Task Order No. HHSP233201300039C
Dental Electronic Clinical Quality Measures (eCQMs)

FINAL REPORT

September 2014

Prepared for the Office of the National Coordinator for Health IT

by the Dental Quality Alliance (Aravamudhan, PI)

and the University of Florida (Herndon, PI)

Table of Contents

Purpose

This project was carried out under a contract between the Dental Quality Alliance (DQA; Aravamudhan, PI) with the Office of the National Coordinator for Health Information Technology (ONC), U.S. Department of Health and Human Services. The purpose of the contract was to support the specification and testing of two oral health electronic clinical quality measures (eCQMs) for inclusion in the Centers for Medicare and Medicaid Services (CMS) Meaningful Use (MU) incentive program for eligible professionals. The two measures are: (1) Oral Health Care Continuity for Children 2-20 Years and (2) Oral Health Sealants for Children 6-9 Years. The purpose of this report is to present the testing protocol, testing results, and finalized measure specifications. This report satisfies requirement under Task 6 of Task Order No. HHSP233201300039C.

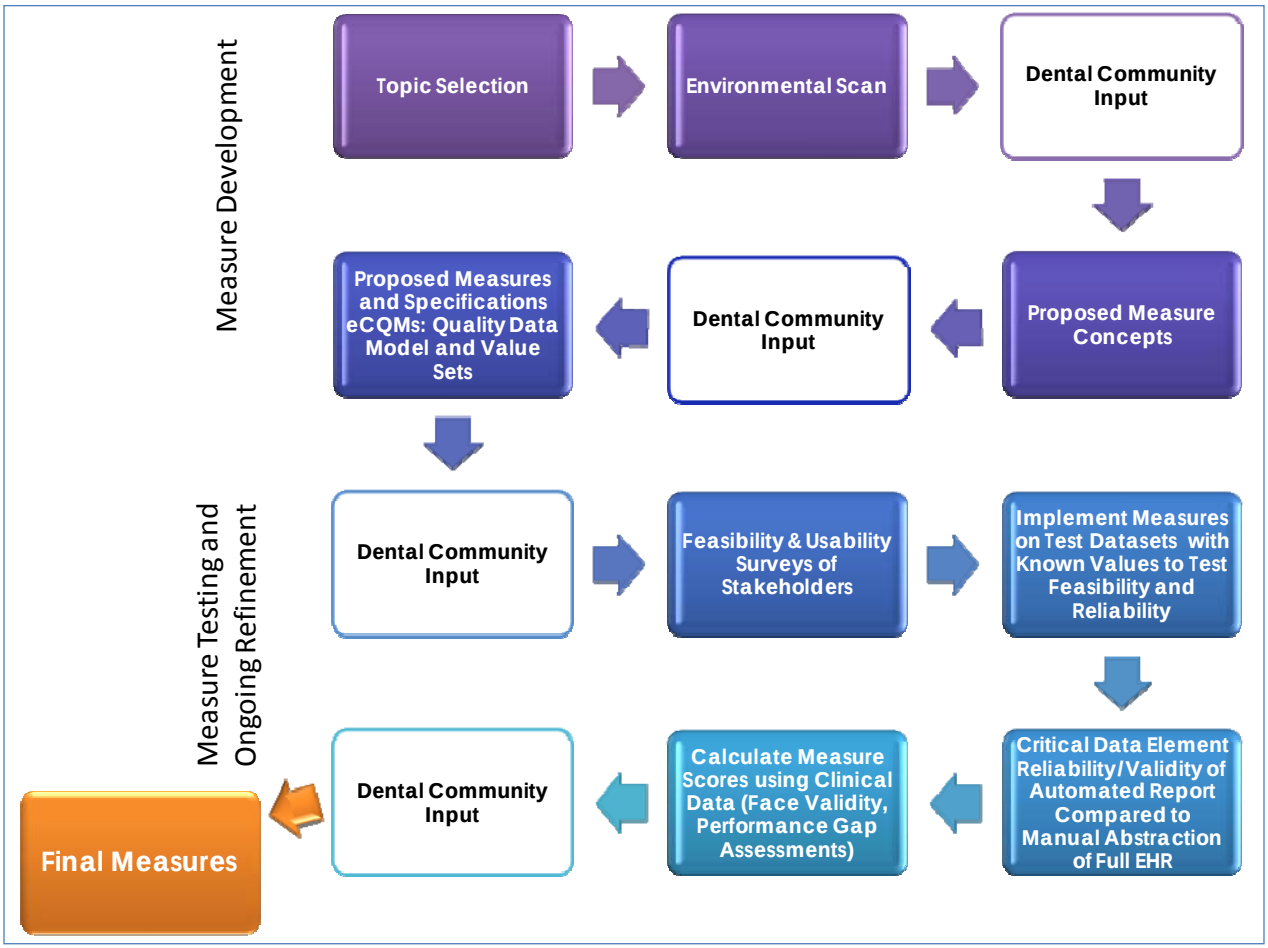
BACKGROUND

The DQA was formed specifically for the purpose of developing oral health quality measures and is comprised of a broad range of stakeholders. Its measure initiatives are guided by principles of stakeholder engagement and transparency throughout all measure development and testing processes, which are summarized in Figure 1.

In 2012, the DQA approved its first fully tested measure set *Dental Caries in Children: Prevention and Disease Management*.[1] These measures were specified for use with administrative claims data. This Starter Set of measures was identified through an in-depth environmental scan;[2] were approved by the full DQA membership after undergoing feasibility, reliability and validity testing; and were accepted into the National Quality Measures Clearinghouse.[3] A subset was recently endorsed by the National Quality Forum.

In 2012, the DQA formed an eCQM Committee to conduct a feasibility assessment of developing standardized pediatric oral health eCQMs, using electronic health record data.[4] Based on the findings of the feasibility assessment, the DQA identified and developed draft specifications for two eCQMs to undergo rigorous measure testing.

Figure 1: Measure Development & Testing Processes



In 2013, the DQA issued an open request for proposal (RFP) process to identify a research team to conduct feasibility, reliability, and validity testing.[5] A team from the University of Florida (UF) was selected through a competitive review process. The UF team partnered with Meaningful Use informatics experts and dental EHR developers to develop the proposal and conduct the testing. This broader group comprises the Project Team (Appendix 1). The contract was awarded effective October 1, 2013, and testing was conducted during the period October 2013 through September 2014. Measure testing processes followed National Quality Forum (NQF) guidance for measure feasibility, reliability and validity.[6] An Interim Report was presented to the project's Oversight Workgroup on April 17, 2014 and released for a one-month public comment period on April 21, 2014. A final presentation of the testing methodology

and results, addressing the NQF criteria for scientific acceptability of measures, was made to the Oversight Workgroup on September 11, 2014. This report presents the detailed methodology and findings.

FEASIBILITY OF ECQMS IN DENTISTRY

In 2012, the DQA conducted a feasibility assessment of developing standardized pediatric oral health eCQMs, building off of its prior development of the Starter Set of pediatric oral health measures calculated using administrative data.[4] To guide eCQM development, the DQA eCQM Committee was expanded to include a broader membership and evolved into the eCQM Oversight Workgroup. This workgroup includes clinicians and representatives from ONC, CMS, Health Resources and Services Administration (HRSA), National Quality Forum (NQF), federally qualified health centers (FQHCs), medical and dental EHR systems, and dental plans (Appendix 2). The Oversight Workgroup served as the Technical Expert Panel for and participated in conference calls throughout the project overseeing the measure development and testing activities and providing key input and subject matter expertise.

Based on the Oversight Workgroup's eCQM feasibility assessment, the two measures Oral Health Care Continuity for 2-20 Year Olds and Sealants for 6-9 Year Olds from the Starter Set were selected for electronic health record specification based on their (1) importance in promoting positive oral health outcomes and (2) feasibility of implementation.

Measure Development and Testing Protocol

This section provides the protocol used to develop and test the measures.

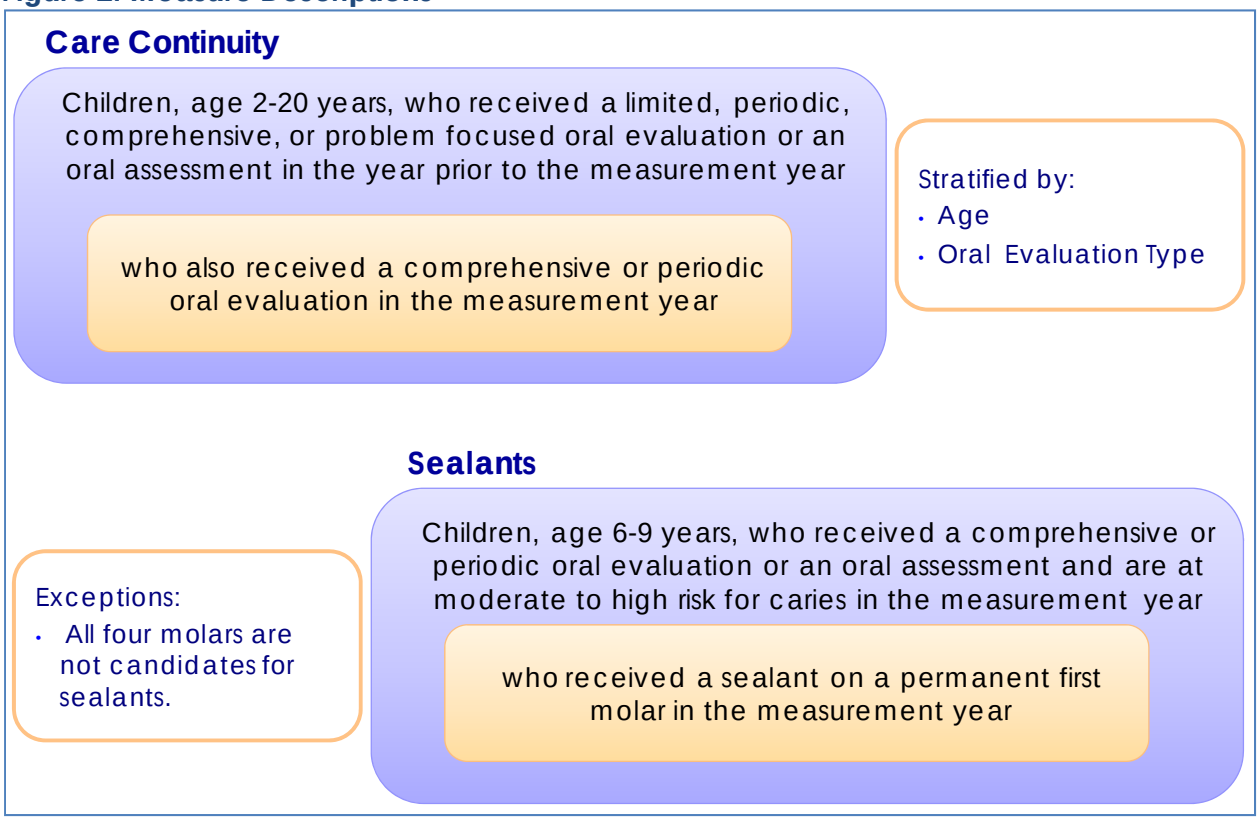
MEASURES

Two oral health eCQMs were developed and tested:

1. Oral Health Care Continuity for Children 2-20 Years
2. Oral Health Sealants for Children 6-9 Years

Figure 2 summarizes the measure descriptions. The text in the larger box for each measure summarizes the denominator criteria, and the text in the inset box summarizes the numerator criteria. The complete specifications - including the xml file, human-readable version with metadata, and complete value sets - were provided to ONC as separate files.

Figure 2: Measure Descriptions

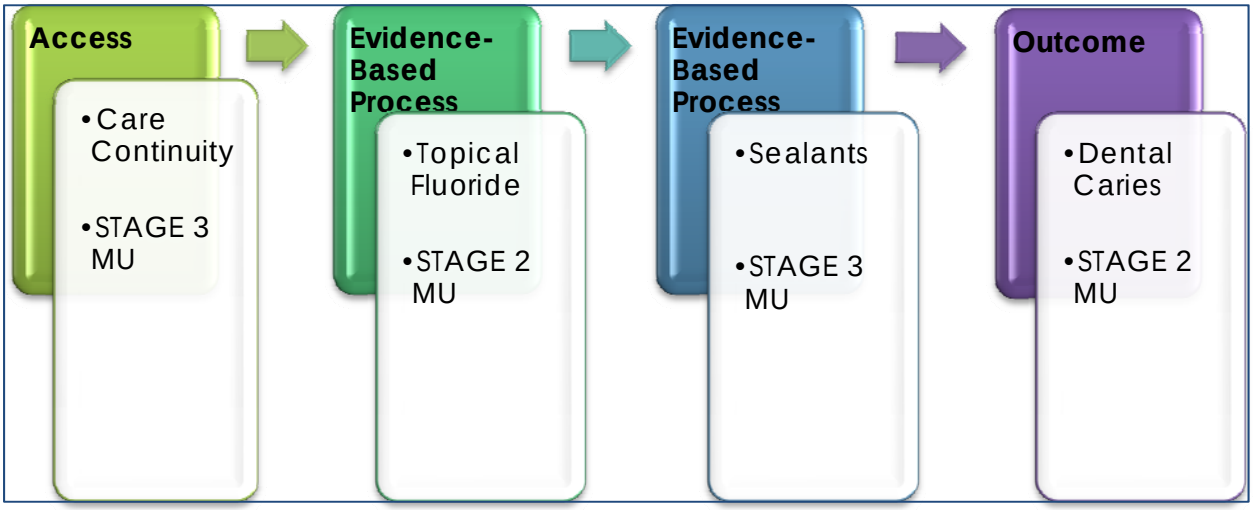


MEASURE IMPORTANCE

Both measures are health care process measures that address a high priority aspect of oral health among children: dental caries. These measures complement oral health eCQMs developed for Stage 2 of Meaningful Use (Figure 3). Dental caries is the most common chronic disease in children in the United States. In 2009–2010, 14% of children aged 3 –5 years, 17% of children 6-9 years, and 11% of children 13-15 years had untreated tooth decay.[7] Dental decay has significant short- and long-term adverse effects on children's oral and overall health.[8] Tooth decay can result in pain, difficulty

eating, school absenteeism, increased risk of future decay, and serious infections leading to emergency department visits, hospitalizations and, in rare cases, death.[9-16]

Figure 3: Meaningful Use Oral Health eCQMs



Oral Health Care Continuity for Children 2-20 Years

Linkage to health outcomes. 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 central to evaluating oral disease and dentition development and assessing risk for developing caries. They include evaluating and recording the patient's dental and medical history and a general health assessment. Clinical oral evaluations also are essential to developing an appropriate preventive oral health regimen and treatment plan tailored to individual patient needs. Disease identification, risk assessment, prevention regimens, and treatment planning are ongoing processes; therefore, evaluating continuity of care over time is an important quality metric.

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.[17, 18] Children who receive problem focused evaluations (episodic users) should also be retained in care. 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.[19]

Performance Gap. Although comprehensive dental benefits are covered under Medicaid and the Children's Health Insurance Program (CHIP), 23% to 63% of children enrolled in Medicaid/CHIP for at least 90 continuous days receive an oral evaluation (referred to as "Dental Diagnostic Services").[20] Even among the highest performing states, more than one-third of publicly-insured children do not receive an oral evaluation as a dental service during the year. Thus, a significant percentage of children are not receiving oral evaluations to assess their oral health status and disease risk and develop an appropriate preventive oral health regimens and treatment plans. A performance gap also was noted in the measure score results for the testing sites (see Testing Results section).

Oral Health Sealants for Children 6-9 Years

Linkage to health outcomes. Evidence-based clinical recommendations recommend that sealants be placed on pits and fissures of children's primary and permanent teeth when it is determined that the tooth, or the patient, is at risk of experiencing caries.[21] The evidence for sealant effectiveness in permanent molars is stronger than the evidence for primary molars. Sealants benefit children across a wide age range; however, for greatest effectiveness in caries prevention, it is recommended that sealants be placed on teeth soon after they erupt.[22]

Oral Health Sealants for 6-9 Year Olds measures whether children at moderate or high caries risk received a sealant on a permanent first molar tooth. Permanent first molars usually erupt between ages 6 and 7 years. Thus, this measure addresses both the tooth type on which sealants are placed and the timeliness of care provision. This measure contributes to the Healthy People 2020 Objective OH 12.2 to increase the percentage

children aged 6 to 9 years who received dental sealants on one or more of their first permanent molars.

Performance Gap. There are documented disparities in dental sealant receipt. For example, using data from the National Health and Nutrition Examination Survey, researchers at the National Center for Health Statistics identified variations in dental sealant prevalence among children 6-9 Years.[7] Specifically: "Dental sealant prevalence was lower among children [6-9 years] living at or below 100% of the federal poverty level (26%) compared with children living above the poverty level (34%)."[7] A performance gap also was noted in the measure score results for the testing sites (see Testing Results section).

PRELIMINARY FEASIBILITY ASSESSMENT

Note: During this project, feasibility was assessed during many phases throughout the project, starting with initial assessment of concept feasibility, conducting critical data element feasibility, and ultimately to implementation feasibility of the measure itself.

An initial feasibility assessment was conducted, following NQF guidance.[23] NQF recommends conducting a feasibility assessment with EHR vendors and measure end users early during the measurement development process after the measure has been conceptualized and the critical data elements have been identified. NQF recommends using a data element feasibility scorecard that evaluates the following four areas using a 3-point scale for both current and future feasibility where 3 is the highest rating and 1 is the lowest rating:[23]

1. Data availability: Is the data element available in a structured format?
2. Data accuracy: Is the data element from an authoritative source and likely to be correct?
3. Data standards: Is the data element coded using nationally accepted terminology standards?

4. Workflow: Is the data element captured during the typical workflow without additional data entry required solely for the quality measure and without requiring changes to the EHR user interface?

The DQA conducted these initial feasibility assessments through semi-structured surveys and phone interviews with EHR vendors, IT programmers in different care settings, and practitioners. Both critical and stratification data elements were assessed (Table 1). Critical data elements are those that are essential in order calculate the measure score. Stratification data elements are used to stratify the measure score by certain characteristics, such as age, race, ethnicity, and gender. Based on the positive findings of the feasibility assessment (reported in Testing Results), the DQA identified and developed draft specifications for the two measures to undergo rigorous measure testing.

Table 1: Critical and Stratification Data Elements

	Care Continuity	Sealants
Critical Data Elements		
Unique Patient Identifier	x	x
Patient Date of Birth	x	x
Date of Service	x	x
Unique Provider Identifier	x	x
Procedure codes (CDT)	x	x
Diagnosis		x
Caries Risk Assessment		x
Tooth number		x
Stratification Elements		
Sex	x	x
Race	x	x
Ethnicity	x	x
Payer Type	x	x
Other Data Elements Considered		
Exclusion reasons	x	x

MEASURE TESTING SITES

CLINICAL TESTING SITES AND EHR SYSTEMS

There were three participating sites, each with multiple clinics included in the testing:

1. University of Florida College of Dentistry (UFCD) dental clinics with three clinic sites in different parts of the state. The electronic record used is axiUm (developed by Exan Group).
2. American Dental Partners (ADP) affiliate practice dental clinics with three different clinic/practice sites in the Midwest. Electronic record used is Improvis (developed by ADP).
3. University of Florida Pediatric primary care clinics with four different primary care clinics in Gainesville. This site was included for the purposes of testing the Oral Health Care Continuity measure in a pediatric medical setting. Electronic record used is Epic (developed by Epic).

Two EHR vendors, Exan and ADP, participated as full project partners throughout the testing process. Additional in-depth interviews were conducted with Dentrix and Epic. These and additional EHR vendors also participated in the eCQM Oversight Workgroup, which addressed and came to consensus around key decisions during testing.

All sites and methodologies were approved by the appropriate Institutional Review Boards. All data were transmitted to the project PI through approved and secure transfer protocols.

TIME PERIOD FOR CLINICAL DATA TESTING

For both measures, the reporting period was calendar year 2013. Data from calendar years 2012 and 2013 were used for Care Continuity, which identifies the initial patient population and denominator based on service use during the year prior to the reporting year.

PATIENT CHARACTERISTICS

Table 2 summarizes the patient characteristics at each test site. For patients with more than one race indicated in the EHR, the primary or first-listed race is indicated. Because patients could have multiple payer types, payer type is reported as a percentage of visits/procedures for the patient population. There was variation between the sites in patient characteristics. There also was variation in patient characteristics between clinics within each site. For example, the pediatric age group distribution differed between the dental clinics within both Site 1 and Site 2. In Site 1, Clinic3 served a much larger proportion of children 13-20 years compared to the other two clinics (71% versus 42%). In Site 2, Clinic 2 served a larger proportion of children 0-5 years compared to the other two sites (58% versus <20%). In Site 1, although there were significant missing race and ethnicity data, patients in Clinic 3 were much more likely to be reported as Hispanic than in the other two clinics, which is consistent with variation in the population demographics in the areas served by each of these clinics. In Site 2, one clinic primarily served patients who had Medicaid coverage, whereas the other two clinics primarily served patients with private coverage. Site 3 was the medical setting, so the focus was on patients younger than 5 years for whom oral assessments (along with topical fluoride application) by physicians are reimbursed by the state Medicaid program.

Table 2: Pediatric Patient Characteristics at Clinical Test Sites

	Site 1				Site 2				Site 3				
	Overall	**Clinic 1**	**Clinic 2**	**Clinic 3**	**Overall**	**Clinic 1**	**Clinic 2**	**Clinic 3**	**Overall**	**Clinic 1**	**Clinic 2**	**Clinic 3**	**Clinic 4**
Total Number of Patients	10,565	8,903	1,113	549	2,186	500	835	851	14,068	1,433	6,714	6,057	753
Age Group Distribution													
0-5 years	20.92%	20.66%	23.09%	4.37%	31.84%	19.00%	58.32%	13.40%	46.84%	48.01%	43.95%	50.90%	50.73%
6-9 years	15.01%	14.88%	16.08%	7.29%	26.62%	28.00%	32.57%	19.98%	34.29%	29.87%	36.27%	32.89%	32.01%
10-12 years	21.53%	21.88%	18.69%	17.67%	13.91%	17.60%	6.11%	19.39%					
13-20 years	42.54%	42.58%	42.14%	70.67%	27.63%	35.40%	2.99%	47.24%	18.87%	22.12%	19.78%	16.21%	17.26%
Race+									**Data Below are 0-5 Years Only**				
American Indian/ Alaskan Native	0.13%	<1%	<1%	<2%	N/A	N/A	N/A	N/A	0.26%	0%	<1%	0.39%	<2%
Asian	0.63%	0.70%	<1%	<2%	N/A	N/A	N/A	N/A	4.67%	<2%	3.83%	6.20%	6.28%
Black/African American	12.35%	12.23%	18.42%	<2%	N/A	N/A	N/A	N/A	37.39%	66.57%	46.46%	21.67%	38.48%
Native Hawaiian/ Pacific Islander	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.39%	0%	0.61%	<1%	<2%
White	18.64%	21.56%	2.25%	4.74%	N/A	N/A	N/A	N/A	45.45%	26.45%	37.17%	58.45%	44.50%
Other	0.90%	1.02%	<1%	<2%	N/A	N/A	N/A	N/A	10.26%	4.80%	10.54%	11.19%	8.64%
Refused/Declined	1.51%	1.76%	0.00%	<2%	N/A	N/A	N/A	N/A	0.53%	<2%	<1%	0.88%	<2%
Missing/Unknown	59.85%	58.93%	51.12%	91.26%	N/A	N/A	N/A	N/A	1.04%	<1%	<1%	<1%	<1%
Ethnicity+													
Hispanic/Latino	5.99%	3.71%	26.50%	<2%	N/A	N/A	N/A	N/A	4.87%	1.60%	6.57%	3.76%	3.93%
NonHispanic/Latino	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	94.48%	98.11%	93.22%	95.17%	95.55%
Refused/Declined	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.65%	<2%	<1%	1.07%	<2%
Gender													
Female	49.49%	49.34%	49.60%	51.37%	49.13%	51.60%	46.23%	50.53%	47.00%	52.62%	43.44%	48.65%	48.17%
Male	48.14%	48.18%	48.16%	48.45%	50.87%	48.40%	53.77%	49.47%	53.00%	47.38%	56.56%	51.35%	51.83%
Unknown	2.38%	2.48%	2.25%	<2%	0%	0%	0	0	0%	0%	0%	0%	0%
Payer#													
Medicaid	76.28%	76.52%	74.36%	76.18%	20.80%	98.02%	4.65%	0.00%	64.49%	84.59%	77.53%	47.55%	57.85%
Medicare	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.52%	<2%	0.41%	0.58%	<2%
Private	5.91%	6.45%	3.72%	0.18%	77.70%	1.32%	92.87%	99.11%	33.34%	13.52%	19.41%	50.93%	39.79%
No Payer/Self Pay	17.81%	17.02%	21.93%	23.64%	1.50%	0.66%	2.48%	0.89%	1.65%	1.45%	2.64%	0.94%	<2%

+If patients had more than one race, primary or first-listed was selected. Site 1 collected ethnicity as a sub-category of race and not as a separate data element; Site 2 did not have race/ethnicity data as structured data elements.

#Because patients may have multiple payer types, payer was run at the visit/procedure level.

TESTING PROTOCOL

The testing protocol involved the following main phases:

1. Conduct Initial Feasibility and Face Validity Assessments with EHR Vendors and End Users
2. Create Test Datasets
3. Implement the eCQMs using the Test Datasets in the EHR Developers' Test Environments
4. Implement the eCQMs using the Test Datasets in Practice Site Test Environment
5. Implement the eCQMs using Clinical EHR Data in Dental Practice Sites
6. Conduct Critical Data Element Validation

Each phase of the process contributed to some aspect of feasibility and reliability/validity testing. **Feasibility, reliability, and validity testing do not occur in a strictly linear fashion; rather, they are inter-related and assessments of each of these measurement domains were ongoing throughout the testing process**.

Throughout the testing process, there was regular communication between the Project Team PI, the DQA, and the ONC through monthly conference calls. There also were regularly scheduled calls with the eCQM Oversight Workgroup. Project calls with the testing sites and EHR vendors were held weekly during measure testing within the sites. Importantly, key stakeholder feedback was solicited regularly during the measure development and testing process so that refinements to the specifications could be made to promote feasible, reliable, and valid measurement.

PHASE 1. CONDUCT INITIAL FEASIBILITY AND FACE VALIDITY ASSESSMENTS

The first phase involved a careful review of the measure specifications by all project team members to identify questions; further assess feasibility related to the data

elements, value sets, and measure logic; and evaluate face validity. In addition, in-depth feasibility assessments were conducted with the project team, EHR vendors, and end users (community health centers and practitioners). This phase was critical to clarifying and refining the measure specifications, informing the development of the synthetic test dataset, and identifying particular areas to be addressed during reliability and validity testing in clinical sites.

The Project Team also reviewed the responses from the initial feasibility assessments and the feasibility instrument itself. The instrument was revised into two different assessment tools: a less technical version to solicit additional feedback from practitioners and a more technical version for vendors and IT programmers. The revised instruments built upon the knowledge gained from the earlier assessments as well as preliminary discussions with the dental EHR vendors partnering on this project. The revised instruments also were designed to explicitly solicit information that would help the project team to make recommendations regarding the measure specifications and to develop the test datasets. Specifically, the revised assessment tools were designed to (1) inform and improve the clarity and reliability of the measure specifications, (2) ensure that the measure descriptions were clear and interpretable to a range of stakeholders, (3) ensure the soundness of the measure logic from both practitioner (clinical meaningfulness/face validity) and vendor (logistical implementation) perspectives, and (4) solicit key stakeholder feedback on specific implementation questions in terms of both clinical workflow and EHR system logistics that arose during the measure development and refinement processes. The instruments are on file with the DQA and provided separately in a supplemental appendix.

The revised feasibility assessment instruments were piloted with project team members who were asked to complete either the practitioner questionnaire (4 team members) or vendor/programmer questionnaire (4 team members) as appropriate. In addition, each member was asked to comment on the assessment instrument itself, noting any suggestions for improvement. In addition, a series of follow-up calls and meetings were held with each project team member to review the responses in depth. Based on the

feedback provided by the team members on the instrument itself, additional revisions were made. The revised instruments were approved by ONC for broader distribution. Given the small size of the vendor community and the participation by both dental and medical EHR vendors in the prior feasibility assessments, additional vendor input through these surveys was used to address information gaps but quantitative assessments were not conducted. EHR vendors also were invited to provide additional input on issues of particular importance from their perspectives, and vendor input was solicited during Oversight Workgroup calls on key issues related to implementation of the measures in different EHR systems. As a value-added component of this project, the DQA solicited clinician feedback from the memberships of the Academy of General Dentistry, American Academy of Pediatric Dentistry, American Academy of Pediatrics, and the National Network for Oral Health Access – the organizations representing the eventual end-users of the Meaningful Use measures.

PHASE 2: CREATE TEST DATASETS

Test datasets with synthetic patient data were created for each measure to test the feasibility and reliability of EHR systems to accurately calculate the measure. The reliability and validity of eCQMs depend not only on having complete and accurate data, but also critically on how the measure specifications are implemented across different EHR products and practice sites. A test dataset allows EHR developers to evaluate their measure reporting processes within their test environments prior to practice site implementation. Practice sites can also use the test datasets to validate their site-specific implementation of the measures, including any local configuration that is required, prior to running the measures on patient data where each measure data element value is not prospectively known.

Synthetic test datasets were created with known values of the critical data elements to test the EHR software's translation of the measure logic and calculated results within both EHR developer and practice site testing environments. Creating the synthetic test datasets consisted of three major development steps: (1) defining and developing the

data schema, (2) defining and developing the programming process, and (3) generating the synthetic data.

Defining the data schema. An initial schema was developed based on the informatics consultants' prior experience working with medical EHR systems and understanding of the similarities and differences in dental EHR systems obtained through the in-depth questionnaires described above. This initial data schema was shared with three dental EHR vendors who provided feedback that helped to refine the schema to a structure that consisted of the essential tables and fields required to successfully import data into each of their systems. Although the processes for generating the Care Continuity and Sealants datasets were different, a single data schema was used to develop the synthetic data for both measures. Microsoft SQL Server was the database platform. The dataset specifications were designed and documented in MS Excel and were translated and scripted into SQL Server tables and fields once the schema was finalized. An initial limited dataset (10 patients) was created to test the import process with Exan and ADP and refine the data schema based on the test import.

Defining and developing the programming process. One of the most important and challenging aspects of developing the synthetic test patients was ensuring that the resulting datasets were both realistic as possible (i.e., that the data simulate real dental practice data) and that they tested all of the main aspects of the measure logic. There were three main utilities used to create the programming and database system infrastructure: (1) Red Gate Software, Inc.'s SQL Data Generator, (2) SQL Server, and (3) Microsoft Access. We used Red Gate to create randomly generated patients and define the content for each data field. Percentages could be defined for patient demographic characteristics (gender, race, ethnicity, and payer type), and the data generated would reflect the defined proportions. The tool also allowed us to define procedure code and visit percentages in order to produce a robust, realistic mix of services. Each data element could be defined with precision, including specifying how many characters a field could contain and setting limits on the range of data values. For example, since we wanted to test age inclusion and exclusion criteria, we could

limit birthdates to be within a set range of beginning and end years. SQL Server served as the database platform for storing and managing the data. Microsoft Access forms and the internal Visual Basic language were used as the interface and programming engine to drive and logically call the processes. Although the final data schemas and structures provided to the vendors for the two measures were the same, each measure required two separate development environments due to differences in the measure specifications. Bringing these utilities together allowed us to create robust datasets to test different aspects of the measure specifications and simulate dental data.

Generating the test datasets. Once the infrastructure of the database system, the programming tools and the synthetic field generator were in place, the final steps were to produce the datasets. The initial datasets were programmed based on specifications developed in consultation with the DQA. The specifications were designed to test all aspects of the measure logic as well as being realistic. The patient populations within each dataset were purposely designed to meet or not meet the inclusion criteria for the initial patient population (IPP), denominator (DEN), numerator (NUM), and exceptions (EXC). For example, we tested the following aspects of the measures:

- Correct age calculations and inclusion/exclusion for IPP and DEN
- Correct procedure inclusion/exclusion for IPP, DEN, NUM, EXC
- Correct provider attribution for IPP, DEN, NUM, EXC
- Correct tooth for sealant placement (Sealants)
- Correct identification of elevated risk (Sealants)
- Correct identification of patients qualifying for exceptions (Sealants)
- Correct required stratifications for age and evaluation type (Care Continuity)
- Correct demographic stratifications (gender, race, ethnicity, payer type)
- Correct implementation across a range and mix of different service use patterns

Thus, the datasets were designed to include patients who qualified and did not qualify for the measures on all of the above dimensions, which meant including age ineligible

patients, including different combinations of procedures so that patients would variously qualify or not qualify for the IPP, NUM, DEN and EXC, including different patient-provider-procedure patterns to ensure correct provider attribution, and so forth.

To ensure the datasets were realistic, additional programming logic was needed. For example, certain procedures are age specific – e.g., a child could not have a sealant on a permanent molar at an age prior to when the molar would be expected to erupt. Therefore, logical age-procedure combinations were identified. Certain procedures, such as restorations, extractions, and sealants require a tooth number. Therefore, logical procedure-tooth number combinations were identified. In addition, different types of diagnoses and findings are associated with different types of procedures (e.g., a finding of elevated caries risk would be associated with a child who received several restorations, and we would not expect to see a finding of low caries risk). Therefore, logical diagnosis/finding-procedure combinations were identified. DQA content experts assisted with the identification of these logical pairings, which were then incorporated into the test dataset programming logic to ensure that illogical combinations were not included. Finally, we also included different code systems where applicable. For example, elevated risk could be captured through CDT or SNOMED codes; diagnoses could be captured through ICD-9, ICD-10, or SNOMED codes.

After the test dataset was generated, the logic was evaluated through a series of tests first in SQL by the informatics lead and then in Stata, Release 13[24] by the project PI to ensure that the datasets would provide robust testing of the measure specifications and that illogical cases were not included (e.g., checking the procedure-age, procedure-tooth number, and procedure-diagnosis combinations described above). Refinements to each test dataset were made until it passed all logic checks. The known values for each measure component, the overall measure score, and measure score stratifications were calculated and verified before providing the datasets to the vendors.

The resulting test datasets were provided to the vendors as relational tables provided in an Excel workbook. For each measure an initial test dataset of synthetic patients was created (50 patients for Care Continuity and 98 patients for Sealants). A second, more complex test dataset (more variations of patient-provider-procedure combinations and visit patterns) for each measure was subsequently created to confirm correct implementation of the measure logic and provide more robust testing (240 patients for Care Continuity and 189 patients for Sealants). Separate test datasets were created for each measure so that they could be tailored to the measure to ensure robust testing of all measure aspects.

PHASES 3 & 4: IMPLEMENT THE ECQMS USING THE TEST DATASETS IN THE EHR DEVELOPERS' AND LOCAL PRACTICE SITE TEST ENVIRONMENTS

The ability to produce the known values in the test dataset is an important form of feasibility and reliability testing – it speaks directly to whether the measure specifications as implemented within the EHR reliably calculate the individual measure components and the overall score. The two project team EHR vendors imported the test dataset into their systems' test environments and implemented the measure logic. They produced provider-level reports on the number of patients meeting the initial patient population, denominator, numerator, and exception criteria along with the measure score calculation for the measure overall and for required and optional stratifications. These reports were compared with the known values. Overall concordance and kappa statistic analyses were used to compare the results. Provider attribution logic was also tested during this process. Patient level files were used to identify the specific patients who were misclassified in order to identify the sources of the discrepancies. Feedback was provided to the vendors regarding identified discrepancies. The vendors then revised their programming, ran the measures again, and resubmitted their reports. This process was iterated until 100% agreement was achieved for each dataset.

With technical assistance from Exan, the UFCD practice site imported the test dataset and ran the measure report. The resulting values were similarly compared to known

values following the process described above. Implementation within the local practice site assesses whether the measure specifications as locally configured reliably calculate the individual measure components as well as the overall rate prior to implementing the measure on clinical data.

PHASE 5: IMPLEMENT THE ECQMS USING CLINICAL EHR DATA IN DENTAL PRACTICE SITES

After the measure logic was successfully implemented using the test dataset, the measures were implemented using clinical data in the practice sites. Prior to measure implementation, practice site-specific feasibility assessments were conducted. Feasibility assessments included confirming the presence and completeness of the critical data elements. Each site also generated summary background data reports describing the patient population characteristics and overall service use for the main procedures included in the measure specifications. These reports were used to provide context for interpreting results, refine and finalize the data element validation methodology, and contribute to face validity assessments of calculated measure scores.

Provider-level measure reports were generated for each measure that included the initial patient population, denominator, numerator, exceptions, and measure score with any required stratifications. Successful implementation of the measure logic on clinical data further demonstrates feasibility.

PHASE 6: CONDUCT DATA ELEMENT RELIABILITY/VALIDITY TESTING

Because newly developed measures often do not have numerous testing sites, NQF advises: "If testing of eMeasures occurs in a small number of sites, it [reliability and validity testing] may be best accomplished by focusing on patient-level data element validity (comparing data used in the measure to the authoritative source)."[25] Therefore, critical data element validation was the primary focus for assuring the reliability and validity of the measures. In addition, we performed validation of each measure score

component (IPP, DEN, NUM, EXC). Further, we performed analyses to address very specific questions related to each measure to enable measure refinement.

Within each of the three clinical testing sites, within one of the clinics, a random sample of 75 patients who met the respective measure's age eligibility criteria was selected for each measure to compare EHR automated reporting with manual abstraction of the full EHR as the referent standard to validate the measure's critical data elements and measure score components. All dates of service between January 1, 2012 and December 31, 2013 were recorded for Care Continuity, and all dates of service between January 1, 2013 and December 31, 2013 were recorded for Sealants. In the medical site, critical data element validation was conducted only for Care Continuity because the sealant measure is not intended for use in medical settings. Further, the sample population in the medical site was restricted to Medicaid patients younger than age five for whom oral assessments and topical fluoride application by physicians are reimbursable.

To assess validity, we calculated overall agreement as well as the kappa statistic, which takes into account agreement by chance. 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:[26]

0.01-0.20 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.

The critical data element validation process involved the following steps:

1. Automated abstraction of critical data elements for each date of service from the electronic record was conducted by the site's clinical informatics specialist.
2. Manual abstraction of critical data elements for each data of service from the electronic record was conducted by a clinician at each site; the clinicians were quality improvement specialists within their sites with record review experience. Manual abstraction noted whether presence of the data element in question was supported by a procedure (CDT or CPT) code, provider documentation, both, or neither. The abstraction form also allowed for recording any inconsistencies between a listed procedure code and documentation.
3. All sites used the same detailed protocol and abstraction forms, which were developed in consultation with the sites.
4. The automated and manual abstraction processes were conducted independently.
5. For each measure and site, 10-15 sample patient records were selected to test the abstraction process. Adjustments were made to the validation process to address issues that would result in inaccurate reporting or false discrepancies. These records were not included in the final validation analyses.
6. The remainder of the records not used for testing the process (60-65 records per measure per site) underwent automated and manual abstraction.
7. The automated and manual abstraction reports were provided to the project PI who conducted the concordance analyses, including calculating kappa statistic values, using Stata, Release 13. Concordance analyses also compared the sensitivity of the manual abstraction findings based on procedure codes in the record versus provider documentation.
8. Discrepancies were analyzed by the information specialist and record reviewer to identify the source of the discrepancy.

9. For discrepancies that were due to abstraction or programming errors, concordance analyses were re-run with the corrected data to evaluate the impact on the results.

Additional Validation Analyses

In addition to establishing critical data element validation, we also conducted the following analyses to assure the validity of the measures.

Validating measure score components between automated and manual abstraction. In addition to validating the individual critical data elements, we also evaluated the measure score components (IPP, DEN, NUM, and EXC) at the patient-provider level between the automated and manual abstraction reports to verify whether automated implementation of the measure logic that identified patients as meeting/not meeting each measure component criteria (specific to individual providers) was supported by the manual record reviews. Because the measure score components are aggregations of the individual critical data elements, it was expected that there would be similar findings between the two levels of analysis. However, there are two benefits to this additional level of reliability/validity testing. First, it assures that the measure score component determinations (qualifies for IPP, DEN, NUM, EXC, respectively) from the automated reports implementing the measure logic on clinical data are validated against the documentation in the full EHR. In addition, it also assures the reliability/validity of the provider attribution for each measure score component.

Evaluating the measure scores with and without exceptions (Sealants). NQF guidance indicates that exceptions or exclusions should be supported by evidence that the exception is important from both a clinical perspective and from a measurement perspective. There should be evidence that not including the exception could meaningfully impact or bias the measure score – i.e., the exceptions are not rare. Therefore, we calculated and compared the measure score for Sealants with and without exceptions.

Comparison of calculated measure scores to overall service use frequencies.

We compared the calculated measure scores to the overall service use frequencies for the relevant procedures provided in the background data reports by each site. Discrepancies were explored and corrections were made to the measure implementation logic as needed. For example, for one measure, the provider attribution logic was accidentally altered between testing on the test dataset and implementation on the clinical data. Comparisons of the measure scores to the service use frequencies revealed that the measure scores were significantly lower than expected and allowed underreporting of the numerator to be identified and the programming logic to be corrected.

Evaluating the impact of tooth-level versus surface-level exceptions. The sealant measure is specific to the four permanent first molars. The anatomical site associated with the procedure codes for identifying sealant placement and for identifying whether exception criteria are met is specified at the tooth level. However, in dental practice, surface level findings and procedures are recorded. During measure development, it was anticipated that the additional measurement burden of going to a surface level detail would outweigh the benefits of increased precision - i.e., that the impact on the measure scores of using tooth versus surface level procedures and findings would be minimal. To assure this was the case, testing specifically explored this issue. The specific concern was related to the exception criteria for Sealants. For a patient to qualify for an exception from the denominator, each of the four first permanent molars must be identified as not being sealable. Reasons that a tooth is not a candidate for a sealant include that the tooth has not yet erupted, it has an existing sealant or restoration, or that there is active caries on an occlusal surface. Because the measure logic was captured at the tooth level, active caries on non-occlusal surfaces would also get captured in the exception measure logic even though a tooth that has caries on only non-occlusal surfaces would still be a candidate for a sealant. Therefore, during the manual record reviews, we evaluated the extent to which capturing active caries on non-occlusal surfaces only resulted in individuals being excepted from the measure who would not qualify for an exception if it were specified at the surface level.

Results: Measure Specification and Refinement from Stakeholder Engagement

Feedback from key stakeholders was solicited in the following ways:

- Initial and follow-up feasibility assessments with EHR vendors using questionnaires and structured interviews
- Survey of key implementers to gain feedback and buy-in through stakeholder organizations
- Transparency in process – public comment on Interim Report widely disseminated and posted online for a one-month comment period
- Regular meetings (conference calls) of the Oversight Workgroup to provide updates and solicit feedback on key issues and decision points
- Regular conference calls between ONC, DQA, and the project team PI to provide updates and solicit feedback on key issues and decision points

This feedback was essential for review and refinement of the measure specifications as well as refining the testing methodologies. The following are examples of key issues that were identified and collaboratively addressed through these processes.

IDENTIFYING UNIQUE "ENCOUNTERS" OR "VISITS" (CARE CONTINUITY & SEALANTS)

Medical EHR systems embed procedures within "encounters," and MU measure logic historically conditioned the initial patient population and denominator criteria on whether certain types of "encounters" were present. [Note: In this context, use of the term "encounters" is not related to how payment is made, e.g., in an FQHC setting. It refers to how data is stored within the EHR's database.] In its initial feasibility assessments, the DQA identified the encounter framework as a challenge to implementing eCQMs within dental systems that do not embed procedures within encounters. Follow-on feasibility assessments during testing further evaluated how dental systems record and relate service provision to patients. The dental EHR vendors

indicated that they typically identify unique "visits" by a posted procedure for a particular date of service or through completed appointments with a unique combination of date of service, patient, and provider. One respondent noted that the use of terminology such as "visits" or "encounters" should be clearly defined in the measure logic.

Implication for Measure Specifications: In addition to standard encounter clauses that are used in the measure logic for medical systems, the measure logic for the Care Continuity and Sealants measures also includes "procedure performed" clauses in the IPP and DEN criteria to address the structures and processes of dental EHR systems.

PATIENT ACTIVE/INACTIVE STATUS (CARE CONTINUITY & SEALANTS)

All of the dental EHR vendors indicated that they had fields within their systems that signify a patient status of "active" or "inactive." We found that the dental EHR vendors have variably used this field in prior MU measures because there was no explicit guidance: some vendors have omitted patients with an inactive status from the initial patient population, while others did not impose this restriction. The medical EHR vendors do not have equivalent fields. We explored whether a status of "inactive" might be considered as a potential exclusion criteria for dental EHR systems. It was determined that this field is used in different ways across the systems and is not sufficiently reliable. There is significant variation in the amount of information contained within this field. For example, in one vendor's system there are only broad active and inactive status categories and historical status changes are not tracked. Another vendor captures a detailed set of reason codes for moving from active to inactive status and maintains a history of the status including date change. The active/inactive status field also is largely under the purview of the individual practice site. Vendors and an IT dental practice programmer noted that there often are not systematic processes at practice sites for updating this field, resulting in variable implementation both within and across practice sites. It was determined that this variability in implementation would compromise the reliability of the measure score generated between systems.

Implication for Measure Specifications: Based on the unreliability of active/inactive status indicators, the DQA measure guidance for Care Continuity and Sealants specifically indicates that active/inactive designations within the EHR should not be used in determining patient eligibility for inclusion in the measure.

MEASURED ENTITY AND PROVIDER DENOMINATOR/NUMERATOR ATTRIBUTION (CARE CONTINUITY & SEALANTS)

Eligibility for participation in the CMS MU Medicaid EHR Incentive Program for Eligible Professionals is at the level of the individual provider and not the practice.[27] Thus, ONC designates the measured entity as the individual clinician (also referred to as "provider" in this report). In our review of the measure logic for existing eCQMs there typically are no specific clauses or guidance regarding how provider attribution occurs for the measure denominator and numerator. Based on feedback from ONC and medical EHR vendors, the most common approach is for the denominator to be provider specific and for the numerator to count qualifying services rendered by any provider. In addition, it is common to assign the denominator attribution based on one preventive care visit/encounter during the measurement period – i.e., if a child has a qualifying preventive care visit/encounter (based on procedure codes) with a provider, then the child will be in the provider's denominator. However, we found that the lack of clear specifications resulted in this approach not being uniformly adopted by dental EHR vendors – for example, some made the numerator provider specific as well as the denominator. Differential implementation by vendors compromises the reliability of the measure. Specific guidance is needed to promote consistent implementation of measures across EHR systems.

Implication for Measure Specifications: The measure guidance includes an explanation of the intended provider attribution for each measures component (IPP, DEN, NUM, and EXC). Figures 4 and 5 illustrate examples of how denominator and numerator determinations would be made for Care Continuity and Sealants.

Figure 4: Provider Attribution Examples, Care Continuity

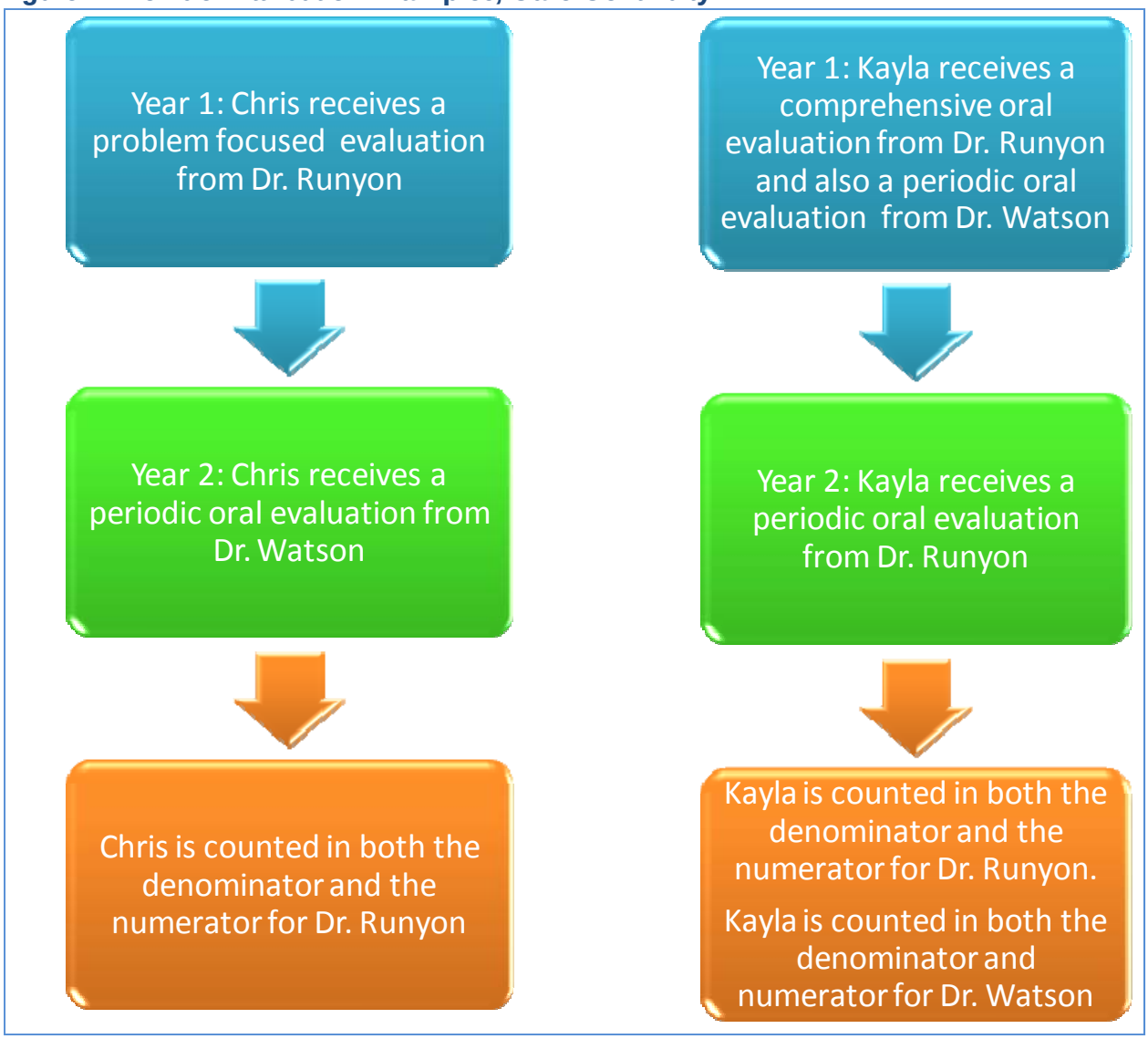
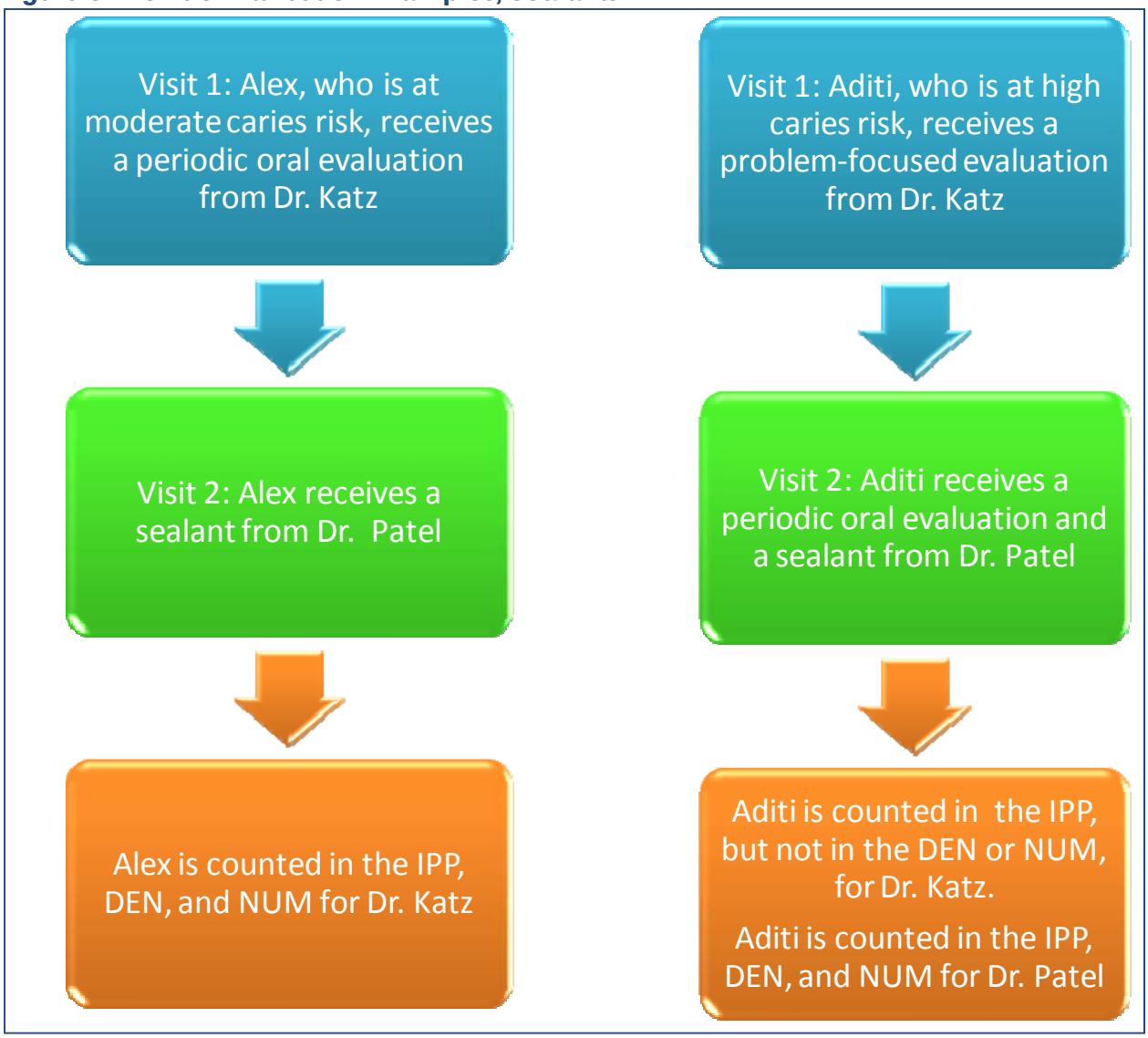


Figure 5: Provider Attribution Examples, Sealants



REQUIRED STRATIFICATIONS

Stakeholders, especially clinicians, were directly requested to provide feedback regarding whether there should be required stratifications for each measure. Given the broad age range for Care Continuity, changes in dentition and caries risk over time, and evidence of performance variation by age for existing oral health measures, it was determined that age stratifications would be beneficial. There was no consensus around specific age groupings; however, most clinicians recommended that children

be grouped into three general categories – early childhood, middle childhood/early adolescence, and middle/late adolescence. Consequently, the Oversight Workgroup elected to use age stratifications consistent with the existing MU measure Primary Caries Prevention of <5 years, 6-12 years and 13-20 years. Feedback from the American Academy of Pediatric Dentistry recommended an additional stratification based on oral evaluation type using the following mutually exclusive classifications: periodic/comprehensive oral evaluations, problem-focused evaluations, and oral assessment. The rationale for the stratifications was to differentiate between "regular" versus "episodic" users of care; i.e., the population of children captured in Year 1 by the measure includes both children who received a comprehensive/periodic evaluation as well as children who received a problem-focused evaluation. The measure then seeks to see whether both groups were retained/brought into regular care in Year 2. The stratifications then allow parsing the measure score by these populations. As noted earlier, 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**.[17, 18] Measure testing confirmed that the measure scores vary by these stratification categories. No stratifications were recommended for Sealants.

Implication for Measure Specifications: The Care Continuity measure includes two required stratification categories: age and oral evaluation type.

SOURCE OF PAYMENT (SOP) STRATIFICATION DATA ELEMENT

Currently, "payer type" as a distinct field is not well represented in EHR vendor systems. In general, payers are represented in the EHR systems as the patients' individual payers, and multiple payers may be associated with a patient. Individual payers may also be associated with individual procedures. There may be hundreds (or thousands) of individual payers. For example, there may be a range of different MetLife health and dental plans, UnitedHealthcare health and dental plans, and so forth. Vendors indicated that they could map individual payers to broader categories, such as Medicare, Medicaid, private, and self-pay categories. However, they also noted

challenges with creating mappings, especially for more detailed and extensive code sets such as those represented within the Source of Payment Value Set. Challenges include codes which may not have an appropriate or clear 1:1 mapping. In addition, dental EHR vendors noted payer types for medical services often do not coincide well with those for dental services.

Implication for Measure Specifications: The DQA submitted recommendations to ONC for a streamlined payer type code set that would be applicable to dental systems.

ASSESSMENT OF CLARITY AND PURPOSE THROUGH SURVEY OF POTENTIAL IMPLEMENTERS

As a value added component for the project, the DQA conducted a survey of potential implementers of the measures to get feedback on clarity and purpose of each measure. This effort was not within the scope of work for the UF project team and not included within the project's IRB. An electronic survey was distributed to clinicians through the American Academy of Pediatric Dentistry, American Academy of Pediatrics and the National Network for Oral Health Access – the primary organizations of potential users. A total of 492 responses were received: 37% of respondents were from solo/small group practices, 32% were from FQHCs/CHCs and 16% were from academic institutions. Additionally, 82% were dental care providers and 12% were medical care providers. For the Care Continuity measure, 89% of respondents indicated that the purpose of and rationale for the measure was clear while 94% indicated clarity for the sealant measure.

Results: Measure Logic and Data Testing

Testing results are organized around the NQF measure acceptability criteria for measure feasibility and reliability/validity.

FEASIBILITY

Feasibility is defined as the "extent to which the required data are readily available or could be captured without undue burden and can be implemented for performance measurement."[28] Both data element feasibility and measure logic feasibility were assessed.

SUMMARY OF FEASIBILITY FINDINGS

The key findings were:

- All but two critical data elements are captured as part of normal clinical workflow as structured data elements with data completeness ranging from 98%-100%.
- Caries risk assessment findings historically were not captured as a structured data element, but this is currently possible with new CDT codes introduced in 2014 to record caries risk assessment findings. This data element's feasibility was specifically reviewed by an NQF expert panel and is included in other oral health performance measures approved for NQF endorsement. It is expected that by 2017, when MU reporting for these measures is anticipated, these codes will be captured routinely as structured data. In general, any CDT code can be easily captured. Feedback from the clinical sites and EHR vendors indicated that they did not foresee problems with this data element.
- Diagnosis and findings codes, needed to identify sealant exceptions, are captured in dental EHR systems within problem and condition lists, and they can be mapped to standardized codes. This was verified and validated against manual record reviews during the testing process. Dental diagnostic codes have previously been reviewed by ONC and are included in one of the existing 2014 MU oral health measures, Children Who Have Dental Decay or Cavities.
- Measure logic implementation feasibility was verified by implementation of the measure logic using both synthetic test data and clinical data from the test sites.

The detailed findings are provided below.

DATA ELEMENT FEASIBILITY

Initial Feasibility Assessments during Measure Development: NQF Data Element Feasibility Scorecard

There were nine participants, representing dental EHR vendors, IT programmers in large group practices and community health centers (including FQHCs), and practitioners in the semi-structured interviews conducted during measure development using a data element feasibility scorecard where 3 is the highest rating and 1 is the lowest rating regarding current and future feasibility. The results of the assessment are summarized in Appendix 3. The following data elements were found to currently and fully meet the four NQF criteria evaluated in the scorecard: patient date of birth, date of service, and identification of specific procedures (e.g., oral evaluation, sealant placement). Tooth numbering also meet all four criteria; however, more than one standard terminology is used and currently no dental systems use SNOMED codes because these are not among the standards used by the profession. EHR vendors indicated that they could map their existing coding systems to SNOMED codes. This was verified during both the synthetic test dataset testing and implementation of the measures with clinical data. Both diagnoses and caries risk assessment data can be captured as structured data elements as part of the clinical workflow, but the feasibility assessment findings indicated that they currently are captured in both structured and unstructured formats across data systems and are not uniformly collected as part of routine care across all practice sites. However, these were identified as issues that could readily be addressed both in the EHR systems (capture as structured data elements) and through quality improvement processes (workflow adaptations) in the near future. These also were identified as areas meriting additional examination during testing. The patient characteristic stratification variables are all currently captured within EHR systems, but race, ethnicity and payer type often are not captured within current dental EHR

systems using national standard terminologies. These also were identified as areas requiring further examination during testing.

Evaluation of Completeness of Critical Data Elements

Each testing site provided a report on the extent to which there were missing and/or invalid data for each critical data element and for each MU demographic stratification element. Tables 3 and 4 summarize the findings for each site. The findings of data completeness for the data elements mirror the findings of the feasibility assessments above. Standard data elements related to patient identifier, birth date, provider identifier, procedure codes, and tooth number have very low rates of missing and invalid data (0-2%).

Table 3: Rates of Complete, Missing and Invalid Data for Critical Data Elements, CY 2013

	Dental Site 1			Dental Site 2			Medical Site		
Data element	% Values Filled & Valid	% Missing	% Filled but Invalid Value	% Values Filled & Valid	% Missing	% Filled but Invalid Value	% Values Filled & Valid	% Missing	% Filled but Invalid Value
Critical Data Elements									
Unique patient identifier	100%	0%	0%	100%	0%	0%	100%	0%	0%
Birthdate	100%	0%	0%	100%	0%	0%	100%	0%	0%
Unique provider identifier	100%	0%	0%	100%	0%	0%	100%	0%	0%
Procedure codes (CDT or CPT)	99.7%	0%	0.3% (research codes)	100%	0%	0%	98.38%	1.62%	0%
Diagnosis (ICD or SNOMED)	ICD used primarily for medical insurance billing			Dental and Medical Problem Lists are used and can be mapped to structured coding systems.			98.38%	1.62%	0%
Diagnosis (other coding system, including proprietary)	Diagnoses recorded if treatment planning module is used; EZ codes are used; otherwise, diagnoses will appear in treatment notes/text fields but not as a structured data element. Dental/medical problems and findings, such as active caries, can be mapped to structured coding systems.			Dental and Medical Problem Lists are used and can be mapped to structured coding systems.			N/A		
Tooth number	Note: Assessed only for dates of service that include CDT Code 1351 100%	0%	0%	Note: Assessed only for dates of service that include CDT Code 1351 100%	0%	0%			
Caries Risk Assessment	Risk assessment forms adapted from standardized tools are used but historically not consistently implemented across all providers and patients. Site is in process of moving to standardized data collection and capturing as structured codes using new CDT codes.			Risk assessment forms adapted from standardized tools are used. Data previously not captured as structured data element. Data are beginning to be captured as structured data elements through Registry and adoption of CDT codes introduced in 2014.					

Table 4: Rates of Complete, Missing and Invalid Data for MU Stratification Data Elements, CY 2013

Stratification Elements									
	Dental Site 1			Dental Site 2			Medical Site		
Sex	98%	2%	0%	100%	0%	0%	100%	0%	0%
Race (CDCRec Codes)	N/A			N/A			>99.9%	<0.1%	0%
Race (Other coding system)	41%	59%	0%	Race and ethnicity data are captured, but were not incorporated as structured data elements into the production system of the EHR until recently.			N/A		
Ethnicity (CDCRec Codes)	N/A			N/A			>99.9%	<0.1%	0%
Ethnicity (Other coding system)	Ethnicity data historically have been captured as a sub-category of race. Site is in process of splitting race and ethnicity into two data fields using categories aligned with CDCRec codes.			Race and ethnicity data are captured, but were not incorporated as structured data elements into the production system of the EHR until recently.			N/A		
Payer Type (Dental/ Medical Benefit plan; Insurance type using SOP codes)	N/A			N/A			N/A		
Payer Type (Dental/ Medical Benefit plan; Insurance type using other coding system)	100% (specific insurance plan recorded)	0%	0%	100%	0%	0%	100%	0%	0%

Feasibility of Caries Risk Assessment and Diagnosis Code Data Elements

As noted above, diagnosis codes and caries risk assessment have not been consistently captured as structured data elements in dental EHR systems. However, these were not viewed by any of the stakeholders to be barriers to measure feasibility for the reasons described below.

Caries Risk Assessment. For Sealants, the denominator is restricted to children identified as being at moderate to high risk for caries. Structured data elements for moderate or elevated caries risk include two CDT codes introduced in 2014. Follow-on feasibility assessments during testing explored caries risk assessment data capture in greater depth. Dental EHR vendors indicated that they have custom-fillable forms used to capture caries risk assessments. These forms are usually adapted from standard risk assessment tools, such as the CAMBRA, ADA and AAPD tools. Practice sites may also have their own customizable forms to meet local objectives. Risk assessments have been recorded similarly across these forms using the categories of low, moderate, and

high, which correspond to the three new CDT codes. The addition of the new CDT codes to record risk assessment findings will allow the finding from caries risk assessment to be captured as a structured data element going forward.

The new CDT caries-risk assessment codes are included in oral health measures approved for NQF endorsement by the Health and Well-Being expert panel and NQF Steering Committee that reviewed the measures. The new CDT codes were included as part of the measure logic for the administrative claims data version of this measure and two other oral health preventive care measures that were approved for NQF endorsement. The expert panel that reviewed this measure specifically discussed the feasibility of including these new codes and determined that this data element was feasible for implementation.[29]

Sealant Exception Criteria and Diagnosis Codes. A permanent molar may not be sealed if it has not yet erupted, is missing, has active caries, was previously sealed, has an existing or planned restoration, or for other clinical reasons such as fracture. An exception may be applied only if all of the four permanent first molars are non-sealable (i.e., if none of the four permanent first molars is a candidate for a sealant). Although diagnoses are not captured as standardized structured data elements at this time, all vendors indicated that their systems allow for the identification of these different exception reasons through "problem lists" or "condition lists", which can be captured as structured data elements and mapped to standard ICD-9, ICD-10 or SNOMED codes. The ability to accurately identify and capture the relevant exception criteria was verified through both testing with the synthetic test dataset and validation against manual record reviews. Further, diagnostic coding currently is required to report one of the existing 2014 MU oral health measures, Children Who Have Dental Decay or Cavities. Inclusion of diagnoses in MU measures will itself promote better data capture in the future.

MEASURE LOGIC IMPLEMENTATION FEASIBILITY

Initial reviews of the proposed measure logic during DQA feasibility assessments indicated that the measures would be feasible to implement. Vendors noted that to date they have not been able to use the MAT-generated HQMF for eCQMs and instead program the measures based on the human readable versions of the measures, and they have done so successfully.

The feasibility of implementing the measure logic was verified through implementation of the measure logic using the synthetic test data sets and clinical data from the test sites.

SCIENTIFIC ACCEPTABILITY: RELIABILITY/VALIDITY

Reliability addresses the repeatability and precision of measurement and the ability to reliably compare measure scores between reporting entities. Validity refers to the "correctness" of measurement - the extent to which a measure captures what it is intended to measure. Reliable and valid measurement is promoted by clear measure specifications and consistent implementation of those specifications across reporting entities. Following NQF guidance, the focus of reliability and validity testing was on critical data element validation. Additional validity assessments also were conducted.

SUMMARY OF RELIABILITY/VALIDITY FINDINGS

- The measures were specified precisely using the Measure Authoring Tool based on the Quality Data Model and value sets.
- Measure specifications were clarified and refined throughout the testing process to ensure clarity and promote consistent implementation.
- Reliability of measure logic implementation was verified through testing with synthetic test datasets.
- All critical data elements for Care Continuity demonstrated the highest level of agreement, "almost perfect" agreement, between manual and automated record abstraction in the two dental sites.

- All critical data elements for Care Continuity demonstrated "almost perfect" agreement between manual and automated record abstraction in the medical site, with the exception of oral assessment, which could not undergo two-way validation due to under-billing. However, 100% of oral assessments identified through automated billing were validated against manual abstraction.
- All date-of-service level critical data elements for Sealants demonstrated "almost perfect" agreement between manual and automated record abstraction in the two dental sites.
- The patient-level data element for exceptions for Sealants demonstrated "substantial" agreement between manual and automated record abstraction. Discrepancy analysis revealed that correcting manual recording errors and refining the automated programming logic to better capture exceptions reduced one-half of the discrepancies bringing this data element to "almost perfect" agreement.
- The patient-level data element for caries risk assessment was not captured as a structured data element in either site in 2013 and, therefore, did not undergo two-way validation. However, 100% of cases identified as being at elevated caries risk using automated text queries were validated against manual abstraction. The 2014 introduction of CDT codes will allow this element to be captured as a structured data element. The EHR vendors and clinicians are confident that this data element will demonstrate similar high levels of reliability/validity as other data elements captured through CDT codes.
- Comparison of calculated measure scores, using the automated reports following the measure specifications, to the overall frequencies of the relevant procedures in the background data reports for each clinic provided face validity support for the measure scores.
- Calculation of the measure scores with and without the exception criteria applied demonstrated that exceptions are sufficiently frequent to impact the measure scores and validated inclusion of the exceptions.
- Tooth-level versus surface-level exceptions were validated through analysis of exception reasons.

WELL-DEFINED AND PRECISE MEASURE SPECIFICATIONS

With respect to the measure specifications, the criteria for measure reliability are that the measure be (1) "well defined and precisely specified so it can be implemented consistently within and across organizations and allow for comparability" and (2) for eCQMs specifically, specified using "the Quality Data Model (QDM) and value sets vetted through the National Library of Medicine's Value Set Authority Center (VSAC)."[6]

Alignment with Quality Data Model

During the initial measure development and implementation process, the DQA worked with ONC to incorporate changes to the Quality Data Model so that it would work better with the way data are structured in dental EHR systems and the nature of dental clinical data. For example, anatomical location (i.e., tooth number) was incorporated as an attribute to the "diagnosis" and "procedure" data elements. In addition, the DQA created several new value sets for the measures that are incorporated in the NLM VSAC.

Measure Specifications

Once the changes were implemented in the QDM and the new value sets were created, the measure was specified using the Measure Authoring Tool (MAT), which is a web-based application to create a quality measure in conformance with the QDM and in a standardized XML file. The MAT ensures that the measure is specified based on the QDM and uses NLM VSAC value sets, and it provides as output both machine-readable and human-readable versions of the measure. The measure specifications were reviewed throughout the testing process and were refined based on testing results, feedback from the project team during testing, and feedback from the dental stakeholder community on the Interim Report. In addition, during testing, several opportunities for providing guidance to clarify appropriate and consistent implementation were identified; this guidance was also included in the human readable version of the measure specifications. Because vendors do not use the MAT generated specifications, additional effort was undertaken to incorporate sufficient

guidance in plain language within the metadata for each measure to ensure uniform implementation. In addition, the Oversight Workgroup calls where all vendors participated were used to ensure that there were no issues with interpretation.

RELIABILITY OF MEASURE LOGIC IMPLEMENTATION

The synthetic test datasets, which contained known values of the critical data elements, were used to test the EHR software's translation of the measure logic and calculated results within both EHR developer and practice site testing environments. Figure 6 illustrates the process used to implement the test datasets and assure reliable measure logic implementation with an example testing iteration for the first Care Continuity test dataset with one of the vendors. Table 5 summarizes the findings of multiple iterations of the first Sealant test dataset with one of the vendors. A complete set of findings for each test dataset is contained in Appendix 4.

Within the vendors' test environments, none of the test datasets successfully passed testing on the first attempt, indicating the value of using test data to ensure not only the feasibility of implementing the measure but also the ability to reliably implement the measure within and across EHR systems. **In addition, engaging in this process early on during measure development in partnership with EHR vendors also provides an early opportunity to clarify and refine the measure logic prior to widespread implementation.** Through this iterative process, the following types of issues were identified: (1) misinterpretation the measure logic and/or programming errors of by vendors and (2) errors in the measure logic specification (compared to the measure intent) by the measure developer. Specific examples of these include: (1) incorrect implementation by vendors of the provider attribution for the IPP, DEN, and NUM, (2) clarification and correction by the measure developer related to age eligibility and calculation, and (3) clarification and refinement of exception criteria by the measure developer based on vendor feedback.

Figure 6: Synthetic Dataset Testing Process

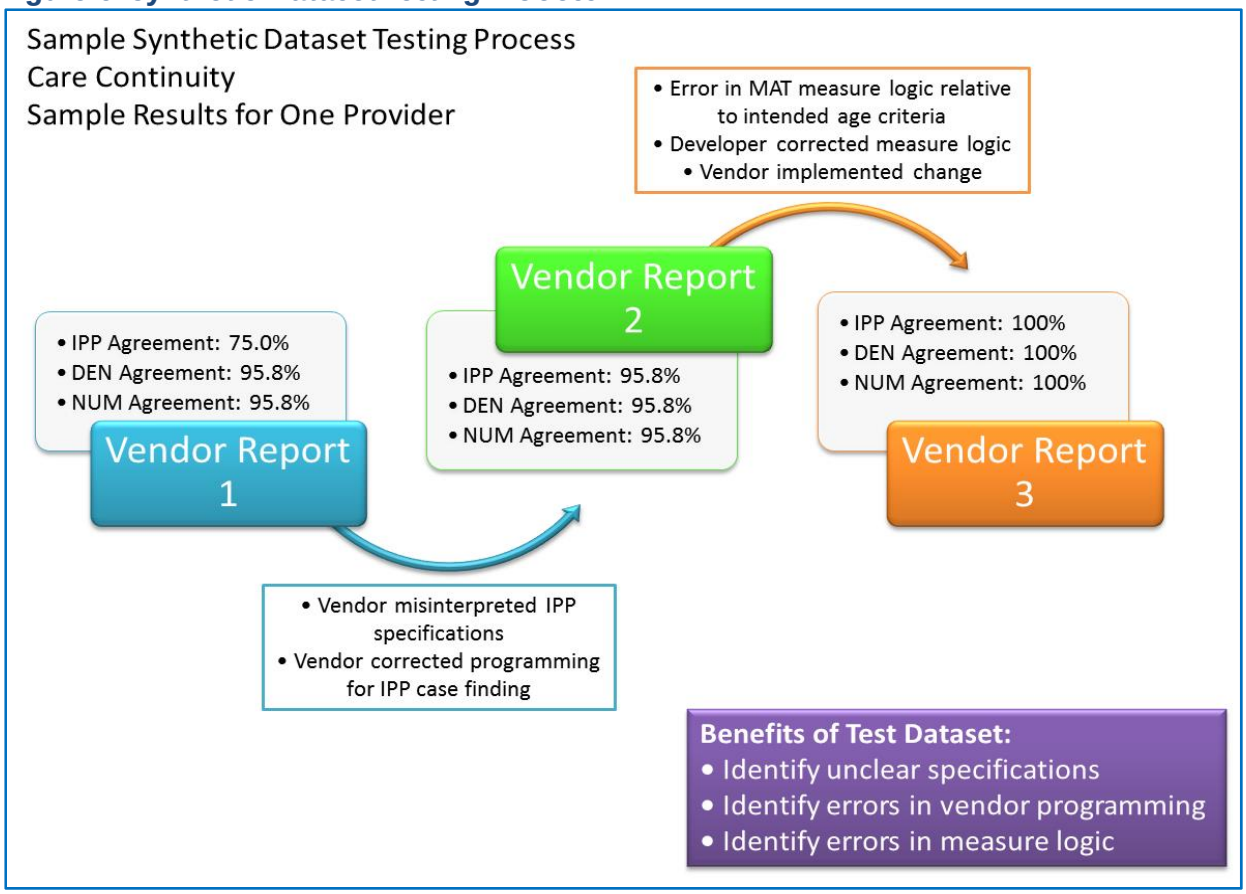


Table 5: Synthetic Test Dataset Testing - Sealants

SYNTHETIC TEST DATASET TESTING: SEALANTS						
Vendor 2, Test Dataset 1	**Agreement between Known Values and Automated Report**					
	Y/Y	**Y/N**	**N/Y**	**N/N**	**Agreement**	**Kappa Statistic**
Report 1	**Provider 1**					
IPP	31	0	0	29	100.00%	1.000
DEN	8	8	0	44	86.67%	0.595
NUM	3	0	2	55	96.67%	0.733
EXC	0	1	0	59	98.33%	Not calculable
	Provider 2					
IPP	35	0	1	24	98.33%	0.966
DEN	11	8	1	40	85.00%	0.615
NUM	4	0	2	54	96.67%	0.783
EXC	0	0	0	60	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	IPP: Age discrepancy. Review of discrepancy identified (1) specification in measure logic did not capture intent and (2) clarification needed about age as continuous versus step function. Measure developer corrected measure logic and provided age calculation guidance in human readable metadata. Vendor implemented changes. DEN: Identification of elevated risk through SNOMED codes (versus CDT codes). Test dataset structure did not mirror how SNOMED risk codes were represented in vendor's system (test dataset associated diagnosis codes with procedures; vendor's system included diagnoses within problem list). Vendor adapted process for implementing test dataset to extract codes and transfer to appropriate place within their system. NUM: Sealant placement not restricted to permanent first molar. Vendor corrected programming logic.					
Report 2	**Provider 1**					
IPP	31	0	0	29	100.00%	1.000
DEN	16	0	0	44	100.00%	1.000
NUM	3	0	0	57	100.00%	1.000
EXC	0	1	0	59	98.33%	Not calculable
	Provider 2					
IPP	35	0	0	25	100.00%	1.000
DEN	19	0	0	41	100.00%	1.000
NUM	3	1	0	56	98.33%	0.849
EXC	0	0	0	60	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	NUM: Vendor programming logic limited numerator procedures to selected denominator provider; vendor corrected logic. EXC: Test dataset associated diagnosis and finding codes with procedures; vendor's system captures this information in problem lists or clinical exam record, so all exceptions were not detected. Vendor created clinical exam for patient so algorithm could detect.					
Report 3	**Provider 1**					
IPP	31	0	0	29	100.00%	1.000
DEN	16	0	0	44	100.00%	1.000
NUM	3	0	0	57	100.00%	1.000
EXC	1	0	0	59	100.00%	1.000
	Provider 2					
IPP	35	0	0	25	100.00%	1.000
DEN	19	0	0	41	100.00%	1.000
NUM	4	0	0	56	100.00%	1.000
EXC	0	0	0	60	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications for each measure component. Measure scores for provider with exception did not take into account exception. Vendor successfully corrected this with submission of final report (Report 4).					

CRITICAL DATA ELEMENT RELIABILITY/VALIDITY

Care Continuity

Dental Sites. Table 6 below summarizes the findings from the two dental sites of the validation between the manual abstraction and automated reports for the critical data elements used to calculate Oral Health Care Continuity for 2-20 Year Olds, including age, provider identifier, and procedures (oral evaluations). There were 197 dates of service included in Site 1 and 221 dates of service in Site 2.

Table 6: Data Element Reliability/Validity, Care Continuity, Dental Sites

CRITICAL DATA ELEMENT VALIDATION: DENTAL SITES	Agreement between Manual Abstraction and Automated Report				Agreement	Kappa Statistic (95% CI)	Concordance Correlation Coefficient (Continuous Variable)
Care Continuity	Y/Y	Y/N	N/Y	N/N			
Dental Site 1 (# Dates of Service: 197)							
Provider_ID					98.98%	0.990 (0.979 - 1.000)	
Age					100.00%		1.000 (1.000-1.000)
	Manual Abstraction: CDT code or Documentation						
Periodic/Comprehensive Evaluation	60	2	0	135	98.98%	0.976 (0.944 - 1.000)	
Problem-Focused Evaluation	24	2	0	171	98.98%	0.954 (0.891 - 1.000)	
	Manual Abstraction: Documentation						
Periodic/Comprehensive Evaluation	59	2	1	135	98.48%	0.964 (0.924 - 1.000)	
Problem-Focused Evaluation	24	2	0	171	98.98%	0.954 (0.891 - 1.000)	
Dental Site 2 (# Dates of Service: 221)							
Provider_ID					100.00%	1.000 (1.000-1.000)	
Age					100.00%		1.000 (1.000-1.000)
	Manual Abstraction: CDT code or Documentation						
Periodic/Comprehensive Evaluation	138	1	0	82	99.55%	0.990 (0.971 - 1.000)	
Problem-Focused Evaluation	10	0	1	210	99.55%	0.950 (0.852 - 1.000)	
	Manual Abstraction: Documentation						
Periodic/Comprehensive Evaluation	138	1	0	82	99.55%	0.990 (0.971 - 1.000)	
Problem-Focused Evaluation	10	0	1	210	99.55%	0.950 (0.852 - 1.000)	

Concordance ranged from 98.48% to 100% for all of the critical data elements. Kappa statistic values ranged from 0.950 – 0.990, indicating "almost perfect" agreement. The results were not sensitive to whether we allowed a match based on a supporting CDT code or documentation versus documentation alone, indicating that there also was

very high agreement between the procedure codes and provider documentation of services performed. Thus, the findings indicate high data element reliability and validity.

Medical Site. Table 7 below summarizes the findings from the medical site. There were 445 dates of service included. Concordance ranged from 97.07% to 100% for the critical data elements used to identify IPP/DEN eligibility. Kappa statistic values ranged from 0.806 – 0.964, indicating "almost perfect" agreement. The results were robust to whether we allowed a match based on a supporting CPT code or documentation versus documentation alone. For oral assessments, during our background data analyses, we found very few procedure codes for oral assessments. The low rates appearing in the data raised face validity concerns among the clinicians who believed these services were performed regularly. Upon further investigation through manual record reviews and tracking the workflow billing processes, it was determined that very few of these services were actually being billed. Therefore, the focus was on validating the cases that had been billed and could be identified through automated reporting against the manual chart abstraction. As a result, we could only compare simple concordance. All cases identified through automated reporting were validated through the manual chart reviews. It is expected that once these services are routinely billed, there will be similarly high levels of agreement and kappa values as is observed with other CPT procedure codes.

Table 7: Data Element Reliability/Validity, Care Continuity, Medical Site

CRITICAL DATA ELEMENT VALIDATION: MEDICAL SITE	Agreement between Manual Abstraction and Automated Report				Agreement	Kappa Statistic (95% CI)	Concordance Correlation Coefficient (Continuous Variable)
Care Continuity	Y/Y	Y/N	N/Y	N/N			
Medical Site (# Dates of Service: 445)							
Provider_ID					100.00%	1.000 (1.000-1.000)	
Age					100.00%		1.000 (1.000-1.000)
	Manual Abstraction: CPT code or Documentation						
Initial Comprehensive Preventive Visit	14	2	3	426	98.88%	0.843 (0.707-0.978)	
Periodic Comprehensive Preventive Visit (443 DOS; 2 indeterminate)	184	8	4	247	97.29%	0.945 (0.914 - 0.976)	
Office Visit (440 DOS, 5 indeterminate)	227	5	3	205	98.18%	0.964 (0.939-0.989)	
	Manual Abstraction: Documentation						
Initial Comprehensive Preventive Visit	13	2	4	426	98.65%	0.806 (0.654-0.957)	
Periodic Comprehensive Preventive Visit (443 DOS; 2 indeterminate)	183	8	5	247	97.07%	0.940 (0.908 - 0.972)	
Office Visit (440 DOS, 5 indeterminate)	226	5	4	205	97.95%	0.959 (0.932-0.986)	

	Oral Assessment: Cases Identified Through Automated Reporting Validated through Manual Abstraction				
	Y/Y	Y/N	N/Y	N/N	Agreement
Oral Assessment (24 Dates of Service; 1 Indeterminate)	24	0	0	0	100.00%

Sealants

Tables 8 and 9 below summarize the findings from the two dental sites of the validation between the manual abstraction and automated reports for the critical data elements used to calculate Oral Health Sealants for 6-9 Year Olds, including age, provider identifier, procedures, tooth number, and exceptions. There were 149 dates of service included in Site 1 and 153 dates of service in Site 2. For the provider identifier, age, oral evaluation procedures, sealant procedures, and tooth number data elements, concordance ranged from 95.97%-100% in Site 1 and from 99.35%-100% in Site 2. Kappa

statistic values ranged from 0.901-1.000 in Site 1 and 0.920-1.000 in Site 2, indicating "almost perfect" agreement for these critical data elements. There were no differences in either site when we allowed a match based on a supporting CPT code or documentation versus documentation alone. Exception validation occurred at the patient level. Concordance was 94.92% in Site 1 and 89.23% in Site 2 with kappa statistic values of 0.742 and 0.767, respectively, indicating "substantial" agreement. Discrepancy analysis revealed in Site 1 that two of three discrepancies were due to standardized reporting of unerupted teeth not detecting how this was locally coded, which was subsequently corrected by revising the programming logic. After this correction, the kappa value increased to 0.931 ("almost perfect" agreement). The third discrepancy was due to a child who had sealants previously placed as part of a research study, resulting in the service not being reflected as a transaction and, therefore, not detected in the automated reporting. In site 2, three of the seven discrepancies were due to manual recording errors that failed to record exceptions; after correction, the kappa statistic value increased to 0.870 ("almost perfect" agreement). Three of the four cases of exceptions detected through manual abstraction but not the automated report were due to prior sealants or restorations that were conducted outside of the system. Site 2 noted that it had identified exceptions related to prior sealants or restorations by querying posted procedures and that revising the query to also include charted findings would further increase accuracy. Caries risk assessment was not captured as a structured data element in either site during 2013. In Site 2, the automated report searched for different text phrases to identify elevated caries risk and we validated those findings against manual record abstraction. All (100%) of the cases identified as being at elevated risk through this automated reporting process were confirmed through manual record abstraction. We did not continue to run different iterations of the automated report to refine the text search strategy because starting in 2014, sites will be capturing these as structured data elements through the use of the newly introduced CDT codes.

Table 8: Data Element Reliability/Validity, Sealants, Dental Site 1

CRITICAL DATA ELEMENT VALIDATION: DENTAL SITE 1	Agreement between Manual Abstraction and Automated Report				Agreement	Kappa Statistic (95% CI)	Concordance Correlation Coefficient (Continuous Variable)
Sealants	Y/Y	Y/N	N/Y	N/N			
Dental Site 1 (# Dates of Service: 149)							
Provider_ID					99.33%	0.993 (0.979 - 1.000)	
Age					100.00%		1.000 (1.000-1.000)
	Date of Service Level Elements						
Periodic/Comprehensive Evaluation	73	0	3	73	97.99%	0.960 (0.915-1.000)	
Problem-Focused Evaluation	5	0	0	144	100.00%	1.000 (1.000-1.000)	
Sealants Placed	39	6	0	104	95.97%	0.901 (0.823-0.978)	
Tooth Number (if sealant placed: n=39)							
3 - Maxillary Right First Molar	24	0	1	14	97.44%	0.945 (0.839-1.000)	
14 - Maxillary Left First Molar	25	0	0	14	100.00%	1.000 (1.000-1.000)	
19 - Mandibular Left First Molar	26	0	1	12	97.44%	0.941 (0.828-1.000)	
30 - Mandibular Right First Molar	25	0	0	14	100.00%	1.000 (1.000-1.000)	
	Patient Level Elements						
Exceptions (n=59)	5	3	0	51	94.92%	0.742 (0.468-1.000)	Two discrepancies due to standardized automated reporting not detecting a local code (correctable); one due to sealants previously placed as part of research study not reflected in procedure codes (because not billed for). Correction of 2 correctable discrepancies brings kappa value to 0.931.

Table 9: Data Element Reliability/Validity, Sealants, Dental Site 2

CRITICAL DATA ELEMENT VALIDATION: DENTAL SITE 2	Agreement between Manual Abstraction and Automated Report				Agreement	Kappa Statistic (95% CI)	Concordance Correlation Coefficient (Continuous Variable)
Sealants	Y/Y	Y/N	N/Y	N/N			
Dental Site 2 (# Dates of Service: 153)							
Provider_ID					99.35%	0.991 (0.963 - 1.000)	
Age					100.00%		1.000 (1.000-1.000)
	Date of Service Level Elements						
Periodic/Comprehensive Evaluation	101	0	0	52	100.00%	1.000 (1.000-1.000)	
Problem-Focused Evaluation	6	0	1	146	99.35%	0.920 (0.763 - 1.000)	
Sealants Placed	20	0	0	133	100.00%	1.000 (1.000-1.000)	
Tooth Number (if sealant placed: n=20)							
3 - Maxillary Right First Molar	14	0	0	6	100.00%	1.000 (1.000-1.000)	
14 - Maxillary Left First Molar	14	0	0	6	100.00%	1.000 (1.000-1.000)	
19 - Mandibular Left First Molar	12	0	0	8	100.00%	1.000 (1.000-1.000)	
30 - Mandibular Right First Molar	13	0	0	7	100.00%	1.000 (1.000-1.000)	
	Patient Level Elements						
Exceptions (n=65)	20	3	4	38	89.23%	0.767 (0.604 - 0.930)	3 discrepancies due to manual recording errors; with correction, agreement=93.85% & Kappa=0.870.

ADDITIONAL VALIDATION ASSESSMENTS

Measure Score Component Reliability/Validity

We also conducted a more aggregated reliability/validity assessment at the measure score component level – that is, validation of the IPP, DEN, NUM and EXC components at the patient-provider level to ensure reliable/valid determinations for each component by the automated reports, including appropriate provider attribution. Tables 10 and 11 provide the results of these analyses for Care Continuity and Sealants, respectively. As expected, the measure score component validation was consistent with that for the individual data elements, with kappa statistic values indicating "almost perfect" agreement – i.e., high reliability/validity – for all components in both sites except for exceptions in Site 2, which was at the high end of the range for "substantial" agreement. For the sealant measure, the elevated risk criteria were not applied to the

denominator eligibility criteria for this analysis for the reasons described above. Because this validation was conducted at the patient-provider level, it also demonstrates the reliability/validity of the provider attribution for each measure score component.

Table 10: Reliability/Validity of Measure Score Components, Care Continuity

MEASURE SCORE COMPONENT VALDATION: CARE CONTINUITY						
Dental Site 1 (141 unique patient-provider observations)	**Agreement between Manual Abstraction and Automated Report**				**Agreement**	**Kappa Statistic (95% CI)**
	Y/Y	**Y/N**	**N/Y**	**N/N**		
IPP/DEN	48	4	0	89	97.16%	0.938 (0.878-0.998)
NUM	26	2	0	24	96.15%	0.923 (0.819-1.000)
Dental Site 2 (130 unique patient-provider observations)	**Agreement between Manual Abstraction and Automated Report**				**Agreement**	**Kappa Statistic (95% CI)**
	Y/Y	**Y/N**	**N/Y**	**N/N**		
IPP/DEN	62	1	1	66	98.46%	0.969 (0.927-1.000)
NUM	50	1	0	12	98.41%	0.950 (0.853-1.000)

Table 11: Reliability/Validity of Measure Score Components, Sealants

MEASURE SCORE COMPONENT VALDATION: SEALANTS						
Dental Site 1 (101 unique patient-provider observations)	**Agreement between Manual Abstraction and Automated Report**				**Agreement**	**Kappa Statistic (95% CI)**
	Y/Y	**Y/N**	**N/Y**	**N/N**		
IPP	72	0	2	27	98.02%	0.951 (0.883-1.000)
DEN	70	0	0	2	100.00%	1.000 (1.000-1.000)
NUM	44	0	0	26	100.00%	1.000 (1.000-1.000)
EXCEPTION	9	1	0	16	96.15%	0.917 (0.759-1.000)
Dental Site 2 (95 unique patient-provider observations)	**Agreement between Manual Abstraction and Automated Report**				**Agreement**	**Kappa Statistic (95% CI)**
	Y/Y	**Y/N**	**N/Y**	**N/N**		
IPP	94	0	1	0	98.95%	Insuffiicent variation to calculate
DEN	90	0	0	4	100.00%	1.000 (1.000-1.000)
NUM	24	0	0	66	100.00%	1.000 (1.000-1.000)
EXCEPTION	32	4	3	27	89.39%	0.787 (0.637-0.936)

Validation of Calculated Clinical Measure Scores against Overall Frequencies of Service Use

Because it would be unwieldy to report individual measure scores for all providers, Tables 12 and 13 provide calculated average provider measure scores by clinic for each of the clinics within the two dental testing sites. We also report average provider measure scores for the required stratifications of the age group and evaluation type (periodic/comprehensive versus problem-focused) for Care Continuity. We used these measure scores to also conduct face validity assessments by comparing the average

provider measure scores with data on service use in each clinic. The findings were consistent. For example, the two clinics in Site 1 with average provider measure scores below10% for Care Continuity (Table 12) had higher percentages of pediatric patients receiving problem-focused evaluations in the year prior to the measurement year and fewer percentages of children who had any type of visit two years in a row. One clinic reported that its pediatric patient population was largely transient (e.g., undocumented, have court-appointed dental screenings but will return to legal guardians, etc.). The other clinic reported that many of its patients are referred as "terminal referrals," often for treatment in the operating room, because of special needs or behavioral issues, and reported challenges with families maintaining follow-up appointments for routine care.

Table 12: Average Provider Measure Scores, by Clinic, Care Continuity

`				Measure Score if DEN>0, (Mean)	Age Stratifications			Evaluation Type Stratifications	
		# Providers with IPP>0	# Providers with DEN>0	Overall	2-5 years	6-12 years	13-20 years	Periodic/ Comp	Problem-Focused
Site 1	Clinic 1	297	297	42.3%	51.5%	50.2%	38.2%	46.3%	23.6%
	Clinic 2	35	35	9.0%	23.1%	5.0%	6.2%	12.0%	2.9%
	Clinic 3	39	39	5.5%	18.2%	9.7%	4.4%	21.6%	2.1%
Site 2	Clinic 1	2	2	47.0%	38.1%	52.6%	43.5%	48.0%	27.8%
	Clinic 2	2	2	63.3%	65.3%	61.5%	79.4%	63.4%	79.7%
	Clinic 3	5	5	79.3%	75.4%	82.9%	77.2%	79.6%	74.3%

The sealant measures are reported with and without exceptions (Table 13). Per above, caries risk findings were not applied. The average provider measure scores ranged from 23% - 38% without exceptions and from 28% - 59% with exceptions. As with Care Continuity, comparison to the frequency of sealant codes in the background data reports for the clinics provided face validity support for the reported measure scores.

Table 13: Average Provider Measure Scores, by Clinic, Sealants

				Measure Score if DEN>0 (Mean, SD)	
		# Providers with IPP>0	# Providers with DEN>0	Without Exceptions	With Exceptions
Site 1	Clinic 1	164	146	38.1%	48.4%
	Clinic 2	31	30	30.3%	39.7%
	Clinic 3	18	11	22.7%	37.5%
Site 2	Clinic 1	2	2	28.5%	33.7%
	Clinic 2	1	1	23.8%	28.0%
	Clinic 3	5	5	35.1%	59.2%

Validation of Inclusion of Exceptions

From a clinical perspective, the exceptions were included to take into account that a child otherwise eligible for the measure may not have any permanent first molars that are candidates for sealants for clinically justified reasons. The practitioner community within the DQA indicated a strong sentiment that this must be properly accounted for, especially in a provider-level measure, in order to establish a valid measure. The testing data from both the data element/measure score component validation and the provider-level measure score reports indicate that exceptions are not rare, they can impact the measure score, and the impact may vary by provider. Thus, the exceptions are also supported from a measurement perspective.

Validation of Tooth-Level versus Surface-Level Exception Specifications

Although a few cases of active caries were observed only on non-occlusal surfaces on individual teeth, these instances had a minor impact on patient-level exceptions (Table 14). Only one patient in Site 1 (and none in Site 2) who qualified for an exception did so, in part, due to active caries on non-occlusal surfaces. The measure score impact was determined to be minimal and outweighed by the additional measurement burden of moving to surface level exceptions.

Table 14: Exception Reasons, Reported as a Percentage of Permanent First Molars among Individuals Qualifying for an Exception

Exception Reason	Site 1 9 patients, 36 permanent 1st molars	Site 2 23 patients, 92 permanent 1st molars	Total (Both Sites)
Active caries - occlusal	8.3%	3.3%	4.69%
Active caries - non-occlusal	5.6%	0.0%	1.56%
Existing restoration	11.1%	8.7%	9.38%
Existing sealant	33.3%	79.3%	66.41%
Fracture	0.0%	0.0%	0.00%
Missing tooth	0.0%	0.0%	0.00%
Pulp involvment	0.0%	0.0%	0.00%
Restoration placed during visit	8.3%	4.3%	5.47%
Unerupted tooth	33.3%	4.3%	12.50%
Other nonsealable indication	0.0%	0.0%	0.0%
Total	100.0%	100.0%	100.0%

Lessons Learned

IMPORTANCE OF STAKEHOLDER ENGAGEMENT THROUGHOUT THE DEVELOPMENT AND TESTING PROCESS

The collaborative, transparent, and stakeholder-engaged approach to measure development and testing provided numerous insights both for the specific measures as well as for the overall development and testing processes. Instead of using a traditional "Technical Expert Panel" solely at the measure conceptualization stage or initial measure specification stage, we used the same group of experts to oversee the entire process. We also ensured that our group had a broad range of subject matter experts including clinicians, EHR vendors, payers, and administrators to ensure a robust interchange.

Release of a survey through national organizations sought to get input from the broader community of implementers who typically are not directly engaged with measure development. Along with the release of an Interim Report to inform and solicit feedback from all stakeholders, these activities were key to build ownership and "buy-in" from the broader practitioner community.

BENEFITS AND LIMITATIONS OF TEST DATASETS

The benefits of the test datasets were substantial. The iterations of testing that were needed to achieve 100% agreement demonstrated that there is a high risk that the measure logic will not be implemented as intended and will not be implemented consistently across EHR systems and practice sites without such testing. Of equal importance, the process of testing with the test dataset before the specifications were finalized brought to light errors in the measure specifications/logic, important areas where guidance was needed, and re-opened dialogue around critical measure aspects. Even though the focus was on the feasibility and reliability aspects of the testing, we also found that some questions addressed face validity as well. Thus, the synthetic test dataset testing became an integral part of the collaborative process to review and refine the measure specifications and logic. The measure specifications are undoubtedly clearer, more precisely defined, and more reflective of the intent of the measure as a result of this process.

One potential limitation of the test datasets is that the test patients need to be carefully designed to ensure robust testing and the patients may need to be customized to the measures being tested. Entering the project, our ideal aim was to create a parameterized tool that would allow us to recreate synthetic datasets on demand using a completely automated process. Ultimately, the customization and complexity required to provide for all of the various scenarios produced in a real-world dataset were more than could be accomplished within the constraints of this particular project. The parameterized elements remained the foundation of the programming and were key drivers of the final data creation process, but a few manual edits were required in order to produce the final datasets. Finally, if measure implementation requires any local customization, then test datasets may not simulate all aspects of actual measure implementation. Local testing and verification is still required.

NEED FOR LOCAL TESTING

As noted above, despite the extensive use of robust test datasets and the numerous benefits that provided, we still found that it was important to do local validation by conducting face validity assessments of the measure scores and comparing automated reporting to manual record review. We also found this process was important for ensuring capture of local codes and data capture related to charted findings or problem/condition lists, which were important for exception determinations.

MU STRATIFICATIONS WITH NON-MUTUALLY EXCLUSIVE CATEGORIES

Sites required guidance on how to address cases in which patients could be classified into more than one category. This was an issue for both the race and payer type categories. Patients may report (and sites may record in their EHR) more than one applicable race. The current categorizations for stratification do not allow for listing multiple categories, nor do they have a multi-racial category. Sites were instructed to use the "primary" race if such a field existed or first-listed. Payer type can vary over time as well as by procedure/service type. Therefore, it is not uncommon for patients to have multiple payer types associated with their service use within a measurement period. When reporting entities are asked to report MU measure scores stratified by payer type, it will be important to have guidance on how to handle patients with more than one payer during the measurement period.

We also provided ONC with specific feedback during the course of this project on the payer type value set that is currently in use for the MU measures. Our project team noted that many of the various existing classifications of payers do not apply to dentistry and such an extensive system may ultimately not result in meaningful information.

COMPARABILITY TO SIMILAR MEASURES SPECIFIED FOR ADMINISTRATIVE CLAIMS DATA

Having previously developed pediatric oral health measures specified for administrative claims data, there was an initial effort to "adapt" the administrative measures for EHR specification. It quickly became apparent, however, that due to both the differences in the nature of the data (historical administrative claims data versus real-time EHR data) and the level of reporting (program/plan versus provider) that these are distinct measures that provide related, but very different information. For example, many administrative data measures include enrollment criteria as a primary denominator inclusion requirement – for example, including all children enrolled at least six months continuously in Medicaid – which captures both children who access the care delivery system and those who do not. By its very nature, EHR data captures only those who access the care delivery system; however, it can also capture data for both insured and uninsured populations and is potentially more reliable for clinician-level measurement.

It is also important to note the limitations that exist with claims data can be overcome with data from patient records and every effort should be made to use the richness of each data source for measurement. EHR data also allows for much more refined information, such as being able to capture charted findings that may not be captured in standard claims data. Thus, simply "re-tooling" measures from one data source to another should not be encouraged. However, in such cases, it should be noted that comparing measure scores from "similar" measures derived from different data sources may not be appropriate or reliable. As the health system evolves and interoperable EHRs become a reality, they likely will become the best source for reliable and valid clinical quality measures. In the interim, efforts should be made to avoid unnecessary duplication and to clarify the differences and relative strengths and limitations of related quality measures. A "library" of measures addressing the same quality improvement goal but specified to different data sources may be needed to engage stakeholders in measurement.

Impact

Ultimately, the goal of measurement is quality improvement that promotes improved health outcomes. An unexpected outcome of this project was the almost immediate impact it had on the test sites involved. Both the clinical informatics and clinical quality specialists within the test sites used this project as an opportunity to advance local quality initiatives. The medical site is developing quality improvement initiatives to increase the number of physicians who conduct oral health screenings, not only locally but with an eye toward interventions that can be shared and implemented on a broader scale. As noted above, one site found that it was under-billing for reimbursable procedures; addressing this has the twofold benefit of both a direct financial benefit for that site as well as a better record of services provided. One dental site is refining its clinical processes to implement a standard caries risk assessment form for documentation in the EHR and to implement the new CDT caries risk finding codes. One dental site that had collected ethnicity as a sub-category of race has already created two separate data elements and aligned the reporting categories with the MU categories. None of these initiatives were required to carry out this project, which relied on existing data at the time the project commenced. Rather, the individuals within these sites indicated that these were all things they felt needed to happen and expressed enthusiasm about the opportunity to move these efforts forward. This proved to be one of the most important outcomes of the projects and illustrates the power of constructively using good data and measurement to examine care delivery.

Measure Maintenance

Measure specifications will be reviewed annually and be revised or retained. Information from user feedback, updates to codes used within the measure, emerging data from measure implementers, emerging evidence, and methodological advances will impact measure updates. The DQA's Measure Development and Maintenance Committee will lead measure review and maintenance.

References

1. Dental Quality Alliance. Dental Quality Alliance Measure Set #1: Dental Caries in Children - Prevention & Disease Management. 2013. Available at: http://www.ada.org/8472.aspx. Accessed April 1, 2014.

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

3. National Quality Measures Clearinghouse Measure Collection: Dental Caries in Children: Prevention & Disease Management. 2014. Available at: http://www.qualitymeasures.ahrq.gov/browse/by-organization-indiv.aspx?orgid=2485&objid=47294. Accessed August 15, 2014.

4. Dental Quality Alliance. Pediatric Oral Healthcare: Exploring the Feasibility for E-Measures. 2012. Available at: http://www.ada.org/sections/dentalPracticeHub/pdfs/ARCHIVE_1214_FINAL_DQA_E-Measure_Working_Document.pdf. Accessed April 1, 2014.

5. Dental Quality Alliance. Request for Proposal: Assuring Feasibility, Reliability, Validity and Usability of eMeasures. 2013. Available at: http://www.ada.org/sections/dentalPracticeHub/pdfs/Dental_eMeasures_RFP.pdf. Accessed May 15, 2013.

6. National Quality Forum. Measure Evaluation Criteria and Guidance. 2013. Available at: http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed September 15, 2014.

7. Dye BA, Li X, Thorton-Evans G. Oral health disparities as determined by selected healthy people 2020 oral health objectives for the United States, 2009-2010. NCHS Data Brief 2012(104):1-8.

8. Tinanoff N, Reisine S. Update on early childhood caries since the Surgeon General's Report. Acad Pediatr 2009;9(6):396-403.

9. Casamassimo PS, Thikkurissy S, Edelstein BL, Maiorini E. Beyond the dmft: the human and economic cost of early childhood caries. J Am Dent Assoc 2009;140(6):650-7.

10. Gift HC, Reisine ST, Larach DC. The social impact of dental problems and visits. Am J Public Health 1992;82(12):1663-8.

11. Gray MM, Marchment MD, Anderson RJ. The relationship between caries experience in the deciduous molars at 5 years and in first permanent molars of the same child at 7 years. Community Dent Health 1991;8(1):3-7.

12. Griffin SO, Gooch BF, Beltran E, Sutherland JN, Barsley R. Dental services, costs, and factors associated with hospitalization for Medicaid-eligible children, Louisiana 1996-97. J Public Health Dent 2000;60(1):21-7.

13. Hollister MC, Weintraub JA. The association of oral status with systemic health, quality of life, and economic productivity. J Dent Educ 1993;57(12):901-12.

14. O'Sullivan DM, Tinanoff N. The association of early dental caries patterns with caries incidence in preschool children. J Public Health Dent 1996;56(2):81-3.

15. Reisine S, Litt M, Tinanoff N. A biopsychosocial model to predict caries in preschool children. Pediatr Dent 1994;16(6):413-8.

16. Sheller B, Williams BJ, Lombardi SM. Diagnosis and treatment of dental caries-related emergencies in a children's hospital. Pediatr Dent 1997;19(8):470-5.

17. American Academy of Pediatric Dentistry Guideline on Periodicity of Examination, Preventive Dental Services, Anticipatory Guidance/Counseling, and Oral Treatment for Infants, Children, and Adolescents. 2013. Available at: http://www.aapd.org/policies/. Accessed December 7, 2013.

18. American Academy of Pediatrics. Policy statement: preventive oral health ntervention for pediatricians. Pediatrics 2008;122(6):1387-94.

19. National Institute for Health and Care Excellence. Clinical Guidelines. CG19: Dental Recall - Recall Interval between Routine Dental Examinations. 2004. Available at: http://guidance.nice.org.uk/CG19. Accessed on February 15, 2014.

20. Centers for Medicare & Medicaid Services Early and Periodic Screening, Diagnosis, & Treatment: 2011 National Data. 2012. Available at: http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Benefits/Early-Periodic-Screening-Diagnosis-and-Treatment.html. Accessed April 29, 2013.

21. Beauchamp J, Caufield PW, Crall JJ, et al. Evidence-based clinical recommendations for the use of pit-and-fissure sealants: a report of the American Dental Association Council on Scientific Affairs. J Am Dent Assoc 2008;139(3):257-68.

22. Centers for Disease Control and Prevention. Dental Sealants. 2013. Available at: http://www.cdc.gov/oralhealth/publications/faqs/sealants.htm. Accessed September 15, 2014.

23. National Quality Forum. Report from the National Quatliy Forum: eMeasure Feasibility Assessment. April 2013. Available at: http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed September 15, 2014.

24. StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP. 2013.

25. National Quality Forum. Review and Update of Guidance for Evaluating Evidence and Measure Testing: Technical Report. 2013. Available at: http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed September 15, 2014.

26. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977;33(2):363-74.

27. Centers for Medicare & Medicaid Services. An Introduction to the Medicaid EHR Incentive Program for Eligible Professionals. 2012. Available at: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Getting_Started.html. Accessed April 1, 2014.

28. National Quality Forum. Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties. January 2011. Available at: http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed September 15, 2014.

29. National Quality Forum. Health and Well-Being Measures: Meeting Transcript, Day 1, April 29, 2014. 2014. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=76558. Accessed September 23, 2014.

Appendix 1: Project Team

THE UNIVERSITY OF FLORIDA PROJECT TEAM

ORAL HEALTH SERVICES RESEARCH TEAM

Jill Boylston Herndon, PhD, Principal Investigator/Team Lead
Department of Health Outcomes & Policy, College of Medicine
Department of Community Dentistry & Behavioral Science, College of Dentistry
University of Florida

Scott L. Tomar, DMD, MPH, DrPH, Co-Investigator
Professor, Department of Community Dentistry and Behavioral Science
University of Florida College of Dentistry

Frank A. Catalanotto, DMD, Co-Investigator
Professor and Chair, Department of Community Dentistry and Behavioral Science
University of Florida College of Dentistry

Elizabeth A. Shenkman, PhD, Co-Investigator
Professor and Chair, Department of Health Outcomes and Policy
Director, Institute for Child Health Policy
University of Florida College of Medicine

Kaitlin Sovich, B.A., Project Coordinator and Research Assistant
Department of Health Outcomes and Policy
University of Florida College of Medicine

COLLEGE OF DENTISTRY – DENTAL INFORMATICS/PRACTICE SITE

Stephen H. Kostewicz, MS, Dental Informatics - Test/Implement Measures in UF Practice
Manager, Application Support & Development
University of Florida College of Dentistry

Richelle Janiec, BS, CDA, RDH, Record Reviewer – Validation of Measures in UF Practice
Quality Assurance Specialist
Coordinator, Clinical Programs Quality Assurance
University of Florida College of Dentistry

COLLEGE OF MEDICINE – MEDICAL SETTING FOR CARE CONTINUITY MEASURE

Lindsay Thompson, MD, MS, Co-Investigator, Validate Care Continuity in Medical Setting
Clinical Associate Professor, Pediatrics
Director, General Pediatrics' Continuity Clinic
Division of General Pediatrics, Department of Pediatrics
University of Florida College of Medicine

Anzeela Schentrup, PharmD, PhD, BCPS, Co-Investigator, Expertise in Medical EHR
Clinical Assistant Professor
Department of Pharmacotherapy and Translational Research
University of Florida College of Pharmacy

CONSULTANTS – MEANINGFUL USE, ECQMS, INFORMATICS

Ron Stephenson, Informatics Lead/Coordinator
Senior Developer
Stephenson Developers Group

Gloria "Gigi" Lipori, MT, MBAA, Informatics/eCQM Consultant
Chief Data Officer
UF Health

DENTAL EHR SYSTEM VENDOR PARTNERS

AMERICAN DENTAL PARTNERS, INC.

Jesley C. Ruff DDS, MPH
Senior Vice President, Chief Professional Officer

Tim Freeman, BA
Senior Systems Business Analyst

Roger A. Horton, DMD, MS
Chief Technology Officer

Michael W. Hoyt, BS
Vice President, Information Services

Victoria Post, DDS
Quality Assurance Manager

Ryan Brandon
Executive Director of Product and Strategy

Paul Delicana
Business Analyst

Joraymond Espiritu
Business Analyst

Appendix 2: DQA eCQM Oversight Workgroup

Dr. Craig W. Amundson, D.D.S.
General Practitioner
HealthPartners Medical and Dental Group
National Association of Dental Plans

Ms. Lauren Barone, M.P.H.
Manager, Oral Health
American Academy of Pediatrics

Mr. Ryan Brandon
Executive Director
Product and Strategy
Exan Group

Howard Bregman, M.D., M.S.
Physician/Clinical Informatics
Epic

Dr. James Crall, D.D.S., M.S., Sc.D.
Professor and Chair, Section of Pediatric Dentistry
Director, National Oral Health Policy Center at UCLA
American Academy of Pediatric Dentistry
DQA Chair-Elect (ex-officio)

Dr. Daniel Hickey
U.S. Public Health Service
Academy of General Dentistry

Dr. Renee W. Joskow, D.D.S., M.P.H., F.A.G.D., CAPT, USPHS
Senior Advisor for Oral Health
Health Resources and Services Administration (HRSA)

Dr. Huong N. Le, D.D.S.
Dental Director
Asian Health Services and National Network for Oral Health Access
UCSF Associate Clinical Professor
A T Still Assistant Clinical Professor
AEGD California Assistant Director, Lutheran Medical Center

Dr. Stephanie Lepsky, D.D.S.
Senior Consultant/Consulting Dental Director
OptumInsight

Mr. Chris Millet
Senior Director of e-Measurement
National Quality Forum

Dr. Lynn Douglas Mouden, D.D.S., M.P.H.
Chief Dental Officer
Centers for Medicare & Medicaid Services

Ms. Jean Narcisi
Director of Dental Informatics
Practice Institute
American Dental Association

Ms. Kathleen Noll
Vice President, Operations
QSI Dental

Dr. W. Ken Rich, D.M.D.
General Practitioner
Medicaid -CHIP State Dental Association
DQA Chair (ex-officio)

Dr. Jesley Ruff, D.D.S.
Senior Vice President
Chief Professional Officer
American Dental Partners (ADPI)

Mr. Max Sessions
Product Manager
Henry Schein Practice Solutions

Ms. Kristen Simmons, M.H.A., B.S.D.H., R.D.H.
Chief Operating Officer
Operations, Human Resources and Clinical Strategy
Willamette Dental Group

Dr. Jordan Sparks, D.M.D.
Open Dental

Mr. Daniel Wilkinson
Dentrix

Appendix 3: Feasibility Scorecard

DATA ELEMENTS

Concept	Data Availability as structured data elements (NQF Score)[1]	Workflow: Typically captured during routine clinical care and documentation (NQF Score)	Data accuracy: Who captures this information? (NQF Score)	Stored using standard taxonomies such as CDT, SNOMED, RxNorm, LOINC (NQF Score)	If interface terminologies are used at the provider end, are validated maps available to standard taxonomies?	Does the system use encounters to embed procedure codes or are they indexed by visit date
Procedure - oral evaluation	Yes **(3)**	Yes **(3)**	Dentist or team member chair side **(3)**	CDT **(3)**	Interface not used	Dental systems index by visit date and don't embed in encounters. It is possible for vendors to **artificially** link a procedure to encounter within the system. FQHC's that are reimbursed based on encounter rates also record procedures by visit date. Each visit date is considered an encounter. This does not fit into the description of "encounter" used in the
Procedure - sealant	Yes **(3)**	Yes **(3)**	Dentist or team member chair side **(3)**	CDT **(3)**	Interface not used	

[1] The National Quality Forum published a document regarding the feasibility of eMeasures at http://www.qualityforum.org/Projects/e-g/eMeasure_Feasibility_Testing/eMeasure_Feasibility_Testing.aspx#t=2&s=&p .

						QDM. One of the respondent participated in MU2 testing. They used the human readable version to program their system and pulled data based on visit date ignoring any written "logic". One vendor requested that if the logic only stipulates the use of "encounters" the DQA should define these appropriately and standardize. The use of the "Procedure performed" clause will alleviate this concern for these two measures.
Diagnosis/ Caries Risk assessment – patient level	**Only Sometimes but can be implemented for 2016 MU 3.** From 2014 this can be captured as a CDT code as well. **(Current: 2; Future 3)**	No but do not anticipate too much burden. **(Current: 2; Future 3)**	Dentist or team member chair side **(3)**	Currently stored as custom codes. (SNOMED codes and CDT under development. Can be implemented for 2016) **(Current: 2; Future 3)**	EZ Code – proposed SNOMED: Validated map available. 1:1 map for this data element	

Exclusion Reasons	Some systems have a specific UI for capturing exclusions. All the concepts are listed in drop down menus against the measure name and providers have the ability to select a reason for exclusion of that patient when closing out the record. Other systems typically include logic to search for specific items in the chart including text search of clinical notes. Providers would like some of the major exclusions addressed without burden to the workflow. Further testing is required to understand impact of exclusions.					
Tooth number	Yes **(3)**	Yes **(3)**	Dentist or team member chair side **(3)**	Tooth Numbering System in itself is considered standard taxonomy. That is universally used although SNOMED codes are available. Again no one uses SNOMED **(Current: 2; Future 3)**		
Visit Date/ Date of Service	Yes **(3)**	Yes **(3)**	Office staff **(3)**			
Patient Date of Birth	Yes **(3)**	Yes **(3)**	Office staff **(3)**			
Race	Only sometimes but can be included **(Current: 2; Future 3)**	Sometimes **(Current: 2; Future 3)**	Office staff **(3)**	So far the dental vendors are not using the ONC value sets except one **(Current: 2; Future 3)**	Custom codes used. Map validity unknown	
Ethnicity	Only sometimes but can be included **(Current: 2; Future 3)**	Sometimes **(Current: 2; Future 3)**	Office staff **(3)**	So far the dental vendors are not using the ONC value sets except one **(Current: 2; Future 3)**	Custom codes used. Map validity unknown	
Dental Benefit plan /Insurance type	Only sometimes but can be included **(Current: 2; Future 3)**	Sometimes **(Current: 2; Future 3)**	Office staff **(3)**	So far the dental vendors are not using the ONC value sets except one **(Current: 2; Future 3)**	Custom codes used. Map validity unknown	

Appendix 4: Synthetic Test Dataset Testing Results

CARE CONTINUITY, TEST DATASET 1, VENDOR TESTING

SYNTHETIC TEST DATASET TESTING: CARE CONTINUITY						
Vendor 1, Test Dataset 1	Agreement between Known Values and Automated Report				Agreement	Kappa Statistic
	Y/Y	Y/N	N/Y	N/N		
Report 1	Provider 1					
IPP	14	0	6	4	75.00%	0.438
DEN	14	0	1	9	95.83%	0.913
NUM	8	0	1	15	95.83%	0.909
	Provider 2					
IPP	16	0	7	3	73.08%	0.345
DEN	16	0	1	9	96.15%	0.917
NUM	9	0	1	16	96.15%	0.917
Sources of Discrepancies & Corresponding Resolution	IPP: Programmed logic captured all appointments and did not restrict by procedure type as provided for in specifications. Vendor corrected programming logic.					
Report 2	Provider 1					
IPP	14	0	1	9	95.83%	0.913
DEN	14	0	1	9	95.83%	0.913
NUM	8	0	1	15	95.83%	0.909
	Provider 2					
IPP	16	0	1	9	96.15%	0.917
DEN	16	0	1	9	96.15%	0.917
NUM	9	0	1	16	96.15%	0.917
Sources of Discrepancies & Corresponding Resolution	All: Upper bound of age range in measure specifications indicated <=20 years, instead of intended <20 years. Measure developer corrected specifications; vendor updated measure logic.					
Report 3	Provider 1					
IPP	14	0	0	10	100.00%	1.000
DEN	14	0	0	10	100.00%	1.000
NUM	8	0	0	16	100.00%	1.000
	Provider 2					
IPP	16	0	0	10	100.00%	1.000
DEN	16	0	0	10	100.00%	1.000
NUM	9	0	0	17	100.00%	1.000
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications.					

SYNTHETIC TEST DATASET TESTING: CARE CONTINUITY						
Vendor 2, Test Dataset 1	**Agreement between Known Values and Automated Report**					
	Y/Y	**Y/N**	**N/Y**	**N/N**	**Agreement**	**Kappa Statistic**
Report 1	**Provider 1**					
IPP	0	14	17	0	0.00%	Not calculable
DEN	0	14	17	0	0.00%	Not calculable
NUM	0	8	9	0	0.00%	Not calculable
	Provider 2					
IPP	0	16	15	0	0.00%	Not calculable
DEN	0	16	15	0	0.00%	Not calculable
NUM	0	9	8	0	0.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	All: Providers were attributed to wrong patients in submitted reports. Vendor corrected provider-patient association.					
Report 2	**Provider 1**					
IPP	14	0	0	10	100.00%	1.000
DEN	14	0	0	10	100.00%	1.000
NUM	8	0	0	16	100.00%	1.000
	Provider 2					
IPP	16	0	0	10	100.00%	1.000
DEN	16	0	0	10	100.00%	1.000
NUM	9	0	0	17	100.00%	1.000
Sources of Discrepancies & Corresponding Resolution	• Complete agreement for measure overall. • Minor discrepancies in the race and gender stratifications related to test dataset import. Vendor corrected.					
Report 3	**Provider 1**					
IPP	14	0	0	10	100.00%	1.000
DEN	14	0	0	10	100.00%	1.000
NUM	8	0	0	16	100.00%	1.000
	Provider 2					
IPP	16	0	0	10	100.00%	1.000
DEN	16	0	0	10	100.00%	1.000
NUM	9	0	0	17	100.00%	1.000
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications.					

CARE CONTINUITY, TEST DATASET 2, VENDOR TESTING

SYNTHETIC TEST DATASET TESTING: CARE CONTINUITY						
Vendor 1, Test Dataset 2	**Agreement between Known Values and Automated Report**				**Agreement**	**Kappa Statistic**
	Y/Y	**Y/N**	**N/Y**	**N/N**		
Report 1	**Provider 1**					
IPP	42	0	0	32	100.00%	1.000
DEN	42	0	0	32	100.00%	1.000
NUM	14	7	0	53	90.54%	0.741
	Provider 2					
IPP	51	0	0	21	100.00%	1.000
DEN	51	0	0	21	100.00%	1.000
NUM	23	4	0	45	94.44%	0.878
	Provider 3					
IPP	46	0	0	29	100.00%	1.000
DEN	46	0	0	29	100.00%	1.000
NUM	19	10	0	46	86.67%	0.700
	Provider 4					
IPP	43	0	0	61	100.00%	1.000
DEN	43	0	0	61	100.00%	1.000
NUM	14	2	0	88	98.08%	0.922
Sources of Discrepancies & Corresponding Resolution	NUM: Provider constraint was erroneously added to numerator calculation (restricting qualifying service to being performed by denominator provider instead of any provider). Vendor removed constraint.					
Report 2	**Provider 1**					
IPP	42	0	0	32	100.00%	1.000
DEN	42	0	0	32	100.00%	1.000
NUM	21	0	0	53	100.00%	1.000
	Provider 2					
IPP	51	0	0	21	100.00%	1.000
DEN	51	0	0	21	100.00%	1.000
NUM	27	0	0	45	100.00%	1.000
	Provider 3					
IPP	46	0	0	29	100.00%	1.000
DEN	46	0	0	29	100.00%	1.000
NUM	29	0	0	46	100.00%	1.000
	Provider 4					
IPP	43	0	0	61	100.00%	1.000
DEN	43	0	0	61	100.00%	1.000
NUM	16	0	0	88	100.00%	1.000
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications.					

SYNTHETIC TEST DATASET TESTING: CARE CONTINUITY						
Vendor 2, Test Dataset 2	**Agreement between Known Values and Automated Report**				**Agreement**	**Kappa Statistic**
	Y/Y	**Y/N**	**N/Y**	**N/N**		
Report 1	**Provider 1**					
IPP	42	0	0	32	100.00%	1.000
DEN	42	0	0	32	100.00%	1.000
NUM	14	7	0	53	90.54%	0.741
	Provider 2					
IPP	51	0	0	21	100.00%	1.000
DEN	51	0	0	21	100.00%	1.000
NUM	23	4	0	45	94.44%	0.878
	Provider 3					
IPP	46	0	0	29	100.00%	1.000
DEN	46	0	0	29	100.00%	1.000
NUM	19	10	0	46	86.67%	0.700
	Provider 4					
IPP	43	0	0	61	100.00%	1.000
DEN	43	0	0	61	100.00%	1.000
NUM	14	2	0	88	98.08%	0.922
Sources of Discrepancies & Corresponding Resolution	NUM: Provider constraint was erroneously added to numerator calculation (restricting qualifying service to being performed by denominator provider instead of any provider). Vendor removed constraint.					
Report 2	**Provider 1**					
IPP	42	0	0	32	100.00%	1.000
DEN	42	0	0	32	100.00%	1.000
NUM	21	0	0	53	100.00%	1.000
	Provider 2					
IPP	51	0	0	21	100.00%	1.000
DEN	51	0	0	21	100.00%	1.000
NUM	27	0	0	45	100.00%	1.000
	Provider 3					
IPP	46	0	0	29	100.00%	1.000
DEN	46	0	0	29	100.00%	1.000
NUM	29	0	0	46	100.00%	1.000
	Provider 4					
IPP	43	0	0	61	100.00%	1.000
DEN	43	0	0	61	100.00%	1.000
NUM	16	0	0	88	100.00%	1.000
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications with the exception of payer type classificaiton for three patients, which was corrected. For the test data, setting up the payer type and attaching it to the patient within the vendor's system was a more complex step in the process (and not completely automated). For these three patinets, the linkage did not occur during loading, resulting in them defaulting to the self-pay category instead of the correct categories.					

SEALANTS, TEST DATASET 1, VENDOR TESTING

SYNTHETIC TEST DATASET TESTING: SEALANTS						
Vendor 1, Test Dataset 1	**Agreement between Known Values and Automated Report**				**Agreement**	**Kappa Statistic**
	Y/Y	**Y/N**	**N/Y**	**N/N**		
Report 1	**Provider 1**					
IPP	31	0	0	29	100.00%	1.000
DEN	8	8	0	44	86.67%	0.595
NUM	3	0	0	57	100.00%	1.000
EXC	0	1	0	59	98.33%	Not calculable
	Provider 2					
IPP	35	0	0	25	100.00%	1.000
DEN	11	7	0	41	86.67%	0.653
NUM	4	0	0	56	100.00%	1.000
EXC	0	0	0	60	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	DEN: Programming error related to identifying SNOMED codes for caries risk. Vendor corrected. EXC: Patient who should have qualified for an exception did not.					
Report 2	**Provider 1**					
IPP	31	0	0	29	100.00%	1.000
DEN	15	1	0	44	98.33%	0.957
NUM	3	0	0	57	100.00%	1.000
EXC	1	0	0	59	100.00%	1.000
	Provider 2					
IPP	35	0	0	25	100.00%	1.000
DEN	19	0	0	41	100.00%	1.000
NUM	4	0	0	56	100.00%	1.000
EXC	0	0	0	60	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	DEN: Patient with exception not represented in denominator. Vendor corrected.					
Report 3	**Provider 1**					
IPP	31	0	0	29	100.00%	1.000
DEN	16	0	0	44	100.00%	1.000
NUM	3	0	0	57	100.00%	1.000
EXC	1	0	0	59	100.00%	1.000
	Provider 2					
IPP	35	0	0	25	100.00%	1.000
DEN	19	0	0	41	100.00%	1.000
NUM	4	0	0	56	100.00%	1.000
EXC	0	0	0	60	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications.					

SYNTHETIC TEST DATASET TESTING: SEALANTS						
Vendor 2, Test Dataset 1	**Agreement between Known Values and Automated Report**				**Agreement**	**Kappa Statistic**
	Y/Y	**Y/N**	**N/Y**	**N/N**		
Report 1	**Provider 1**					
IPP	31	0	0	29	100.00%	1.000
DEN	8	8	0	44	86.67%	0.595
NUM	3	0	2	55	96.67%	0.733
EXC	0	1	0	59	98.33%	Not calculable
	Provider 2					
IPP	35	0	1	24	98.33%	0.966
DEN	11	8	1	40	85.00%	0.615
NUM	4	0	2	54	96.67%	0.783
EXC	0	0	0	60	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	IPP: Age discrepancy. Review of discrepancy identified (1) specification in measure logic did not capture intent and (2) clarification needed about age as continuous versus step function. Measure developer corrected measure logic and provided age calculation guidance in human readable metadata. Vendor implemented changes. DEN: Identification of elevated risk through SNOMED codes (versus CDT codes). Test dataset structure did not mirror how SNOMED risk codes were represented in vendor's system (test dataset associated diagnosis codes with procedures; vendor's system included diagnoses within problem list). Vendor adapted process for implementing test dataset to extract codes and transfer to appropriate place within their system. NUM: Sealant placement not restricted to permanent first molar. Vendor corrected programming logic.					
Report 2	**Provider 1**					
IPP	31	0	0	29	100.00%	1.000
DEN	16	0	0	44	100.00%	1.000
NUM	3	0	0	57	100.00%	1.000
EXC	0	1	0	59	98.33%	Not calculable
	Provider 2					
IPP	35	0	0	25	100.00%	1.000
DEN	19	0	0	41	100.00%	1.000
NUM	3	1	0	56	98.33%	0.849
EXC	0	0	0	60	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	NUM: Vendor programming logic limited numerator procedures to selected denominator provider; vendor corrected logic. EXC: Test dataset associated diagnosis and finding codes with procedures; vendor's system captures this information in problem lists or clinical exam record, so all exceptions were not detected. Vendor created clinical exam for patient so algorithm could detect.					
Report 3	**Provider 1**					
IPP	31	0	0	29	100.00%	1.000
DEN	16	0	0	44	100.00%	1.000
NUM	3	0	0	57	100.00%	1.000
EXC	1	0	0	59	100.00%	1.000
	Provider 2					
IPP	35	0	0	25	100.00%	1.000
DEN	19	0	0	41	100.00%	1.000
NUM	4	0	0	56	100.00%	1.000
EXC	0	0	0	60	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications for each measure component. Measure scores for provider with exception did not take into account exception. Vendor successfully corrected this with submission of final report (Report 4).					

SEALANTS, TEST DATASET 2, VENDOR TESTING

SYNTHETIC TEST DATASET TESTING: SEALANTS						
Vendor 1, Test Dataset 2	**Agreement between Known Values and Automated Report**				**Agreement**	**Kappa Statistic**
	Y/Y	**Y/N**	**N/Y**	**N/N**		
Report 1	**Provider 1**					
IPP	48	0	0	31	100.00%	1.000
DEN	21	0	0	58	100.00%	1.000
NUM	9	0	0	70	100.00%	1.000
EXC	0	0	0	79	100.00%	Not calculable
	Provider 2					
IPP	52	0	0	31	100.00%	1.000
DEN	33	0	0	50	100.00%	1.000
NUM	12	0	0	71	100.00%	1.000
EXC	1	0	1	81	98.80%	0.661
	Provider 3					
IPP	55	0	0	34	100.00%	1.000
DEN	27	0	0	62	100.00%	1.000
NUM	11	0	0	78	100.00%	1.000
EXC	0	0	0	89	100.00%	Not calculable
`	EXC: Discrepancy due to timing clause in exception for nonsealable teeth. Vendor corrected. This discrepancy led to a broader discussion and review of the timing for all nonsealable reasons. Measure developer re-defined exception value sets and adjusted timing as a result.					
Report 2	**Provider 1**					
IPP	48	0	0	31	100.00%	1.000
DEN	21	0	0	58	100.00%	1.000
NUM	9	0	0	70	100.00%	1.000
EXC	0	0	0	79	100.00%	Not calculable
	Provider 2					
IPP	52	0	0	31	100.00%	1.000
DEN	33	0	0	50	100.00%	1.000
NUM	12	0	0	71	100.00%	1.000
EXC	1	0	0	82	100.00%	1.000
	Provider 3					
IPP	55	0	0	34	100.00%	1.000
DEN	27	0	0	62	100.00%	1.000
NUM	11	0	0	78	100.00%	1.000
EXC	0	0	0	89	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications.					

SYNTHETIC TEST DATASET TESTING: SEALANTS						
Vendor 2, Test Dataset 2	**Agreement between Known Values and Automated Report**				**Agreement**	**Kappa Statistic**
	Y/Y	**Y/N**	**N/Y**	**N/N**		
Report 1	**Provider 1**					
IPP	48	0	0	31	100.00%	1.000
DEN	21	0	0	58	100.00%	1.000
NUM	9	0	0	70	100.00%	1.000
EXC	0	0	0	79	100.00%	Not calculable
	Provider 2					
IPP	52	0	0	31	100.00%	1.000
DEN	33	0	0	50	100.00%	1.000
NUM	12	0	0	71	100.00%	1.000
EXC	1	0	1	81	98.80%	0.661
	Provider 3					
IPP	55	0	0	34	100.00%	1.000
DEN	27	0	0	62	100.00%	1.000
NUM	11	0	0	78	100.00%	1.000
EXC	0	0	0	89	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	EXC: Discrepancy due to timing clause in exception for nonsealable teeth. Vendor corrected. This discrepancy led to a broader discussion and review of the timing for all nonsealable reasons. Measure developer re-defined exception value sets and adjusted timing as a result.					
Report 2	**Provider 1**					
IPP	48	0	0	31	100.00%	1.000
DEN	21	0	0	58	100.00%	1.000
NUM	9	0	0	70	100.00%	1.000
EXC	0	0	0	79	100.00%	Not calculable
	Provider 2					
IPP	52	0	0	31	100.00%	1.000
DEN	33	0	0	50	100.00%	1.000
NUM	12	0	0	71	100.00%	1.000
EXC	1	0	0	82	100.00%	1.000
	Provider 3					
IPP	55	0	0	34	100.00%	1.000
DEN	27	0	0	62	100.00%	1.000
NUM	11	0	0	78	100.00%	1.000
EXC	0	0	0	89	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications.					

CARE CONTINUITY, TEST DATASETS 1&2, PRACTICE SITE TESTING

SYNTHETIC TEST DATASET TESTING: CARE CONTINUITY						
Clinical Practice Site, Test Dataset 1	Agreement between Known Values and Automated Report				Agreement	Kappa Statistic
	Y/Y	Y/N	N/Y	N/N		
Report 1	Provider 1					
IPP	14	0	0	10	100.00%	1.000
DEN	14	0	0	10	100.00%	1.000
NUM	8	0	0	16	100.00%	1.000
	Provider 2					
IPP	16	0	0	10	100.00%	1.000
DEN	16	0	0	10	100.00%	1.000
NUM	9	0	0	17	100.00%	1.000
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications.					
Clinical Practice Site, Test Dataset 2	Agreement between Known Values and Automated Report				Agreement	Kappa Statistic
	Y/Y	Y/N	N/Y	N/N		
Report 1	Provider 1					
IPP	42	0	0	32	100.00%	1.000
DEN	42	0	0	32	100.00%	1.000
NUM	21	0	0	53	100.00%	1.000
	Provider 2					
IPP	51	0	0	21	100.00%	1.000
DEN	51	0	0	21	100.00%	1.000
NUM	27	0	0	45	100.00%	1.000
	Provider 3					
IPP	46	0	0	29	100.00%	1.000
DEN	46	0	0	29	100.00%	1.000
NUM	29	0	0	46	100.00%	1.000
	Provider 4					
IPP	43	0	0	61	100.00%	1.000
DEN	43	0	0	61	100.00%	1.000
NUM	16	0	0	88	100.00%	1.000
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, including all stratifications.					

SEALANTS, TEST DATASET 2, PRACTICE SITE TESTING

SYNTHETIC TEST DATASET TESTING: SEALANTS						
Clinical Practice Site, Test Dataset 2	Agreement between Known Values and Automated Report				Agreement	Kappa Statistic
	Y/Y	Y/N	N/Y	N/N		
Report 1	Provider 1					
IPP	48	0	0	31	100.00%	1.000
DEN	21	0	0	58	100.00%	1.000
NUM	0	9	0	70	88.61%	Not calculable
EXC	0	0	9	70	88.61%	Not calculable
	Provider 2					
IPP	Report was not generated for this provider.					
DEN						
NUM						
EXC						
	Provider 3					
IPP	55	0	0	34	100.00%	1.000
DEN	27	0	0	62	100.00%	1.000
NUM	0	11	0	78	87.64%	Not calculable
EXC	0	0	11	78	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	NUM & EXC: Formatting error in report led to the values being reversed. Vendor corrected. Missing provider report: Provider report was run using incorrect provider id. Clinical site informatics specialist corrected.					
Report 2	Provider 1					
IPP	48	0	0	31	100.00%	1.000
DEN	21	0	0	58	100.00%	1.000
NUM	9	0	0	70	100.00%	1.000
EXC	0	0	0	79	100.00%	Not calculable
	Provider 2					
IPP	52	0	0	31	100.00%	1.000
DEN	33	0	0	50	100.00%	1.000
NUM	12	0	0	71	100.00%	1.000
EXC	1	0	0	82	100.00%	1.000
	Provider 3					
IPP	55	0	0	34	100.00%	1.000
DEN	27	0	0	62	100.00%	1.000
NUM	11	0	0	78	100.00%	1.000
EXC	0	0	0	89	100.00%	Not calculable
Sources of Discrepancies & Corresponding Resolution	Complete agreement reached, on measure score components, including all stratifications. Calculated measure score for provider wh had an exception was incorrect due to bracket being placed incorrectly in calculation. Vendor corrected.					