62%	1.16%
10-11 yrs	0.00%	1.93%	3.41%	0.84%
12-14 yrs	0.00%	3.64%	2.44%	0.53%
15-18 yrs	0.00%	4.10%	1.13%	0.11%
19-20 yrs	0.00%	0.21%	0.57%	0.00%

Table showing age distribution

*Cell left intentionally blank

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

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

Table 2b.03-2 Agreement between Record and Administrative Data for Specific Services

*	Concordance	Prevalence	Sensitivity	Specificity	PPV	NPV	Kappa
Fluoride	89.91%	0.647	0.907	0.884	0.935	0.839	0.782
95% confidence intervals indicated in parentheses	*	*	(0.857-0.942)	(0.806-0.934)	(0.888-0.963)	(0.757-0.898)	(0.710-0.853)
Dates of Service: 317	*	*	*	*	*	*	*
# indeterminate: 0	*	*	*	*	*	*	*

Table showing results of statistical tests of agreement between records and administrative data for specific services.

*Cell left intentionally blank

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

[Response Ends]

2b.04. Provide 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?)

[Response Begins]

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. Our analyses of CPT code 99188 were as expected with the expected provider types rendering CPT 99188 services and services being concentrated within the age ranges eligible for reimbursement. Overall, we interpret these findings as evidence that validates the accuracy of administrative claims data for performance measurement purposes.

[Response Ends]

2b.05. 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 in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

(1) We evaluated performance score data with 95% confidence intervals to assess whether there were statistically significant differences in the measure scores between programs.

(2) We calculated the mean, median, standard deviation, and percentile distribution (10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th}).

(3) To illustrate meaningful differences in performance, we calculated the interquartile range for the measure rates. We used the chi-square test to evaluate statistically significant differences in measure scores between programs in the lowest and highest quartiles.

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include 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.

[Response Begins]

1) Performance Scores with 95% CIs:

CY 2018

Program	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
DE, 2018	0.16%	(	0.0001	,	0.0013	,	0.0019	)
OK, 2018	0.29%	(	0.0001	,	0.0027	,	0.0031	)
WA, 2018	0.32%	(	0.0001	,	0.0031	,	0.0034	)
ID, 2018	0.37%	(	0.0002	,	0.0034	,	0.0041	)
NV, 2018	0.38%	(	0.0001	,	0.0035	,	0.0040	)
OR, 2018	0.53%	(	0.0001	,	0.0050	,	0.0055	)
WY, 2018	0.55%	(	0.0005	,	0.0046	,	0.0064	)
AZ, 2018	0.59%	(	0.0001	,	0.0057	,	0.0060	)
MI, 2018	0.61%	(	0.0001	,	0.0059	,	0.0062	)
MS, 2018	0.71%	(	0.0002	,	0.0068	,	0.0074	)
NM, 2018	0.74%	(	0.0002	,	0.0071	,	0.0078	)
AK, 2018	0.90%	(	0.0004	,	0.0083	,	0.0097	)
SC, 2018	1.38%	(	0.0002	,	0.0135	,	0.0141	)
NC, 2018	3.60%	(	0.0002	,	0.0356	,	0.0364	)

Table showing performance scores with standard deviation and 95% confidence intervals for 14 state Medicaid programs in 2018.

*Cell left intentionally blank

CY 2017

Program	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
DE, 2017	0.17%	(	0.0001	,	0.0014	,	0.0020	)
ID, 2017	0.20%	(	0.0001	,	0.0018	,	0.0022	)
OK, 2017	0.25%	(	0.0001	,	0.0023	,	0.0027	)
WA, 2017	0.33%	(	0.0001	,	0.0031	,	0.0034	)
OR, 2017	0.50%	(	0.0001	,	0.0047	,	0.0053	)
AZ, 2017	0.50%	(	0.0001	,	0.0048	,	0.0052	)
MS, 2017	0.63%	(	0.0001	,	0.0060	,	0.0065	)
NM, 2017	0.64%	(	0.0002	,	0.0061	,	0.0067	)
MI, 2017	0.71%	(	0.0001	,	0.0069	,	0.0072	)
WY, 2017	0.74%	(	0.0005	,	0.0064	,	0.0084	)
AK, 2017	0.80%	(	0.0003	,	0.0073	,	0.0087	)
SC, 2017	1.17%	(	0.0001	,	0.0114	,	0.0120	)
NV, 2017	1.33%	(	0.0002	,	0.0128	,	0.0138	)
NC, 2017	3.53%	(	0.0002	,	0.0350	,	0.0357	)

Table showing performance scores with standard deviation and 95% confidence intervals for 14 state Medicaid programs in 2017.

*Cell left intentionally blank

CY 2016

Program	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
OK, 2016	0.23%	(	0.0001	,	0.0021	,	0.0025	)
AK, 2016	0.23%	(	0.0002	,	0.0020	,	0.0027	)
WA, 2016	0.38%	(	0.0001	,	0.0037	,	0.0040	)
DE, 2016	0.40%	(	0.0002	,	0.0036	,	0.0044	)
OR, 2016	0.40%	(	0.0001	,	0.0038	,	0.0043	)
SC, 2016	0.44%	(	0.0001		0.0042		0.0046	)
MI, 2016	0.46%	(	0.0001	,	0.0045	,	0.0048	)
AZ, 2016	0.48%	(	0.0001	,	0.0046	,	0.0049	)
NM, 2016	0.59%	(	0.0001	,	0.0056	,	0.0062	)

Program	Score	*	SD	*	95% CI, lower bound	*	95% CI, upper bound	*
NV, 2016	0.66%	(	0.0002	,	0.0063	,	0.0070	)
WY, 2016	0.68%	(	0.0005	,	0.0058	,	0.0077	)
NC, 2016	3.48%	(	0.0002	,	0.0344	,	0.0351	)

Table showing performance scores with standard deviation and 95% confidence intervals for 12 state Medicaid programs in 2016.

*Cell left intentionally blank

(2) Performance Distribution and Chi-Square Results, CY2016—CY2018

*	2018 (n=14)	2017 (n=14)	2016 (n=12)
Mean	0.0079	0.0082	0.0073
SD	0.0086	0.0085	0.0092
Minimum	0.0016	0.0017	0.0023
10th Percentile	0.0029	0.0020	0.0023
25th Percentile	0.0037	0.0033	0.0038
Median	0.0057	0.0063	0.0046
75th Percentile	0.0074	0.0080	0.0066
90th Percentile	0.0138	0.0133	0.0068
Maximum	0.0360	0.0353	0.0348
Interquartile Range	0.0037	0.0047	0.0028
p-value	<0.0001	<0.0001	<0.0001

Table showing descriptive statistics of performance scores for each year.

*Cell left intentionally blank

[Response Ends]

2b.07. Provide 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.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

The performance scores with 95% confidence intervals demonstrates the ability of the measure to detect statistically significant differences in performance between measured entities. In CY 2018, the overall measure scores ranged from 0.16% to 3.6%, reflecting significant variation between states. The state with the highest scores, North Carolina, has long-established programs aimed at providing topical fluoride to young children through medical primary care providers, affirming the ability of the measure to capture the effect of improvement efforts on performance. There were statistically significant differences in measure rates between programs in the lowest and highest quartiles in each of the three years ($p<0.0001$). Consequently,

the measure enables identification of both statistically significant and clinically meaningful differences in performance.

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

We used data from 14 Medicaid programs submitted by the states to the Centers for Medicaid and Medicare Services contained within the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAFs). We assessed data quality and missing/invalid data through two methods.

1. CMS T-MSIS Data Quality Atlas. The Medicaid and CHIP Business Information Solutions (MACBIS) conducted data quality assessments of T-MSIS enrollment, claims, expenditures and service use for each state and for each year/release. (Centers for Medicare & Medicaid Services. *DQ Atlas*. Available at https://www.medicaid.gov/dq-atlas/ as of November 2021.) There is a background and methodology report for each topic assessed.

For each state, the Atlas assigns one of the values listed below to indicate the extent to which a state's TAF data are usable, reliable, and accurate for analyzing a particular topic.

- **Low concern**: No major problems were identified that would affect the usability of the TAF data for analyzing a given topic.
- **Medium concern**: Some problems were identified that may affect the usability of the TAF data for analyzing a topic.
- **High concern**: Major problems in the completeness or reliability of the TAF data are likely to impede an analysis of a topic.
- **Unusable**: Extreme problems in the completeness or reliability of the TAF data will prevent a topic from being analyzed.
- **Unclassified**: The topic is either not applicable to a state, or there were not enough TAF or benchmark data for a reliable analysis, or a methodological issue prevented a state's data from being classified into one of the four categories above.

We reviewed the results of these assessments for the following topics (with their descriptions contained within the Quality Atlas) that are relevant to the calculation of Topical Fluoride for Children:

- **Age.** This analysis examines the completeness and distribution of beneficiary age information in the TAF.
- **Medicaid enrollment.** This analysis examines how well the TAF data on the number of total Medicaid beneficiaries align with an external benchmark, the Performance Indicators data set.
- **Claims file completeness: Claims Volume - other services (includes outpatient).** Examining the volume of service use records adjusted for program size can identify outlier states that may have incomplete claims, encounter records, or eligibility data in the TAF. This analysis examines the volume of OT header records, the volume of OT line records, and the average number of lines per header.
- **Claims file completeness: Service Users - other services (includes outpatient).** Examining the overall percentage of beneficiaries with any service use can identify outlier states that may have incomplete claims, encounter, or eligibility data in the TAF. Low rates of service use may also indicate problems in linking service use and eligibility records. This analysis examines the percentage of beneficiaries in

each state with an OT record indicating the receipt of ambulatory, physician, or other medical services during the year.

- **Service use - Procedure Codes - other services (includes outpatient).** This analysis examines how often the procedure code is missing on professional claims in the OT file and how often the non-missing values on these claims represent valid national or state-specific codes.

2. **Additional Evaluations.** We conducted our own assessments of the following data fields:
   - **Date of Birth**. We evaluated how frequently date of birth was missing.
   - **Beneficiary ID.** We evaluated how frequently beneficiary ID was missing among children <21 years.
   - **Dental Procedure Codes (CDT codes).** For each year, we used the list of active and valid procedure codes for each year available from the American Dental Association to evaluate how often non-missing values represent non-valid or non-active codes.

   For consistency with the cut-points used by MACBIS for the Data Quality Atlas, we defined the following categories based on the percentage of missing data:

   -**Low concern:** Missing $\leq$ 10%

   -**Medium concern:** 10% < Missing $\leq$ 20%

   -**High concern:** 20% < Missing $\leq$ 50%

   -**Unusable** Missing > 50%

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

Data Quality Atlas Assessments

The tables below provide the results of the assessments of data completeness, missing data, and invalid data. For the data content areas addressed by MACBIS, as reported in the T-MSIS Quality Atlas, most states were assessed as having data of "low concern" for all content areas in all three years. For CY 2016, there were two states for which all fields were not low concern.

(1) New Mexico was assessed as having outpatient claims volume of "medium concern" in 2017. However, Service Users was "low concern" which is another indicator of claims data quality completeness and specifically assesses the percentage of beneficiaries with any service use. Examination of the measure scores found that performance was within expected ranges and similar between years (2018: 0.74%; 2017: 0.64%; 2016: 0.59%). Consequently, we are comfortable with the data quality for measure reporting purposes.

(2) In 2016, Mississippi was assessed as having outpatient claims volume of “high concern” and Medicaid-only enrollment of “medium concern.” Examination of the measure scores indicates that these data deficiencies significantly impact the measure scores (2018: 0.71%; 2017: 0.63%; 2016: 0.02%). Consequently, we excluded 2016 reporting for Mississippi from our testing.

Additional Evaluations

For the additional data fields assessed (beneficiary ID, age, dental procedure codes, and dental provider taxonomy codes), the rates of missing and invalid data were commonly <1%.

Because there was generally low concern/low rates of missing/invalid data for the critical data elements used to calculate the measure scores, no rules were developed for handling missing data. In addition, because these are standard data fields contained within administrative claims data, when state Medicaid or CHIP programs are identified as having incomplete or poor quality data, they are encouraged to improve data collection and quality as part of their quality improvement efforts rather than using statistical methods to address missing data.

Percentage of Missing and Invalid Values for Critical Data Elements, State Medicaid program, <21 Years Old, CY2018, Release 2 (From T-MSIS Data Quality Atlas)

*	AK	AZ	DE	ID	MI	MS	NV
Age	LC	LC	LC	LC	LC	LC	LC
Medicaid-Only Enrollment	LC	LC	LC	LC	LC	LC	LC
Claims Volume - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Service Users - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Procedure Codes - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Additional Checks	*	*	*	*	*	*	*
Beneficiary ID	0.01%	0.01%	0.01%	0.00%	0.02%	0.01%	0.01%
Dental Procedure Codes - CDT (% invalid)	0.01%	0.00%	0.01%	0.00%	0.00%	0.02%	0.00%
*	**NM**	**NC**	**OK**	**OR**	**SC**	**WA**	**WY**
Age	LC	LC	LC	LC	LC	LC	LC
Medicaid-Only Enrollment	LC	LC	LC	LC	LC	LC	LC
Claims Volume - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Service Users - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Procedure Codes - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Additional Checks	*	*	*	*	*	*	*
Beneficiary ID	0.00%	0.01%	0.01%	0.01%	0.02%	0.01%	0.04%
Date of Birth	0.00%	0.00%	0.00%	0.00%	0.54%	0.00%	1.00%
Dental Procedure Codes - CDT (% invalid)	0.00%	0.01%	0.00%	0.01%	0.00%	0.00%	0.05%

Table showing percentage of missing and invalid values for critical data elements for state Medicaid programs for patients under age twenty-one in CY 2018 (T-MSIS Data)

LC=Low concern, MC=Medium concern, HC=High concern

*Cell left intentionally blank

Percentage of Missing and Invalid Values for Critical Data Elements, State Medicaid program, <21 Years Old, CY2017, Release 2 (From T-MSIS Data Quality Atlas)

*	AK	AZ	DE	ID	MI	MS	NV
Age	LC	LC	LC	LC	LC	LC	LC
Medicaid-Only Enrollment	LC	LC	LC	LC	LC	LC	LC
Claims Volume - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Service Users - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Procedure Codes - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Additional Checks	*	*	*	*	*	*	*
Beneficiary ID	0.01%	0.01%	0.01%	0.00%	0.02%	0.01%	0.01%
Date of Birth	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Dental Procedure Codes - CDT (% invalid)	0.02%	0.00%	0.01%	0.00%	0.00%	0.03%	0.11%
*	**NM**	**NC**	**OK**	**OR**	**SC**	**WA**	**WY**
Age	LC	LC	LC	LC	LC	LC	LC
Medicaid-Only Enrollment	LC	LC	LC	LC	LC	LC	LC
Claims Volume - Other Services (Outpatient)	MC	LC	LC	LC	LC	LC	LC
Service Users - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Procedure Codes - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Additional Checks	*	*	*	*	*	*	*
Beneficiary ID	0.01%	0.01%	0.01%	0.01%	0.02%	0.01%	0.04%
Date of Birth	0.00%	0.00%	0.00%	0.00%	0.61%	0.00%	1.35%
Dental Procedure Codes - CDT (% invalid)	0.00%	0.01%	0.00%	0.01%	0.00%	0.00%	0.06%

Table showing percentage of missing and invalid values for critical data elements for state Medicaid programs for patients under age twenty-one in CY 2017 (T-MSIS Data)

LC=Low concern, MC=Medium concern, HC=High concern

*Cell left intentionally blank

Percentage of Missing and Invalid Values for Critical Data Elements, State Medicaid program, <21 Years Old, CY2016, Release 2 (From T-MSIS Data Quality Atlas)

*	AK	AZ	DE	ID	MI	MS	NV
Age	LC	LC	LC	LC	LC	LC	LC
Medicaid-Only Enrollment	LC	LC	LC	LC	LC	MC	LC

*	AK	AZ	DE	ID	MI	MS	NV
Claims Volume - Other Services (Outpatient)	LC	LC	LC	LC	LC	HC	LC
Service Users - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Procedure Codes - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Additional Checks	*	*	*	*	*	*	*
Beneficiary ID	0.00%	0.01%	0.02%	0.00%	0.05%	0.01%	0.01%
Date of Birth	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%
Dental Procedure Codes - CDT (% invalid)	0.02%	0.00%	0.11%	0.00%	0.00%	0.07%	0.25%
*	**NM**	**NC**	**OK**	**OR**	**SC**	**WA**	**WY**
Age	LC	LC	LC	LC	LC	LC	LC
Medicaid-Only Enrollment	LC	LC	LC	LC	LC	LC	LC
Claims Volume - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Service Users - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Procedure Codes - Other Services (Outpatient)	LC	LC	LC	LC	LC	LC	LC
Additional Checks	*	*	*	*	*	*	*
Beneficiary ID	0.01%	0.01%	0.01%	0.02%	0.02%	0.01%	0.54%
Date of Birth	0.00%	0.00%	0.00%	0.01%	0.53%	0.00%	0.00%
Dental Procedure Codes - CDT (% invalid)	0.02%	0.01%	0.00%	0.01%	0.00%	0.01%	0.05%

Table showing percentage of missing and invalid values for critical data elements for state Medicaid programs for patients under age twenty-one in CY 2016 (T-MSIS Data)

LC=Low concern, MC=Medium concern, HC=High concern

*Cell left intentionally blank

[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]

As noted above, this measure relies on standard data elements in administrative claims. These data are readily available and can be easily retrieved because they are routinely used for billing and reporting purposes. Despite the fact that these are routine data elements, we undertook due diligence to ensure the data completeness within the specific databases that we used for measure testing. Overall, there was low concern/low rates of missing/invalid data for the critical data elements used. Because these are routine data elements that are already collected for other important purposes, particularly claims processing, no rules were developed for handling missing data other than not reporting on the performance measure when data quality is poor. As noted above, programs and plans are instead encouraged to improve their data quality rather than developing statistical techniques to overcome poor data quality.

[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

Describe the steps—do not just name a method. Indicate what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]

N/A or no exclusions

[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

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?

[Response Begins]

Not applicable.

[Response Ends]

2b.17. Provide 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.

[Response Begins]

Not applicable.

[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, 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.

[Response Begins]

Not applicable.

[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]

No risk adjustment or stratification

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

Not applicable.

[Response Ends]

2b.21. 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 (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]

Not applicable.

[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

Other (specify)

[Other (specify) Please Explain]

Not applicable.

[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]

Not applicable.

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

Not applicable.

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]

Not applicable.

[Response Ends]

2b.26. 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). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

Not applicable.

[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]

Not applicable.

[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]

Not applicable.

[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

Not applicable.

[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]

Not applicable.

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]

Not applicable.

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]

Not applicable.

[Response Ends]

Criterion 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.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

ALL data elements are in defined fields in electronic claims

[Response Ends]

3.03. 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 data elements not from electronic sources.

[Response Begins]

Not applicable.

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

Not applicable.

[Response Ends]

3.06. 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.

[Response Begins]

This measure relies on standard fields in administrative claims data (e.g., patient ID, patient birthdate, enrollment information, date of service, procedure codes) that are routinely collected for billing and other purposes. Consequently, the time and cost of data collection for performance measurement purposes are relatively low. As noted elsewhere in this application, we found low rates of missing data for the critical data elements used to calculate the measure.

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail 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),

Attach the fee schedule here, if applicable.

[Response Begins]

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

[Response Ends]

Criterion 4: Use and Usability

4a. 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.

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

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 demonstrating performance improvement.

4a.01. Check all current uses. For each current use checked, please provide:

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]

Not in use

[Not in use Please Explain]

New measure submission

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Public reporting

[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

1. Topical Fluoride for Children, Oral Health Services, has been adopted by the Centers for Medicare and Medicaid Services (CMS) for Child Core Health Care Quality Measurement for federal fiscal year 2022 reporting by state Medicaid and the Children's Health Insurance Program (CHIP) programs (https://www.medicaid.gov/medicaid/quality-of-care/downloads/2022-child-core-set.pdf). The CMS Core Set measures are foundational to assessing access to and quality of health care being provided to Medicaid and CHIP beneficiaries. The two related measures proposed to be "grouped" with this measure (Topical Fluoride for Children, Dental Services, and Topical Fluoride for Children, Dental or Oral Health Services) also have been adopted for inclusion in the Child Core Set. Reporting on the Child Core Set measures will change from voluntary to mandatory reporting in federal fiscal year 2024.
2. This measure has been included in the Center for Oral Health Systems Integration and Improvement (COHSII) Oral Health Quality Indicators for the Maternal and Child Health Population, which is funded by Health Services and Resources Administration (HRSA) Maternal and Child Health Bureau, for 2022 reporting. The measure will be reported with the two related measures proposed to be "grouped" with this measure (Topical Fluoride for Children, Dental Services, and Topical Fluoride for Children, Dental or Oral Health Services).

[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A 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.

[Response Begins]

We anticipate widespread adoption of this measure within 3 years. As noted above, this measure has been included in the CMS Child Core Set for use by Medicaid programs and their contracted MCOs. This measure is also under consideration by NCQA for plan-level reporting. The measure has also been included in the COHSII Oral Health Quality Indicators for the Maternal and Child Health Population.

[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail 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.

[Response Begins]

During measure development, performance results and data were shared with key stakeholders, including the CMS Oral Health Technical Advisory Group, NCQA, and the Center for Oral Health Systems Integration and Improvement.

For all measures, the DQA provides technical assistance to these and other users of DQA measures through webinars, resource document development, and one-on-one staff support. In addition to these activities, the DQA has now released a State Oral Healthcare Quality Dashboard on key oral healthcare measures, including Topical Fluoride for Children, Oral Health Services, that provides a snapshot of oral healthcare quality. .

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]

During measure development, performance results and data were shared with key stakeholders through webinar presentations including time for question/answer and stakeholder feedback. In addition to these activities, the DQA has now released a State Oral Healthcare Quality Dashboard on key oral healthcare measures, including Topical Fluoride for Children, Oral Health Services, that provides a snapshot of oral healthcare quality.

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]

Stakeholders responded very positively to this measure and its usefulness for quality improvement. The DQA's measure Topical Fluoride for Children, Oral Health Services, was developed based on feedback from the measure implementers, such as the CMS, an examination of the evidence supporting the measure, and testing data. As noted above, DQA staff also gave webinar presentations to key stakeholder groups and solicited feedback.

In addition, the DQA has a structured ongoing process for reviewing and updating all measures. 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) month-long 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.

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]

Please see above responses.

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

Please see above responses.

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

Please see above responses.

[Response Ends]

4b. Usability

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, 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, provide an explanation. 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.

[Response Begins]

Testing data presented above demonstrate variations in performance between measured entities and disparities in performance among sub-populations within entities but such characteristics as age, race and ethnicity, and geographic location.

Topical Fluoride for Children, Oral Health Services, along with two related measures proposed to be "grouped" with this measure (Topical Fluoride for Children, Dental Services and Topical Fluoride for Children, Dental or Oral Health Services) has been adopted by the Centers for Medicare and Medicaid Services (CMS) for Child Core Health Care Quality Measurement for federal fiscal year 2022 reporting by state Medicaid and the Children's Health Insurance Program (CHIP). The CMS Core Set measures are foundational to assessing access to and quality of health care being provided to Medicaid and CHIP beneficiaries. Reporting on the child core set of measures will change from voluntary to mandatory reporting in federal fiscal year 2024. Mandatory reporting as part of the CMS Child Core Set will greatly facilitate the ability to identify baseline performance, establish improvement goals, evaluate changes in performance over time, and assess how improvement varies across measured entities.

As noted above, this measure is also under consideration by NCQA for plan-level reporting. The measure has also been included in the COHSII Oral Health Quality Indicators for the Maternal and Child Health

Population. These provide additional opportunities for measure implementation to reach a broad population of children.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]

No unintended or negative consequences have been identified. The measure seeks to improve receipt of an evidence-based process of care among children, that currently has significant performance gaps (i.e., most children do not receive at least two professionally applied topical fluoride applications during the year). As noted in the evidence section, the potential for harm is minimal. No negative unintended impacts have been identified.

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]

We expect the benefits of implementation to be consistent with existing evidence as described in the evidence section.

[Response Ends]

Criterion 5: Related and 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.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)

[Response Begins]

2689: Ambulatory Care Sensitive Emergency Department Visits for Dental Caries in Children

2511: Utilization of Services, Dental Services

2695: Follow-Up after Emergency Department Visits for Dental Caries in Children

2517: Oral Evaluation, Dental Services

[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

Not applicable.

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

Yes

[Response Ends]

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

[Response Begins]

The NQF-endorsed DQA measures (#2511, #2517, #2689 and #2695) all address broadly the same population - children enrolled in Medicaid and CHIP. But the denominators are specified differently. The measures are complementary to one another but distinct.

As noted above, this measure is proposed to be "grouped" with NQF #2528 (existing endorsed measure) and NQF #3700 (new measure).

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]

Topical Fluoride for Children, Dental or Oral Health Services, along with two related measures proposed to be "grouped" with this measure (Topical Fluoride for Children, Dental Services, and Topical Fluoride for Children, Oral Health Services) has been adopted by the Centers for Medicare and Medicaid Services (CMS) for Child Core Health Care Quality Measurement for federal fiscal year 2022 reporting by state Medicaid and the Children's Health Insurance Program (CHIP). The CMS Core Set measures are foundational to assessing access

to and quality of health care being provided to Medicaid and CHIP beneficiaries. Reporting on the Child Core Set of measures will change from voluntary to mandatory reporting in federal fiscal year 2024. Mandatory reporting as part of the CMS Child Core Set will greatly facilitate the ability to identify baseline performance, establish improvement goals, evaluate changes in performance over time, and assess how improvement varies across measured entities.

[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.:

No appendix

Contact Information

Measure Steward (Intellectual Property Owner): American Dental Association

Measure Steward Point of Contact: Colangelo, Erica, colangeloe@ada.org

Ojha, Diptee, ojhad@ada.org

Alliance, Dental, dqa@ada.org

Measure Developer if different from Measure Steward: American Dental Association

Measure Developer Point(s) of Contact: Herndon, Jill, jill.herndon@keyanalyticsconsulting.com

Colangelo, Erica, colangeloe@ada.org

Ojha, Diptee, ojhad@ada.org

Alliance, Dental, dqa@ada.org

Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]

No appendix

[Response Ends]

2. List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

[Response Begins]

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

Measures Development and Maintenance Committee (MDMC)

1. Craig W. Amundson, DDS, Senior Dental Advisor, HealthPartners. Dr. Amundson serves as chair for the Committee.
2. Frederick Eichmiller, DDS, Vice President & Science Officer Emeritus, Delta Dental of Wisconsin
3. Chris Farrell, RDH, BSDH, MPA, Oral Health Program Director, Michigan Department of Health and Human Services
4. An Nyugen, Chief Dental Officer, Clinica Family Health
5. Chris Okunseri, B.D.S., M.Sc., Director, Predoctoral Program, Dental Public Health, Marquette University

1. Bob Russell, DDS, MPH, MPA, CPM, FACD, FICD, Previous State Public Health Dental Director, Chief, Bureau of Oral and Health Delivery Systems, Iowa
2. Tim Wright, DDS, MS, Distinguished Professor, University of North Carolina School of Dentistry, Editor-in-Chief, Journal of American Dental Association
3. Rob Margolin, DDS, Executive Committee Liaison to the MDMC
4. Paul Casamassimo, DDS, MS, Dr. Casamassimo is the current chair of the DQA
5. Ralph Cooley, DDS, Dr. Cooley is the current Chair-Elect of the DQA.

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]

2013 [earlier version of measure with different specifications]

[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]

January 2022

[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]

Annual

[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]

2023

[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".

[Response Begins]

[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A".

[Response Begins]

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 © 2021 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).

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.

[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".

[Response Begins]

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

[Response Ends]