CBE ID

2517

Title

Oral Evaluation, Dental Services

Project

Primary Prevention

Endorsement Status

Endorsed with Conditions

E&M Committee Rationale/Justification

When the measure comes back for maintenance in 5 years, the developer should have: explored with accountable entities some nuances in the improvement data, including any challenges with and reasons for any limited improvement over time, and have identified facilitators to overcome those identified challenges.

Is Under Review

No

Next Maintenance Cycle

Spring 2030

Previous Endorsement Cycle

Spring 2025

Initial Endorsement

Thu, 09/18/2014 - 08:44

Steward

American Dental Association

1.0 New or Maintenance

Maintenance

1.1 Measure Structure

Single Measure

1.3 Electronic Clinical Quality Measure (eCQM)

No

1.6 Measure Description

Percentage of enrolled children under age 21 years who received a comprehensive or periodic

oral evaluation within the reporting year.

1.7 Measure Type

Process

1.8 Level of Analysis

Health Plan, Program (e.g., Medicaid Home Health and Community-Based Services, Part D)

1.9 Care Setting

Clinician Office/Clinic

1.10 Measure Rationale

Dental caries is the most common chronic disease in children in the United States (CDC 2024a). In 2015–2016, the prevalence of total caries (untreated and treated) was 45.8% and untreated caries was 13.0% among youth aged 2–19 years (Fleming and Afful 2018). Children from lower-income households are more likely to have untreated cavities than children from higher income household (Fleming and Afful 2018; CDC 2024b). Dental decay among children has significant short- and long-term adverse consequences. Childhood caries is associated with increased risk of future caries, difficulty eating and poor nutrition (Casamassimo et al. 2009), missed school days and poorer academic performance (Ruff et al. 2019), hospitalization and emergency room visits (Allareddy 2014), and, in rare cases, death (Otto 2017; National Institutes of Health 2021).

Identifying dental caries early is important to reverse the disease process, prevent progression of caries, and reduce incidence of future lesions. Comprehensive and periodic clinical oral evaluations are diagnostic services that are critical to evaluating current oral disease, risk for future disease, and dentition development.* Clinical oral evaluations also are essential to developing an appropriate preventive oral health regimen and treatment plan. Thus, clinical oral evaluations play an essential role in caries identification, prevention and treatment, thereby promoting improved oral health, overall health, and quality of life.

National guidelines from the American Academy of Pediatric Dentistry (AAPD) and the American Academy of Pediatrics (AAP) recommend that children receive oral health services by 1 year of age and have regular visits thereafter. The most common recall interval is six months. However, evidence-based guidelines indicate that the recall schedule for routine oral evaluations should be tailored to individual needs based on assessments of existing disease and risk of disease (e.g., caries risk) with a recommended recall frequency ranging from 3 months to no more than 12 months for individuals younger than 18 years of age (National Institute for Health and Care Excellence (NICE), Clinical Guideline 19, 2004).

However, there are significant performance gaps and variations in care. More than one-half (56%)

of children living in poverty have dental caries compared with 35% of children in household with income greater than 300% of the federal poverty level. Untreated dental caries occurs among 19% of children living in poverty compared with 7% of children in households with incomes greater than 300% of the federal poverty level (Fleming and Afful 2018). Although comprehensive dental benefits are covered under Medicaid and the Children's Health Insurance Program (CHIP), fewer than half of children enrolled in Medicaid and CHIP for at least 6 months receive oral examinations (CMS 2023). Even among the highest performing states, more than one-fourth of publicly-insured children do not receive an oral evaluation as a dental service during the year (CMS 2023). Thus, a significant percentage of children are not receiving oral evaluations to assess their oral health status and disease risk and to develop an appropriate preventive oral health regimen and treatment plan tailored to individual needs.

The proposed measure, Oral Evaluation - Dental Services, captures whether children receive a comprehensive or periodic oral evaluation as a dental service during the reporting year. In addition, this measure also includes important stratifications by the children's age. Oral Evaluation allows plans and programs to assess whether children are receiving at least one oral evaluation during the reporting year as recommended by evidence-based guidelines.

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

* A Comprehensive Oral Evaluation may be performed on new or established patients and is "a thorough evaluation and recording of the extraoral and intraoral hard and soft tissues" and includes "an evaluation for oral cancer where indicated, the evaluation and recording of the patient's dental and medical history and a general health assessment. It may include the evaluation and recording of dental caries, missing or unerupted teeth, restorations, existing prostheses, occlusal relationships, periodontal conditions (including periodontal screening and/or charting), hard and soft tissue anomalies, etc." A Periodic Oral Evaluation is performed "on a patient of record to determine any changes in the patient's dental and medical health status since a previous comprehensive or periodic evaluation." In addition, there is a code for Oral Evaluation for a Patient under Three Years of Age and Counseling with Primary Caregiver, which includes "[d]iagnostic services performed for a child under the age of three, preferably within the first six months of the eruption of the first primary tooth, including recording of the oral and physical health history, evaluation of caries susceptibility, development of an appropriate preventive oral health regimen and communication with and counseling of the child's parent, legal guardian and/or primary caregiver." American Dental Association. 2025. "CDT 2025: Current Dental Terminology." Chicago, IL: American Dental Association.

Allareddy V, Nalliah RP, Haque M, Johnson H, Rampa SB, Lee MK. Hospital-based emergency department visits with dental conditions among children in the United States: nationwide epidemiological data. Pediatr Dent. 2014 Sep-Oct;36(5):393-9. PMID: 25303506.

American Academy of Pediatric Dentistry. Periodicity of examination, preventive dental services, anticipatory guidance/ counseling, and oral treatment for infants, children, and adolescents. The Reference Manual of Pediatric Dentistry. Chicago, Ill.: American Academy of Pediatric Dentistry; 2024:293-305.

Casamassimo PS, Thikkurissy S, Edelstein BL, Maiorini E. Beyond the DMFT: the human and economic cost of early childhood caries. *Journal of the American Dental Association.* 2009;140(6):650–7.

Centers for Disease Control and Prevention. Oral Health Tips for Children. U.S. Dept of Health and Human Services; 2024(a) https://www.cdc.gov/oral-health/prevention/oral-health-tips-for-childre...

Centers for Disease Control and Prevention. Oral Health Surveillance Report: Dental Caries, Tooth Retention, and Edentulism, United States, 2017–March 2020. U.S. Dept of Health and Human Services; 2024(b) https://www.cdc.gov/oral-health/php/2024-oral-health-surveillance-repor...

Center for Medicaid and CHIP Services, Division of Quality and Health Outcomes. 2023. Medicaid & CHIP Beneficiaries at a Glance: Oral Health. Centers for Medicare & Medicaid Services. Baltimore, MD.

Fleming E, Afful J. Prevalence of Total and Untreated Dental Caries Among Youth: United States, 2015-2016. NCHS Data Brief. 2018 Apr;(307):1-8. PMID: 29717975. https://www.cdc.gov/nchs/products/databriefs/db307.htm

National Institute for Health and Care Excellence (NICE). 2004. Clinical Guidelines. "CG19: Dental Recall – Recall Interval between Routine Dental Examinations." Available at: http://guidance.nice.org.uk/CG19.

National Institutes of Health. Oral Health in America: Advances and Challenges. Bethesda, MD: US Department of Health and Human Services, National Institutes of Health, National Institute of Dental and Craniofacial Research,
2021. https://www.nidcr.nih.gov/sites/default/files/2021-12/Oral-Health-in-Am...

Otto M. *Teeth: The Story of Beauty, Inequality, and the Struggle for Oral Health in America.* New York: The New Press; 2017.

Ruff RR, Senthi S, Susser SR, Tsutsui A. Oral health, academic performance, and school absenteeism in children and adolescents: a systematic review and meta-analysis. *Journal of the American Dental Association.* 2019;150(2):111–21.

1.11 Measure Webpage

https://www.ada.org/-/media/project/ada-organization/ada/ada-org/files/resource...

1.13 Data Dictionary

Attached

1.13a Attach Data Dictionary

CBEID_2517_113a_DataDictionary.xlsx

1.14 Numerator

Unduplicated number of enrolled children under age 21 years who received a comprehensive or periodic oral evaluation as a dental service.

1.14a Numerator Details

The numerator is the unduplicated number of enrolled children under age 21 years (subset of denominator) who received a comprehensive or periodic oral evaluation as a dental service during the reporting year.

The numerator calculation relies on standard data fields and coding systems contained within administrative claims data. The reporting period is a single calendar year.

The coding systems needed to calculate the measure are:

-Dental procedure codes (American Dental Association CDT codes)

-Provider taxonomy codes (NUCC maintained provider taxonomy codes)

Please see the data dictionary for the specific codes.

Please see 1.18 below for the detailed measure calculation logic.

1.15 Denominator

Unduplicated number of enrolled children under age 21 years.

1.15a Denominator Details

The denominator is the unduplicated number of children <21 years enrolled at least 180 continuous days during the reporting year.

The denominator calculation requires only age and enrollment information that is standard in administrative eligibility/enrollment data. The reporting period is a single calendar year.

Please see 1.18 below for the detailed measure calculation logic.

1.15b Denominator Exclusions

None.

1.15c Denominator Exclusions Details

None.

1.15d Age Group

Other

1.15e Age Range in Years

0 through 20 years (<21 years)

1.16 Type of Score

Rate/proportion

1.17 Measure Score Interpretation

Better performance = Higher score

1.18 Calculation of Measure Score

Oral Evaluation, Dental Services Calculation

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

2. Check if the subject meets age criterion at the last day of the reporting year:

a. If age criterion is met, then proceed to next step.

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

3. Check if subject is continuously enrolled for at least 180 days during the reporting year:

a. If subject meets continuous enrollment criterion, then include in denominator; proceed to next step.

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

YOU NOW HAVE THE DENOMINATOR (DEN) COUNT: Subjects who meet age and enrollment criteria

4. Check if subject received an oral evaluation as a dental service during the reporting year:

a. If [CDT CODE] = D0120 or D0150 or D0145, AND

b. If [RENDERING PROVIDER TAXONOMY] code = any of the NUCC maintained Provider Taxonomy Codes in Table 1 below.

c. If both a AND b are met, then include in numerator; proceed to next step.

d. If either a OR b is not met, then a dental service was not provided; STOP processing. This subject is already included in the denominator but will not be included in the numerator.

Note: In this step, all claims with missing or invalid CDT CODE, missing or invalid NUCC maintained Provider Taxonomy Codes, or NUCC maintained Provider Taxonomy Codes that do not appear in Table 1 should not be included in the numerator.

YOU NOW HAVE NUMERATOR (NUM) COUNT: Subjects who received an oral evaluation as a dental service

5. Report:

a. Unduplicated number of subjects in numerator

b. Unduplicated number of subjects in denominator

c. Measure Rate (NUM/DEN)

d. Rate stratified by age

Table 1: NUCC maintained Provider Taxonomy Codes classified as “Dental Service”*

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

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

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

1.19 Measure Stratification Details

This measure is stratified by age using the following categories:

<1; 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 specifications for complete measure details.

1.20 Types of Data Sources

Administrative Data, Claims Data

1.21a Data Collection Tool URL(s)

http://example.com

1.25 Data Source Details

The measure is specified for use with administrative enrollment and claims data for children with private or public insurance coverage. The measure relies on standard fields contained within these databases.

1.26 Minimum Sample Size

There is no minimum sample size. Because the measure is calculated at the program and plan levels, small sample size is not a concern.

2.1 Attach Logic Model

CBEID_2517_2_1_LogicModel.pdf

2.2 Evidence of Measure Importance

Overview

- Oral Evaluation, Dental Services measures whether children receive a comprehensive or periodic oral evaluation as a dental service during the reporting year. Dental caries is the most common chronic disease in children in the U.S. (CDC 2024), and a significant percentage of children have untreated dental caries (Fleming and Afful 2018). Dental decay causes significant short- and long-term adverse consequences for children's health, functioning, and quality of life (Allareddy, 2014; Casamassimo et al. 2009; National Institutes of Health 2021; Ruff et al. 2019). Identifying caries early is important to reverse the disease process, prevent progression of caries, and reduce incidence of future lesions.

- Comprehensive and periodic clinical oral evaluations are diagnostic services that are critical to evaluating oral disease, risk of future disease, and dentition development. Clinical oral

evaluations also are essential to developing an appropriate preventive oral health regimen and treatment plan. Thus, clinical oral evaluations play an essential role in caries identification, prevention and treatment, thereby promoting improved oral health, overall health, and quality of life.

- Regular oral evaluations are important given changes in tooth eruption and development of the primary, mixed and permanent dentitions throughout childhood as well as changes in caries risk over time.

- **Evidence-based guidelines** recommend clinical oral evaluations with a regular recall schedule that is tailored to individual needs based on assessments of existing disease and risk of disease (e.g., caries risk).

    - The American Academy of Pediatric Dentistry recommends that children have their first clinical oral evaluation at the eruption of the first tooth and no later than by age 1 year. AAPD recommends a recall frequency of 6 months for children "who have no contributory medical conditions and are developing normally" with more frequent recall based on caries risk and disease susceptibility. AAPD recommendations are developed by the Clinical Affairs Committee and are based on regularly-updated searches of electronic databases of articles in the medical and dental literature, restricted to clinical trials and ages birth through 18 years. When data are insufficient or inconclusive, recommendations are based upon expert/consensus opinion. The last revision was in 2022.

    - The National Institute for Health and Care Excellence (NICE) Clinical Guidelines recommended recall frequency ranges from 3 months to no more than 12 months for individuals younger than 18 years of age (National Institute for Health and Care Excellence (NICE), Clinical Guideline 19, 2004. NICE built upon an existing systematic review conducted by Davenport et al. (2003) that addressed the focus of the guidelines. Davenport et al.'s review covered the literature through February 2001. NICE updated that search through July 2003.

Age of First Visit: Clinical Recommendations by Professional Associations

- The American Academy of Pediatrics, American Academy of Pediatric Dentistry, American Dental Association, and American Public Health Association recommend that children have a dental visit by age 1 to allow for timely prevention and identification of dental disease and to enable more conservative approaches to early childhood caries management.

- The American Academy of Pediatrics (AAP) notes the importance of establishing care with a

dental provider in early childhood through medical-dental coordination in addition to conducting oral health screenings and providing basic preventive services and anticipatory guidance within medical settings. An AAP clinical report specifically comments: "With early referral to a dental provider, there is an opportunity to maintain good oral health, prevent disease, treat disease early, and potentially decrease cost." (Krol & Whelan 2023)

- Delays in the first dental visit are associated with an increase in dental caries, treatment needs, and number of dental procedures with a consequent increase in the likelihood of using general anesthesia for treatment and caries-related ED visits (Ahmed et al. 2021; Nowak et al. 2014).

- A study using Medicaid claims data for children 6 months-6 years of age found a significant increase in dental caries when: (1) the first oral health exam occurred at age 4 compared with age 1 (hazard ratio: 5.4) and (2) the oral exam was with a physician compared with a general dentist (Ahmed et al. 2021).

Recommended Recall Interval for Oral Evaluations

NICE Guidelines

(Terminology note: "Oral health reviews" in the UK are equivalent to "oral evaluations" in the US)

- "The recommended interval between oral health reviews should be determined specifically for each patient and tailored to meet his or her needs, on the basis of an assessment of disease levels and risk of or from dental disease." (NICE Guidelines, 2004, p. 40)

- "The longest interval between oral health reviews for patients younger than 18 years should be 12 months." (NICE Guidelines, 2004, p. 41) The Guidelines further explain: "There is evidence that the rate of progression of dental caries can be more rapid in children and adolescents than in older people, and it seems to be faster in primary teeth than in permanent teeth (see Chapter Three, Section 3.1.2.) Periodic developmental assessment of the dentition is also required in children. Recall intervals of no longer than 12 months give the opportunity for delivering and reinforcing preventive advice and for raising awareness of the importance of good oral health. This is particularly important in young children, to layout the foundations for life-long dental health." (NICE Guidelines, 2004, p. 41) "For practical reasons, the patient should be assigned a recall interval of 3, 6, 9, or 12 months if he or she is younger than 18 years, or 3, 6, 9, 12, 15, 18, 21, or 24 months if he or she is aged 18 years or older." (NICE Guidelines, 2004, p. 41)

AAPD Clinical Periodicity Guidelines:

- "The developing dentition and occlusion should be monitored throughout eruption at regular clinical examinations." (p. 294)

- "The interval of examination should be based on the child's individual needs or risk status/susceptibility to disease; some patients may require examination and preventive

services at more or less frequent intervals, based upon historical, clinical, and radiographic findings." (p. 294)

- The AAPD recommends clinical oral examinations every six months for "children who have no contributing medical conditions and are developing normally." Recall frequency may be shorter intervals based on a child's risk status and susceptibility to disease. (p.305)

Benefits Obtained

- Disease prevention, including reduction in risk of future disease. Clinical oral evaluations include diagnosis, caries risk assessment, prevention and treatment planning, all of which contribute to disease prevention and management. A systematic review concluded that prior caries experience is an important predictor of future risk (Zero et al. 2010). A more recent study reaffirmed this for children, finding that caries in primary teeth predicted future caries in permanent teeth (Lin et al 2021).

- Improved oral and overall health. "Evidence-based early detection and management of caries/oral conditions can improve a child's oral and general health, well-being, and school readiness." (p. 114 of AAPD Clinical Guidelines)

- Every visit provides the opportunity to provide anticipatory guidance, which "is the process of providing practical and developmentally-appropriate information about children's health to prepare parents for the significant physical, emotional, and psychological milestones." (AAPD Clinical Guidelines, p. 116) "Individualized discussion and counseling [anticipatory guidance] should be an integral part of each visit. Topics should include oral hygiene practices, oral/dental development and growth, speech/language development, nonnutritive habits, diet and nutrition, injury prevention, tobacco/nicotine product use, substance misuse, and intraoral/perioral piercing and oral jewelry/accessories." (AAPD Clinical Guidelines, p. 116).

Guideline Citations

American Academy of Pediatric Dentistry. Periodicity of examination, preventive dental services, anticipatory guidance/ counseling, and oral treatment for infants, children, and adolescents. The Reference Manual of Pediatric Dentistry. Chicago, Ill.: American Academy of Pediatric Dentistry; 2024:293-305.

National Institute for Health and Care Excellence (NICE). 2004. Clinical Guidelines. "CG19: Dental Recall – Recall Interval between Routine Dental Examinations." Available

at: http://guidance.nice.org.uk/CG19.

Other Citations

Ahmed I, McGivern S, Beymer MR, Okunev I, Tranby EP, Frantsve-Hawley J, Tseng CH, Ramos-Gomez F. Age of First Oral Health Examination and Dental Treatment Needs of Medicaid-Enrolled Children. JDR Clin Trans Res. 2021 Dec 18:23800844211057793. doi: 10.1177/23800844211057793. Epub ahead of print. PMID: 34923877.

Allareddy V, Nalliah RP, Haque M, Johnson H, Rampa SB, Lee MK. Hospital-based emergency department visits with dental conditions among children in the United States: nationwide epidemiological data. Pediatr Dent. 2014 Sep-Oct;36(5):393-9. PMID: 25303506.

American Academy of Pediatric Dentistry. 2024. "Periodicity of Examination, Preventive Dental Services, Anticipatory Guidance/Counseling, and Oral treatment for Infants, Children and Adolescents. The Reference Manual of Pediatric Dentistry. Chicago, IL: American Academy of Pediatric Dentistry; 2024:293-305. " Available at: https://www.aapd.org/globalassets/media/policies_guidelines/bp_periodic....

Casamassimo PS, Thikkurissy S, Edelstein BL, Maiorini E. Beyond the DMFT: the human and economic cost of early childhood caries. Journal of the American Dental Association. 2009;140(6):650–7.

Centers for Disease Control and Prevention. Oral Health Tips for Children. U.S. Dept of Health and Human Services;
2024(a) https://www.cdc.gov/oral-health/prevention/oral-health-tips-for-childre...

Fleming E, Afful J. Prevalence of Total and Untreated Dental Caries Among Youth: United States, 2015-2016. NCHS Data Brief. 2018 Apr;(307):1-8. PMID:
29717975. https://www.cdc.gov/nchs/products/databriefs/db307.htm

Krol DM, Whelan K, Section On Oral H. Maintaining and Improving the Oral Health of Young Children. *Pediatrics*. Jan 1 2023;151(1)doi:10.1542/peds.2022-060417

Lin YT, Chou CC, Lin YJ. Caries experience between primary teeth at 3-5 years of age and future caries in the permanent first molars. J Dent Sci. 2021 Jul;16(3):899-904. doi: 10.1016/j.jds.2020.11.014. Epub 2020 Dec 15. PMID: 34141103; PMCID: PMC8189882.

National Institutes of Health. Oral Health in America: Advances and Challenges. Bethesda, MD: US Department of Health and Human Services, National Institutes of Health, National Institute of Dental and Craniofacial Research,
2021. https://www.nidcr.nih.gov/sites/default/files/2021-12/Oral-Health-in-Am...

Nowak AJ, Casamassimo PS, Scott J, Moulton R. Do early dental visits reduce treatment and treatment costs for children? Pediatr Dent. 2014 Nov-Dec;36(7):489-93. PMID: 25514078.

Ruff RR, Senthi S, Susser SR, Tsutsui A. Oral health, academic performance, and school absenteeism in children and adolescents: a systematic review and meta-analysis. Journal of the American Dental Association. 2019;150(2):111–21.

Zero D, Fontana M, Lennon AM. Clinical applications and outcomes of using indicators of risk in caries management. J Dent Educ. 2001 Oct;65(10):1126-32. PMID: 11699989.

Additional resources and evidence in AAPD Periodicity Guidelines:

American Academy of Pediatric Dentistry. Perinatal and infant oral health care. The Reference Manual of Pediatric Dentistry. Chicago, Ill.: American Academy of Pediatric Dentistry; 2022:277-81.

American Academy of Pediatric Dentistry. Policy on the dental home. The Reference Manual of Pediatric Dentistry. Chicago, Ill.: American Academy of Pediatric Dentistry; 2022:21-2.

American Academy of Pediatric Dentistry. Management of the developing dentition and occlusion in pediatric dentistry. The Reference Manual of Pediatric Dentistry. Chicago Ill.: American Academy of Pediatric Dentistry; 2022:424-41.

American Academy of Pediatric Dentistry. Policy on early childhood caries (ECC): Consequences and preventive strategies. The Reference Manual of Pediatric Dentistry. Chicago, Ill.: American Academy of Pediatric Dentistry; 2022:90-3.

American Academy of Pediatric Dentistry. Prescribing dental radiographs for infants, children, adolescents, and individuals with special health care needs. The Reference Manual of Pediatric Dentistry. Chicago, Ill.: American Academy of Pediatric Dentistry; 2022:273-6.

American Academy of Pediatric Dentistry. Risk assessment and management of periodontal diseases and pathologies in pediatric dental patients. The Reference Manual of Pediatric Dentistry. Chicago, Ill.: American Academy of Pediatric Dentistry; 2022:466-83.

American Academy of Pediatric Dentistry. Management of the developing dentition and occlusion in pediatric dentistry. The Reference Manual of Pediatric Dentistry. Chicago Ill.: American Academy of Pediatric Dentistry; 2022:424-41.

American Academy of Pediatric Dentistry. Policy on early childhood caries (ECC): Unique challenges and treatment options. The Reference Manual of Pediatric Dentistry. Chicago, Ill.: American Academy of Pediatric Dentistry; 2022:94-5.

American Academy of Pediatric Dentistry. Policy on prevention of sports-related orofacial injuries. The Reference Manual of Pediatric Dentistry. Chicago, Ill.: American Academy of Pediatric Dentistry; 2022:121-6.

American Academy of Pediatric Dentistry. Policy on intraoral/perioral piercing and oral jewelry/accessories. The Reference Manual of Pediatric Dentistry. Chicago, Ill.: American Academy of Pediatric Dentistry; 2022:119-120.

American Academy of Pediatric Dentistry. Policy on tobacco use. The Reference Manual of Pediatric Dentistry. Chicago Ill.: American Academy of Pediatric Dentistry; 2022:103-7.

American Dental Association, U.S. Department of Health and Human Services. Dental Radiographic Examinations: Recommendations for Patient Selection and Limiting Radiation Exposure. Revised 2012. Available at: “https://www.fda.gov/media/84818/download”.

Beil HA, Rozier RG. Primary health care providers’ advice for a dental checkup and dental use in children. Pediatr 2010;126(2):435-41.

Berg JH, Stapleton FB. Physician and dentist: New initiatives to jointly mitigate early childhood oral disease. Clin Pediatr 2012:51(6):531-7.

Clarke M, Locker D, Berall G, Pencharz P, Kenny DJ, Judd P. Malnourishment in a population of young children with severe early childhood caries. Pediatr Dent 2006;28(3):254-9.

Dye BA, Shenkin JD, Ogden CL, Marshall TA, Levy SM, Kanellis MJ. The relationship between healthful eating practices and dental caries in children ages 2-5 years in the United States, 1988-1994. J Am Dent Assoc 2004;135(1):55-6.

Fontana M. Noninvasive caries risk-based management in private practice settings may lead to reduced caries experience over time. J Evid Based Dent Pract 2016;16(4): 239-42.

Jackson SL, Vann WF, Kotch J, Pahel BT, Lee JY. Impact of poor oral health on children's school attendance and performance. Amer J Publ Health 2011;10(10):1900-6.

Li H, Zou Y, Ding G. Dietary factors associated with dental erosion: A meta-analysis. PLoSOne 2012;7(8):e42626. doi:10.1371/journal.pone.0042626. Epub2012 Aug 31.

Mobley C, Marshall TA, Milgrom P, Coldwell SE. The contribution of dietary factors to dental caries and disparities in caries. Acad Pediatr 2009;9(6):410-4

Patel S, Bay RC, Glick M. A systematic review of dental recall intervals and incidence of dental caries. J Am Dent Assoc 2010;141(5):527-39.

Sigurdsson A. Evidence-based review of prevention of dental injuries. Pediatr Dent 2013;35(2):184-90.

2.4 Performance Gap

Note: Please see Attachment CBEID_2517_2_4_Performance_Gap for detailed tables.

Maintenance Testing: Program Level

Data Source

Maintenance evaluations were conducted using Medicaid enrollment and claims data contained within the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAFs) available from the Centers for Medicare and Medicaid Services (CMS) through a data use agreement. We used data from calendar year 2022 for 47 states assessed as having sufficiently complete data for all critical data elements in the T-MSIS database. A total of 33,381,516 denominator eligible patients were identified across these 47 states.

Performance Score Summary

The measure scores ranged from 26.0% to 56.8%, a significant variation in performance. Both the mean and median performance was 42%, demonstrating an overall performance gap. Even in the highest performing program, 43% of Medicaid-enrolled program did not receive a periodic or comprehensive oral evaluation during the year. The bottom three deciles represent 40% of all denominator-eligible beneficiaries and include three of the four largest Medicaid programs. These data indicate both an overall performance gap and variation between programs, demonstrating substantial opportunities for improvement.

Please see attachment CBEID_2517_2_4_Performance_Gap for the complete scores and 95% confidence intervals (Tables 2.4.A and 2.4.B).

The measure can also be used to evaluate variations in care by patient characteristics. Please see Section 3: Equity for more detail.

Maintenance Testing: Plan Level

Data Source

Our original testing, the basis for initial endorsement and prior continued maintenance, included plan-specific data (please see below). The DQA measure of Oral Evaluation has been implemented by dental plans operating in commercial, Health Insurance Marketplace, and Medicaid/CHIP markets. The T-MSIS data used for our updated testing currently do not enable reliable identification of which dental services are provided by which dental plans. This does not represent a feasibility issue for Medicaid programs and their participating plans to calculate the measure at the plan level. We know from working with state Medicaid programs and state Health Insurance Marketplaces that it is highly feasible to have plan-level reporting of dental quality measures. Rather, this reflects a limitation of the database that we used for testing.

There is some limited plan-level reporting that is publicly available that we have included:

- Plans participating in the Texas Medicaid and CHIP programs for 2013 through 2022.
  - 2 plans serving Texas Medicaid and Texas CHIP (same two plans serving each program), 2013-2020.
  - 3 plans serving Texas Medicaid and Texas CHIP (same three plans serving each program), 2021-2022.
  - Denominators, numerators, and scores are available for each plan.
- CCOs participating in Oregon Medicaid and CHIP (Oregon Health Plan) for 2018 through 2023.
  - 13 CCOs, 2018-2019
  - 15 CCOs, 2020
  - 16 CCOs, 2021-2023
  - Only measure scores are available at the plan level. Denominators are available only at the statewide level.

Performance Score Summary

In 2022, performance scores ranged:

- 46% to 66% across the three plans participating in the Texas Medicaid program;
- 44% to 68% across the three plans participating in the Texas CHIP program; and
- 22% to 46% across the 16 CCOs participating in the Oregon Health Plan.

Thus, in 2022, even for the highest performing plan in the Texas Medicaid and CHIP programs, approximately 1/3 of children did not receive at least one comprehensive or periodic oral evaluation during the reporting year. The measure scores also demonstrate significant variation between plans.

Please see Table 2.4.C in attachment CBEID_2517_2_4_Performance_Gap for the complete scores and 95% confidence intervals for the plans participating in the Texas Medicaid and CHIP programs for years 2013 through 2022. With only 2-3 plans participating in the program, deciles are not reported.

Please see Table 2.4.D in attachment CBEID_2517_2_4_Performance_Gap for reporting performance among the Oregon CCOs by quartile for years 2018 through 2023. Quartiles were selected instead of deciles because there are only 16 plans. Confidence intervals were not calculated for individual plan performance due to lack of sample size information.

Original Testing: Program and Plan Levels

Data Source

We used data from five sources for program and plan level evaluation. We included data for publicly insured children in the Texas Medicaid, Texas CHIP, Florida CHIP, and Florida Medicaid programs as well as national commercial data from Dental Service of Massachusetts, Inc. Data from calendar years 2010 and 2011 were used for all programs except Florida Medicaid. Full-year data for CY 2011 were not available for Florida Medicaid. Therefore, we reported only CY 2010 data for Florida Medicaid.

In the data summaries, “Programs” refer to population data from (1) Texas Medicaid, (2) Texas CHIP, (3) Florida CHIP, (4) Commercial Data, and (5) Florida Medicaid. “Plans” refer to data from the two dental plans that served Florida CHIP members in both 2010 and 2011. [Technically, there were three plans represented in the data because Texas CHIP was served by a single dental plan. Since the program=plan in that case, we included it in the “program” level data.]

Performance Score Summary

The measure rate range of 26% to 67% in CY 2010 (year in which data were available for all programs) indicated a significant performance gap overall. Even in the highest performing program, one-third of children did not receive a comprehensive or period oral evaluation during the year. The rates for the two plans in CY 2010 were 43.7% and 41.7%, respectively, and these scores were statistically significantly different. Please see Table 2.4.E in attachment CBEID_2517_2_4_Performance_Gap for the complete scores and 95% confidence intervals.

Table 1. Performance Scores by Decile

Performance Gap

	Overall	Minimum	Decile_1	Decile_2	Decile_3	Decile_4	Decile_5	Decile_6	Decile_7	Decile_8	Decile_9	Decile_10	Maximum
Mean Performance Score	41.6%	26.0%	28.8%	34.9%	37.4%	39.5%	41.7%	43.3%	44.3%	46.7%	48.7%	51.7%	56.8%
N of Entities	47	1	5	4	5	5	5	5	4	5	4	5	1
N of Persons / Encounters / Episodes	33,381,516	116,719	4,347,043	3,223,300	5,744,725	3,254,334	3,210,395	1,988,564	2,518,937	1,414,415	2,117,726	5,562,077	3,691,533

2.4a Attach Performance Gap Results

CBEID_2517_2_4_PerformanceGap.pdf

2.6 Meaningfulness to Target Population

As noted above, dental caries is the most common chronic condition of childhood. Dental caries

and its sequelae can negatively and substantially impact children's health and quality of life. These adverse effects may include pain and difficulty eating, chewing, smiling, talking, and learning. Lack of timely identification and disease management can result in hospitalizations and emergency department visits. Periodic and comprehensive oral evaluations (dental "check-ups") are procedures during which diagnosis, risk assessment, and prevention and treatment planning occur. These visits are important for early identification, prevention and treatment of dental caries.

A national survey of 1,071 parents of children ages 12 and younger that was conducted December 28, 2020 to January 8, 2021 (commissioned by the Delta Dental Plans Association) found that 96% of parents consider oral health to be very or extremely important to overall health. 94% of parents believed that "throughout the pandemic maintaining their child's oral health is essential to protecting their child's overall health." 77% of parents whose children had oral health issues were concerned about negative impacts on their child's overall confidence, and 22% reported impacts on school attendance, reporting an average of 4.5 missed days of school in the prior 12 months. (Delta Dental Plans Association 2021)

Similar themes came out of a series of community listening sessions held by Community Catalyst with community members in three states (Colorado, Nebraska, and Ohio). A major theme that emerged was: "Communities care about and understand how important their oral health is." The report noted: "Not only did participants continually name oral health as a priority for them, but many naturally named connections between oral health and other areas of health and life. For example, participants named the role dental problems have in interrupting children's ability to learn in school . . . ." The report also noted: "Participants described both the trauma associated with the chronic pain that comes with having untreated oral health problems – without good access to regular dental care, many participants described having to live with ongoing dental pain and the acute pain of undergoing invasive dental procedures." The concept of focusing on care that is oriented to avoiding oral health problems and consequent invasive care extended to children: "For example, one participant from Nebraska stated, 'My baby just had a cavity filled and if there's a less invasive way to do that that's less painful, that should be available to everybody regardless of your insurance status'" (Community Catalyst 2024).

Listening sessions conducted by Strategic Concepts in Organizing and Policy Education (SCOPE) with residents in South Los Angeles found that "most participants felt that oral health care is a high priority in their lives and many take their children to the dentist every 6 months even if they themselves do not always access services due to the high costs" (SCOPE 2016).

Although the measure focuses specifically on children, these findings are similar to surveys conducted among adults regarding their own oral health. A Delta Dental Plans Association counterpart national survey of 1,067 adults found that 93% of adults considered oral health to be

very or extremely important to overall health (Delta Dental Plans Association 2021). A national survey of 14,962 adults conducted by the Health Policy Institute of the American Dental Association found that 97% of adults agreed that "I value keeping my mouth healthy", and 95% agreed that "Regular visits to the dentist will help keep me healthy" (Health Policy Institute, American Dental Association, 2015). SCOPE conducted door-to-door surveys of 246 South LA residents (in addition to conducting the listening sessions noted above), which found that 79% of residents agreed "going to the dentist for a check-up is as important as gong to the doctor for a check-up" and 87% agreed that "taking care of my teeth is important for my overall health" (SCOPE 2016) .

References

Community Catalyst. 2024. Community Perspectives on Access, Quality, and Invasiveness of Dental Care. Available at: https://communitycatalyst.org/wp-content/uploads/2024/03/3.18-Community….

Delta Dental Plans Association. 2021. State of America's Oral Health. Available at: https://www.deltadental.com/content/dam/ddpa/us/en/press-releases/The%2….

Health Policy Institute, American Dental Association. 2015. Oral Health and Well-Being in the United States. Available at: https://www.ada.org/resources/research/health-policy-institute/coverage….

Strategic Concepts in Organizing and Policy Education (SCOPE). 2016. The State of Oral Healthcare in South Los Angeles. Available at: https://scopela.org/the-state-of-oral-healthcare-in-south-los-angeles/.

3.1 Contributions Towards Closing Care Gaps

(1) Evidence of Known Disparities

There is extensive oral health surveillance and literature documenting variations in dental service use among children by age, race/ethnicity, and geographic region, including within vulnerable populations. For example, using data from the National Health and Nutrition Examination Survey, researchers at the National Center for Health Statistics identified variations in untreated dental caries among children by race and ethnicity and poverty level during 2017 through March 2020

(CDC 2024). Specifically, they found:

- 11% of children aged 2-5 years had untreated decay in their primary teeth with higher prevalence (18%) among children in high poverty (<100% FPL) compared with 6.6% in low poverty (>=200% FPL).
- 17% of children aged 6-9 years had untreated decay in their primary or permanent teeth: 26.3% of children in high compared with 10.0% in low poverty.
- 10% of adolescents children aged 12-19 years had untreated decay in their permanent teeth: 14.1% in high poverty compared with 8.2% in low poverty.

Dental decay also varies by race and ethnicity. For example, 18.5% of Mexican American children had tooth decay compared with 8.1% of non-Hispanic white children (CDC 2024). Phipps and Ricks (2017) found that American Indian and Alaskan Native children, 6-8 years old, were twice as likely to have untreated dental caries in their primary teeth and five times as likely to have untreated in their permanent teeth compared with U.S. children overall.

Using data from the Medical Expenditure Panel Survey, Edelstein and Chinn (2009, p. 417) noted variations in dental utilization (any dental visit) by age, family income, race and ethnicity, and education: "Stepwise disparities in dental utilization by income remained as strong in 2004 as in 1996, with 30.8% of poor children, 33.9% of low-income children, 46.5% of middle income children, and 61.8% of high income children having at least 1 dental visit in 2004. One third of minority children (34.1% black and 32.9% of Hispanic children) obtain dental care in a year compared with half (52.5%) of white children. Children whose parents attained less than high school education were less than half as likely to obtain a dental visit in 2004 as children whose parents are college graduates (25% vs 54%)." An analysis by Bouchery (2013) of the Medicaid Analytic eXtract files for nine states, examined dental utilization for preventive services and found variations in dental service use by age, race, and geographic area. Specifically, relative to the reference group of 9 year olds, the percentage point change in the probability of having a dental preventive services was -27.6 for 3 years old; -8.6 for 6 years, -2.2 for 12 years and -15.4 for 15 years (all significant at $p<0.0001$); relative to the reference group of white, non-Hispanic, the percentage point change was -1.8 for black non-Hispanic and 7.8 for Hispanic ($p<0.0001$ for both); relative to the reference group of small metro area, the percentage point change was 5.9 for large metro area ($p<0.0001$). Disparities in the use of dental services have also been noted in other literature and summarized in national reports, including the NIH report Oral Health in America: Advances and Challenges; the IOM report, Improving Access to Oral Health Care for Vulnerable and Underserved Populations; and the IOM report, Advancing Oral Health in America.

References

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

Centers for Disease Control and Prevention. Oral Health Surveillance Report: Dental Caries, Tooth Retention, and Edentulism, United States, 2017–March 2020. U.S. Dept of Health and Human Services; 2024(b) https://www.cdc.gov/oral-health/php/2024-oral-health-surveillance-repor…

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

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

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

National Institutes of Health. Oral Health in America: Advances and Challenges. Bethesda, MD: US Department of Health and Human Services, National Institutes of Health, National Institute of Dental and Craniofacial Research, 2021.
https://www.nidcr.nih.gov/sites/default/files/2021-12/Oral-Health-in-Am…

Phipps, K. and Ricks, T. The Oral Health of American Indian and Alaska Native Children Aged 6–9 years: Results from the 2016–2017 IHS Oral Health Survey. *Indian Health Service Data Brief*. Rockville, MD: USDHHS, IHS; 2017.
https://www.ihs.gov/doh/documents/Data%20Brief%20IHS%206-9%20Year%20Old…. Accessed May 8, 2025.

(2) Methodology: Variations in Performance Scores

Original measure testing found variations in performance by child age, geographic location, and race and ethnicity. Chi-square tests were used to detect whether there were statistically significant differences in performance between groups.

Results are provided below:

Program 1

Overall performance score: 66.55%

Scores by Age

Age <1 years: 18.66%

Age 1-2 years: 58.83%

Age 3-5 years: 73.56%

Age 6-7 years: 76.26%

Age 8-9 years: 76.24%

Age 10-11 years: 75.12%

Age 12-14 years: 71.46%

Age 15-18 years: 61.99%

Age 19-20 years: 36.71%

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

Scores by Geographic Location

Urban: 67.60%

Rural: 60.10%

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

Scores by Race/Ethnicity

Non-Hispanic White: 55.80%

Non-Hispanic Black: 62.72%

Hispanic: 72.32%

p-value from Chi-square test <.0001

Program 2

Overall performance score: 54.18%

Scores by Age

Age <1 years: 7.17%

Age 1-2 years: 45.38%

Age 3-5 years: 56.93%

Age 6-7 years: 61.33%

Age 8-9 years: 60.98%

Age 10-11 years: 59.03%

Age 12-14 years: 53.37%

Age 15-18 years: 44.80%

Age 19-20 years: n/a

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

Scores by Geographic Location

Urban: 55.40%

Rural: 46.75%

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

[Race/ethnicity data were not available/sufficiently complete.]

Program 3

Overall performance score: 46.43%

Scores by Age

Age <1 years: n/a

Age 1-2 years: n/a

Age 3-5 years: 39.34%

Age 6-7 years: 50.37%

Age 8-9 years: 53.29%

Age 10-11 years: 50.66%

Age 12-14 years: 46.29%

Age 15-18 years: 39.79%

Age 19-20 years: n/a

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

Scores by Geographic Location

Urban: 46.56%

Rural: 45.39%

p-value from Chi-square test: 0.0191

[Race/ethnicity data were not available/sufficiently complete.]

Program 4

Overall performance score: 63.26%

Scores by Age

Age <1 years: 0.80%

Age 1-2 years: 11.88%

Age 3-5 years: 62.25%

Age 6-7 years: 75.01%

Age 8-9 years: 75.53%

Age 10-11 years: 73.50%

Age 12-14 years: 70.16%

Age 15-18 years: 63.11%

Age 19-20 years: 52.32%

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

Scores by Geographic Location

Urban: 63.61%

Rural: 55.29%

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

[Race/ethnicity data were not available/sufficiently complete.]

Program 5

Overall performance score: 26.25%

Scores by Age

Age <1 years: 0.27%

Age 1-2 years: 5.84%

Age 3-5 years: 27.99%

Age 6-7 years: 37.32%

Age 8-9 years: 40.10%

Age 10-11 years: 36.69%

Age 12-14 years: 32.31%

Age 15-18 years: 27.06%

Age 19-20 years: 15.73%

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

Scores by Geographic Location

Urban: 25.56%

Rural: 34.89%

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

Scores by Race

Non-Hispanic White: 25.00%

Non-Hispanic Black: 24.18%

Hispanic: 30.35%

p-value from Chi-square test <.0001

Note: N/A for age indicates that those ages are not within the program's age eligibility.

Updated evaluation was not conducted for this maintenance review. However, we note that the DQA Oral Healthcare Quality Dashboard (https://www.ada.org/resources/research/dental-quality-alliance/dqa-impr...) includes stratifications of this measure by age, geographic location, language, race and ethnicity, and sex.

(3) Results/Interpretation

Both existing evidence and testing data provide evidence of variations in care by patient characteristics. In our original testing data that included five programs, we consistently found lower performance scores among the youngest children and the oldest children. In four of the five programs, there were lower performance scores for children living in rural areas compared to urban areas, but one program had lower performance scores for children living in urban areas. In the two programs that had sufficiently complete race and ethnicity data, the highest performance was among Hispanic children compared with non-Hispanic white and non-Hispanic black children. We note, however, that variations may be different across programs and plans and emphasize that it is important for each measured entity to evaluate variations in the context of the population that it serves.

(4) Anticipated Impacts

Measured entities can use information related to variations in care to focus and inform their root cause analysis as well as to target outreach and improvement efforts. As noted above, it is important that each measured entity evaluates variations of care specifically for the population that it serves since dental service use patterns may differ across programs and plans. The DQA User Guide provides guidance on how to implement optional stratifications in addition to the required age stratifications. In addition, the Dental Quality Alliance's Oral Healthcare Quality Dashboard (https://www.ada.org/resources/research/dental-quality-alliance/dqa-impr...) reports the measure with the following stratifications: age, geographic location, language, race/ethnicity, and sex. Thus, state Medicaid programs can use this resource to understand how performance varies by beneficiary characteristics.

4.1a Data Structure and Availability

All data elements are available in electronic data sources as structured data elements. This measure relies on standard fields routinely captured and used in administrative enrollment and claims data. The critical data elements required for this measure are:

- Age
- Enrollment
- Current Dental Terminology procedure codes
- NUCC maintained provider taxonomy codes

These data elements are routinely captured as structured data in plan and program enrollment and claims databases. This measure meets the feasibility requirements for reporting in the CMS Medicaid Child Core Set. This measure also was the only oral health measure selected for inclusion in the CMS Core Set of Health Care Quality Measures for 1945A Health Home Programs.

4.1b Implementation Costs and Burden

There is minimal burden for measure implementation. As noted above, it relies on standard data fields contained within administrative enrollment and claims data. The measure calculation logic is not complex. This is a straightforward measure to report.

4.1c Confidentiality

Implementation would follow standard approaches for data protection and confidentiality as would be used with any administrative enrollment and claims data for reporting claims-based measures. There are no special considerations for this measure.

4.3 Feasibility Informed Final Measure

As noted above, this is a claims-based measure that relies on standard, commonly used data fields. Because of the high feasibility of implementation, there was no need to adjust the measure based on feasibility considerations.

4.4 Proprietary Information

Proprietary measure or components (e.g., risk model, codes), without fees

4.4a Fees, Licensing, or Other Requirements

This measure is intended to be transparent and available for widespread adoption. As such, it was purposefully designed to avoid using software or other proprietary materials that would require licensing fees. The measure specifications, including a companion User Guide, are accessible through the DQA website and available free of charge for non-commercial (direct revenue generating) purposes: “Use by individuals or other entities for purposes consistent with the DQA’s mission and that is not for commercial or other direct revenue generating purposes is permitted without charge.”

5.1.1 Data Used for Testing

Maintenance Testing :

Program Level Assessments

Maintenance evaluation performance evaluations were conducted using Medicaid enrollment and claims data contained within the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAFs) available from the Centers for Medicare and Medicaid Services (CMS) through a data use agreement. We used data from calendar year 2022 for 47 state Medicaid programs assessed as having sufficiently complete data for all critical data elements in the T-MSIS database.

Plan-Level Assessments

This measure is specified for reporting at both the program (e.g., Medicaid and CHIP) and plan (e.g., dental and health plans) levels for both public and private/commercial reporting. Our original testing, the basis for initial endorsement and prior continued maintenance, included plan-specific data (please see below). The DQA measure of Oral Evaluation has been implemented by dental plans operating in commercial, Health Insurance Marketplace, and Medicaid/CHIP markets. The T-MSIS data used for our updated testing currently do not enable reliable identification of which dental services are provided by which dental plans. This does not represent a feasibility issue for Medicaid programs and their participating plans to calculate the

measure at the plan level. We know from working with state Medicaid programs and state Health Insurance Marketplaces that it is highly feasible to have plan-level reporting of dental quality measures. Rather, this reflects a limitation of the database that we used for testing.

There is some limited plan-level reporting that is publicly available that we have included:

- Plans participating in the Texas Medicaid and CHIP programs for 2013 through 2022.
  - 2 plans serving Texas Medicaid and Texas CHIP (same two plans serving each program), 2013-2020.
  - 3 plans serving Texas Medicaid and Texas CHIP (same three plans serving each program), 2021-2022.
  - Denominators, numerators, and scores are available for each plan.

- CCOs participating in Oregon Medicaid and CHIP for 2018 through 2023.
  - 13 CCOs, 2018-2019
  - 15 CCOs, 2020
  - 16 CCOs, 2021-2023
  - Only measure scores are available at the plan level. Denominators are available only at the statewide level. However, there is a flag for each CCO as to whether there are small denominator size concerns (<30), and there were no small denominators for any of the CCOS in any of the years.

Because plan-level data are limited, we were also limited in the extent of testing. However, the original testing data and publicly available data indicate that plan performance gaps and variation exist.

Based on prior testing of this and other administrative claims-based dental measures, 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.

The main potential concern would be if the plan level denominators were too small to yield reliable results. However, the denominator requirements for this measure capture a broad population with thousands of patients in the denominator of individual plans (please see sample sizes in reliability reporting).

This measure has been in use for more than 10 years, and we have not encountered issues with small denominator sizes in our testing or in feedback from the user community. The DQA

membership includes the National Association of Dental Plans as well as representatives from specific dental plans that serve both commercial and publicly insured populations throughout the United States, and insufficient denominator sizes have not arisen as an issue. The DQA is also in frequent communication with the Centers for Medicare & Medicaid Services (CMS), as well as with Mathematica which serves as CMS's technical resource to state Medicaid programs for quality measure implementation; there have been no issues raised related to concerns about small denominator sizes related to plan-level implementation.

Original Measure Testing: Program and Plan Levels

We used data from five sources. We included data for publicly insured children in the Texas Medicaid, Texas CHIP, Florida CHIP, and Florida Medicaid programs as well as national commercial data from Dental Service of Massachusetts, Inc. Testing data also included separate analyses for the two dental plans that served the Florida CHIP program. Florida and Texas represent two of the largest and most diverse states. The two states also represent the upper and lower bounds of dental utilization based on dental utilization data available from the Centers for Medicare and Medicaid Services. The five programs collectively represent different delivery system models. The Texas Medicaid data represented dental fee-for-service, and Texas CHIP data reflected a single dental managed care organization (MCO). The Florida CHIP data included data from two dental MCOs. The Florida Medicaid data include dental fee-for-service and prepaid dental data. The commercial data included members in indemnity and preferred provider organization (PPO) product lines. We used data from calendar years 2010 and 2011 for all programs except Florida Medicaid. Full-year data for 2011 were not available for Florida Medicaid.

5.1.1a Dates of Testing Data

Field not required for Spring 2025.

5.1.2 Differences in Data

Maintenance Testing

As noted above, there are differences in the data sources based on level of analysis:

Program Level

T-MSIS data for 47 state Medicaid programs were used for testing at the program-level of analysis.

Plan Level

- Plans participating in the Texas Medicaid and CHIP programs for 2013 through 2022.
  - 2 plans serving Texas Medicaid and Texas CHIP (same two plans serving each program), 2013-2020.
  - 3 plans serving Texas Medicaid and Texas CHIP (same three plans serving each program), 2021-2022.
  - Denominators, numerators, and scores are available for each plan.
- CCOs participating in Oregon Medicaid and CHIP for 2018 through 2023.
  - 13 CCOs, 2018-2019
  - 15 CCOs, 2020
  - 16 CCOs, 2021-2023
  - Only measure scores are available at the plan level. Denominators are available only at the statewide level. However, there is a flag for each CCO as to whether there are small denominator size concerns (<30), and there were no small denominators for any of the CCOS in any of the years.

There were no differences in data sources across the different aspects of testing.

Original Testing

The following additional data sources (in addition to the administrative claims data previously described) were used for face validity assessments and data element validation:

1. **Data element validation using medical chart reviews** included patient record data from the Texas Medicaid and CHIP programs to validate the data elements used to calculate the measure. Texas has the third largest Medicaid program and second largest CHIP in the U.S., both with significant diversity represented. In addition, the research team that conducted the testing was the External Quality Review Organization for Texas and had years of experience conducting medical chart audits for the Texas Medicaid and CHIP programs for ongoing quality assurance purposes. Thus, an established infrastructure and expertise was in place to conduct chart reviews for these programs.

2. **Face validity** assessments included expert consensus processes, including conducting an environmental scan of measure concepts and using the RAND-UCLA modified Delphi process to rate the importance, feasibility and validity. Please see section 5.3.3 for a complete description.

5.1.3 Characteristics of Measured Entities

Maintenance Testing

Program Level

Maintenance evaluation performance evaluations were conducted using Medicaid enrollment and claims data contained within the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAFs) available from the Centers for Medicare and Medicaid Services (CMS) through a data use agreement. We used data from calendar year 2022 for 47 state Medicaid programs assessed as having sufficiently complete data for all critical data elements in the T-MSIS database.

Plan Level

- Plans participating in the Texas Medicaid and CHIP programs for 2013 through 2022.
    - 2 plans serving Texas Medicaid and Texas CHIP (same two plans serving each program), 2013-2020.
    - 3 plans serving Texas Medicaid and Texas CHIP (same three plans serving each program), 2021-2022.
    - Denominators, numerators, and scores are available for each plan.
- CCOs participating in Oregon Medicaid and CHIP for 2018 through 2023.
    - 13 CCOs, 2018-2019
    - 15 CCOs, 2020
    - 16 CCOs, 2021-2023
    - Only measure scores are available at the plan level. Denominators are available only at the statewide level. However, there is a flag for each CCO as to whether there are small denominator size concerns (<30), and there were no small denominators for any of the CCOS in any of the years.

Original Testing

We included data for publicly insured children in the Texas Medicaid, Texas CHIP, Florida CHIP, and Florida Medicaid programs as well as national commercial data from Dental Service of Massachusetts, Inc. Testing data also included separate analyses for the two dental plans that served the Florida CHIP program. Florida and Texas represent two of the largest and most diverse states. The two states also represent the upper and lower bounds of dental utilization based on dental utilization data available from the Centers for Medicare and Medicaid Services. The five programs collectively represent different delivery system models. The Texas Medicaid data represented dental fee-for-service, and Texas CHIP data reflected a single dental managed care organization (MCO). The Florida CHIP data included data from two dental MCOs. The Florida Medicaid data include dental fee-for-service and prepaid dental data. The commercial

data included members in indemnity and preferred provider organization (PPO) product lines. We used data from calendar years 2010 and 2011 for all programs except Florida Medicaid. Full-year data for 2011 were not available for Florida Medicaid.

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

Programs

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

Texas Medicaid, 2011: 3,544,247

Texas Medicaid, 2010: 3,393,963

Florida CHIP, 2011: 317,146

Florida CHIP, 2010: 315,975

Commercial, 2011: 184,152

Commercial, 2010: 189,968

Florida Medicaid, 2010: 2,068,670

Plans

Within these programs, we had claims data available in both years for two dental managed care plans in Florida CHIP.

Plan 1, 2010: 77,255

Plan 2, 2010: 116,388

Plan 1, 2011: 140,986

Plan 2, 2011: 168,191

5.1.4 Characteristics of Units of the Eligible Population

Note: Please see Attachment CBEID_2517_5_1_4_PatientCharacteristics for detailed data summaries.

Maintenance Testing

Program Level

The first table (5.1.4.A) in the attachment summarizes the characteristics of the 33,381,516 denominator-eligible children across the 47 Medicaid programs by age, race and ethnicity, geographic location, sex, and language (CY 2022). The highest percentages of children by age are in the categories of 15-18 years (18%) followed by 3-5 years (17%) and 12-14 years (15%). By race and ethnicity, the highest percentages were non-Hispanic white (30%), Hispanic (28%), and non-Hispanic black (20%); 15% of children had missing data for race and ethnicity. 83% resided in urban areas, and 17% in urban areas. 51% were female and 49% were male. 71% spoke English as the primary language at home, and 11% spoke Spanish as the primary language at home; 18% of children had missing data for language.

Plan Level

The second table (5.1.4.B) in the attachment summarizes the characteristics of the 547,585 children in the Oregon Health Plan in December 2023 by age, race and ethnicity, and language. These data come directly from the Oregon Health Plan website and reflect the patient demographic categories as defined by OHP. 27% of children were aged 0-5 years, 25% were aged 6-10 years, 35% were aged 11-17 years, and 13% were aged 18-20 years. By race and ethnicity, the highest percentages were non-Hispanic white (36%) and Hispanic (28%); 16% of children had missing or declined data for race and ethnicity. 65% spoke English as the only language at home, and 13% spoke Spanish as the only language at home; 18% of children were excluded due to age less than 5 and an additional 1.1% of children had missing language data.

The third table (5.1.4.C) in the attachment summarizes the characteristics of the 4,406,154 children enrolled in the Texas Medicaid and CHIP programs in CY 2022 by age, race and ethnicity, geographic location, sex, and language. Breakouts of enrollment by patient characteristics were not publicly available directly from the Texas Medicaid and CHIP website. Thus, we relied on the same T-MSIS data source used for program-level testing. The highest percentages of children by age are in the categories of 15-18 years (18%), 3-5 years (16%), and 12-14 years (15%). By race and ethnicity, the highest percentages were Hispanic (51%), followed by non-Hispanic white (15%), and non-Hispanic black (13%); 18% of children had missing data for race and ethnicity. 88% resided in urban areas, and 12% in urban areas. 49% were female and 51% were male. 74% spoke English as the primary language at home, and 20% spoke Spanish as the primary language at home; 6% of children had missing data for language.

Original Testing

The patient characteristic tables provided for original testing using data from the Florida and Texas Medicaid and CHIP at the program level and plan levels, as well as commercial program data, are also included in Tables 5.1.4.D and 5.1.4.E of the attachment.

5.2.1 Level(s) of Reliability Testing Conducted

Person or encounter level (i.e., data element) (e.g., inter–abstractor reliability), Accountable entity level (i.e., measure score) (e.g., signal-to-noise analysis)

5.2.2 Method(s) of Reliability Testing

Maintenance Testing

Program Level

(1) Accountable Entity Measure Score Reliability Testing using a Permutation Random Split-Sample Methodology with Spearman-Brown Correction:

Reliability indicates the extent to which repeated measurements yield consistent results. Nieser and Harris (2024) compared 14 different methods of evaluating reliability of health care quality measures. They recommended the split-sample approach as a preferred method because it does not rely on parametric assumptions about the data and is calculable for any type of quality measure.

Split Sample Methodology

To obtain an overall program-level reliability estimate, we used a split sample methodology and calculated the intraclass correlation coefficient (ICC) of the measure scores. For each of the 47 state Medicaid programs, we randomly split the population of children aged <21 years present in the T-MSIS demographic and eligibility files without replacement. The denominator, numerator, and measure score were calculated for each sample within each state. 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 a one-way random effects model to calculate the ICC to evaluate the agreement between the two 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.

Permutation Split-Sample

Nieser and Harris (2023) observed that a limitation of the split sample approach is that the reliability estimates may vary based on the random split of the data; that is, estimate may vary across different split sample draws. They consequently recommend taking many splits of the data, estimating reliability for each split, and then averaging the reliability estimates to arrive at a more stable estimate.

However, they **noted that this potential limitation of the split sample approach is more likely to arise when there are smaller sample sizes:**

"According to our simulation study, which was limited to only a single data-generating process, single-SSR estimates might suffice in cases where there are more than 250 observations per provider. However, even in these cases, researchers can use the code mentioned above to repeat the sample-splitting at least several times (e.g., 10) to gain some stability in their estimate."

In our program level data, the smallest denominator size was 47,522. Thus, the concern about unstable estimates related to small sample sizes is minimal. We repeated the sample splitting 10 times to confirm the stability of our estimate (see results below). We did not conduct additional splits given clear stability in the results and in light of the time required to conduct additional runs. Neiser and Harris (2024) note that the main drawback of the permutation split sample approach is the computational time required.

Spearman-Brown Correction

The Spearman-Brown formula was used for sample size correction to obtain reliability estimates for each of the 47 Medicaid programs.

We report the mean, minimum and maximum values, and values by decile for the Spearman-Brown adjusted ICCs calculated on 10 sets of random split samples without replacement within each Medicaid program.

(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) Data Element Reliability/Validity

The original data element validity/reliability testing using chart reviews holds for the updated measure specifications. The critical data elements have not changed. Consequently, we did not repeat data element reliability and validity testing since the reliability and validity of the data elements used are well-established from prior testing and in the peer-reviewed literature as reported in our prior submission.

Plan Level

As noted above, the T-MSIS data used for our updated testing currently do not enable reliable identification of which dental services are provided by which dental plans. This does not represent a feasibility issue for Medicaid programs and their participating plans to calculate the measure at the plan level. We know from working with state Medicaid programs and state Health Insurance Marketplaces that it is highly feasible to have plan-level reporting of dental quality measures.

As described above, there is publicly available plan-level reporting for Texas Medicaid and CHIP and the Oregon Health Plan (Oregon's Medicaid and CHIP program). However, complete data to enable in-depth reliability testing is limited.

- The Texas reporting includes denominator sizes, but only two plans are represented in most years.
- The Oregon Health Plan includes more reporting entities (>=13) but denominator/sample size data are not available.

Consequently, we conducted the following assessments:

- **Reported measure scores for all years for each plan, including 95% confidence intervals** for the plans in the Texas and Medicaid CHIP programs.

- **Evaluated relative rankings of the overall measure scores between pairs of consecutive years using the Spearman rank correlation coefficients** to evaluate whether there were substantial changes that could suggest a threat to reliability. Because of the greater number of reporting entities, we used the Oregon Health Plan program for this evaluation. Specifically, using the measure scores for the Oregon Health Plan CCOs, we calculated the Spearman rank correlation coefficients for each pair of consecutive years.

- **Examined plan-level denominator sizes.** The main potential threat to plan-level reliability compared with program-level reliability is if the plan level denominators are too

small to yield reliable results. However, the denominator requirements for this measure capture a broad population (children <21 years enrolled for at least six months continuously during the reporting year). To confirm this, we report the denominator sizes for each plan included in our testing. For plans participating in Texas Medicaid and CHIP, we had specific information on sample size. For the Oregon Health Plan, we did not have the sample size for each CCO, so we reported the average sample size based on the state-wide denominator.

Original Testing

(1) Data Element Reliability/Validity

Clarification note: Following CBE guidance at the time of original testing, reliability for original endorsement was established through:

1. Data element reliability was demonstrated through data element validation using chart audits as described in Section 5.3.3 below.
2. Evaluation and refinement of measure specifications to ensure that they were detailed, clear, and complete.
3. Voting by the DQA membership on whether reliability criterion was met.

Data Elements:

- See section 5.3.3 for validity testing of data elements.
- Note: Unlike measures that rely on medical record data for which issues such as inter-rater reliability are likely to introduce measurement concerns or measures that rely on survey data for which issues such as internal consistency may be a concern, this measure relies on standard data fields commonly used in administrative data for a wide range of billing and reporting purposes.

(2) Measure Score – 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 initial measure specifications were developed by the Dental Quality Alliance (DQA). The Dental Quality Alliance includes 30 members, representing a broad range of stakeholders, including federal agencies involved with oral health services, dental professional associations, medical professional associations, dental and medical health insurance commercial plans, state Medicaid and CHIP programs, quality accrediting bodies, and the general

public. The initial specifications were developed based on (1) evidence, (2) an environmental scan, and (3) face validity assessments of the measure concept. These specifications were contained in the competitive Request for Proposals to conduct measure testing; a research team was selected to conduct testing. The research team independently carefully evaluated whether the measure specifications identified all necessary data elements to calculate the numerators and denominators for each measure. In addition, the research team carefully reviewed the logic flow and made revision recommendations to improve the reliability of the resulting calculations. The DQA also solicited public comment on an Interim Report and posted the measurement specifications online for public comment. The research team worked with the DQA to evaluate and address all comments provided. Throughout the testing period, there were numerous reviews and revisions of the specifications conducted jointly by the research team and the DQA to ensure clear and detailed measure specifications.

Since original measure development and endorsement, the DQA has undergone an Annual Measure Review process that includes a public comment period and review by the DQA membership. No refinements to the measure have been made.

2. Other Threats to Reliability - Sample Size

The measured entities include very 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.

Nieser KJ, Harris AHS. Comparing methods for assessing the reliability of health care quality measures. Stat Med. 2024;43(23):4575-4594. PMID: 39145538.

Nieser KJ, Harris AHS. Split-sample reliability estimation in health care quality measurement: Once is not enough. Health Serv Res. 2024;59(4):e14310. PMID: 38659301.

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

5.2.3 Reliability Testing Results

Note: Please see Attachment CBEID_2517_5_2_3_ReliabilityTestingResults for detailed tables.

MAINTENANCE TESTING: PROGRAM LEVEL, ACCOUNTABLE ENTITY

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

Permutation Split Sample ICC Estimates

Table 5.2.3.A. in the attachment provides the split sample overall ICC estimates with 95% confidence intervals for each of the 10 repeated random split samples (without replacement) for the 47 Medicaid programs. To summarize:

Mean ICC: 0.999678

Minimum ICC: 0.9993478

Maximum ICC: 0.9997328

Following the guidance in Koo and Li (2016), this indicates "excellent" reliability given that even the lower bound of each of the 95% CIs is greater than 0.90. Moreover, there is significant stability across the 10 estimates with all ICC values greater than 0.999.

Spearman-Brown Corrected Entity-Level Reliability Estimates

For the 47 Medicaid programs, the reliability estimates ranged from a minimum of 0.9953 to a maximum of 0.9999, with mean reliability of 0.9990. Even the lowest reliability estimate is within the “excellent” range.

Please see Table 5.2.3.B in the attachment for detailed results by decile.

Performance Scores with 95% CI

To further illustrate the reliability of the measure scores, Table 5.2.3.C in the attachment provides the overall performance scores and the performance scores for one set of the split samples with their 95% confidence intervals. There were overlapping 95% CIs within each state Medicaid program for all three scores (overall population, sample 1, and sample 2), demonstrating consistency of measurement.

(2) Relative Rankings: Split Samples

Table 5.2.3.D contains the performance score statistics for the overall population and one set of split samples by decile. The results were consistent between the split samples. Moreover, the same states remained in the same deciles. States did not move between deciles when comparing the overall population, sample 1, and sample 2.

MAINTENANCE TESTING: PLAN LEVEL, ACCOUNTABLE ENTITY

(1) Oregon Health Plan (Medicaid and CHIP) CCOs: Correlation between Years

The relative rankings based on measure scores between pairs of years were as follows:

Year-to-Year Comparison, Spearman's rank correlation coefficient, p-value

2018 & 2019 (n=13), 0.9615, <0.0001

2019 & 2020 (n=13),* 0.7912, 0.0013

2020 & 2021 (n=15),* 0.9393, <0.0001

2021 & 2022 (n=16), 0.9235, <0.0001

2022 & 2023 (n=16), 0.9353, 0.0001

*We hypothesized potentially lower correlations for the evaluations between 2019 & 2020 and 2020 & 2021 because (1) The number of CCOs between the two years were different. There were 13 CCOs in 2019, 15 in 2020, and 16 in 2021. The relative rankings were based only on those plans present in both of the two years being compared. (2) Plans may have experienced relatively different magnitudes of reduced service use from COVID-19 impacts.

The correlation coefficients ranged from 0.7912 to 0.9615 indicating a “strong” to “very strong” correlation (Akoglu, 2018). Data are also presented in Table 5.2.3.E in the attachment.

(2) Texas Medicaid and CHIP: Performance Scores with 95% CI

With only 2-3 plans within a program, we were limited in the ability to conduct robust reliability testing. Table 5.2.3.F in the attachment provides the overall performance scores with their 95% confidence intervals for each of the plans in Texas Medicaid and in Texas CHIP for years 2013, 2022. These scores demonstrate an overall gradual improvement in performance in 2013 through 2019 for each of the two plans that participated in the programs for all years. As expected, there was a performance disruption (decline) associated with the COVID-19 impacts on service use with a subsequent gradual increase in performance again from 2020 through 2022. The overall results are consistent with expectations and do not demonstrate signs of unexpected variations in performance over time that could suggest threats to reliability.

(3) Denominator Size Summary

Table 5.2.3.G in the attachment summarizes the sample sizes for the plan-level entities used in testing. For plans participating in Texas Medicaid and CHIP, we have specific information on sample size. For the Oregon Health Plan, we do not have the sample size for each CCO, so we report the average sample size based on the overall state denominator. The Oregon Health Plan does flag for each CCO whether there are small denominator size concerns (<30), and there were no small denominator flags for any of the CCOS in any of the years for this measure.

Denominator sizes ranged from a low of 8,860 (UnitedHealthCare Dental, Texas CHIP, 2022) to a high of 2,153,439 (DentaQuest, Texas Medicaid, 2022). The average denominator size for the Oregon CCOs ranged from 27,779 in 2020 to 31,841 in 2023. The smallest sample size was 8,860 members for UnitedHealthCare Dental in the Texas CHIP program in 2022, which began participating in the program in 2021. Thus, even the lowest denominator has thousands of children represented. These values indicate little concern about low plan denominator sizes. This finding is consistent with historical experience. This measure has been in use for more than 10 years, and we have not encountered issues with small denominator sizes in our testing or in feedback from the user community.

ORIGINAL TESTING: ENCOUNTER LEVEL

Our original testing demonstrated the reliability of the data elements used to calculate the measures by establishing data element validity). Please see section 5.3.4 for data element reliability/validity results from the original testing based on chart audits.

Reference

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

5.2.3a Attach Additional Reliability Testing Results

CBEID_2517_5_2_3_ReliabilityTestingResults.pdf

5.2.4 Interpretation of Reliability Results

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

1. Program-level reliability estimates calculated using a permutation split sample methodology with Spearman-Brown correction ranged from a minimum of 0.9953 to a maximum of 0.9999, with a mean value of 0.9990. Following the guidance of Koo and Li (2016), this indicates "excellent" reliability.

2. The program-level measure scores between the split samples for each state were similar and have overlapping 95% CIs.

3. The program-level relative rankings based on measure scores were stable.

4. Plan-level relative rankings based on measure scores are stable between years. The Spearman's rank correlation coefficient ranged from 0.7912 to 0.9615, indicating a "strong" to "very strong" 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 (Akoglu, 2018), we consider it is a fair characterization to classify the association as "strong" to "very strong."

5. Plan-level denominators include thousands of patients, mitigating concerns about small sample size as a threat to reliability.

Our original testing demonstrated the reliability of the data elements used to calculate the measures (by establishing data element validity). This updated testing demonstrates the reliability of the performance measure scores at the accountable entity level.

Table 2. Accountable Entity Level Reliability Testing Results by Denominator, Target Population Size

Accountable Entity-Level Reliability Testing Results

	Overall	Minimum	Decile_1	Decile_2	Decile_3	Decile_4	Decile_5	Decile_6	Decile_7	Decile_8	Decile_9	Decile_10	Maximum
Reliability	0.9990	0.9953	0.9965	0.9979	0.9987	0.9993	0.9995	0.9996	0.9997	0.9997	0.9998	0.9999	0.9999
Mean Performance Score	41.6%	26.0%	42.0%	39.6%	41.3%	41.4%	45.4%	41.8%	44.2%	42.6%	38.6%	40.7%	56.8%
N of Entities	47	1	5	5	4	5	4	5	4	5	5	5	1
N of Persons / Encounters / Episodes	33,381,516	47,522	336,027	532,212	718,168	1,580,683	1,691,977	2,758,678	2,588,969	3,946,354	5,715,304	13,513,144	3,844,834

5.3.1 Level(s) of Validity Testing Conducted

Person or encounter level (i.e., data element) (e.g., sensitivity and specificity), Accountable entity level (i.e., measure score) (e.g., criterion validity)

5.3.2 Type of Accountable Entity Level Validity Testing Conducted

Empirical validity testing at the accountable entity-level (e.g., criterion validity, construct validity, known groups analysis), Systematic assessment of face validity of the measure's performance score as an indicator of quality or resource use

5.3.3 Method(s) of Validity Testing

TESTING OVERVIEW

Maintenance Testing – Program and Plan Levels:

- Accountable entity: empirical validity testing
- Missing data evaluation

Original Testing:

- Person/encounter: critical data element testing
- Accountable entity: face validity assessment
- Missing data evaluation

Maintenance Testing: Accountable Entity Empirical Validity Testing

Previous testing focused on data element validity and accountable entity measure score face validity. Therefore, we focused additional testing on accountable entity level validity testing. Ideally, to do so, we would identify a "gold standard" outcomes indicator of quality that is hypothesized to be associated with this process measure. However, we are limited in available dental quality measures. The DQA was formed at the request of CMS precisely because there was a lack of validated dental quality measures. Consequently, the other rigorously tested quality

measures against which this measure can be validated are other DQA measures. Moreover, these are mainly process measures. Dental claims data do not contain diagnosis codes, which significantly limits the ability to develop outcome measures, and the field of dentistry is in nascent stages of developing PRO-PMs.

(1) Program Level

Comparators for Correlation Analysis and Rationale for Hypothesized Relationships

We selected two other DQA measures that we hypothesized to be positively correlated with Oral Evaluation, Dental Services:

1. Utilization of Services, Dental Services (CBE 2511). This is a broad measure that evaluates whether a child received any dental service during the reporting period, regardless of the type of service. It serves as an indirect indicator for access, indicating whether the child has had any contact with the dental care delivery system. State Medicaid programs in which a greater percentage of children access the dental care delivery system (i.e., receive some type of dental service) were expected to have higher measure scores for Oral Evaluation than those in which fewer children are using the dental care system. Thus, we hypothesized a positive moderate to strong correlation between the Oral Evaluation and Utilization of Services measures. Thus, we hypothesized a positive moderate to strong correlation between the Oral Evaluation and Utilization of Services measures

2. Topical Fluoride, Dental Services (CBE 2528). This measures whether a child received at least two topical fluoride applications during the reporting period. Children who receive a comprehensive or periodic oral evaluation are more likely to be established into routine care (versus seeking only problem- or emergency-based dental care) that includes risk assessment and care planning that includes prevention. Thus, we hypothesized a positive moderate to strong correlation between the Oral Evaluation and Topical Fluoride measures. Thus, we hypothesized a positive moderate to strong correlation between the Oral Evaluation and Topical Fluoride measures.

Methodology

Using the measure scores for the 47 state Medicaid programs, we calculated the Spearman's rank correlation coefficient. We assessed concordance between: (1) Oral Evaluation and Utilization of Services and (2) Oral Evaluation and Topical Fluoride.

(2)Plan Level

Comparators for Correlation Analysis and Rationale for Hypothesized Relationships

We used performance scores for the 16 CCOs participating in the Oregon Health Plan in 2022 and

in 2023. We did not conduct this analysis for the Texas Medicaid and CHIP programs due to the small number of plans in those programs.

The Oregon Health Plan has a counterpart to Utilization of Services, but it is reported for children and adults combined (and not children only) so it is not an appropriate comparator.

The Oregon Health Plan does report Topical Fluoride for children. As above, we hypothesized a positive moderate to strong correlation between the Oral Evaluation and Topical Fluoride measures.

Methodology

Using the measure scores for the 16 Oregon CCOs, we calculated the Spearman's rank correlation coefficient to assess concordance between Oral Evaluation and Topical Fluoride. We calculated this separately for performance years 2023 and 2022.

Maintenance Testing: Missing data evaluation, program level

Updated testing used data from 47 Medicaid programs using data 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. (Centers for Medicare & Medicaid Services. *DQ Atlas*. Available at https://www.medicaid.gov/dq-atlas/) 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 Oral Evaluation:

- **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:

- **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.
- **Rendering Provider Taxonomy for Dental Procedure Codes**. We evaluated how often valid, active dental procedure codes had missing data on the rendering provider type (i.e., did not have an accompanying provider taxonomy code).
- **Dental Claims Volume Quality Assessment.** We evaluated claims volume relative adjusted for program size to identify outliers that may signify incomplete data.

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%

Original Testing: Person/encounter critical data element validity

Oral evaluation measures the percentage of children who received a comprehensive or periodic oral evaluation using procedure codes in administrative claims data to identify clinical oral evaluations. Thus, assessing the accuracy of procedure codes reported in the claims data is essential. The critical data elements for this measure include: (1) member ID (to link between claims and enrollment data), (2) date of birth, (3) monthly enrollment indicator, (4) date of service, and (5) Current Dental Terminology (CDT) codes. The first four items are core fields used in virtually all measures relying on administrative data and essential for any reporting or billing purposes. As such, it was determined that these fields have established reliability and validity. Thus, critical data element validity testing focused on assessing the accuracy of the dental procedure codes reported in the claims data as the data elements that contribute most to the measure score. To evaluate data element validity, we conducted reviews of dental records for the Texas Medicaid and CHIP programs. Validation of clinical codes in administrative claims data are most often conducted using manual abstraction from the patient's full chart as the authoritative source. As described in detail below, we evaluated agreement between the claims data and dental charts by calculating the sensitivity, specificity, positive predictive value, and negative predictive value as well as the kappa statistic.

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. The number of eligible records received (414) exceeded the total targeted number of 400 records.

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 **oral evaluation 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.

Methodology: 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 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.

Methodology: Oral Evaluation Procedures Codes

Data Extraction. For the second component of validation, assessing whether oral evaluations are accurately captured by claims data, chart abstraction forms were developed by the research team to document evidence in the dental record that an oral evaluation had been performed. The chart abstraction forms and process were reviewed and approved by the DQA Measure 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 an oral evaluation being performed 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 CDT codes for oral evaluations. 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.)

Original Testing: Accountable entity face validity assessment

Face Validity Assessment – Measure Development

Face validity was systematically assessed by recognized experts. The Dental Quality Alliance (DQA) was formed at the request of the Centers of Medicare and Medicaid Services (CMS) specifically for the purpose of bringing together recognized expertise in oral health to develop quality measures through consensus processes.

During the measurement development process, the DQA Measure Development and Maintenance Committee, purposely comprised of individuals with recognized and appropriate expertise in oral health to lead quality measure development, undertook an environmental scan of existing pediatric oral health performance measures, which involved the following: (1) Literature Search, (2) Measure Solicitation, (3) Review of Measure Concepts, (4)Delphi Ratings of Measure Concepts, (5) Scan Results Analysis, (6) Gap Analysis, (7) Identification of Measures. A more detailed description of this process, the findings and the resulting measure concepts that were pursued is provided in reports published by the DQA. (Dental Quality Alliance. Pediatric Oral Health Quality and Performance Measures: Environmental Scan. 2012; Dental Quality Alliance. Pediatric Oral Health Quality & Performance Measure Concept Set: Achieving Standardization & Alignment. 2012).

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

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

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

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

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

process, 38 were deemed “low-scoring measure concepts” and 21 were deemed “high-scoring measure concepts.”

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

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

Face Validity Assessment – Measure Testing

The research team and the DQA Committee continued to assess face validity throughout the testing process. Face validity also was gauged through feedback solicited through public comment periods. In March 2013, an Interim Report describing the measures, testing process, and preliminary results was sent to a broad range of stakeholders, including representatives of federal agencies, dental professionals/professional associations, state Medicaid and CHIP programs, community health centers, and pediatric medical professionals/professional associations. Each comment received was carefully reviewed and addressed by the research team and DQA, which entailed additional sensitivity testing and refinement of the measure specifications. Draft measure specifications were subsequently posted on the DQA’s website in a public area and public comment was invited. National presentations, including presentations at the National Oral Health Conference, were made by the research team and DQA in the spring and summer of 2013, which included reference to the website containing the measure specifications and invitations to provide feedback. All comments received were reviewed and addressed by the research team and DQA, including additional sensitivity testing and refinement of the measure specifications.

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

Original Testing: Missing data evaluation

The administrative databases were queried to assess the extent of missing or invalid data for the critical data elements used to calculate the measure.

References

Brook RH. The RAND/UCLA appropriateness method. In: McCormick KA, Moore SR, Siegel R, United States. Agency for Health Care Policy and Research. Office of the Forum for Quality and Effectiveness in Health Care., editors. Clinical practice guideline development: methodology perspectives.

Dental Quality Alliance. Pediatric Oral Health Quality and Performance Measures: Environmental Scan. 2012; Dental Quality Alliance. Pediatric Oral Health Quality & Performance Measure Concept Set: Achieving Standardization & Alignment. 2012.

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.

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.

5.3.4 Validity Testing Results

Note: Please Attachment CBEID_2517_5_3_4_ValidlityTestingResults for tables containing validity testing results data.

TESTING OVERVIEW

Maintenance Testing:

- Accountable entity: empirical validity testing
- Missing data evaluation

Original Testing:

- Person/encounter: critical data element testing
- Accountable entity: face validity assessment
- Missing data evaluation

Maintenance Testing: Accountable Entity Empirical Validity Testing

Program Level

Below are the Spearman rank correlation coefficients evaluating associations between (1) Oral Evaluation and Utilization of Services and (2) Oral Evaluation and Topical Fluoride:

- Oral Evaluation and Utilization of Services (CBE 2511), CY 2022: 0.9352 ($p<0.0001$)
- Oral Evaluation and Topical Fluoride (CBE 2528), CY, 2022: 0.8615 ($p<0.0001$)

These values 0.9352 and 0.8615 indicate “strong” to “very strong” associations. Please also see Table 5.3.4.A in the attachment.

Plan Level

Below are the Spearman rank correlation coefficients evaluating associations between Oral Evaluation and Topical Fluoride among the Oregon CCOs for 2022 and 2023:

- Oral Evaluation and Topical Fluoride, CY 2022: 0.6794 ($p=0.0048$)
- Oral Evaluation and Topical Fluoride, CY 2023: 0.6667 ($p=0.0038$)

These values indicate “moderate” to “strong” associations. Please also see Table 5.3.4.B in the attachment.

Maintenance Testing: Data Completeness and Quality Evaluation

Table 5.3.4.C in the attachment contains the results of the assessments of data completeness, missing data, and invalid data. With the exception of two states, all 47 states had data quality of “low concern” or “medium concern” with low rates of missing data. Massachusetts and New Jersey had a “high concern” assessment for “Claims Volume – Other Services (Outpatient)” in the

CMS Data Quality Atlas. Further evaluation revealed that this assessment was based on having a relatively high claims volume. Because services are deduplicated in the measure calculation (the measure requires “at least one” service), even if there were duplication of services, it would not inflate the measure score. There is a greater concern when claims volume is relatively low because of the risk of undercounting services due to incomplete claims data.

These findings are consistent with original testing data, which found very low rates of missing and invalid data (see below). These findings are consistent with expectations of generally low rates of missing data for these data elements because they are routinely used for a range of payment and reporting purposes. However, claims data are not always perfect and there may be various factors that periodically impact data completeness and quality. Because these are standard data fields contained within administrative claims data, when a program or plan is identified as having incomplete or poor quality data, they are encouraged to improve data collection and quality as an important component of their quality improvement efforts rather than relying on statistical methods to address missing data.

Prior Testing: Critical Data Element Validity

1. Critical Data Element Validity

A. Encounter Data Validation – Overall Assessment

Please see Table 5.4.3D in the attachment. Encounter data validation of 1,135 procedure codes in the claims data against dental charts found agreement for 94% of the procedure codes. Only 4.2% of procedure codes reported in the administrative data were not supported by evidence in the dental record. For 1.8% of the records reviewed, the documentation was insufficient to determine whether the service indicated by the procedure code had been rendered or not.

B. Critical Data Element Validation – Dental Service Procedure Codes for Oral Evaluations

To assess whether oral evaluations performed are accurately captured by claims data, the 414 records, representing 631 dates of service, were reviewed. Table 5.3.4.E in the attachment summarizes the agreement between the dental records and administrative data.

Agreement (concordance) between the dental records and administrative claims data was 86.6%. Sensitivity was 85.1% and specificity was 92.5%. The positive predictive value was 97.9%, and the negative predictive value was 59.7%.

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 0.642, indicating "substantial" agreement.

Collectively, these findings indicate strong concordance with a greater likelihood of false negatives than false positives. Evaluating dental records for documented evidence oral evaluations was more challenging than identifying whether other specific procedures were performed, such as topical fluoride application or restorative procedures, because oral evaluations encompass a set of services and there is greater variability in charting practices related to documenting oral evaluations. The RHITs erred on the side of being over-inclusive in recording evidence of an oral evaluation, which may have contributed to the finding of a greater likelihood of false positives.

We compared our findings 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. They evaluated "oral examinations," which were broadly defined. For oral examinations, they found lower sensitivity (42%), similar specificity (96%), and a lower kappa value (0.44). They noted, however, that the categories in the form they used to identify oral examinations through observation were general in nature and "included any activity that was used to determine the oral health or status of a patient from simple mouth mirror examinations to Diagnodent evaluation." (Demko CA, Victoroff KZ, Wotman S. 2008. "Concordance of chart and billing data with direct observation in dental practice" Community Dent Oral Epidemiol. 36(5):466-74.)

2. Face Validity

Oral Evaluation, and specifically a comprehensive or periodic oral evaluation, was identified through the Delphi rating process as a high-scoring measure concept with a mean importance score of 8, mean feasibility score of 8, and mean validity score of 8, all out of a 9-point scale. [Rating of 1-3: not scientifically sound and invalid; 4-6 – uncertain scientific soundness and uncertain validity; 7-9 – scientifically sound and valid.] Median score ratings were equal to the mean ratings. Thus, the measure has face validity. However, gaps were identified with existing measures, including defining "diagnostic services" or "examination" too broadly, lack of clear specifications, and lack of standardization.

Missing Data (Original Testing)

Missing data evaluations from prior testing:

PROGRAM 1

Member ID: 0.00%

Date of Birth: 0.00%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.00%

Date of Service: 0.01%

Rendering Provider ID: 0.28%

PROGRAM 2

Member ID: 0.00%

Date of Birth: 0.00%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.00%

Date of Service: 0.00%

Rendering Provider ID: 0.00%

PROGRAM 3

Member ID: 0.27%

Date of Birth: 0.00%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.28%

Date of Service: 0.00%

Rendering Provider ID: 0.18%

PROGRAM 4

Member ID: 0.00%

Date of Birth: 0.00%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.01%

Date of Service: 0.00%

Rendering Provider ID: 0.61%

PROGRAM 5

Member ID: 0.43%

Date of Birth: 0.02%

Monthly enrollment indicator: 0.00%

Dental Procedure Codes - CDT: 0.00%

Date of Service: 0.00%

Rendering Provider ID: 0.67%

References

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

5.3.4a Attach Additional Validity Testing Results

CBEID_2517_5_3_4_ValidityTestingResults.pdf

5.3.5 Interpretation of Validity Results

Maintenance Testing

The findings were as hypothesized and summarized below.

Program Level

1. Program-level correlations were in the expected direction. Both correlations demonstrated a positive relationship. Higher measure scores for Utilization of Services were associated with higher measure scores for Oral Evaluation. Higher measure scores for Topical Fluoride were associated with higher measures scores for Oral Evaluation.

2. Program-level correlations were statistically significant ($p<0.0001$).

3. Program-level correlations demonstrated a “strong” to “very strong” strength of association: 0.9352 for Utilization of Services and Oral Evaluation; and 0.8615 for Topical Fluoride and Oral Evaluation. 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 “strong” to “very strong.”

Plan Level

1. Plan-level correlations were in the expected direction. Higher measure scores for Topical Fluoride were associated with higher measure scores for Oral Evaluation.

2. Plan-level correlations were statistically significant ($p<0.01$).

3. Plan-level correlations demonstrated a “moderate” to “strong” strength of association: 0.6667 between Oral Evaluation and Topical Fluoride in 2022; 0.6794 between Oral Evaluation and Topical Fluoride in 2023. Categorization of strength of association is based on Akoglu (2018).

Collectively, these results complement our prior testing, which established data element validity and measure score face validity, by lending further support of the validity of the measure scores at the accountable entity level through empirical testing.

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

5.4.1 Methods Used to Address Risk Factors

No risk adjustment or stratification

6.1.1 Current Status

In use

6.1.3 Current Use(s)

Public Reporting, Payment Program, Quality Improvement with Benchmarking (external benchmarking to multiple organizations), Other

6.1.3 Program Details

Name of the program and sponsor
Medicaid and CHIP Children's Health Care Quality Measures (Child Core Set); 1945A Home Health Core Set – Sponsored by the Centers for Medicare & Medicaid Services (CMS)
URL of the program
https://www.medicaid.gov/medicaid/quality-of-care/downloads/dentaloralhealth-ta...
Purpose of the program
Public reporting and quality improvement. To monitor and improve the quality of dental services provided. to children enrolled in Medicaid and CHIP by tracking the receipt of comprehensive or periodic oral evaluations.
Geographic area and percentage of accountable entities and patients included
Nationwide - all U.S. Medicaid and CHIP programs; enrolled children under 21. As of October 2024, 37.6 million children were enrolled.
https://www.medicaid.gov/resources-for-states/downloads/eligib-oper-and...)
Applicable level of analysis and care setting

Program-level analysis focusing on dental services provided in various care settings, including dental clinics and private dental practices.

,
Name of the program and sponsor
Texas CHIP and Medicaid Dental Services Pay-for-Quality (P4Q) program and Texas Healthcare Learning Collaborative. Sponsor - Texas Health and Human Services Commission.
URL of the program
https://www.hhs.texas.gov/sites/default/files/documents/laws-regulations/handbo...
Purpose of the program
This measure is publicly reported via the Texas Healthcare Learning Collaborative since 2013 and is part of Texas Medicaid/CHIP Dental P4Q since 09/01/2017, promoting quality improvement. (https://thlcportal.com/measures/dental) (https://www.hhs.texas.gov
Geographic area and percentage of accountable entities and patients included
Statewide in Texas, this measure applies to all 3 dental plans serving CHIP and Medicaid. As of December 2024, 3,171,414 children were enrolled.
(https://hhs.texas.gov/about-hhs/records-statistics/data-statistics/heal...)
Applicable level of analysis and care setting

Program-level and plan level dental services provided in various care settings, including dental clinics and private dental practices.

,
Name of the program and sponsor

Covered California, the California Health Benefit Exchange
URL of the program
https://hbex.coveredca.com/stakeholders/2025-Amend_QHP-IND_Att-2_8-1-24_Clean-F...
Purpose of the program
Part of Covered California QHP and QDP contracts since 2017, this measure is now tied to financial incentives via Performance Standards with Penalties focused on equity, data, and oral health. (https://hbex.coveredca.com/stakeholders/2025-Amend_QDP_Att-2_
Geographic area and percentage of accountable entities and patients included
Statewide. As of February 1, 2025 were 7,680 children (18 years and younger) enrolled in 5 Qualified Dental Plans. (http://hbex.coveredca.com/data-research/)
Applicable level of analysis and care setting

Program-level and plan level dental services provided in various care settings, including dental clinics and private dental practices.

,
Name of the program and sponsor
Oregon Health Authority Coordinated Care Organization Performance Metrics
URL of the program
https://www.oregon.gov/oha/HPA/ANALYTICS/MetricsScoringMeetingDocuments/4b.-Ora...
Purpose of the program
Public reporting and quality improvement. Reported since 2018 to increase transparency and track CCO performance in care quality.
(https://visual-data.dhsoha.state.or.us/t/OHA/views/CCOPerformanceMetric...
Geographic area and percentage of accountable entities and patients included
Statewide. In 2023, the measure was reported for 16 CCOs, and the statewide denominator for the measure was 509,459 children ages 0 through 20 years.
Applicable level of analysis and care setting

Program-level and plan level dental services provided in various care settings, including dental clinics and private dental practices.

6.1.3a Other Current Use

Medicaid/CHIP

6.2.1 Actions of Measured Entities to Improve Performance

In addition to illustrating ways that measured entities can improve performance, we also note that the Dental Quality Alliance partnered with the Institute for Healthcare Improvement to develop IHI-DQA Open School Course focused specifically on how to improve dental care using the IHI Model for Improvement, which can serve as a resource for developing and implementing improvement plans.

Several examples of improvement activities were included in the logic model. As noted in the logic model, there is not a one-size-fits all approach to improving performance on this measure.

We view the logic model as demonstrating possible pathways of attaining improvement. Different programs and plans will have different types of improvement opportunities as well as face different types of barriers (e.g., differences in reimbursement policies, resources to implement QI strategies, etc.). QI strategies will need to be carefully considered in the context of the setting in which it is implemented and with investigation of the root causes of sub-optimal performance.

What is more challenging to implement in one Medicaid program or managed care organization may be less challenging to implement in another. Not all of the suggested activities and resources will be available to all programs; and the resources and activities needed will vary based on the reasons for children not receiving oral evaluations. Thus, in this section, we provide examples of QI actions that have been successful to illustrate ways that improvement can be made.

Dental Provider Recruitment

The Center for Evidenced Based Policy at Oregon Health & Science University (2020) identified a set of non-financial strategies that Medicaid programs can use to recruit new dentists. The strategies were identified through case studies and interviews with Medicaid administrators in five states. Each strategy is accompanied by implementation examples. The strategies broadly include dentist outreach, developing stakeholder partnerships and engaging stakeholders, administrative streamlining (e.g., prior authorization, credentialing, reimbursement), beneficiary outreach and care coordination, and managed care contracting requirements, among others.

Patient Outreach

Borrelli et al. (2019) conducted a feasibility randomized controlled trial to test the effect of oral health text messages (compared with child wellness text messages) related to oral health, such as brushing, dental visits, and diet on improving pediatric oral health behaviors and parent attitudes. The researchers found that parents who received the oral health text messages were more likely to have improved attitudes toward dental check-ups.

Multifaceted Strategies

Nietert et al. (2005) evaluated the effects of a reform in 2000 to South Carolina's Medicaid dental program. There were several components to the program, which included increasing reimbursement rates, developing a children's oral health collation, recruitment by the state's dental association to encourage Medicaid participation, and streamlining Medicaid billing procedures. The authors found that the downward trend prior to the reform program was "dramatically reversed" after the program was implemented

The Missouri Foundation for Health and the Health Care Foundation of Greater Kansas City commissioned a report, “State Strategies to Improve Dental Compliance in Missouri’s Medicaid Population.” The report includes a range of strategies that states have successfully used, including approaches for addressing broken appointments, increasing dentist reimbursement, developing outreach materials for dentists to use with their patients, developing partnerships with other pediatric care providers (e.g., HeadStart programs, obstetricians, pediatricians), implementing case management/care facilitation programs that help to address barriers to care such as lack of transportation or having to miss work. The report also noted specific results associated with multi-faceted improvement strategies. For example, Virginia’s Medicaid program increased provider participation by 62% and children’s use of dental services increased from 24% to 35%. Alabama’s Medicaid program doubled its provider participation and had a 76% increase in dental service use.

References

Borrelli B, Henshaw M, Endrighi R, Adams WG, Heeren T, Rosen RK, Bock B, Werntz S. 2019. An Interactive Parent-Targeted Text Messaging Intervention to Improve Oral Health in Children Attending Urban Pediatric Clinics: Feasibility Randomized Controlled Trial. JMIR Mhealth Uhealth. 11;7(11):e14247. PMID: 31710306.

Center for Evidence-Based Policy, Oregon Health & Science University. 2020. Nonfinancial Strategies to Increase Dentist Participation in Medicaid. Available at: https://centerforevidencebasedpolicy.org/wp-content/uploads/2020/02/Ora....

Health Management Associates. 2008. State Strategies to Improve Dental Compliance in Missouri’s Medicaid Population. Prepared for The Missouri Foundation for Health and The Health Care Foundation of Greater Kansas City.

Nietert PJ, Bradford WD, Kaste LM. 2005. The impact of an innovative reform to the South Carolina dental Medicaid system. Health Serv Res. 40(4):1078-91. PMID: 16033493.

6.2.2 Feedback on Measure Performance

Since original measure development and endorsement, the DQA has undergone an Annual Measure Review process, overseen by its Measure Development and Maintenance Committee, that includes a one-month public comment period and review by the DQA membership. All comments are carefully reviewed and considered. An Annual Measure Review report is published each year. In addition, this measure is included in the CMS Child Core Set. The DQA is in

frequent communication with the Centers for Medicare & Medicaid Services, as well as with Mathematica which serves as CMS's technical resource to state Medicaid programs for quality measure implementation related to the Child Core Set. Any measure specification or technical assistance questions related to Child Core Set implementation also are considered during the DQA's Annual Measure Review processes.

The feedback received on this measure since the last maintenance review included:

During the 2022 Annual Measure Review, clarification was sought about the eligible provider types and if oral screenings by medical providers would be captured.

Response: The measure intent is to capture whether children are receiving a periodic or comprehensive oral evaluation as these services are defined by the Code on Dental Procedures and Nomenclature. These oral evaluation services include diagnosis and treatment planning, extending beyond the oral health screenings conducted by non-dental health care professionals. Including such screenings would deviate from the measure's intent. The DQA recognizes and appreciates the important role played by medical primary care providers in promoting oral health, which includes screenings, topical fluoride application, and referrals to dental care. Consequently, there are other DQA measures, such as Topical Fluoride for Children, that capture oral health services provided by medical primary care providers.

During the 2021 Annual Measure Review, there was a suggestion to exclude children six months of age or younger.

Response: Although the age criteria do not explicitly exclude children younger than six months of age, the enrollment requirement of at least six months effectively ensures that children are at least six months of age. Evidence-based guidelines recommend clinical oral evaluations with a regular recall schedule that is tailored to individual needs based on assessments of existing disease and risk of disease (e.g., caries risk) with the recommended recall frequency ranging from 3 months to no more than 12 months for individuals younger than 18 years of age. Clinical guidelines and literature support the recommended age for the first oral evaluation to be at the time of the eruption of the first tooth and no later than 12 months of age. Consequently, the DQA maintains the measure is applicable to all children under the age of 21 years. The DQA also notes that the age stratifications include the age band of 0-1 years of age to allow implementers to understand measure performance across age groups.

6.2.3 Consideration of Measure Feedback

The DQA carefully evaluates all comments received during its Annual Measure Review process. There have been no changes made to this measure. The rationale is described in the previous response in 6.2.2.

6.2.4 Progress on Improvement

We report (1) national score at the program level using two different data source, (2) state-specific scores for Texas Medicaid and CHIP at both the program and the plan levels, and(3) state-specific scores for Oregon at the program and plan (CCO) levels.

Across these data sources, there was observed improvement on the measure from the time of reporting through 2019. Performance dropped in 2020 as expected due to COVID-19 impacts on health service use. Subsequently, performance scores have increased to varying degrees across entities. The increase from measure implementation through 2019 clearly demonstrates improvement.

National Scores, State Medicaid Programs, DQA Dashboard

The DQA's Oral Healthcare Quality Dashboard (https://www.ada.org/resources/research/dental-quality-alliance), which calculates measures using the same T-MSIS database as was used for Maintenance testing, reports national scores (population-weighted averages) based on states with data quality of low or medium concern.

The scores from 2016 through 2022 were:

2016 (27 states): 45.0%

2017 (31 states): 45.2%

2018 (35 states): 47.1%

2019 (35 states): 47.5%

2020 (43 states): 38.0%

2021 (41 states): 41.4%

2022 (47 states): 41.7%

This pattern – an increase through 2019, a drop in 2020, and subsequent increase in 2021 and 2022 – is generally observed for individual states as well. The dashboard contains time trend

reports for each state.

National Scores, State Medicaid Programs, CMS Child Core Set Reporting

Oral Evaluation was first reported as part of the Child Core Set in in FFY 2022. Prior to FFY 2024, Child Core Set Reporting was not mandatory so not all states are included. The mean and median scores were based on Mathematica analysis of Quality Measure Reporting system reports.

FFY 2022 (27 states): Mean: 42.6%; Median: 43.2%

FFY 2023 (38 states): Mean 42.8%; Median: 42.8%

Data are currently available only for FFY 2022 and FFY 2023, which does not allow sufficient time to observe improvement trends. Source: https://www.medicaid.gov/medicaid/quality-of-care/performance-measureme….

Texas Medicaid and CHIP: Program and Plan Performance

The Texas Healthcare Learning Collaborative (https://thlcportal.com/home) has reported this measure since 2013. It reports the measure for 2 programs (Medicaid and CHIP) and 2-3 dental plans (a third plan was added in 2021).

Summary:

Medicaid program performance increased from 68% in 2013 to 71% in 2019; dropped to 60% in 2020; increased to 63% in 2022.

CHIP program performance increased from 59% in 2013 to 68% in 2019; dropped to 58% in 2020; increased to 65% in 2022.

Individual plan performance in both programs also increased through 2019, dropped in 2020, and increased again after 2020.

The detailed scores are below (reported for the program overall and then for each plan):

2013 Medicaid: 68.4% overall; 70.1% DentaQuest; 65.8% MCNA Dental

2013 CHIP: 58.6% overall, 63.8% DentaQuest, 58.5% MCNA Dental

2014 Medicaid: 67.4% overall; 69.%2 DentaQuest; 65.4% MCNA Dental

2014 CHIP: 59.4% overall, 62.9% DentaQuest, 58.2% MCNA Dental

2015 Medicaid: 69.1% overall; 71.2% DentaQuest; 66.5% MCNA Dental

2015 CHIP: 63.4% overall, 68.8% DentaQuest, 63.6% MCNA Dental

2016 Medicaid: 70.1% overall; 71.4% DentaQuest; 68.6% MCNA Dental

2016 CHIP: 65.6% overall, 70.1% DentaQuest, 67.3% MCNA Dental

2017 Medicaid: 71.0% overall; 72.6% DentaQuest; 68.9% MCNA Dental

2017 CHIP: 68.0% overall, 72.1% DentaQuest, 68.8% MCNA Dental

2018 Medicaid: 71.0% overall; 72.5% DentaQuest; 69.0% MCNA Dental

2018 CHIP: 67.7% overall, 72.1% DentaQuest, 68.2% MCNA Dental

2019 Medicaid: 70.7% overall; 72.0% DentaQuest; 69.0% MCNA Dental

2019 CHIP: 68.4% overall, 72.4% DentaQuest, 69.1% MCNA Dental

2020 Medicaid: 60.0% overall; 61.7% DentaQuest; 57.7% MCNA Dental

2020 CHIP: 59.0% overall, 61.0% DentaQuest, 57.8% MCNA Dental

2021 Medicaid: 63.3% overall; 66.0% DentaQuest; 62.6% MCNA Dental; 40.7% UHC Dental

2021 CHIP: 60.5% overall, 62.7% DentaQuest, 59.4% MCNA Dental; 38.7% UHC Dental

2022 Medicaid: 62.8% overall; 66.2% DentaQuest; 62.1% MCNA Dental; 45.6% UHC Dental

2022 CHIP: 64.6% overall, 68.5% DentaQuest, 65.3% MCNA Dental; 43.5% UHC Dental

Oregon CCO

Oregon reports this measure statewide and for each of 16 CCOs.

Data are available from 2018 through 2023.

Statewide performance was:

2018: 44.4%

2019: 46.4%

2020: 32.7%

2021: 38.7%

2022: 38.9%

2023: 39.8%

The pattern is similar to above with improvement prior to 2020, drop in 2020, and then an increase.

Performance for each CCO is below:

Advanced Health

2018 30.8%

2019 32.2%

2020 16.4%

2021 25.2%

2022 31.7%

2023 31.1%

AllCare CCO

2018 41.3%

2019 44.9%

2020 34.4%

2021 37.5%

2022 41.1%

2023 41.0%

Cascade Health Alliance

2018 27.7%

2019 33.7%

2020 27.5%

2021 30.2%

2022 36.0%

2023 36.0%

Columbia Pacific

2018 35.7%

2019 39.0%

2020 23.0%

2021 28.7%

2022 31.5%

2023 31.9%

Eastern Oregon CCO

2018 41.9%

2019 42.1%

2020 32.6%

2021 37.1%

2022 37.1%

2023 37.1%

Health Share of Oregon

2018 48.5%

2019 50.0%

2020 31.4%

2021 39.9%

2022 38.8%

2023 40.1%

InterCommunity Health Network

2018 40.5%

2019 42.7%

2020 30.3%

2021 36.6%

2022 35.3%

2023 38.5%

Jackson Care Connect

2018 47.1%

2019 49.8%

2020 38.0%

2021 39.7%

2022 43.0%

2023 43.5%

PacificSource Central

2018 39.1%

2019 40.8%

2020 30.1%

2021 35.3%

2022 35.1%

2023 38.1%

PacificSource Gorge

2018 54.8%

2019 52.1%

2020 36.5%

2021 44.2%

2022 39.0%

2023 37.7%

PacificSource Lane

2020 36.3%

2021 43.1%

2022 43.4%

2023 46.2%

PacificSource Marion Polk

2020 41.4%

2021 45.4%

2022 45.7%

2023 46.3%

Trillium North

2021 21.0%

2022 21.6%

2023 19.2%

Trillium South

2018 43.7%

2019 46.1%

2020 27.3%

2021 32.5%

2022 32.0%

2023 35.5%

Umpqua Health Alliance

2018 24.2%

2019 31.5%

2020 18.7%

2021 26.6%

2022 29.9%

2023 33.7%

Yamhill Community Care

2018 48.8%

2019 49.7%

2020 38.4%

2021 42.2%

2022 42.6%

2023 43.3%

6.2.5 Unexpected Findings

There have not been any unintended impacts or unexpected findings.

7.1 Supplemental Attachment

CBEID_2517_5_1_4_PatientCharacteristics.pdf

Developer POC email

aravamudhank@ada.org

Measure Developer POC

Krishna Aravamudhan
Dental Quality Alliance - American Dental Association

Chicago, IL
United States

The measure developer is different from the measure steward

No

Steward Address

Krishna Aravamudhan
Chicago, IL
United States

Steward Organization

American Dental Association

Steward Organization URL

https://www.ada.org/resources/research/dental-quality-alliance

Steward POC email

aravamudhank@ada.org