NOVEMBER 2018

DENTAL QUALITY ALLIANCE™

Improving Oral Health Through Measurement

GUIDANCE ON PRACTICE BASED MEASURES IMPLEMENTATION

The development of this guidance document has been informed by the DQA's Project on Pediatric Practice-Based Measures Testing:

Alignment and Harmonization in Reporting Quality: Establishing Reliability Across Reporting Levels

KEY RECOMMENDATIONS

- Comparisons should be made within a data source (i.e., one practice's measure score calculated using claims data should not be compared with another practice's measure score calculated using billing data or electronic dental record, EDR, data).
- Recommended minimum denominator sizes when used in accountability applications that are based on relative comparisons between practices are:

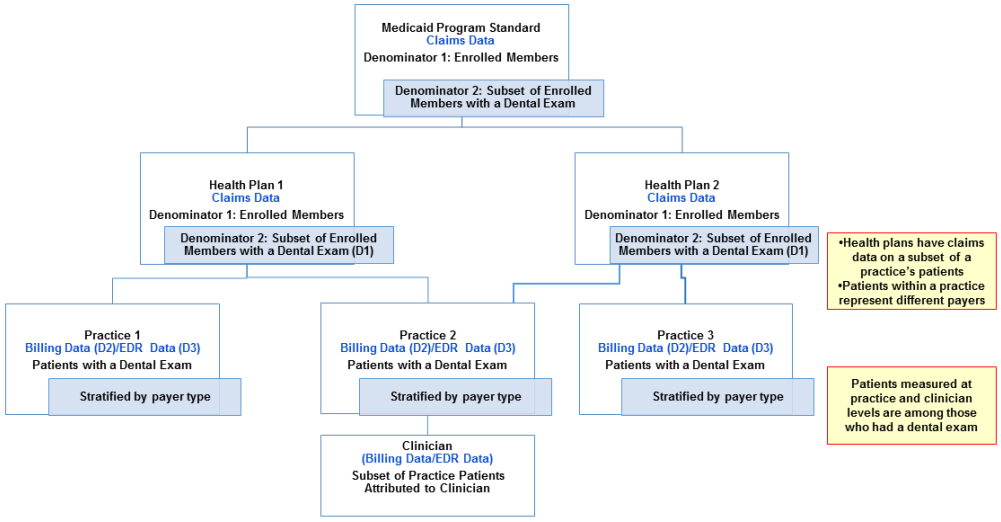
  (1) 100 when using payer claims data, (2) 50 when using practice billing data, and (3) 50 when using practice EDR data.
- When practice measure scores are clustered closely together (i.e., there is low practice-to-practice variation), accountability applications should focus on overall improvement across practices rather than relative comparisons between practices.
- Before incorporating any quality measures in accountability applications, those applications should be tested using multiple years of measure data to evaluate whether the application achieves the intended goals and whether there are unintended consequences that may undermine quality improvement efforts. It is incumbent upon the users of performance measures to carefully evaluate these impacts prior to implementing the accountability application.

Background

Quality of care is assessed at multiple levels, such as practices, managed care organizations or medical/dental benefits administrators, public insurance programs, and public health programs. There often are different measurement considerations at different "levels" of care or "reporting unit" as well as across different types of data sources (e.g., administrative claims, EHRs, or surveys). Measures should be reported at the level (e.g., program, plan, or practice) and using the data source (e.g., administrative claims or EHR) for which they were developed and validated. Implementation of measures at different levels or with different data sources than those for which the measure was intended may not be reliable.

The DQA has developed dental quality measures for adults and children to evaluate Medicaid programs and dental plans. As Medicaid programs and managed care organizations are increasingly held accountable for performance on these measures, they in turn hold their contracted practices accountable. Because practice-level measurement is often driven vertically (from program to plan to practice), practice-level measures will be most effective when aligned with program- and plan-level measurement. Standardized measurement that is aligned across public and private sectors and harmonized across different levels of reporting aggregation can help pave the way to improvement. Starting with broad populations, national goals guide the development of program-level measures, which are then used to derive practice- and clinician-level measures such that the underlying care improvement goals remain unchanged and the derivative measures are relevant to the populations served at each level.

Figure 1. Conceptual Relationship between Program Standard and Measure Derivatives



Lack of such standardization is problematic because practices and other users of the information may receive conflicting performance information from different sources. Lack of standardization also compromises reliability. When reliability of measurement is poor, it increases the likelihood that

distinctions in performance do not reflect real differences in performance. It also increases the likelihood of misclassification – e.g., low performing practices being classified as high performing and vice versa.

PURPOSE

The purpose of this guidance document is to inform the stakeholders on key implementation considerations recommended by the Dental Quality Alliance (DQA) for the use of practice-based measures for accountability applications, external reporting, and internal quality improvement. The recommendations have been informed through the DQA's Project on Pediatric Practice-Based Measures Testing: Alignment and Harmonization in Reporting Quality: Establishing Reliability Across Reporting Levels[1] using the following measures:

- Topical Fluoride Application for children ages 1-20 years at elevated caries risk
- Care Continuity for children ages 2-20 years
- Dental Sealants for children ages (6-9 years and 10-14 years) at elevated caries risk

Dental practices are identified using Taxpayer Identification Numbers (TINs), which are standard data elements in administrative databases. Recognizing that a TIN can represent multiple physical locations, measure implementers may need to use local data elements to identify individual practice locations.

Before using a measure for accountability purposes, the DQA strongly recommends that the ***accountability application*** be preceded by a period during which reporting entities gain experience with measure implementation, data are collected to establish baseline values, and appropriate benchmarks for comparison and performance goals are identified. **An accountability application is defined as: "use of performance results about identifiable, accountable entities to make judgments and decisions as a consequence of performance, such as reward, recognition, punishment, payment, or selection (e.g., public reporting, accreditation, licensure, professional certification, health information technology incentives, performance-based payment, network inclusion/exclusion)"** (NQF Glossary).

RECOMMENDATION 1: Minimum denominator recommendations

Based on testing data, the DQA recommends the following minimum denominator thresholds that should be used for practice-based measurement specifically in cases of **accountability applications** in order to promote measurement that is reliable for drawing conclusions when making comparisons between practices:

- 100 when using payer claims data,
- 50 when using practice billing data, and
- 50 when using practice EDR data.

The requirements for measurement used solely for internal quality improvement purposes are not as strict. However, it is important to be cognizant and transparent of the limitations of any measurement results presented in the context of internal quality improvement so that the information provided is not misleading. For example, sample sizes should be noted as should small variation between practices.

[1] A final report with testing results from this project is available from the DQA upon request. Please contact the DQA@ada.org.

RECOMMENDATION 2: Baseline performance assessment

Measure implementers should recognize that for comparison purposes reliability is affected by baseline performance and practice-to-practice variation as well as minimum denominator size. At any given denominator size, the smaller the practice-to-practice variation, the lower the reliability of the measure scores for comparison purposes. **If measure scores are generally clustered very closely together, it will be more difficult to make reliable and meaningful distinctions in performance between the measured entities.**

Consequently, when using practice-based measures in accountability applications, implementers should consider the extent to which there are differences in baseline performance between practices. **When performance for a measure is generally clustered within a small absolute range, that measure is best suited for use in accountability applications that drive all practices toward a higher benchmark, but the accountability application should not be based on relative performance between practices because between-entity differences in performance may not be meaningful.**

RECOMMENDATION 3: Data source considerations

Any comparisons or categorization of practice measure scores should be conducted within a data source; i.e., one practice's measure score calculated using claims data should not be compared with another practice's measure score calculated using billing data or EDR data. Even though the specifications *may be* similar for billing data and EDR data, measure scores for one practice calculated using billing data should only be compared with measure scores for another practice calculated using billing data.

When reporting measure scores using claims data, it should be recognized that the payer subset of the practice's patients may not reflect the overall performance of the practice, particularly when the payer covers a small percentage of the practice's patients.

RECOMMENDATION 4: Selecting measures

A balanced measurement approach that evaluates multiple aspects of care is essential to promoting improved outcomes, understanding disparities, and planning for improved performance. Although it would be ideal to measure all aspects of care, resource constraints and measurement limitations may require prioritization. Thus, organizations engaged in oral healthcare performance measurement must first define their care goals, assess the capability to implement reliable measurement, and then select an appropriate set of measures for implementation. In addition to selecting measures based on the care goals, measure selection should also take into account whether there are sufficient numbers of eligible patients to achieve reliable measurement and the extent of practice-to-practice variation.

RECOMMENDATION 5: Assessing internal readiness for measurement

The use of standardized specifications is also a pre-requisite to reliable measure implementation. Equally important is appropriate implementation of the measures. Measure users should verify that they can feasibly, reliably and validly implement the measures within their own systems of care. This includes assessing the completeness and accuracy of the critical data elements used to calculate the measures, implementing the measures following the detailed measure specifications, and evaluating face validity of the resulting measure scores with individuals who have appropriate local expertise.

RECOMMENDATION 6: Establishing a baseline and time trends prior to implementing accountability applications

Quality assessment and performance improvement are ongoing processes, and iterative measurement is essential for identifying, implementing, evaluating, monitoring, and sustaining quality improvement initiatives. Initial measurement can be used to establish baseline performance and to identify potential areas for improvement. Subsequent measurements further inform the identification of performance gaps and disparities in care and can be used to evaluate the effect of improvement efforts and monitor performance over time.

Current applications of practice-level measurement in dentistry frequently include classifying practices into groups for comparative purposes. For example, practices may be identified as high or low performing based on whether they fall above or below certain percentiles, respectively; it should be recognized that a certain proportion of practices will always fall above or below these cutoffs regardless of whether there are real differences in performance.[2] Even the use of statistical tests to identify practices that are different from the average performance can result in misclassifications of performance when reliability is low.[3]

To reduce the potential for misclassification, in addition to following recommendations regarding minimum denominators, before using a measure for accountability purposes, it is strongly recommended that the accountability application be preceded by a period during which reporting entities gain experience with measure implementation, data are collected to establish baseline values and appropriate benchmarks for comparison, and performance goals are identified. The DQA **recommends that incorporating any quality measures for accountability applications be tested using multiple years of measure data to evaluate whether the application achieves the intended goals and whether there are unintended consequences that may undermine quality improvement efforts. It is incumbent upon the users of performance measures to carefully evaluate these impacts prior to implementing the accountability application. Development of benchmarks for quality measures used in any reporting applications should be guided by historical data evaluation for the**

[2] Landon BE, Normand SLT, Blumenthal D, Daley J. Physician clinical performance assessment: Prospects and barriers. *JAMA*. 2003;290:1183-1189.

[3] Adams J. The Reliability of Provider Profiling: A Tutorial. RAND Corporation. 2009; https://www.rand.org/pubs/technical_reports/TR653.html. Accessed March 15, 2018.

population being served. When using practice-level measures in accountability applications, the Medicaid program, plan, or other organization should develop benchmarks using historical data based on the same definition of the measure that practices will be held accountable to and test the application prior to implementation. Additionally, benchmarks need to be evaluated for each re-measurement period to avoid undermining the strides in quality improvements. The application and underlying methodology should be transparent, and practice feedback should be sought on the validity of the measure scores and the methodology. Implementing measures initially in non-accountability quality improvement initiatives can inform the development of accountability applications. Only after there is experience with measure implementation, careful review and interpretation of the resulting measure rates, and an evaluation of the measure's effectiveness in promoting identified quality improvement and care goals, should accountability applications be considered.

RECOMMENDATION 7: Reporting performance

Performance reporting should be transparent and clear. Practices that do not meet minimum denominator thresholds for reliable measurement should be noted as having "low denominators." The data source used for calculating the measure scores should be noted in reporting. Reporting measure scores using claims data should explicitly recognize that the payer subset of the practice's patients may not reflect the overall performance of the practice, particularly when the payer covers a small percentage of the practice's patients.

RECOMMENDATION 8: Assessing performance for small-denominator practices

A significant proportion of practice locations (especially those located in rural areas) within groups may be excluded from measurement at the denominator thresholds required for reliable measurement, particularly for the two sealant measures that apply to restricted age ranges. While reporting measures below the recommended denominator thresholds may not be reliable for accountability purposes, the results used internally can still provide insight for process improvement efforts in these practices and geographic areas.

Identify a core set of measures that can be reported on for most practices. Not all measures need to be reported for all practices. Measures such as Care Continuity that capture most of the practice's pediatric population may be reliably implemented for most practices. Other measures, such as the sealant measures, may be used in accountability applications only among the subset of practices for which they can be reliability measured. Public reporting should note "low denominator" for those practices that do not meet the minimum denominator requirements for a measure.

Provide information to smaller practices that can be used for quality improvement purposes. Reports

can be provided to practices that do not meet minimum denominator requirements as informational to guide quality improvement efforts, but not for mandatory public reporting or comparisons to other practices. Qualitative assessments of trends over time can be used to support practices' quality improvement efforts.

Do not use in accountability applications. If the recommended minimum denominator thresholds are not met, the measures should not be used in any type of accountability application, including mandatory public reporting and reward, recognition, payment, selection or other incentive programs.