Development of Effectiveness of Care and Use of Services Measures for Dental Care Plans

James D. Bader, DDS, MPH; Daniel A. Shugars, DDS, PhD; B. Alex White, DDS, DrPH; D. Brad Rindal, DDS

Abstract

Objectives: *Standardized measures to assess clinical aspects of the performance of managed dental care plans are not available. This project sought to develop and evaluate measures for effectiveness of care and use of services that could be calculated using a plan's administrative data.* ***Methods:*** *Two panels of stake holders representing dental plans, purchasers, and dental providers participated in a modified Delphi process to refine initial sets of effectiveness of care and use of services measures modeled after HEDIS® measures for medical care. The refined measures were then pilot tested in two dental health maintenance organizations.* ***Results:*** *The development process resulted in specification of seven effectiveness of care measures assessing disease activity classification, and prevention and outcomes for caries, periodontal disease, and tooth loss. Six use of services measures focusing on prophylaxes, third molar surgery, preventive, restorative, prosthetic, surgical, and endodontic care also were specified. Pilot testing of the measures indicated reasonable reliability and sensitivity, but also demonstrated the need for supervision or auditing of the process.* ***Conclusions:*** *These standardized measures for dental care plan performance are available for immediate use. However, because the measures depend on diagnostic information (periodontal probing data and diagnoses associated with restorative treatments) in the administrative data set, their adoption will require changes in most plans' data systems and data collection policies. [J Public Health Dent 1999;59(3):142-49]*

Key Words: effectiveness of care; use of services; performance measures; quality of health care; health benefit plans—standards, outcomes, and process assessment.

Arrangements for the provision of dental care have been changing rapidly in the past few years. In 1994 an estimated 32.2 million persons participated in managed dental care plans, including all dental HMOs and PPOs, as well as dental referral networks. By 1997, estimated participation was 56.4 million, a 75 percent increase that resulted in managed care plans representing 38 percent of the total dental benefits market (1). The growth in this sector may be partially attributed to purchasers' (employers and government) perceptions that such plans offer cost savings when compared to traditional indemnity plans. However, if, as expected, expansion of dental managed care replicates that of medical managed care (2,3), purchasers may soon begin to expect "proof" that the costs savings are not occurring at the price of reduced quality.

Purchasers traditionally have been said to have a "bottom line" orientation, caring principally about the cost of a dental program. The only reported exploration of purchasers' attitudes would suggest that they tend to select dental plans initially on the basis of price and beneficiary access, and tend to continue with these plans unless frequent cost or premium increases or a high volume of employee complaints lead to reconsideration (4). Benefits managers indicated that most employee complaints involved clinician and plan communications, convenience, and claims performance, rather than quality (4). However, commentary in the human resource management literature indicates that purchasers are beginning to pay attention to issues of quality, chiefly by selection of plans that "credential" their clinicians and to some extent through use of plans that are accredited (5). Again, if developments in dentistry parallel the experience in medicine, purchasers will soon expect to have objective information available that goes beyond credentialing or accreditation to describe dental plan performance, i.e., a dental health plan "report card" (6).

Health plan report cards are sets of standardized measures that reflect the performance of health care plans. Typically, these standardized measures capture information at the population level, i.e., plan performance as it pertains to all plan enrollees whether they use services or not, or all enrollees with a specific condition. The most well-known and widely used set of measures is the Health Plan Employer Data and Information Set (HEDIS®), developed by the National Committee for Quality Assurance (NCQA) (7). These measures originally were developed in the early 1990s, and currently are in their third release, having undergone a thorough revision in 1997. HEDIS® divides the performance spectrum into eight domains, two of which—effectiveness of care and use of services—deal with clinical issues. The remaining six domains address access/availability of care, satisfaction with the experience of care, cost of care, health plan financial stability, health care choices, and other health plan descriptive information. No such set of performance measures exists for

Send correspondence to Dr. Bader, Sheps Center for Health Services Research, University of North Carolina, CB #7590, Chapel Hill, NC 27599-7590. E-mail: jim_bader@unc.edu. Web site: http://www.shepscenter.unc.edu. Dr. Shugars is with the Sheps Center for Health Services Research; Dr. White is with Kaiser Permanente Center for Health Research, Portland, OR; and Dr. Rindal is with HealthPartners Research Foundation, Minneapolis, MN. Reprints will not be available. Supported by the Agency for Health Care Policy and Research, HS09453. Manuscript received: 4/6/99; accepted for publication: 5/17/99.

dental care. While it is conceivable that for nonclinical domains, the HEDIS® measures might be adapted for use in dentistry, it is immediately clear that the clinical domains essentially are not transferable, given the specific clinical questions addressed. Thus, any set of clinical performance measures for dentistry must be based either on existing measures in dentistry, or on completely new measures.

Unfortunately, the availability of clinical performance measures in dentistry is limited, and the few available measures generally are associated with the use of services, i.e., process measures, as opposed to the effects of those services, i.e., outcomes measures (6). The problem with process measures is that in many instances, it is not clear that the service is needed. Also, instances where needed services are not provided typically are not detected. Unless the use of a service can be measured only among that portion of the enrolled population for which it is appropriate, the measurement will be meaningless. In addition, the effectiveness of many common dental procedures is essentially untested; thus, the link between process and outcome is not established (8). The development of dental performance measures has been impeded by this lack of knowledge of outcomes of dental treatment and, hence, a lack of consensus among dentists about what constitutes appropriate treatment (8). This lack of knowledge of outcomes occurs in part because dentistry does not have a tradition of formally recording specific diagnoses or associating such diagnoses with specific services (9), and because most dental practices and dental plans lack information systems capable of capturing the information necessary to calculate performance measures.

These limitations notwithstanding, it is important that dentistry begin to develop clinical performance measures. Purchasers will soon begin to demand objective, standardized measures that reflect the effectiveness of the care they purchase as they begin to select plans on the basis of their value as well as their price. Purchasers also will likely expect objective standardized information describing how dental plans choose to expend their resources in terms of the use of services because that information will help define plans' care management philosophies. This paper describes the development and initial testing of a set of effectiveness of care and use of services measures.

TABLE 1
Original Specification of the Effectiveness of Care Measures (Ref 6)

1. Proportion of enrollees with risk assessment within past two years
2. Proportion of high caries risk 8-year-olds with sealants on four first molars
3. Proportion of high caries risk enrollees receiving supplemental fluoride therapy
4. Proportion of adults enrolled more than a year who receive treatment for caries
5. Proportion of 14-year-olds enrolled more than a year who receive treatment for caries
6. Proportion of high periodontal risk enrollees who receive at least two periodontal treatment sequences
7. Proportion of enrollees with one or more PSR scores of 4 who have at least one score of 4 decrease
8. Proportion of enrollees with any PSR score increasing to 4
9. Proportion of enrollees experiencing tooth loss
10. Mean score of all enrollees completing dental "quality of life" instrument
11. Mean change in dental quality of life score from initial enrollment to third year

Methods

The performance measures were developed in three steps. First, a preliminary set of measures was specified. Second, two stake-holder panels participated in refinement of the preliminary measures. Third, the refined measures were pilot tested in two dental managed care plans.

Specification of Preliminary Measures. The development of the preliminary measures has been described in a previous publication (6). For the effectiveness of care measures, we adopted criteria used by the NCQA in developing the HEDIS® 2.0 measures (10). According to these criteria, measures must address outcomes of importance or processes linked to such outcomes by strong empirical evidence, and outcomes where appropriate treatment can have a substantial beneficial effect. Whenever possible, measures should be based on administrative data, i.e., data used by the plan for management purposes. They should be population based, patient centered, and permit risk adjustment. We reviewed the existing literature and also collected unreported measures we knew to be in use. We identified no effectiveness of care measures that met these criteria. We did find several use of services measures. However, when we examined measures from different sources that quantified use of the same service, we found that the specifications for calculation varied widely.

In the absence of existing measures, we specified 11 effectiveness of care measures *de novo* that met most of the HEDIS® criteria (6) (Table 1). The measures focused on dentistry's *raison d'être*, the extent to which caries and periodontal disease are prevented and managed effectively. To establish this focus, two exceptions to the criteria had to be made. One measure relied on the identification of treatment necessitated by caries, which required that either a reason for treatment or a diagnostic code be linked to every restorative procedure. Such information is not commonly a part of dental plans' administrative data. In addition, patients' caries and periodontal activity classifications were the subject of a rate measure, and were also to be used to stratify analyses to permit comparisons across plans. Here, not only is disease activity classification uncommon in administrative data, but also little empirical evidence supports the predictive validity of the available classification schemes.

We also specified 10 use of services measures from among those we had identified. Our approach was to select measures that reflected both the rate at which enrollees received several specific services (examination, prophylaxis, full-coverage crowns, endodontic treatment, periodontal scaling) and the overall distribution of service provision by type of service (diagnostic, preventive, operative, prosthetic, other) within the plan. No information concerning the validity or reliability of the measures was available.

Refinement of Preliminary Measures. We refined the preliminary

measures with the help of two stakeholder panels. We first formed a steering committee consisting of senior managers (dental director, executive director) of 11 dental plans, including not-for-profit dental health care organizations, preferred provider networks, and indemnity plans. We discussed the preliminary measures at the first meeting of this committee and the attributes of the measures for which we wished the stake holders' opinions. Prior to the second meeting, we distributed summaries of all preliminary measures together with rating sheets with which to assess each measure along five dimensions (Table 2). The rating sheets also probed for detailed comments explaining the ratings and for suggested alternative measures. At the second meeting, each committee member received feedback materials listing for each measure the member's own ratings, the group's mean ratings, and all open-ended comments and suggestions. This information served as the basis for discussion about changes in the measure. Prior to the third meeting, members again rated the effectiveness of care measures and also participated in two additional rounds of ratings for the use of services measures. Prior to the fourth meeting, considerably more detailed descriptions of the revised measures, including most of the information necessary to actually calculate each measure, were circulated to the committee members. At this meeting the specification of each measure was finalized.

We also sought comments from an advisory committee composed of benefits managers or health benefits specialists for several medium- to large-sized employers, directors of a state dental program and a state dental Medicaid program, and practicing dentists. This committee met once after the initial two meetings of the steering committee, where it reviewed and discussed the effectiveness of care measures. It then participated in parallel ratings exercises for the use of services measures and in determining the final specifications for the effectiveness of care measures.

Pilot Testing the Refined Measures. The refined measures were pilot tested using administrative data from two group model dental health maintenance organizations (DHMOs). The two plans included approximately 115,000 and 90,000 eligible enrollees, respectively. The latter plan offers two distinctly different coverage options, with roughly equal numbers of enrollees in each. One coverage is a preventive dental option available through an enrollee's medical plan, and is frequently used to supplement comprehensive dental benefits held through other plans, while the other coverage is a full-service DHMO benefit. We tested the measures separately for these two coverage options. These plans were selected for participation in the development project partially because their data systems were expected to supply all of the necessary data elements for calculation of the measures. In particular, both plans had standardized protocols for classification of caries and periodontal disease activity. One plan had been using a complete set of approximately 500 diagnostic codes (11) for several years, and the other plan was scheduled to introduce a set of 11 reason-for-treatment codes prior to the test period.

In both plans, our approach to pilot testing the measures was similar, and involved working with personnel from the plan and from the plan's affiliated health research center. In both instances, these personnel included a programmer familiar with the plan's data systems. We provided the programmer with a manual containing complete draft specifications for each of the measures. We asked the programmer to calculate each measure initially based on the information in the manual, and to contact us whenever the information was unclear. We deliberately did not directly supervise or assist in the initial calculation because we wished to assess the feasibility of unassisted compilation of standardized measure reports and to determine which steps in the calculation process were the most problematic. The initial calculations of the measures were reviewed by plan personnel and by us to identify anomalies and internal inconsistencies, whereupon recalculations were requested.

The reliability of the data used to calculate the measures was assessed in two separate evaluations in one plan. In one assessment, the receipt of service information in the administrative data system was compared to the same information recorded in the patient chart. Entries in 59 patient records documenting 1,170 patient visits over the entire eight-year history of the electronic data set were evaluated. In the other assessment of administrative data, which has been reported separately (12), the reliability of classifications of caries and periodontal risk, and of dentists' reasons for treatment was explored by asking a second clinician to conduct replicate examinations using the same criteria for a convenience sample of patients and treatments.

TABLE 2
Explanation of Assessment Dimensions for Stake-holder Ratings

Assessment Dimension	Assessment Explanation
Validity	To what extent do you think this measure assesses an aspect of oral health that should be a concern of a dental care plan?
Accuracy	Assuming that they are collected, to what extent do you think the data necessary to calculate this measure are likely to be valid, truthful, and reliable?
Sensitivity	How much do you think this measure will change in response to changes in clinical treatment protocols?
Interpretability	How well do you think purchasers will understand the implications of this measure in terms of the oral health of their employees?
Future likelihood	How likely is it that your plan will begin to collect the necessary data for this measure within the next five years?

Results

Figure 1 outlines the structural relationship of the set of effectiveness of care measures, and Table 3 shows the numerator and denominator for each measure. All of the measures in this

TABLE 3
Final Effectiveness of Care Measures

1. Current Disease Activity Assessment

(All enrollees in denominator between ages 6 and 17 years, inclusive, at end of reporting year having caries activity assessment within 2 years of end of reporting year) + (All enrollees in denominator 18 years or older at end of reporting year having caries activity and periodontal disease status assessments within two years of end of reporting year)

All enrollees aged 6+ at end of reporting year with continuous enrollment for current and previous reporting years

2a. Preventive Treatment for Caries-active Children

Enrollees in denominator receiving fluoride treatment or dental sealant during reporting year

All enrollees between ages 6 and 17 years, inclusive, at end of reporting year with:
- continuous enrollment during reporting year
- caries-active classification for entire reporting year

2b. Preventive Treatment for Caries-active Adults

Enrollees in denominator receiving fluoride treatment during reporting year

All enrollees in denominator 18 years or older at end of reporting year with:
- continuous enrollment during reporting year
- caries-active classification for entire reporting year

3. New Caries

Enrollees in denominator receiving restorative, prosthetic, endodontic, or oral surgery treatment for caries during reporting year

All* enrollees with continuous enrollment during both reporting year and previous reporting year

4. Periodontal Treatment for Perio-present Adults

Enrollees in denominator receiving periodontal therapy or at least 2 prophylaxes during reporting year

All adult enrollees with:
- continuous enrollment for entire reporting year
- perio-present classification for entire reporting year

5. Improvement in Periodontal Status

Enrollees in denominator where at least one 5+ mm probing depth from previous assessment has resolved to <5 mm, or where at least one sextant with a PSR score of 4 from previous assessment has resolved to a score of 3 or less

All* adult enrollees with:
- continuous enrollment for entire reporting year
- perio status assessment performed in reporting year
- previous perio-present assessment within previous 2 years

6. Deterioration in Periodontal Status

Enrollees in denominator where at least one site with previous probing depth of 4 mm or less has increased to 5+ mm, or where at least one sextant with previous PSR score < 4 is now scored as 4

All* adult enrollees with:
- continuous enrollment for entire reporting year
- perio status assessment performed in reporting year
- previous perio-present assessment within previous 2 years

7. Tooth Loss

Enrollees in denominator receiving an extraction during reporting year for erupted permanent tooth other than third molar or premolar removed for orthodontic reasons

All* enrollees with continuous enrollment for current and preceding reporting years

*These measures may be calculated separately by caries and/or periodontal disease activity status. Measures 3 and 7 (new caries, tooth loss) also may be calculated separately for children and adults.

domain are proportions, i.e., the proportion of all enrollees meeting certain criteria who have experienced a certain clinical outcome, or have received a certain service. The first measure evaluates an aspect of patient assessment, whether patients have current disease activity classifications. As shown in Figure 1, these activity classifications are used to stratify all further measure calculations. No criteria for designation of "high caries activity" and "periodontal disease present" are stated. Promulgation and application of criteria for these classifications are the responsibility of individual plans and can include considerations of extent of current disease as well as risk indicators for future disease. The measure assesses only whether an enrollee has been classified. The second and third measures address dental caries, with a process measure assessing receipt of appropriate preventive services and an outcome measure assessing caries experience among enrollees. This pattern is repeated for periodontal disease, although two outcome measures are used to examine improvement and deterioration in periodontal status. A final outcome measure assesses the ultimate sequela of both diseases, tooth loss.

Table 4 shows the numerators and denominators for the use of services measures. Of the six measures, three are ratio measures comparing the provision of services that could be considered as alternative therapies. Two measures examine issues related to third molar extractions, one reporting the proportion of young adults who have one or more third molars removed, and the other reporting the mean number of third molars removed among those young adults receiving this service. The one measure remaining from the preliminary set is a traditional assessment of the proportion of enrollees receiving a prophylaxis.

Table 5 presents the values obtained in the pilot testing of the measures. For some measures, disease status categories have been combined. The reporting year—i.e., the year for which the measures were calculated—was calendar year 1997. Plan A had not implemented the reason-for-treatment codes in time for all 1997 restorative procedures to be associated with a reason for treatment, so values for the third measure, which is dependent on

FIGURE 1
Structural Relationship of Effectiveness of Care Measures

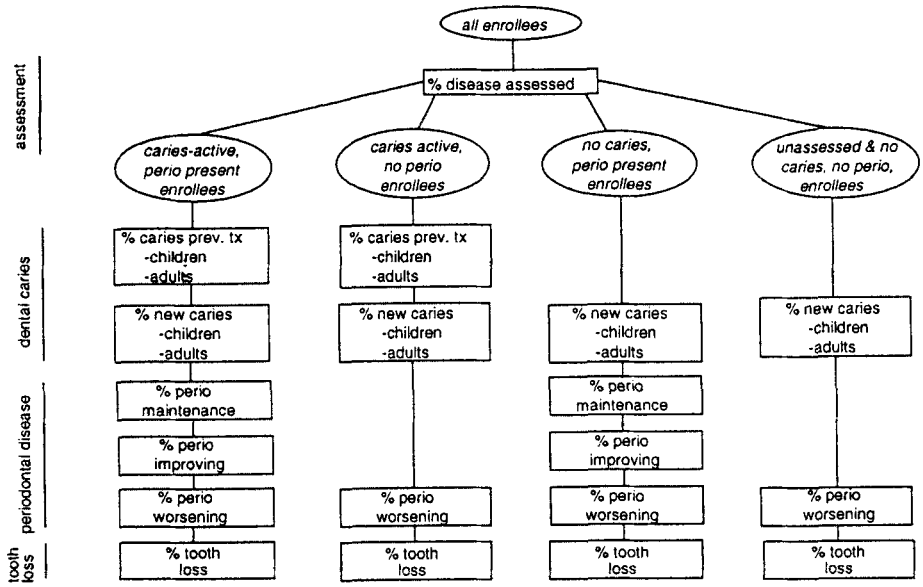

TABLE 4
Final Use of Services Measures

1. Receipt of Prophylaxes	
All enrollees in denominator who receive prophylaxis during reporting year	
All* enrollees with continuous enrollment for entire reporting year	
2. Preventive Treatment : Restorative Treatment Ratio	
Total* number of preventive procedures provided during reporting year	
Total* number of direct restorative procedures provided during reporting year	
3. Casting : Large Direct Fillings Ratio	
Total* number of casting procedures provided to adult enrollees during reporting year	
Total* number of large direct filling procedures provided to adult enrollees during reporting year	
4. Endodontic Treatment : Extraction Ratio	
Total* number of teeth treated endodontically during reporting year for adult enrollees	
Total* number of nonthird molar teeth extracted during reporting year for adult enrolllees	
5. Receipt of Third Molar Extractions	
Total* number of enrollees in denominator who receive at least 1 third molar extraction during reporting year	
Enrollees aged 16–24 (inclusive) at any time during reporting year	
6. Mean Number of Third Molars Extracted	
Total number of teeth extracted among enrollees in denominator	
Enrollees aged 16–24 (inclusive) at any time during reporting year who receive at least 1 third molar extraction	

*These measures may be calculated separately by caries and/or periodontal disease activity status.

this information, are missing. Neither plan included periodontal data, either probing depths or PSR scores, in their electronically based administrative data sets. The values appearing for these measures are based on follow-up chart audits of random samples of 402 and 102 adults, respectively, in plans A and B1. The values shown in the table represent the final results of several iterations of a calculate-and-review cycle that occurred for both plans. The proportions of adults classified as having "high caries activity" among those with assessments were 16 percent, 17 percent, and 18 percent for plans A, B1, and B2, respectively, with 29 percent, 25 percent, and 39 percent classified as having "periodontal disease present."

Table 6 shows the results of two assessments of the reliability of the data used to construct the measures in Plan A. Administrative data were in good agreement with chart data, with a majority of errors involving the separate specification of fluoride application in the administrative record, but not the patient chart. Dentists' agreement with a nominal standard examiner for caries and periodontal disease activity and for classification of the reason for restorative treatment was lower, perhaps reflecting the subjective nature of the criteria used for these classifications or the lack of formal calibration of the plan's dentists on these criteria (12).

Discussion

Our intent in asking stake holders to participate in the measure development process was both to gain some degree of "buy-in" through their involvement in the process, and to use their specialized knowledge of aspects of the dental plan industry to tailor the measures' focus and specifications to this environment. We used a modified Delphi process to obtain the feedback (13), and found that this approach was effective in garnering useful suggestions. Throughout the process, stake holders' comments guided, but did not dictate, the refinement effort. Due to the nature of the development effort, we relied on the steering committee, composed of dental plan executives, in the early stages of the refinement effort, which concentrated on the effectiveness of care measures. However, as development proceeded, the input from the advisory committee as-

TABLE 5
Pilot-test Values for Performance Measures by Plan

Performance Measure	Plan A	Plan B1	Plan B2
Effectiveness of care			
Proportion with disease status assessed	.64	.54	.26
Proportion with appropriate preventive tx for high caries			
Children	.81	.73	.69
Adults	.13	.11	.18
Proportion with new caries			
Caries-active children	*	.29	.17
Caries-inactive children	*	.11	.03
Caries-active adults	*	.27	.28
Caries-inactive adults	*	.13	.07
Proportion with appropriate preventive tx for high perio	.81	.79	.86
Proportion with deterioration in perio status	.50†	.29†	*
Proportion with improvement in perio status	.41†	.25†	*
Proportion with tooth loss			
Children	.04	.06	.01
Adults	.05	.05	.02
Use of services			
Proportion receiving prophylaxis	.57	.62	.34
Castings : large direct fillings ratio	.14	.93	.65
Preventive : restorative services ratio	1.8	3.1	4.3
Endodontic treatment : extraction ratio	.47	.72	.51
Proportion receiving third molar extraction	.02	.03	<.01
Mean number of third molars extracted	2.1	3.6	3.7

*Data to calculate these measures are unavailable.
†These values determined through chart audit.

TABLE 6
Reliability Assessments of Administrative Data Used to Calculate Performance Measures

	%	Kappa Score
Percent distribution of matches between patient chart and administrative data set*		
Precedures and date identical	95.3	
Procedures identical, but dates different	0.9	
Procedures in data base, but not in chart	3.9	
Procedure in chart, but not in data base	0.0	
Percent agreement of dentists' classifications with those of nominal standard examiner†		
Caries activity classification (n=66)	76	0.56
Periodontal disease classification (n=66)	83	0.70
Reason for treatment classification (n=73)	74	0.69

*n=1,170 patient visits.
†Ref 12.

sumed more importance. This committee's comments and recommendations concerning both types of measures reflected a depth of scrutiny similar to that shown by the steering committee, but with differing emphases. While the steering committee's concerns emphasized the issues of validity, accuracy, and future likelihood of use of the measures (Table 2), the advisory committee's comments focused on the measures' accuracy, sensitivity, and interpretability. Both types of perspectives benefited the development effort.

More substantial changes were made in the preliminary use of services measures than in the effectiveness of care measures. For the latter, two of the preliminary measures were dropped, targeting of specific ages for preventive assessments was eliminated, and definitions of several key terms were narrowed. For the former, the number of measures was reduced from 10 to six, and only one of these was represented in the original set. The difference in the extent of change in the two sets of measures might be related to the face validity of the concepts addressed by the effectiveness of care measures. The application of preventive measures to those at highest risk of disease, and the actual rates at which disease is experienced are obvious choices for evaluating the effectiveness of care, especially when the bulk of dental care is provided to prevent, manage, or repair the sequelae of two diseases (14). In contrast, selecting the "right" use of services measures is more difficult not only because dental care includes so many types of services, but also because the effectiveness of the various services is often uncertain, especially in the absence of diagnostic information. Thus, in three instances, committee members' comments resulted in the replacement of simple distributions of services with measures that contrasted alternative treatment approaches. These ratio measures identified by the development process tend to provide information that reflects the "practice style" of the providers in a plan, which in turn may be influenced by the plan's benefit structure. Although there is insufficient knowledge of the outcomes of dental care to permit any supported statements about what ratio values are preferred, the measures do permit purchasers to gain some insight into the extent to which specific treatment approaches are emphasized in particular plans.

Programmers at both plans initially produced calculated values by following the manual of procedures with a

minimum of calls for clarification. However, these initial calculations contained a number of obvious errors in the denominators. Working with the programmers and other staff to obtain consistent denominators across measures led to the identification of specification errors in several numerators. Although no attempt was made to fully analyze the reasons for these errors, several were the result of defining the numerator and denominator of a rate independently, rather than specifying the numerator as a subset of the denominator. Another reason for errors specific to one plan was related to the manner in which historical data were stored, which necessitated the creation and eventual merging of multiple data sets. A clear implication of this experience is that the programming initially created for a dental plan's first calculation of performance measures is likely to be flawed, perhaps because the instructions and criteria were not clear, or because the programmer was unfamiliar with the types of calculations or data system manipulations required. Intensive logic and error checking, independent parallel development and comparison by other programmers, and program auditing should all be considered as means to ensure that the measures are being calculated correctly.

The results shown in Table 5 must not be taken as criterion standards. These are the first such performance data to be reported for any dental plan. Thus, some might be tempted to use them as benchmarks for comparison with other plans. However, without broader knowledge of performance levels across various types of managed dental delivery organizations, reference to these values may set unrealistically high or low standards. The values for the effectiveness of care measures do have implied goals of zero and/or 100 percent; however, actual operational or achievable ranges have yet to be determined. Values for the use of services measures do not have implied goals, and should not be ranked on a "good-bad" continuum.

The pilot test does provide partial evidence that the measures are reliable and sensitive to differences among plans. Using the patient chart as a comparison, administrative data overreported services by 4 percent in the one system in which a systematic comparison was made. While not perfect, this level of agreement is good for such systems (15,16). The further observation that the majority of overreported services traditionally were included in bundles that were automatically disaggregated by the data system suggests that administrative data may be a more accurate record of the actual services provided than the patient chart. The reliabilities of the disease activity classifications and the reason-for-treatment classification were lower, but fell in the upper part of the range of typical values for agreement among dentists for subjective diagnostic decisions (17). Hence, these simple classifications seem to be as reliable as is reasonable to expect.

Although the overall accuracy of the measures was not assessed directly in the pretest, there is informal evidence for one measure that the values reported do reflect the clinical situation. Personnel from both DHMOs were surprised by the low rates of fluoride treatment for caries-active adult enrollees, and at first both plans doubted the accuracy of these values. After reviewing clinical operations, however, personnel from both plans acknowledged that the values probably did reflect clinical practice. This anecdotal observation illustrates the potential offered by these measures for internal quality improvement activity.

The values for some measures reflect known or expected differences between the plans, suggesting that the measures are sensitive, i.e., that they capture differences in plan design and practitioner behavior. For example, clinical protocols in both plan A and plan B1 encourage disease activity assessment for all enrollees, whereas plan B2 protocols do not. Plan B2 assessment might be still lower if the same practitioners did not see patients from both B1 and B2 plans. Plan A has a higher copayment rate for cast restorations than plans B1 and B2, and strict protocols limiting cast restorations to teeth with healthy periodontal and pulpal tissues. The effect of these differences is evident in the ratios of castings to large direct restorations.

The decision not to define criteria for caries-active and periodontal disease-active individuals stems from the current lack of validated criteria for these classifications, as well as known differences in caries experience among regions, among plans within a region and even among clinics within a local plan. One plan's high caries level might be another plan's mean level. Thus, we believe that the classification criteria should be set by individual plans. We have designed the measure set to discourage "gaming the system" by deliberately limiting or expanding the pool of high-disease-activity individuals. If a plan excludes enrollees with moderately high disease activity, the plan's rates for new disease among nonhigh-disease enrollees will increase. If a plan deliberately includes enrollees with relatively low rates of disease activity in the high-activity classification, the plan will either reduce the rate at which it reports appropriate prevention for high disease enrollees, or suffer the economic consequences of provision of unnecessary preventive treatment.

These clinical performance measures, and especially the effectiveness of care measures, represent the second generation of such measures. The first generation is characterized by two types of measures, unstandardized process measures reflecting the use of specific services, and epidemiologic measures of dental disease that require clinical examination for collection (6). This new generation of measures offers several advantages that should help promote their implementation. They can be calculated directly from a dental plan's administrative data system, thereby minimizing data collection costs and related recording errors. They are standardized through complete documentation, facilitating comparisons across plans. They include a means of risk adjustment to account for differing oral disease status among enrollees of different plans. They encourage initial assessment of all enrollees both by measuring the assessment rate, and by classifying all unassessed enrollees as having low disease activity, which will have the effect of inflating disease experience rates among "low disease activity" enrollees. They also avoid the misclassifications of new caries that inevitably accompany use of the DMF index (18).

Despite these advantages, it is certain that these measures will not be implemented immediately by most dental care plans, for no other reason than because their administrative data systems do not contain the necessary information and the plans choose not to invest in the systems unless and

until asked by the purchasers. Diagnostic information and disease activity and disease risk classifications are not common components of plan data. Nevertheless, the rationale for developing the measures in the light of this limitation is clear. Diagnostic information—including both procedure-linked diagnostic codes or reasons for treatment as well as current disease status information—is absolutely necessary if valid assessments of a plan's effectiveness are to be made. The presence of these measures may stimulate movement toward adoption of administrative systems that do provide such information. In truth, without knowledge of disease status and activity, plans cannot "manage" care appropriately, nor be compared fairly.

The path toward inclusion of diagnostic information in administrative data systems is an uncertain one; nevertheless, the forthcoming release of a set of diagnostic codes by the American Dental Association should help facilitate this process. Pressure from purchasers for performance information exemplified by the measures described here should also help drive adoption. But a true transformation to informed management of care will require dental care plans to decide that they cannot afford not to have detailed performance information for reasons of quality assessment and improvement.

Acknowledgments

We thank Paul Cheek, Mike Goodman, Sally Jo Little, Oralee Olson, Bill Rush, and Jeff Showell for their assistance in obtaining and analyzing the data; Jamie Allison, John Kopacz, and Proctor & Gamble for assistance with steering committee arrangements; and the members of the steering and advisory committees for their willingness to participate in the project.

References

1. National Association of Dental Plans. 1998 dental HMO/PPO industry profile. Dallas, TX: NADP, 1988.
2. Capilouto E. Market forces driving health care reform. J Dent Educ 1995;59: 480-3.
3. Mulvihill J. Insights on a new era under a reforming health care system. J Dent Educ 1995;59:620-7.
4. Bader J, Tresolini C, Shugars D. Benefits managers' expectations for dental care plans [Abstract]. J Dent Res 1997;76:207.
5. Harris L. Smile; dental benefits industry. HRMagazine 1997;42:110-12.
6. Bader J, Shugars D, Hayden W, White B. A health plan report card for dentistry. J Am Coll Dent 1996;63:29-38.
7. National Committee for Quality Assurance. HEDIS® 3.0. Vol 1. Narrative. Washington, DC: NCQA, Jan 1997.
8. Bader J, Shugars D. Variation, treatment outcomes, and practice guidelines in dental practice. J Dent Educ 1995;59:61-96.
9. Bader J, Shugars D. A case for diagnoses. J Am Coll Dent 1997;64:44-6.
10. Corrigan J, Nielsen D. Toward the development of uniform reporting standards for managed care organizations: the health plan employer data and information set. Version 2.0. J Qual Improve 1993;19:366-73.
11. Orlowski R. Dental diagnosis codes. Winston-Salem, NC (mimeo).
12. Bader J, White A, Olson O, Senft G, Shugars D. Clinician reliability in classifying disease risk and reasons for treatment [Abstract]. J Dent Res 1999;78:168.
13. Fink A, Kosecoff J, Chassin M, Brook R. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984;74:979-83.
14. Hayden W. Dental health services research utilizing comprehensive clinical databases and information technology. J Dent Educ 1997;61:47-55.
15. Kashner T. Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. Med Care 1998;36:1324-36.
16. Steinwachs D, Stuart M, Scholle S, Staffield B, Fox M, Weiner J. A comparison of ambulatory Medicaid claims to medical records: a reliability assessment. Am J Med Qual 1998;13:63-9.
17. Bader J, Shugars D. Variation in dentists' clinical decisions. J Public Health Dent 1995;55:181-8.
18. Bader J, Shugars S, Rozier R. Relationship between epidemiologic coronal caries assessments and practitioners' treatment recommendations in adults. Community Dent Oral Epidemiol 1993;21:96-101.