

Using a stakeholder-engaged approach to develop and validate electronic clinical quality measures

RECEIVED 18 May 2016
REVISED 28 July 2016
ACCEPTED 16 August 2016



Jill Boylston Herndon,¹ Krishna Aravamudhan,² Ronald L Stephenson,³ Ryan Brandon,⁴ Jesley Ruff,⁵ Frank Catalanotto,⁶ and Huong Le⁷

ABSTRACT

Objective To describe the stakeholder-engaged processes used to develop, specify, and validate 2 oral health care electronic clinical quality measures.

Materials and Methods A broad range of stakeholders were engaged from conception through testing to develop measures and test feasibility, reliability, and validity following National Quality Forum guidance. We assessed data element feasibility through semistructured interviews with key stakeholders using a National Quality Forum–recommended scorecard. We created test datasets of synthetic patients to test measure implementation feasibility and reliability within and across electronic health record (EHR) systems. We validated implementation with automated reporting of EHR clinical data against manual record reviews, using the kappa statistic.

Results A stakeholder workgroup was formed and guided all development and testing processes. All critical data elements passed feasibility testing. Four test datasets, representing 577 synthetic patients, were developed and implemented within EHR vendors' software, demonstrating measure implementation feasibility. Measure reliability and validity were established through implementation at clinical practice sites, with kappa statistic values in the “almost perfect” agreement range of 0.80–0.99 for all but 1 measure component, which demonstrated “substantial” agreement. The 2 validated measures were published in the United States Health Information Knowledgebase.

Conclusion The stakeholder-engaged processes used in this study facilitated a successful measure development and testing cycle. Engaging stakeholders early and throughout development and testing promotes early identification of and attention to potential threats to feasibility, reliability, and validity, thereby averting significant resource investments that are unlikely to be fruitful.

Keywords: health care quality indicators, quality measure development, quality measure testing, electronic health records, meaningful use

BACKGROUND AND SIGNIFICANCE

Reporting clinical quality measures (CQMs) using electronic health record (EHR) data is part of national efforts designed to improve patient care and population health.¹ The Medicare and Medicaid EHR Incentive Programs have spurred widespread adoption of EHRs, with the goal of using EHR technology to positively impact patient care (meaningful use).^{2,3} Although the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) will eventually result in evolution of these incentive programs, it retains a focus on quality measurement and continues to require meaningful use of certified EHR technology.^{4,5}

Quality measures historically have been calculated using administrative claims data or data from manual reviews of patient records.^{6,7} Administrative data offer high feasibility at relatively low cost, particularly for reporting at aggregated levels (eg, health plan, program, or population), but they lack detailed clinical information.^{7,8} Detailed clinical information, which is important for measuring patient outcomes, can be captured through resource-intensive manual record review. Patient data from interoperable EHRs offer the potential to obtain the joint benefits of automated reporting and detailed clinical data.^{6,7} Thus, electronic clinical quality measures (eCQMs) have been a focus of initiatives to use EHR technology to promote health care quality improvement. Amster et al.⁹ commented that “[a]ccessing the rich clinical data in EHRs requires specifying quality measures so that EHRs implemented in a wide array of settings and across multiple provider groups

can be used to completely and accurately extract required data and accurately compute performance results.” Kern et al.⁷ have recommended validation against manual chart reviews and testing measures using simulated patient records with known values as strategies for ensuring the accuracy of eCQMs. As these authors suggest, the processes used to develop, specify, and validate eCQMs require careful attention. However, there is little information in the published literature about these processes. This article specifically focuses on measure development and validation processes to inform future eCQM development rather than concentrating on measure-specific validation results.

The purpose of this study is to describe the stakeholder-engaged processes used to develop, specify, and validate 2 pediatric oral health care eCQMs. There is a significant need for these measures. Only 2 of 64 eCQMs in the 2014 edition of the EHR Incentive Programs related to oral health. As a result, eligible oral health care providers have had to implement processes to monitor and collect medical-related data to demonstrate meaningful use. As of February 2016, more than 23 000 dentists were registered for the EHR Incentive Programs.¹⁰ Although the Health Resources and Services Administration (HRSA) Health Center Program requires HRSA-funded health centers to report quality-of-care measures,¹¹ there were no oral health care measures, even though centers must provide preventive dental services on site or through referral.¹² Consequently, the Dental Quality Alliance, which is a multistakeholder group formed in 2010 at the request of the Centers for Medicare and Medicaid Services to advance oral health care

Correspondence to Jill Boylston Herndon, PhD, Key Analytics and Consulting, 4111 NW 16th Blvd #357724, Gainesville, FL 32635, USA; jill.herndon@keyanalyticsconsulting.com; Tel: 352-240-3736.

©The Author 2016. Published by Oxford University Press on behalf of the American Medical Informatics Association. All rights reserved.

For Permissions, please email: journals.permissions@oup.com For numbered affiliations see end of article.

performance measurement,¹³ undertook development of oral health care eQMs.

Following the recommendation by the Collaborative for Performance Measure Integration with EHR Systems that measure developers “collaborate with EHR vendors to field test measures for feasibility of implementation,”⁹ EHR vendors, along with other stakeholders, were engaged early and throughout an iterative measure development and testing process to ensure feasibility and validity. Another novel aspect of our approach was to create and evaluate the usefulness of simulated patient data for testing feasibility and reliability prior to validation with clinical data. We also identified special considerations in eQm development for dentistry that may have implications for other health care domains and specialties, such as pharmacy and behavioral health, that have different delivery and information systems compared with traditional medical care systems. These processes led to the successful validation of 2 eQMs, Oral Health Care Continuity for Children 2–20 Years (Care Continuity) and Oral Health Sealants for Children 6–9 Years (Sealants), which are published in the United States Health Information Knowledgebase.¹⁴ HRSA included the Sealants eQm as its first oral health care quality measure for reporting in 2016.¹¹

These evidence-based, process-of-care measures focus on dental caries, which is the most common chronic disease among children in the United States, with significant effects on oral and overall health.^{15,16} Evidence-based guidelines recommend clinical oral evaluations with a regular recall schedule that is tailored to individual needs based on current oral health status and disease risk assessment (eg, caries risk), with the recommended recall frequency ranging from 3 to 12 months for individuals younger than 18 years of age.¹⁷ Clinical oral evaluations play an essential role in identification, prevention, and treatment of caries. Because disease identification, risk assessment, prevention regimen development, and treatment planning are ongoing processes, evaluating continuity of care over time is an important quality metric. National data indicate that only 48% of children in the United States received any dental services during 2013¹⁸. Evidence-based guidelines also recommend that sealants be placed on the pits and fissures of teeth for children at risk for dental caries, with stronger evidence of

effectiveness for permanent molars than for primary molars.¹⁹ Prior research found variations in program-level performance on sealant quality measures.²⁰

MATERIALS AND METHODS

Building off its prior development of oral health care measures calculated using administrative data,^{13,20} the Dental Quality Alliance (DQA) conducted initial feasibility assessments and identified Care Continuity and Sealants to develop as eQMs (Figure 1). After developing draft measure specifications, the DQA used a competitive request-for-proposal process to select a research team to conduct testing (Figure 2). Two dental EHR vendors participated as project partners: Exan Group (Exan) and American Dental Partners (ADP). Exan’s axiUm EHR and ADP’s Improvis EHR are certified by an Office of the National Coordinator for Health Information Technology (ONC) Authorized Certification Body. The clinical practice sites included ADP practice affiliates and University of Florida College of Dentistry dental clinics and College of Medicine pediatric clinics. The pediatric clinics were included to validate Care Continuity, because oral health screenings may be done by medical providers. Testing occurred from October 2013 to September 2014 and followed National Quality Forum (NQF) guidance for measure feasibility, reliability, and validity.²¹ This study was approved as expedited by the appropriate institutional review boards, with a full waiver of informed consent. Figure 3 summarizes the measure development and testing processes.

Formation and engagement of stakeholder workgroup

The DQA eQm Workgroup included clinicians and representatives from medical and dental EHR systems, federal agencies, NQF, community health centers, and dental plans (Table 1). The workgroup conducted initial feasibility assessment and measure development and subsequently oversaw measure testing. Stakeholder feedback also was solicited through surveys with clinicians and EHR vendors. An interim report that included the measure specifications, testing methodology, and results to date was presented to the workgroup and the full DQA membership, disseminated electronically to a broad range of stakeholders, and posted online for a 1-month public comment period.

Figure 1: DQA oral health eQMs.

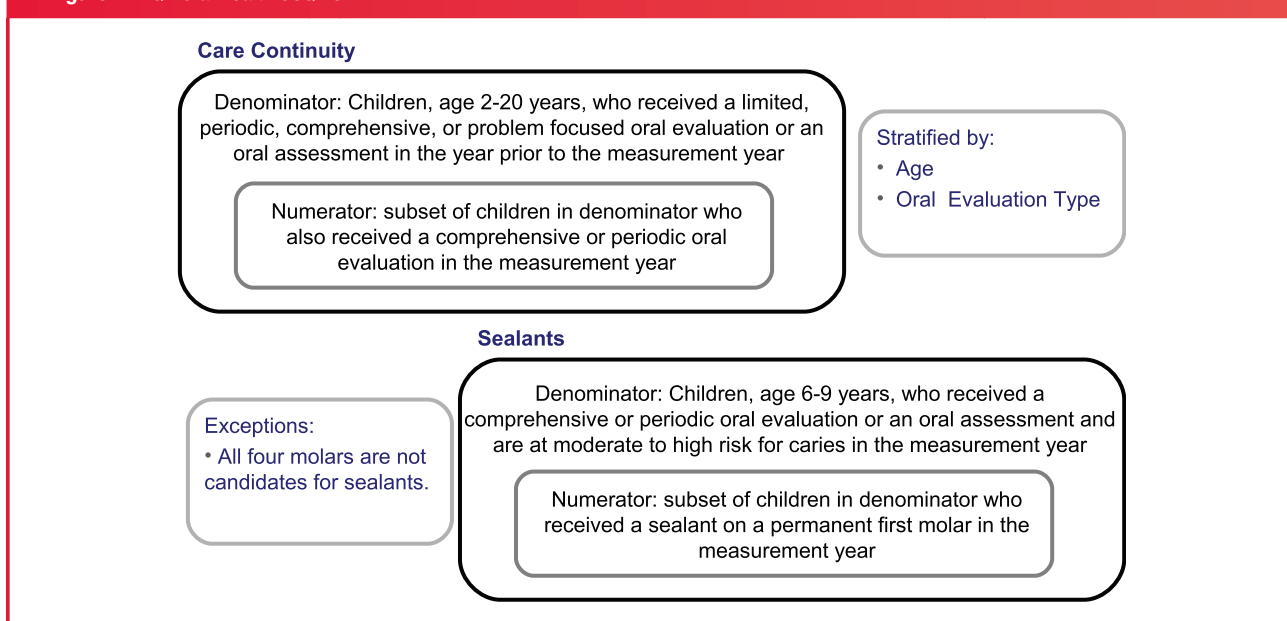


Figure 2: Project team.

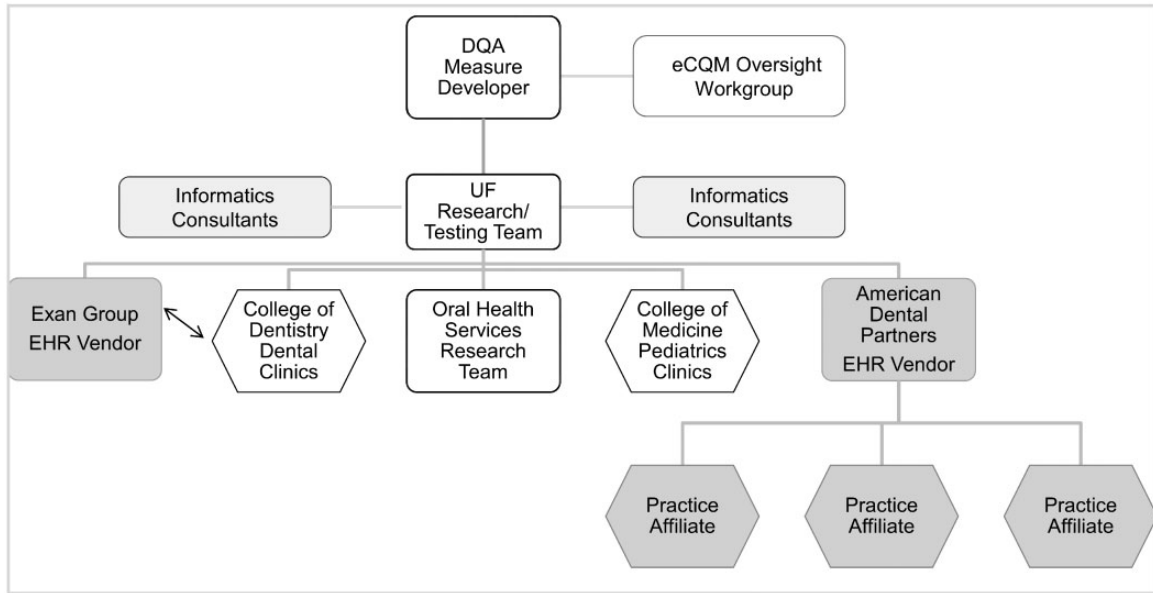


Figure 3: Measure development and testing processes.

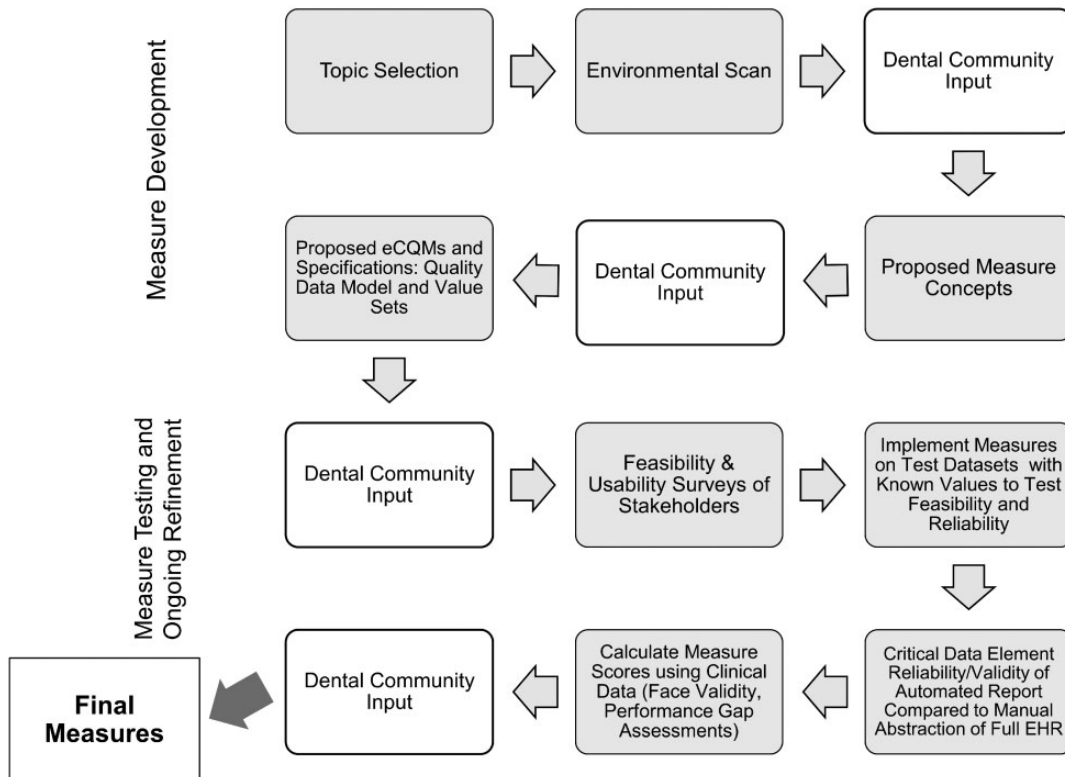


Table 1: Dental Quality Alliance eCQM Workgroup: stakeholder agencies and organizations

Academy of General Dentistry	Health Resources and Services Administration
American Academy of Pediatrics	Henry Schein Practice Solutions (Dentrix)
American Academy of Pediatric Dentistry	Medicaid-CHIP State Dental Association
American Dental Association, Dental Informatics	National Network for Oral Health Access
American Dental Partners	National Quality Forum
Centers for Medicare and Medicaid Services	Office of the National Coordinator for HIT
Dental Quality Alliance	Open Dental
Exan Group	OptumInsight
Epic	QSIDental
HealthPartners Medical and Dental Group	Willamette Dental Group

Assessment of data element and measure specification feasibility

In 2012, the workgroup evaluated electronic data element feasibility using an NQF-recommended scorecard with a 3-point scale (3 = highest rating) based on whether the data element was: (1) available in a structured format, (2) derived from an authoritative source and likely to be accurate, (3) coded using a nationally accepted terminology standard, and (4) captured during typical workflow without additional data entry or required EHR interface changes.²² The DQA conducted semistructured surveys and phone interviews with EHR vendors, practice site information systems specialists, and clinicians. Two categories were assessed: critical data elements needed to calculate the measure score (eg, date of birth, date of service, and procedure codes), and supplemental data elements (sex, race, ethnicity, and payer type) that are used to stratify the score by patient subgroups. The goal was to identify measures that could be reliably calculated using data already collected within electronic records. The DQA also solicited clinician feedback regarding the clarity, face validity, and usability of the proposed measures from the memberships of the Academy of General Dentistry, the American Academy of Pediatric Dentistry, the American Academy of Pediatrics, and the National Network for Oral Health Access, organizations representing end users. In 2013, the research team conducted additional in-depth interviews with dental and medical EHR vendors to obtain a deeper understanding of the data elements, value sets, calculation logic, and issues related to differences in medical and dental clinical workflows and EHR systems.

Creation of synthetic patient test datasets

Synthetic test datasets were created for each measure with known values of the data elements to test implementation feasibility and whether the EHR software's translation of the measure logic could reliably calculate the individual measure components and overall score. A test dataset allows EHR vendors to evaluate the measure programming logic and configuration within their test environments prior to practice site implementation. Practice sites can use test datasets to validate their site-specific implementation, 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.

The measure components include the initial patient population (IPP), denominator (DEN), numerator (NUM), and exceptions (EXC). The measure score is equal to the NUM divided by (DEN – EXC). The datasets were designed to include patients who variously did or did not qualify for each measure component. We tested age inclusion/exclusion for IPP and DEN; procedure inclusion/exclusion for IPP, DEN, NUM, EXC; provider attribution for IPP, DEN, NUM, EXC; tooth number for sealant placement (Sealants); identification of elevated caries risk (Sealants); identification of patients qualifying for exceptions (Sealants); stratification by age and oral evaluation type (Care Continuity); demographic stratification (gender, race, ethnicity, payer type); and implementation across a range and mix of service use patterns. We also tested different patient-provider-procedure patterns to ensure correct provider attribution.

We developed a data schema that was pilot tested and refined based on vendor feedback. We randomly generated patients and defined data fields using Red Gate Software's SQL Data Generator. Percentages were defined for such data elements as patient demographic characteristics and procedure codes. We used Microsoft SQL Server for data management and Microsoft Access as the interface, with additional programming logic to ensure representative data that could be imported into the vendors' systems. Each test dataset was verified with logic tests by 2 individuals, first in SQL and then in Stata, Release 13. The known values for each measure component, overall measure scores, and measure score stratifications were verified before providing the datasets to the vendors. For each measure, an initial test dataset was created, followed by a more complex dataset (greater variation of patient-provider-procedure combinations and visit patterns) to confirm correct measure implementation.

eCQM implementation with test datasets

Exan and ADP imported the synthetic datasets into their test environments, implemented the measure logic, and produced provider-level reports on the number of patients meeting the IPP, DEN, NUM, and EXC criteria and the overall measure scores with stratifications. Concordance analyses were used to compare the results with the known values. Patient-level reports were used to identify test patients who were misclassified, and feedback was provided to the vendors regarding identified discrepancies. The vendors revised their programming and configuration and resubmitted their reports. This process was iterated until 100% agreement was achieved. Practice site testing followed the same iterative process.

eCQM implementation with clinical data

After successful implementation with the test datasets, the measures were implemented with practice site clinical electronic record data for calendar years 2012 and 2013 after verifying implementation feasibility, which included confirming the presence and completeness of the critical data elements. Provider-level measure reports were generated that included the individual components and the overall scores with required stratifications.

Validation of automated reports against manual record reviews

We validated each critical data element and measure component generated by the automated reports against manual reviews of a random sample of 60–65 patient records per measure per site for patients who met the measure's age eligibility criteria, using the kappa statistic to evaluate the strength of agreement (Stata, Release 13). All sites used the same detailed protocol and abstraction forms, which were developed in consultation with the sites and pilot-tested. Discrepancies were analyzed jointly by the site's information systems specialist, who ran the automated reports, and the record reviewer.

We validated measure components at the patient-provider level to verify not only the validity of the measure score, but also the accuracy of provider attribution for each measure component.

RESULTS

Initial feasibility assessment

Nine participants, representing dental EHR vendors, IT programmers in clinical settings, and practitioners, completed the semistructured interviews. Three critical data elements met all 4 NQF criteria: date of birth, date of service, and identification of specific procedures. Tooth numbering also met all criteria; however, more than 1 standard terminology was used and no dental systems used the Systematized Nomenclature of Medicine (SNOMED) to record tooth numbers, because it was not a standard used by the profession. EHR vendors' mappings of their coding systems to SNOMED codes were verified during testing. During testing, we also confirmed that all but 2 critical data elements (caries risk assessment and diagnosis) were routinely captured as part of normal clinical workflow as structured data elements, with data completeness ranging from 98% to 100%. Caries risk assessment findings were not captured as structured data, but new dental procedure codes were introduced in 2014 enabling structured data capture. Dentistry in general has not routinely captured diagnosis codes (which were needed to identify exceptions for the Sealant measure) as standardized structured data in EHRs or billing systems.^{23,24} However, dental EHR systems capture diagnoses within problem and condition lists and map those to standardized codes, which we validated through manual record review during testing. Identifying measures that could be feasibly and reliably calculated using data elements already collected within EHRs facilitated implementation, since the data collection processes did not require modification. All 4 supplemental data elements were captured within EHR systems, but race, ethnicity, and payer type often were not captured using national standard terminologies in noncertified products.

Alignment between Quality Data Model and measure specifications

Initial feasibility assessments identified critical data elements that could not be specified using the Quality Data Model (QDM), which represents clinical concepts in a standardized format.²⁵ Many dental procedures

are tooth-specific. Although the QDM includes both anatomic location and procedure code data elements, they could not be directly associated. This was problematic because we could not associate sealant placement with permanent molars specifically (for which there is the strongest evidence of effectiveness in caries prevention¹⁹). Similarly, diagnoses required to identify exceptions also are tooth-specific. The DQA worked with the ONC to modify the QDM to incorporate anatomic location (tooth number) as an attribute to the "procedure" and "diagnosis" data elements and thereby better align with how clinical data are represented in dental EHR systems (Figure 4).

Measure specifications and logic

We specified the measures using the Measure Authoring Tool (MAT), a web-based application to create eQMs that exports machine-readable and human-readable specifications. During feasibility assessments, EHR vendors indicated that they had not been able to use the MAT-generated Health Quality Measures Format Extensible Markup Language (HQMFL XML) for eQMs and instead programmed measures using human-readable specifications. Consequently, we incorporated additional plain-language guidance in the specifications' metadata to promote uniform implementation. We also used the workgroup's conference calls, which included EHR vendors, to verify the clarity of the specifications.

Refinement of measure specifications from stakeholder engagement

Feedback from key stakeholders throughout measure development and testing was essential for refining the specifications and addressing potential implementation barriers. Examples below illustrate how stakeholder engagement revealed assumptions and practices that would have adversely affected feasibility and reliability had they not been identified and collaboratively addressed.

Identifying unique "encounters" or "visits". Issue: Medical EHR systems embed procedures within "encounters," and eQDM measure logic historically conditioned the IPP and DEN criteria on whether certain types of encounters were present. During feasibility assessments, the DQA identified the encounter framework as an implementation

Figure 4: Measure specification excerpt illustrating modifications to quality data model and measure logic to accommodate dental clinical concepts and their representation in dental electronic records

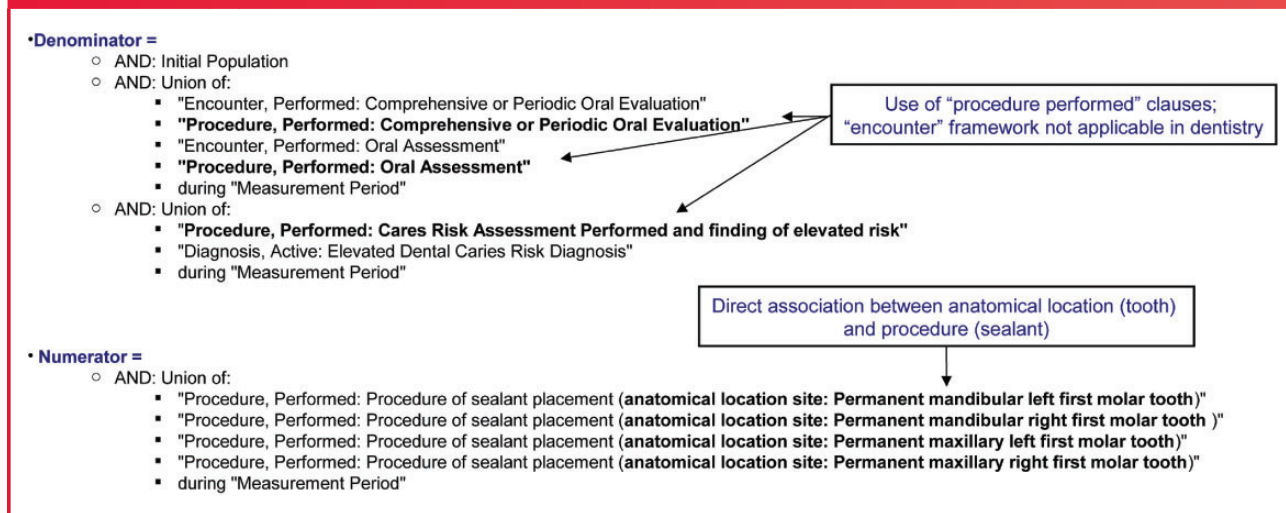
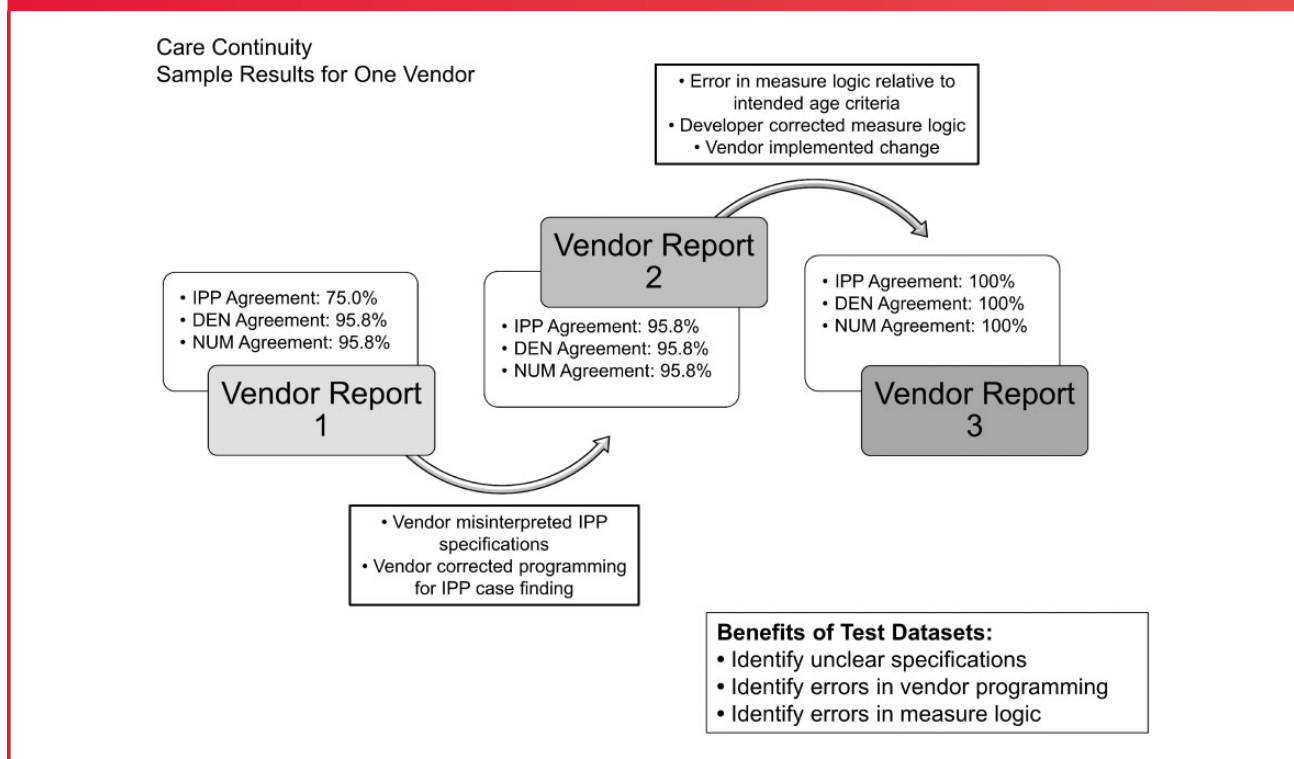


Figure 5: Sample synthetic dataset testing process.

challenge within dental EHR systems. Dental EHR vendors indicated that they typically identify unique “visits” by a posted procedure for a particular date of service or by completed appointments with a unique combination of date of service, patient, and provider.

Response: In addition to standard encounter clauses that are used for medical systems, the measure logic also includes “procedure performed” clauses in the IPP and DEN criteria (Figure 4).

Measured entity and provider attribution. Issue: In the Medicaid EHR Incentive Program, the measured entity was the individual clinician.²⁶ In our review of the measure logic for existing eCQMs, there typically was no specification or guidance regarding provider attribution. Feedback from ONC and medical EHR vendors indicated that, commonly, the denominator is provider-specific and the numerator counts qualifying services rendered by any provider. However, we found that dental vendors did not uniformly adopt this approach in the absence of clear specifications; for example, some made both the numerator and the denominator provider-specific.

Response: To avoid differential interpretation and promote measurement reliability, the measure guidance includes an explanation of the intended provider attribution.

Feasibility and reliability of measure implementation: test datasets

Two datasets of synthetic test patients for each measure were developed, totaling 4 datasets and 577 test patients (50 and 240 patients for Care Continuity; 98 and 189 patients for Sealants). The number of test patients was influenced by the number of aspects tested, sufficient variation in the randomized patient generation process, and sufficient patient-provider-procedure combinations to test provider attribution. None of the test datasets successfully passed testing on the first attempt by either vendor. Figure 5 illustrates the iterative

process from testing the first Care Continuity synthetic dataset with 1 vendor. Table 2 summarizes the findings of multiple iterations of the first Sealant test dataset with 1 vendor. Through this iterative process, we identified misinterpretations of the measure logic, vendor programming errors, and developer measure specification errors. Specific examples include: incorrect implementation by the vendor of the provider attribution for the IPP, DEN, and NUM; incorrect specification by the developer of the age eligibility measure logic (compared with what was intended); and refinement of the exception criteria by the developer based on vendor feedback. We received positive feedback from the EHR vendors about the utility of these test datasets to confirm appropriate measure logic implementation prior to implementation with clinical data.

Validation against manual record reviews

For both measures, except EXC for Sealants, validity across practice sites was high, with >96% overall agreement for all critical data elements and kappa statistic values for measure components in the “almost perfect” agreement range of 0.80–0.99.²⁷ The Sealant EXC component had kappa statistic values of 0.74 or higher, signifying “substantial” agreement. Reasons for discrepancies between manual and automated identification of exceptions included manual recording errors (corrected upon review), incomplete capture of “findings” codes used in local charting practices (corrected in programming logic and configuration), and procedures previously performed outside the system that were reflected in manual records but not automated reporting. After the manual recording errors and programming logic were corrected, the kappa statistic values increased to “almost perfect” agreement. Because measure component validation was conducted at the patient-provider level, it also demonstrated the reliability of the provider attribution.

Table 2: Synthetic test dataset testing: Sealants

Vendor 2, Test Dataset 1	Agreement between Known Values and Automated Report				Agreement (%)
	Y/Y	Y/N	N/Y	N/N	
REPORT 1	Provider 1				
IPP	31	0	0	29	100.00
DEN	8	8	0	44	86.67
NUM	3	0	2	55	96.67
EXC	0	1	0	59	98.33
	Provider 2				
IPP	35	0	1	24	98.33
DEN	11	8	1	40	85.00
NUM	4	0	2	54	96.67
EXC	0	0	0	60	100.00
Sources of Discrepancies and Corresponding Resolution	IPP: Age discrepancy. Review of discrepancy identified (1) specification in measure logic did not capture intent and (2) clarification was needed about age as continuous vs step function. Developer corrected measure logic and provided age calculation guidance in human-readable metadata. Vendor implemented changes. DEN: Identification of elevated risk through SNOMED codes (vs Current Dental Terminology 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 the system. NUM: Sealant placement was not restricted to permanent first molar. Vendor corrected programming logic.				
REPORT 2	Provider 1				
IPP	31	0	0	29	100.00
DEN	16	0	0	44	100.00
NUM	3	0	0	57	100.00
EXC	0	1	0	59	98.33
	Provider 2				
IPP	35	0	0	25	100.00
DEN	19	0	0	41	100.00
NUM	3	1	0	56	98.33
EXC	0	0	0	60	100.00
Sources of Discrepancies and 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 captured 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
DEN	16	0	0	44	100.00
NUM	3	0	0	57	100.00
EXC	1	0	0	59	100.00
	Provider 2				
IPP	35	0	0	25	100.00
DEN	19	0	0	41	100.00
NUM	4	0	0	56	100.00
EXC	0	0	0	60	100.00
Sources of Discrepancies and Resolution	Complete agreement reached, including all stratifications for each measure component. Measure scores for provider with exception did not take exception into account. Vendor successfully corrected this with submission of final report (Report 4).				

DISCUSSION

This study developed and validated 2 pediatric oral health care eQMs that are now included in the United States Health Information Knowledgebase.¹⁴ A novel aspect of this project was broad stakeholder engagement early and throughout the development and testing process, with ongoing assessments of both measure concept feasibility and measure implementation feasibility. Another novel aspect was the use of comprehensive synthetic patient test datasets, which established measure implementation feasibility and promoted implementation reliability. We share several lessons learned.

Importance of stakeholder engagement throughout development and testing

The collaborative, transparent, stakeholder-engaged approach provided numerous insights for the specific measures and the overall development and testing processes. Instead of using a traditional technical expert panel at the initial measure conceptualization and specification stages, we used the same group of experts to oversee the entire process from conception through testing. The wide range of subject matter experts fostered a robust interchange, which allowed us to identify and address potential threats to feasibility, reliability, and validity early in the process. Collaboration with ONC was essential for aligning dental clinical data systems with the QDM and MAT measure specifications. Collaboration with EHR vendors that had prior experience with meaningful use eQM implementation was integral to ensuring that the measures were accurately specified and could be reliably implemented. Feedback solicited from members of national organizations representing end users provided input from the broader community of implementers who typically are not directly engaged with measure development. Along with the public release and dissemination of an interim report to solicit feedback from all stakeholders, these activities established ownership and buy-in among the broader stakeholder community.

Benefits and limitations of test datasets

The reliability and validity of eQMs 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. The test datasets, with prospectively known values, allowed vendors to check the accuracy of their measure logic implementation prior to implementation with clinical data. The iterations of testing that were needed to achieve 100% agreement demonstrated that, without such testing, there is a high risk that the measure logic will not be implemented as intended or it will be implemented inconsistently across EHR systems and practice sites. Of equal importance, with synthetic data before finalizing the specifications brought to light errors in the measure specifications and logic, identified areas where guidance was needed, and reopened dialogue around critical aspects. Even though the focus was on feasibility and reliability, some issues addressed face validity. Thus, the synthetic test datasets were an integral part of the collaborative process to review and refine the measures. The resulting specifications were clearer, more precisely defined, and more reflective of the measure intent.

A potential limitation is that test patients require careful design and may need to be customized for individual measures to ensure robust testing. In addition, test datasets may not simulate all measure aspects if local customization is required or critical data are in unstructured fields. Although we hoped to create a parameterized tool that would allow us to recreate synthetic datasets on demand using completely automated processes, the customization required to incorporate all the desired scenarios was more than could be accomplished

through automated processes within the resource constraints of the project; therefore, some manual edits were required to produce the final datasets. Broader initiatives to automatically generate test patients are under way and should facilitate their use.²⁸

Need for local testing

Despite the use of comprehensive test datasets, we found it important to do local validation by conducting face validity assessments of the measure scores and comparing automated reporting to manual record review. Local validation also was important to ensure appropriate exclusion/inclusion of structured data elements that had local modification within the EHR and correct capture of clinical data contained in unstructured data fields.

Supplemental data elements with non-mutually exclusive categories

For race and payer type, practice sites required guidance on how to address cases where patients could be classified into more than 1 category. Practice sites may record more than 1 applicable race, but the implementation of race did not allow for listing multiple categories, nor was there a multiracial category within the value set. Sites were instructed to use the “primary” race if such a field existed or the first-listed category. Payer type can vary over time as well as by procedure type; therefore, patients may have multiple payer types within a measurement period. If reporting entities are asked to stratify measure scores by payer type, they will require guidance on how to handle patients with multiple payers.

Advancing quality measurement

EHR data offer the potential to realize the joint benefits of automated reporting (reducing the resource burden of measurement) and clinical detail (promoting provider-level, outcomes-focused measurement). However, EHR-based quality measurement is not a panacea. eQMs encounter some of the same challenges with provider-level measurement as measures using other data sources, such as identifying the minimum sample size for reliable measurement and developing valid methods of attributing patients to clinicians.²⁹ As noted here and elsewhere, eQMs have greater reliability and validity when they rely on structured data elements recorded using standard national terminologies.³⁰ Currently, much clinical detail is captured as unstructured data or is inconsistently captured as structured data.³¹ Within dentistry, there has been a significant increase in the amount of clinical information collected within electronic data systems, enhancing the potential for use of electronic dental records in quality improvement and research applications; however, similar to medicine, challenges persist regarding variability in data capture, data accuracy across systems, and clinical data captured in unstructured formats.^{32,33} The benefits of automated reporting are limited by the lack of comprehensive interoperability within and between systems of care.^{30,34,35} Interoperability is needed to track patients through health care systems and obtain a more complete record of their care.^{31,35} Consequently, lack of interoperability may adversely affect performance measurement when services are provided in other settings and are not captured in the electronic record from which the measure is calculated. Interoperability also is necessary for aggregating data across populations. Despite increased adoption of EHRs, EHR-based measurement remains a resource issue, particularly for solo and small practices that continue to face barriers to adoption.^{34,35} Implementation of the Medicare Access and CHIP Reauthorization Act recognizes these challenges and prioritizes interoperability while also allowing more flexibility in health IT implementation to address practice needs.^{4,5} Despite

current limitations, eQMs offer significant potential for engaging in more meaningful, outcomes-based measurement at greater efficiency and scale than was possible in the past.

CONCLUSIONS

Multistakeholder engagement is essential to a successful measure development and testing cycle. A diverse stakeholder team, along with the use of test datasets, promotes early identification of and attention to potential threats to feasibility, reliability, and validity and thereby averts resource investments that are unlikely to be fruitful. Ultimately, the coordinated input of a wide range of stakeholders, resulting well-tested measures, and detailed implementation guidance are expected to reduce the ambiguity and costs of measure implementation.

CONTRIBUTORS

JBH led the research team that did the validation testing; contributed to study design/methodology, data acquisition, and data analysis and interpretation; led the initial draft of the manuscript; and contributed to critical revisions. KA led the overall project and contributed to measure conception and development, testing study design, results interpretation, manuscript drafting, and critical review/revision of the manuscript for important intellectual content. RLS developed the synthetic patient data and contributed to study design, results interpretation, manuscript drafting, and critical review/revision of the manuscript for important intellectual content. RB led the Exan team that implemented the measures using synthetic and clinical patient data. RB contributed to study design, results interpretation, and critical review/revision of the manuscript for important intellectual content. JR led the ADP team that implemented the measures using synthetic and clinical patient data and that conducted ADP practice site measure validation. JR contributed to data acquisition, study design, results interpretation, and critical review/revision of the manuscript for important intellectual content. FAC contributed to data acquisition, study design, results interpretation, and critical review/revision of the manuscript for important intellectual content. HL led the stakeholder workgroup that guided and oversaw measure development and testing processes. HL contributed to study conception and critical review/revision of the manuscript for important intellectual content.

FUNDING

This project was supported by the Office of the National Coordinator for Health Information Technology (#HHS233201300039C).

COMPETING INTERESTS

The authors have no competing interests to declare.

ACKNOWLEDGEMENTS

We would like to express our deepest gratitude to the following individuals within our organizations for their substantial contributions to this project. The American Dental Partners team included Mike Hoyt, information technology team leader; Roger Horton, who led the programming for automated reporting and provided extensive feedback on measure specifications; Victoria Post, who contributed to development of the record review abstraction process and conducted record reviews; and Tim Freeman and Jill Jansky, who provided feedback on the test dataset schema and led programming related to critical data element evaluation and patient characteristic reporting. The Exan Group team included Paul Delicana and Joraymond Espiritu, who provided feedback on the test dataset data schema, led programming for the automated reporting, provided extensive feedback on measure specifications, and facilitated practice site implementation of the measures. The University of Florida team included

Richelle Janiec and Lindsay Thompson, who contributed to development of the record review abstraction process and conducted pediatric dental record reviews and pediatric medical record reviews, respectively; Stephen Kostewicz and Gloria Pflugfelder Lipori, who led programming related to critical data element evaluation, patient characteristic reporting, and automated reporting in pediatric dental clinics and pediatric medical clinics, respectively; Anzeela Schentrup, who contributed to data reports for the pediatric medical clinics; and Scott Tomar, who assisted with interpretation of data reports for the pediatric dental clinics. Finally, we appreciate the support and guidance of Kevin Larsen, medical director of Meaningful Use, and LaVerne Perlie, senior nurse consultant, with the Office of the National Coordinator for Information Technology when these measures were developed.

REFERENCES

1. U.S. Department of Health and Human Services. *2015 Annual Progress Report to Congress: National Strategy for Quality Improvement in Health Care*. 2015. <http://www.ahrq.gov/workingforquality/reports.htm>. Accessed March 23, 2016.
2. Centers for Disease Control and Prevention. *Meaningful Use*. 2012. <http://www.cdc.gov/ehrmmeaningfuluse/introduction.html>. Accessed May 9, 2015.
3. Centers for Medicare and Medicaid Services. *An introduction to Medicaid EHR Incentive Program for Eligible Professionals*. 2014. http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/EHR_Medicaid_BegGuide_Stage1.pdf Accessed May 9, 2014.
4. Slavitt A, DeSalvo K. EHR incentive programs: Where we go next. *The CMS Blog*. Centers for Medicare and Medicaid Services. January 19, 2016. <https://blog.cms.gov/2016/01/19/ehr-incentive-programs-where-we-go-next/>. Accessed March 23, 2016.
5. Miliard M. Meaningful use will still be part of MIPS reimbursement, CMS official says. *Healthcare IT News*. March 2, 2016. <http://www.healthcareitnews.com/news/meaningful-use-will-still-be-part-mips-reimbursement-cms-official-says>. Accessed March 23, 2016.
6. Garrido T, Kumar S, Lekas J, et al. E-measures: Insight into the challenges and opportunities of automating publicly reported quality measures. *JAMIA*. 2014;21:181–84.
7. Kern LM, Malhotra S, Barron Y, et al. Accuracy of electronically reported “meaningful use” clinical quality measures: a cross-sectional study. *Ann Intern Med*. 2013;158:77–83.
8. Harris SB, Glazier RH, Tompkins JW, et al. Investigating concordance in diabetes diagnosis between primary care charts (electronic medical records) and health administrative data: a retrospective cohort study. *BMC Health Serv Res*. 2010;10:347.
9. Amster A, Jentzsch J, Pasupuleti H, et al. Completeness, accuracy, and computability of National Quality Forum–specified e-measures. *JAMIA*. 2015;22:409–16.
10. Centers for Medicare and Medicaid Services. EHR incentive program: Active registrations, February 2016. 2016. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/February2016_SummaryReport.pdf. Accessed March 23, 2016.
11. Health Resources and Services Administration. *Health center program: Uniform Data System resources*. 2016. <http://bphc.hrsa.gov/datareporting/reporting/index.html>. Accessed February 25, 2016.
12. Jones E, Shi L, Hayashi AS, et al. Access to oral health care: the role of federally qualified health centers in addressing disparities and expanding access. *Am J Public Health*. 2013;103:488–93.
13. Herndon JB, Crall JJ, Aravamudhan K, et al. Developing and testing pediatric oral healthcare quality measures. *J Public Health Dent*. 2015;75:191–201.
14. Agency for Healthcare Research and Quality. *United States Health Information Knowledgebase: Draft Measures*. 2014. <http://ushik.org/QualityMeasuresListing?draft=true&system=dqcm&sortField=570&sortDirection=ascending&enableAsynchronousLoading=true>. Accessed May 12, 2015.
15. Centers for Disease Control and Prevention. *Hygiene-related Diseases: Dental Caries*. 2014. http://www.cdc.gov/healthywater/hygiene/disease/dental_caries.html. Accessed March 23, 2015.

16. Tinanoff N, Reisine S. Update on early childhood caries since the surgeon general's report. *Acad Pediatr*. 2009;9:396–403.
17. National Institute for Health and Care Excellence. Clinical guidelines. *CG19: Dental Recall – Recall Interval Between Routine Dental Examinations*. 2004. <https://www.nice.org.uk/Guidance/CG19>. Accessed July 23, 2016.
18. Nasseh K, Vujicic M. Dental care utilization rate continues to increase among children, holds steady among working-age adults and the elderly. *Health Policy Resources Center Research Brief*. American Dental Association; 2015. Available at: http://www.ada.org/~media/ADA/Science%20and%20Research/HPI/Files/HPIBrief_1015_1.pdf?la=en. Accessed September 10, 2016.
19. Beauchamp J, Caulfield 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;139:257–68.
20. Herndon JB, Tomar SL, Catalanotto FA, et al. Measuring quality of dental care: caries prevention services for children. *J Am Dent Assoc*. 2015;146:581–91.
21. National Quality Forum. *Measure Evaluation Criteria and Guidance*. 2013. <http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=73365>. Accessed September 23, 2014.
22. National Quality Forum. *Report from the National Quality Forum: Emeasure Feasibility Assessment*. April 2013. http://www.qualityforum.org/Publications/2013/04/eMeasure_Feasibility_Assessment.aspx. Accessed April 1, 2014.
23. Bader JD. Challenges in quality assessment of dental care. *J Am Dent Assoc*. 2009;140:1456–64.
24. Kalenderian E, Ramoni RL, White JM, et al. The development of a dental diagnostic terminology. *J Dent Educ*. 2011;75:68–76.
25. Centers for Medicare and Medicaid Services Office of the National Coordinator for Health Information Technology. *Quality Data Model*. 2015. <http://www.healthit.gov/ecqi-resource-center/qdm/index.html>. Accessed May 12, 2015.
26. Centers for Medicare and Medicaid Services. *An introduction to the Medicaid EHR Incentive Program for Eligible Professionals*. 2012. http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Getting_Started.html. Accessed April 1, 2014.
27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
28. Centers for Medicare and Medicaid Services Office of the National Coordinator for Health Information Technology. Bonnie tool website. 2015. <https://bonnie.healthit.gov/>. Accessed May 12, 2015.
29. Scholle SH, Roski J, Dunn DL, et al. Availability of data for measuring physician quality performance. *Am J Manag Care*. 2009;15:67–72.
30. Parsons A, McCullough C, Wang J, et al. Validity of electronic health record-derived quality measurement for performance monitoring. *JAMIA*. 2012;19:604–09.
31. Bailey LC, Mistry KB, Tinoco A, et al. Addressing electronic clinical information in the construction of quality measures. *Acad Pediatr*. 2014;14:S82–89.
32. Schleyer T, Spallek H, Hernandez P. A qualitative investigation of the content of dental paper-based and computer-based patient record formats. *JAMIA*. 2007;14:515–26.
33. Liu K, Acharya A, Alai S, et al. Using electronic dental record data for research: a data-mapping study. *J Dent Res*. 2013;92:90S–6S.
34. Lehmann CU, O'Connor KG, Shorte VA, et al. Use of electronic health record systems by office-based pediatricians. *Pediatrics*. 2015;135:e7–15.
35. Furukawa MF, King J, Patel V, et al. Despite substantial progress in EHR adoption, health information exchange and patient engagement remain low in office settings. *Health Aff*. 2014;33:1672–79.

AUTHOR AFFILIATIONS

¹Key Analytics and Consulting, LLC, Gainesville, Florida, USA (Dr Herndon was with the Department of Health Outcomes and Policy, University of Florida College of Medicine, Gainesville, Florida, USA, when this study was conducted)

²Dental Quality Alliance, American Dental Association, Chicago, Illinois, USA

³Stephenson Developers Group, Clearwater, Florida, USA

⁴Exan Group, Vancouver, Canada

⁵American Dental Partners, Inc., Wakefield, Massachusetts, USA

⁶Department of Community Dentistry and Behavioral Science, University of Florida College of Dentistry, Gainesville, Florida, USA

⁷Asian Health Services, Oakland, California and National Network for Oral Health Access, Denver, Colorado, USA