Measuring up
Implementing a dental quality measure in the electronic health record context

Aarti Bhardwaj, BDS, MPH; Rachel Ramoni, DMD, ScD; Elsbeth Kalenderian, DDS, MPH, PhD; Ana Neumann, DDS, MPH, PhD; Nutan B. Hebbali, BDS, MPh; Joel M. White, DDS, MS; Lyle McClellan, DDS; Muhammad F. Walji, PhD

Self-evaluation will ensure that dentistry as a profession can provide evidence to the community at large that its members are responsible stewards of oral health.1 Dentistry has made strides in developing dental quality measures in the wake of increasing adoption of electronic health records (EHR) and enhanced focus on accountability in health care. This is a critical, and perhaps overdue,2 step forward. According to President Clinton’s Advisory Commission on Consumer Protection and Quality in the Health Care Industry, “A key element of improving health care quality is the nation’s ability to measure the quality and provide easily understood, comparable information on the performance of the industry.”3

In the years since 1998, the clinical quality measure ecosystem has evolved, and the evolution has accelerated with the uptake of EHRs. The Medicare and Medicaid EHR incentive programs that are part of the Health Information Technology for Economic and Clinical Health (HITECH) Act provide financial incentives for the “meaningful use” of certified EHR technology. An eligibility criterion for the incentives is the use of the EHRs to report clinical quality measures4 (Box 1). Meaningful Use Stage 2 was the first set of guidelines to incorporate oral health measures that were developed by the Centers for Medicare and Medicaid Services (CMS) through a contract with Booz Allen Hamilton.5 In parallel, the Dental Quality Alliance (DQA) also has been defining electronic clinical quality measures. In response to a request from CMS, the American Dental Association established the

ABSTRACT

Background. Quality improvement requires using quality measures that can be implemented in a valid manner. Using guidelines set forth by the Meaningful Use portion of the Health Information Technology for Economic and Clinical Health Act, the authors assessed the feasibility and performance of an automated electronic Meaningful Use dental clinical quality measure to determine the percentage of children who received fluoride varnish.

Methods. The authors defined how to implement the automated measure queries in a dental electronic health record. Within records identified through automated query, the authors manually reviewed a subsample to assess the performance of the query.

Results. The automated query results revealed that 71.0% of patients had fluoride varnish compared with the manual chart review results that indicated 77.6% of patients had fluoride varnish. The automated quality measure performance results indicated 90.5% sensitivity, 90.8% specificity, 96.9% positive predictive value, and 75.2% negative predictive value.

Conclusions. The authors’ findings support the feasibility of using automated dental quality measure queries in the context of sufficient structured data. Information noted only in free text rather than in structured data would require using natural language processing approaches to effectively query electronic health records.

Practical Implications. To participate in self-directed quality improvement, dental clinicians must embrace the accountability era. Commitment to quality will require enhanced documentation to support near-term automated calculation of quality measures.

Key Words. Dental care for children; dental public health; informatics.

Copyright © 2016 American Dental Association. All rights reserved.
DQA to frame oral health care quality measures using a consensus-building process among a broad base of stakeholders. For Meaningful Use Stage 3, the DQA and Meaningful Use efforts may intersect, as the DQA is poised to contribute to Meaningful Use clinical quality measures.

Before the advent of EHRs, stakeholders could gather information related to quality measures by reviewing administrative claims data, which are proprietary to the payers and are aggregated only at the administrative claims level. The Meaningful Use program has the potential to transform EHRs into a catalyst for what the National Academy of Medicine (formerly the Institute of Medicine) has called a learning health care system, which is defined as a “healthcare system designed to generate and apply the best evidence for the collaborative healthcare choices of each patient and provider; to drive the process of discovery as a natural outgrowth of patient care; and to ensure innovation, quality, safety, and value in health care.” The learning health care system has implications for quality improvement, as well as clinical research and the analysis of the comparative effectiveness of different treatments. Quality measurement is fundamental to quality improvement, and successful implementation of quality measures depends on the availability of timely, accurate, and reliable data sources. For this reason, we believed it was an opportune time to implement one of these Meaningful Use measures in a real-world setting, in which clinicians use EHRs, to evaluate the feasibility and performance of this measure. In particular, our objectives were to evaluate whether the measure could be implemented on the basis of structured data available in the record and to compare the performance of an EHR-based query with the findings derived from an in-depth manual chart review.

METHODS

We assessed the feasibility and performance of Meaningful Use Stage 2 measure CMS74V3,10 which measured the percentage of children, aged 0 to 20 years, who received a fluoride varnish application. The site chosen was the School of Dentistry at the University of Texas Health Science Center at Houston. Representatives from different dental institutions met to frame the exact definition of the query by reviewing the measure specifications and determining how to retrieve them from the EHR. Our group represents a range of areas of expertise, including dental public health (A.B., N.B.H., A.N., and R.R.); oral surgery, health policy and management, and data standards (E.K.); general dentistry, caries risk management (J.M.W.); general dentistry (L.M.); and informatics (M.F.W.). At each of the institutions represented, clinicians used axiUm (Exan Group, Henry Schein) software for dental EHRs. We received institutional review board approval from the University of Texas Health Science Center at Houston before conducting the retrospective chart reviews. In particular, we used the following 5-step procedure:

Step 1—determine the total patient population that meets the denominator criteria. By querying the dental EHRs, we determined the total number of eligible patients who had an encounter (visit) for oral evaluation in the 2013 measurement period. We used evaluation-related Current Dental Terminology (CDT)11 treatment codes to determine an encounter (as shown in Box 2) for patients who were 20 years or younger before the start of the measurement year.

Step 2—determine the sample size necessary for manual review. We determined the sample size requirement using precision of 5% around the estimate at the 95% confidence interval (CI) level. We obtained data from the Texas Medicaid database, estimating the fluoride varnish application rate at 44.27% (N. Yang, PhD, e-mail communication, April 2014).

Step 3—determine the population within the sample that meets the numerator criteria through EHR query. We determined the total number of eligible patients who received a fluoride varnish application in the measurement period by querying the set of patients identified in step 1 (denominator). We used fluoride-
related CDT treatment codes to determine fluoride application (Box 2).

Step 4—determine the population within the sample that meets the numerator criteria through manual chart review. Two authors (A.B. and N.B.H.) conducted independent manual reviews on the sample size determined in step 2 and came to consensus for situations for which they had differences. First, these authors identified discrepancies, and then they referred back to the definitions of the numerator and denominator and discussed the discrepancies. They were able to resolve all their differences in this way. They assessed interrater reliability using the κ statistic.

Step 5—compare the concordance between query and manual review. We compared the performance of the dental EHR query with an in-depth manual chart review, which served as the criterion standard with respect to information contained within the record. We calculated sensitivity, specificity, positive predictive value, and negative predictive value.

RESULTS
In the paragraphs that follow, we describe the results of each step involved in creating and evaluating the dental quality measure outlined in the Methods section.

Step 1—identify the denominator. We identified 4,376 patient charts that met the requisite criteria. The mean age (standard deviation [SD]) of the patients was 8.81 years (4.80 years), and 51.2% were male. At the time of our review, the term “Hispanic” was used as 1 of 1 of the categories of race; currently, Hispanic is considered to be an ethnicity that should be categorized separately. Thus, we categorized the population as 50.3% Hispanic, 15.8% African American, 14.3% white, 4.4% Asian, and 0.4% American Indian or Alaskan. We noted that 1.8% of the patients were categorized as “other,” and 13.0% had unknown or unspecified race, or they had not answered the question regarding their race. Since this review was completed, race and ethnicity have been recorded separately.

Step 2—identify the manual review sample size. Our sample size calculations determined that we needed a minimum of 381 charts to have a precision of 5% around the estimate at the 95% CI level. Conservatively, we randomly selected 500 charts for review from the denominator records using a random number generator. The mean age (SD) of these patients was 9.05 years (4.90 years), and 50.2% were male. The race and ethnicity distribution was as follows: 51.8% Hispanic, 13.2% African American, 12.6% white, 5.3% Asian, 0.2% American Indian or Alaskan, 2.0% other race or ethnicity, and 15.0% race or ethnicity unknown, unspecified, or not provided. We excluded 10 patients’ records from the manual review because they belonged to test patients (that is, they were not the records of actual patients). We did not exclude these patients’ records from the automated review, however, as their records would have been captured in a real-world automated review.

Step 3—identify the numerator. Among the 4,376 charts that met the denominator criteria, 3,209 met the numerator criteria (that is, the patient had received a fluoride varnish) through the dental EHR query.

Step 4—identify the numerator through a manual chart review of sampled records. The 2 chart reviewers (A.B. and N.B.H.) had excellent interrater agreement with respect to the denominator (κ = 0.907). These same reviewers also had excellent interrater reliability (κ = 0.984) determining whether the patient had received an application of fluoride varnish (numerator). Manual chart review indicated fluoride varnish application in 380 of the 490 charts.

Step 5—evaluate the concordance between query and manual review. When we used the dental EHR-based query, we noted that 355 of 500 patients (71.0%; 95% CI, 67.2-75.2%) had received fluoride varnish, and when we used the manual chart review, we noted that 380 of 490 patients (77.6%; 95% CI, 73.9-81.2%) had received fluoride varnish. The table shows how the automated EHR-based query performed compared with the manual chart review.

Because the procedure code for fluoride varnish had not always been entered, those missing data were not identifiable by the automated query, which was worded to search for structured data. We found that although the fluoride varnish procedure code was documented 90.3% of the time as structured data, in 9.7% of the cases, the

**Box 2**

Current Dental Terminology procedure codes used to query electronic health records as part of the fluoride varnish quality measure.*

<table>
<thead>
<tr>
<th>CDT PROCEDURE CODES QUERIED IN THE DENOMINATOR TO DETERMINE AN ENCOUNTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>D0120: Periodic oral evaluation</td>
</tr>
<tr>
<td>D0140: Limited oral evaluation</td>
</tr>
<tr>
<td>D0145: Oral evaluation for a patient under 3 years of age and counseling with primary caregiver</td>
</tr>
<tr>
<td>D0150: Comprehensive oral evaluation</td>
</tr>
<tr>
<td>D0160: Detailed and extensive oral evaluation—problem focused</td>
</tr>
<tr>
<td>D0170: Re-evaluation—limited</td>
</tr>
<tr>
<td>D0180: Comprehensive periodontal evaluation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CDT PROCEDURE CODES QUERIED IN THE NUMERATOR TO DETERMINE FLUORIDE APPLICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1203: Topical application of fluoride—child</td>
</tr>
<tr>
<td>D1204: Topical application of fluoride—adult</td>
</tr>
<tr>
<td>D1206: Topical fluoride varnish</td>
</tr>
<tr>
<td>D1208: Topical application of fluoride</td>
</tr>
</tbody>
</table>

* Source: American Dental Association.† CDT: Current Dental Terminology.
application of varnish was noted only in a free text clinical note. Reviewers also found the following reasons documented in the note regarding why fluoride varnish had not been applied: fluoride had been provided earlier during a visit with another dental clinic, the patient had a heart murmur and the dentist wanted to contact the patient’s cardiologist first, the patient had been transferred to another clinic after having been accepted to the Children’s Health Insurance Program, and the patient’s mother had declined the fluoride application.

**DISCUSSION**

Our preliminary results after implementing a single quality measure in the EHR are promising and show the feasibility of using EHR data to determine dental quality measures. We felt that the time was right to focus on this, as the momentum builds for widespread adoption of EHR use in dental practices. Indeed, some states now require dental practices to implement EHRs. We chose to implement a Meaningful Use measure that relied on data that were collected in a standardized way across sites and EHR installations. A measure that requires data that not all sites collect or data that were collected in different ways at different sites would be more challenging to implement broadly. In the course of our work, we identified best practices for defining quality measures that should be followed as dentistry expands its quality measure set. First, the population defined in the denominator should be appropriate to the numerator being assessed. In the measure we implemented, we followed the Meaningful Use specification and therefore did not exclude those patients who were seen only for an emergency visit, although it is unlikely that fluoride varnish would be appropriate in the emergency setting. Likewise, we did not explicitly exclude infants younger than 6 months; however, those patients would be unlikely to have teeth available for fluoride varnish application. Second, measures should be unambiguously definable. In the case of the measure we assessed, the denominator is specified as “children, age 0-20 years, with a visit during the measurement period.” In our cross-site discussions, we noted that the term visit is ambiguous because it is not clear, for example, whether it applies to 1-time visits such as a single focused periodontal evaluation or to emergency visits. Ambiguity in measure definition leads to cross-site variation in implementation, which inhibits interpretation and comparison.

A great promise of oral health care quality measures is the ability to compare and share data across sites. When we can transform these data into knowledge, we establish the foundation of a learning health care system. In the case of quality improvement, quality measures are the vehicle by which raw data are first transformed into actionable knowledge. In addition to the importance of consistent specification, there is a need to aggregate measures across sites for shared learning to occur. For the most part, aggregation occurs only with administrative claims data at the payer level, as noted by the DQA. This status quo is being challenged by initiatives to aggregate clinical data, such as the BigMouth dental data repository, which the authors of this article have developed. Thus far, the BigMouth repository is amassing data extracted from 6 dental schools’ EHRs. One of our goals for BigMouth is that it will either allow the calculation of quality measures on the basis of raw data deposited into BigMouth or it will be a repository for quality measures that have been preprocessed at the originating clinical site, likely via something akin to the Health Quality Measures Format. The Health Quality Measures Format is becoming a standard system for representing a health quality measure as an electronic document.

One of the challenges often mentioned in the framing of dental clinical quality measures is the absence of a standardized dental diagnostic terminology. The lack of such a terminology has limited a number of bodies, including the DQA, in terms of the types of measures they can propose. For example, the DQA proposed a

---

**TABLE**

Results showing performance of the electronic health record–based query compared with the manual chart review (which served as the criterion standard) for determining patients who had received fluoride varnish.

<table>
<thead>
<tr>
<th>STATISTICAL MEASURE</th>
<th>PERFORMANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>90.5% (95% CI, 87.1-93.3%) of the time that manual review revealed fluoride varnish application.</td>
</tr>
<tr>
<td>Specificity</td>
<td>90.8% (95% CI, 84.2-95.3%) of the time that the manual review did not reveal fluoride varnish application.</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>96.9% (95% CI, 94.5-98.4%) of the time that the query identified fluoride varnish application, the manual review revealed fluoride varnish application.</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>75.6% (95% CI, 67.3-82.0%) of the time that the query did not identify fluoride varnish application, the manual review did not reveal fluoride varnish application.</td>
</tr>
</tbody>
</table>

* CI: Confidence interval.
measure that identified the “percentage of adults treated for periodontitis who received comprehensive oral evaluation…”15; however, if a patient was diagnosed with periodontitis but had not received treatment, his or her data would not be captured by such a measure, creating a quality “blind spot.” In response to this need, our research group has established the Dental Diagnostic System (formerly called EZCodes), which has been designed for use within EHRs in the dental setting.19–22 The Dental Diagnostic System is available free and is installed at 16 academic institutions and several large dental group practices in the United States, Caribbean and Latin American countries, and Europe. The Dental Diagnostic System has been designed to serve as a clinician-friendly interface to rich reference terminologies such as SNOMED, into which SNODENT23 has been integrated. Broader use of standardized diagnostic terminologies will give the profession and the public a more complete picture of dental health care quality. With that said, our own experience with dental diagnostic terms confirms that there may be some gaps in documentation.21 This serves as a reminder that quality measures on the basis of secondary data analysis represent what was documented, which does not always align with what occurred, during the clinical visit.

CONCLUSIONS

Clearly, the solution is not to forgo clinical data–based quality measures. Instead we must encourage, and sometimes enforce, the entry of important structured data. More sophisticated approaches would include natural language processing (NLP)24 to identify measure elements from free text notes. We should not, however, underestimate the work that will be required to accomplish this goal. Beyond the simplest rule-based NLP (for example, looking for the phrase “fluoride varnish”), NLP is analytically complex and requires access to large and varied sources of real-world free text data.25 In addition to NLP, there also exists the potential to infer missing data from elements that are present.26 Ultimately, clinicians should benefit from their documentation efforts, although dental appointments might require additional time so that clinicians can complete the thorough documentation needed to correctly and completely capture the data necessary. By conducting measurement efforts and sharing the lessons we learn, we can work together to live up to the high standards that the dental profession and our patients expect.

Dr. Bhardwaj was a graduate research assistant, Department of Diagnostic and Biomedical Sciences, School of Dentistry, University of Texas Health Science Center at Houston, Houston, TX, when this article was written. She now is a dental student, University of California Los Angeles, Los Angeles, CA.

Dr. Ramoni is an assistant professor, Department of Oral Health Policy and Epidemiology, Harvard School of Dental Medicine, Boston, MA.

Dr. Kalenderian is the chair, Department of Oral Health Policy and Epidemiology, Harvard School of Dental Medicine, Boston, MA.

Dr. Neumann is an assistant professor, Department of General Practice and Dental Public Health, School of Dentistry, University of Texas Health Science Center at Houston, Houston, TX.

Dr. Hebballi is a research associate, Department of Diagnostic and Biomedical Sciences, School of Dentistry, University of Texas Health Science Center at Houston, Houston, TX.

Dr. White is a professor, Department of Preventive and Restorative Dental Sciences, School of Dentistry, University of California, San Francisco, CA.

Dr. McClellan is the director of doctor development, Willamette Dental Group, Portland, OR.

Dr. Walji is an associate dean, Technology Services and Informatics, and an associate professor, Department of Diagnostic and Biomedical Sciences, School of Dentistry, University of Texas Health Science Center at Houston, Houston, TX. When this article was written, she was a graduate research assistant, Department of Diagnostic and Biomedical Sciences, University of California Los Angeles, Los Angeles, CA. Dr. Walji is the author of the book, “The Hidden Practice of Health Care: The Developing Role of Information Technology in the Realization of Health Care Quality.”

Disclosure. None of the authors reported any disclosures.

Research reported in this publication was supported in part by award RO1DE28466 from the National Institute of Dental and Craniofacial Research of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The authors thank Krishna Kumar Kookal, MS, Office of Technology Services and Informatics, School of Dentistry, University of Texas Health Science Center at Houston, for assistance in extracting data from the electronic health records.


