Community Dent Oral Epidemiol 2008; 36: 466–474

COMMUNITY DENTISTRY AND ORAL EPIDEMIOLOGY

Concordance of chart and billing data with direct observation in dental practice

Catherine A. Demko, Kristin Zakariasen Victoroff and Stephen Wotman

Department of Community Dentistry, Case Western Reserve School of Dental Medicine, Cleveland, OH, USA

Demko CA, Victoroff KZ, Wotman S. Concordance of chart and billing data with direct observation in dental practice. Community Dent Oral Epidemiol 2008; 36: 466–474. © 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

Abstract – *Objectives:* The commonly used methods of chart review, billing data summaries and practitioner self-reporting have not been examined for their ability to validly and reliably represent time use and service delivery in routine dental practice. A more thorough investigation of these data sources would provide insight into the appropriateness of each approach for measuring various clinical behaviors. The aim of this study was to assess the validity of commonly used methods such as dental chart review, billing data, or practitioner self-report compared with a 'gold standard' of information derived from direct observation of routine dental visits. *Methods:* A team of trained dental hygienists directly observed 3751 patient visits in 120 dental practices and recorded the behaviors and procedures performed by dentists and hygienists during patient contact time. Following each visit, charts and billing records were reviewed for the performed and billed procedures. Dental providers characterized their frequency of preventive service delivery through self-administered surveys. We standardized the observation and abstraction methods to obtain optimal measures from each of the multiple data sources. Multi-rater kappa coefficients were computed to monitor standardization, while sensitivity, specificity, and kappa coefficients were calculated to compare the various data sources with direct observation. *Results:* Chart audits were more sensitive than billing data for all observed procedures and demonstrated higher agreement with directly observed data. Chart and billing records were not sensitive for several prevention-related tasks (oral cancer screening and oral hygiene instruction). Provider self-reports of preventive behaviors were always over-estimated compared with direct observation. Inter-method reliability kappa coefficients for 13 procedures ranged from 0.197 to 0.952. *Conclusions:* These concordance findings suggest that strengths and weaknesses of data collection sources should be considered when investigating delivery of dental services especially when using practitioner survey data. Future investigations can more fully rely on charted information rather than billing data and provider self-report for most dental procedures, but nonbillable procedures and most counseling interactions will not be captured with routine charting and billing practices.

Key words: concordance; dental charts; direct observation; validity

Prof. Catherine Demko, 10900 Euclid Ave., Cleveland, OH 44106, USA.
Tel: +1 216 368 8804
Fax: +1 216 368 5888
e-mail: cad3@case.edu

Submitted 18 January 2007;
accepted 15 May 2007

Valid data measuring what actually happens in dental practice is critically important to practitioners, policy makers, and researchers, serving as a reference point for important decisions involving individual practices, dental public health policy, and dental insurance performance measures. Accurate measures of the type and frequencies of services delivered in the practice setting are useful for quality improvement assessment, practice improvement goals, comparison with professional educational objectives, and benchmarking for practitioners performing self-assessment. Data for these purposes are generally obtained from dental claims data, chart reviews, national surveys of dentists'

doi: 10.1111/j.1600-0528.2007.00415.x

self-reported activities, or patient report of services received (1–3). There is little empirical evidence, however, supporting the reliability or validity of these sources for measuring what has actually transpired in the provider–patient interaction. Further, the accuracy and completeness of these sources comes into question when they are used for evaluation, quality improvement or monitoring of service delivery because each method is subject to certain limitations (4).

Dental claims and self-report data have been previously used as single-method data sources in productivity and time studies to estimate workload requirements and efficiencies in service delivery (5, 6). Frequency distribution of dental services and procedures are routinely collected from surveys of dental professionals, but self-report is subject to both over- and under-reporting (7–9) and has not been compared with any other data source. Gilbert et al. (10) concluded that the validity of self-reported dental care from patients ranged from poor to excellent depending upon the type of service when compared with dental charts as the 'gold standard'. One explanation for differences by service type may be that dental charts themselves are subject to recording biases, missing data, and variable charting behaviors among providers and may also reflect a bias toward billable procedures. Direct observation was first used in the dental setting in a Swedish study with trained observers recording the frequency but not the duration of procedures in public health dental clinics (11). However, this study did not compare the direct observation findings to other data sources. Thus, a comparison of multiple data sources in a single study has not been performed for dental care.

Direct observation as the gold standard has been used in numerous studies in medical practice to validate information from multiple data sources including assessing the delivery of preventive services (12), emergency room services for asthma care (13), and time–motion studies in rural health centers (14, 15). Despite its cost and potential intrusiveness, direct observation may be an appropriate standard to establish the validity and inter-method reliability of common data sources such as chart audit (CA), dental claims and self-reported behaviors.

We undertook the Direct Observation Study of Dental Practices (DOS) to better understand the content and context of dental visits, especially concerning the delivery of preventive services, using a multi-method approach to data collection. Within the larger DOS study, this report compares the content of directly observed patient visits with data from chart reviews, billing data and provider self-reported behaviors, to ascertain the concordance among various methods and compute the sensitivity and specificity of chart and billing data when compared with direct observation as a gold standard. Comparison of these multi-method approaches will provide insight into the appropriateness of each approach for measuring various clinical behaviors.

Methods

Practice and patient characteristics

The Ohio Practice-Based Research Network, formed in 1998 (16) was expanded in 2004 with the funding of the DOS and the creation of CROWN, the Community Research for Oral Wellness Network. Dentists on the state of Ohio's licensure list who practiced within a 100-mile radius of the Case Western Reserve University campus in Cleveland received letters of invitation to participate in the DOS. Recruitment letters outlined the multi-method data collection from various sources to characterize the dental visit, but did not divulge the specific aims of the study. Out of 2500 invitations, 166 general practice dentists volunteered, and we enrolled 120 offices based on power calculations to test the main hypotheses of the overall study. Information gathered from study decliners who completed a recruitment survey (n = 306) revealed no substantial differences in age, gender, or practice years between those dentists who participated and those who did not, but participating offices were more likely to employ a hygienist. Sixteen percent of practices were located in five urban areas, another 16% in rural areas as defined using Census 2000 data (17) and the remainder in suburban settings. The study coordinator scheduled practice visitation at mutually convenient times and requested that patients be scheduled in the usual manner for both dentists and hygienists. Data were collected on 4 days in each practice between June 2004 and September 2005. In total, 3751 patient visits were observed and 92% of the patients approached about the study agreed to participate.

Multi-method data collection and instruments

Two trained observation teams each included two dental hygienists, with at least 10 years of professional experience, and a patient coordinator. One

hygienist observer followed the dentist and the other followed the office dental hygienist (DH) in each practice. Observers were unobtrusively positioned in the dental operatory to both see and hear visit interactions but not to participate in any way. The patient coordinator remained in the waiting room to recruit patients and obtain informed consent in accordance with human subjects' protection and privacy requirements. In addition, signed informed consent was obtained from participating dentists and hygienists. The Case Institutional Review Board approved the study.

Three data collection methods, *direct observation*, *self-administered surveys*, *and data abstraction* were used to collect data using the following instruments: (1) *direct observation* of dentist and hygienist patient visits used the dental DOC, a modified version of the medically oriented Davis Observation Code, which prompted the recording of 24 practitioner behaviors at 30-s intervals (9) and the postvisit checklist (PVC) for recording the occurrence of dental procedures from among a list of 65 procedures; (2) s*elf-administered surveys* completed by patients, dentists, and hygienists; (3) *abstraction* of chart information (chart audit, CA) and billing records (BR) from each observed patient visit. Observed and abstracted data were entered directly into a tablet computer, while the paper-based surveys were optically scanned (18).

A predefined list of 65 items on the PVC and CA were identical to permit comparison of the observed occurrence and the charted occurrence of behaviors and procedures. The 65 procedures included and extended the 25 most-frequent procedures reported from a dental workload study which defined 246 tasks/procedures (19). The procedure definitions were reviewed by the network steering committee members and tested in practices during the initial pilot study (16). Modifications to the procedure list and definitions were made during consensus review by study investigators, network steering committee members, dentists in the pilot study and input from the observers during the training period. Billing codes were abstracted from patient BRs and translated using the Common Dental Terminology billing codes, versions 4 and 5 (CDT-4 and CDT-5) to identify the billed procedures. From abstracted codes, 7% were internal office codes with no associated charges and 2% were erroneous codes that could not be identified. Seventy-four charts were unavailable for abstraction. Both dentists and hygienists completed surveys that included demographic information, education, years in practice, self-assessed service delivery practices and attitudes regarding preventive services.

Training and standardization for data collection

The observation teams completed a 4-week training period for the use of the direct observation instruments prior to the start of the study, using videotaped patient encounters, followed by field experience in the dental school faculty practice and offices of the Network Steering Committee that would not be part of the DOS. The entire methodology was reviewed by a National Advisory Committee consisting of experts in health services research who offered suggestions and comments on validity early in the collection process.

Observer variability was minimized by repeated standardization of observers through the use of videotapes of routine dental encounters in practice situations. The four research hygienists remained for the entire data collection period. For inter-rater reliability, the four hygienists watched video-taped patient/dentist encounters and individually scored these encounters using a preset list of behaviors adapted from the Davis Observation Code (9). These standardizations were repeated at 4- to 6-week intervals throughout data collection, in addition to monthly debriefing sessions to discuss on-going data collection issues. We calculated multi-rater kappa coefficients from 13 videotape segments totaling 181 min using STATA (20). Kappa coefficients for codes common to both the DOC and PVC were treatment planning, 0.74; oral examination, 0.92; preventive procedures (fluoride treatment and sealants), 0.83; oral health instruction, 0.77; and oral cancer examination, 0.69. Reliability of defining and recognizing standard dental procedures (e.g. extractions, amalgams, and crowns/bridges) was assured by the experience and retention of the observer hygienists and discussions of definitions at each debriefing and standardization session. Inter-rater reliability of items abstracted from CAs was determined by having all six team members perform chart abstraction on the same sets of charts from a nonparticipating private office (multi-rater kappa coefficient for all procedures was 0.87).

Concordance of multi-source data measures

To measure the ability of chart and billing data to capture the occurrence of common dental procedures, we calculated the sensitivity and specificity of these methods by comparing them with the

'true' occurrence of the procedures as recorded on the PVC during direct observation (the gold standard). Sensitivity is the probability that the chart (or BR) includes the procedure when it was performed (as detected by direct observation), while specificity is the probability that the chart (BR) correctly excludes the procedure when it was not performed by the dental provider. High sensitivity indicates that the data source is likely to accurately report the occurrence of performed procedures, while low sensitivity would suggest the source will underestimate the delivery of the service. The probability that a procedure was performed, given its presence in the chart (or BR), is reported as the positive predictive value (PPV). Similarly, the negative predictive value (NPV) is the probability that a negative chart (or BR) correctly identifies a visit that did not include the procedure. Calculating predictive values provides information for how accurately the data sources measure the procedures when direct observation is not available for comparison. The predictive values, however, are influenced by the underlying prevalence rate of the procedure.

To quantify agreement between the abstracted data and direct observations, a kappa statistic was calculated for the same common dental procedures. The kappa statistic quantifies the extent to which the observed agreement is beyond chance (21). kappa values greater than 0.75 are considered to represent excellent agreement beyond chance, values 0.40–0.75 represent moderate agreement, while values below 0.40 represents poor agreement (22).

After the completion of the observation visit, providers were asked to characterize their preventive service delivery practices on a scale of general frequency (almost always, often, sometimes, rarely, and almost never). For each provider, we calculated the percent of observed patients who received preventive services. Because we could not compare delivery of observed services with self-report on an individual patient basis, we compared the mean percent of patients who received preventive services within three combined categories of dentist and hygienist self-reported frequency of delivery (almost always/often; sometimes; and rarely/almost never). We also compared the overall observed, charted and provider-reported percentages of patients who received each of the following six services; oral hygiene instruction, oral cancer screening examination, caries susceptibility examination, bacteriostatic mouthwash prescription, diet/nutrition discussion, and smoking discussion.

Calculations of means and percents (sensitivity, specificity, and predictive values) were performed in SPSS (23), while multi-rater kappa coefficients were obtained using STATA v8 (20).

Results

Participating dentists were similar in distribution to all Ohio dentists in terms of age (mean ± SD; 49.0 ± 9.7 years) and years in practice (22.1 ± 9.8 years), while female dentists were slightly over-represented in the study (16.8%) compared with the state (12%). For the observed patients, the distribution of ages mirrors the US population, with almost as many patients over 65 as children 5–18 years (15% in each category) (24). Eighteen percent of the observed dentist visits were for acute care (emergency visits), 27.5% for rehabilitative procedures (crown/bridge, partial/full dentures) and 54.5% for primary care (all other procedures) (25).

From the dentist and hygiene visits combined, we observed 13 073 occurrences of the procedures on the PVC, over half of which were accounted for by the 13 unique procedures included in Tables 1 and 2. The frequency of these directly observed procedures was compared with the frequency obtained from CA and BR abstraction to compute the sensitivity, specificity, kappa coefficients and predictive values shown in Table 1 for the dentist visit and Table 2 for the hygiene visit. The frequency of services recorded by the observer was higher than the frequency recorded from the CA or billing data for all procedures and measures of validity and agreement varied based on the specific dental service. Sensitivity was high for single-visit procedures such as extractions, amalgams, and composites restorations (81.5–93.7%) but low for counseling or screening behaviors such as oral hygiene instruction (21.9%) and oral cancer examinations (44.3%). For multi-appointment prosthetic procedures, sensitivity of CAs was moderate, but BRs were very insensitive to these types of procedures, as might be expected (19–27%). Sensitivity of chart and billing data was generally higher among all hygienist-provided procedures, except oral cancer screenings and oral hygiene instruction. Overall, CAs showed higher levels of sensitivity than BRs when compared with direct observation.

Specificity was high for all dentist procedures (range 94–100%) resulting in very few false

Table 1. Validity and reliability of directly observed common procedures with abstracted data from chart audits (CA) and billing records (BR)

DDS procedures	Sensitivity[a] (%)	Specificity[b] (%)	Kappa	PPV[a] (%)	NPV[b] (%)
Oral examination (717)					
CA	60	94	0.57	82	82
BR	42	96	0.44		
Oral cancer screening (110)					
CA	44	99	0.49	70	97
BR	0	100	0		
Amalgam restoration (158)					
CA	81	99	0.81	83	98
BR	78	99	0.79		
Composite restoration (506)					
CA	89	95	0.83	85	97
BR	79	96	0.77		
Partial denture (103)					
CA	75	98	0.69	69	97
BR	21	99	0.31		
Full denture (118)					
CA	68	99	0.76	92	96
BR	30	99	0.40		
Prophylaxis (139)					
CA	87	95	0.65	55	99
BR	79	96	0.65		
Single/multiple crown (368)					
CA	86	97	0.83	87	97
BR	53	97	0.66		
Extraction (110)					
CA	94	99	0.90	89	99
BR	84	99	0.86		
Oral hygiene instruction (64)					
CA	22	99	0.20	37	98
BR	3	99	0.05		

The number of procedures observed is in parentheses after the procedure name.
[a]Sensitivity is the extent to which an observed service was also recorded in the chart or billing data. PPV is positive predictive value which is the percent of charted or billed data that was performed.
[b]Specificity is the extent to which a service that was not observed was also not recorded in the chart or billing records. Negative predictive value (NPV) is the percent of negative charts or billing records in which the procedure was not performed.

Table 2. Validity and reliability of directly observed common procedures in the hygiene visit with abstracted data from chart audits (CA) and billing records (BR)

DH procedures	Sensitivity[a] (%)	Specificity[b] (%)	Kappa	PPV[a] (%)	NPV[b] (%)
Oral examination (1366)					
CA	93	72	0.48	98	41
BR	81	76	0.25		
Oral cancer screening (682)					
CA	41	89	0.31	77	63
BR	0	100	0		
Prophylaxis (1374)					
CA	97	87	0.70	99	62
BR	83	96	0.33		
Radiographs (625)					
CA	90	89	0.79	86	92
BR	81	92	0.74		
Fluoride treatment (328)					
CA	89	97	0.87	91	97
BR	80	98	0.81		
Sealants (11)					
CA	91	100	0.95	100	99
BR	73	100	0.84		
Oral hygiene instruction (553)					
CA	50	75	0.26	55	71
BR	5	96	0.01		

The number of procedures observed is in parentheses after the procedure name.
[a]Sensitivity is the extent to which an observed service was also recorded in the chart or billing data. PPV is positive predictive value which is the percent of charted or billed data that was performed.
[b]Specificity is the probability that a nonperformed service is correctly absent from the chart or billing record. Negative predictive value (NPV) is the percent of negative charts or billing records in which the procedure was not performed.

positives; if a service was not performed, it was not charted or billed. Specificity for the hygiene-provided procedures had a wider range, 72–100%; specificity for the oral examination CA (72%) and oral hygiene instruction CA (75%) were slightly lower possibly related to more stringent definitions of these procedures in the study than practicing hygienists use when recording tasks in the chart.

Kappa values, representing chance-corrected agreement between the data sources and direct observation, were generally moderate to high particularly for charted information. Lower agreement was seen between billing data and observation for multi-appointment procedures, such that the visit we observed was not the visit that generated the final total billing. Not surprising, kappa values between observation and billing for nonreimbursable procedures were extremely low, so that claims data would not adequately reflect the level of these frequent services.

Calculating predictive values provides an estimate of how accurately the data sources measure the procedures when direct observation is not available for comparison. Overall, the PPV of the CA were very good except for prophylaxis by the dentist and oral hygiene instruction by the dentist or hygienist. The predictive values, however, are influenced by the underlying prevalence rate of the procedure as well as the specificity of the method (CA or BRs). For example, the PPV falls but the NPV rises as prevalence increases, demonstrated by the higher PPV but lower NPV of CAs for oral cancer screenings among hygiene visits compared

Table 3. Mean percent of patients (±SD) who received an observed preventive service within response categories that represent how often the provider self- reported delivering that service

	Categories of provider self-reported service delivery		
	Always/often	Sometimes	Rarely/never
DDS preventive behaviors			
Oral cancer screening	16.0 ± 14.7 (98)[a]	1.2 ± 1.7 (2)	0
Smoking discussion[b]	8.6 ± 13.6 (25)	11.1 ± 14.5 (37)	10.1 ± 15.1 (38)
Nutrition discussion	3.9 ± 5.8 (28)	3.4 ± 3.6 (43)	2.2 ± 4.0 (30)
Oral hygiene instruction	20.2 ± 13.7 (98)	5.6 ± 3.3 (2)	0
DH preventive behaviors			
Oral cancer screening	25.3 ± 33.7 (93)	1.1 ± 2.7 (6)	0 (1)
Smoking discussion[b]	27.1 ± 29.0 (47)	26.6 ± 36.8 (39)	40.3 ± 41.7 (14)
Nutrition discussion	13.7 ± 13.4 (46)	6.2 ± 9.8 (36)	4.2 ± 9.2 (18)
Oral hygiene instruction	79.1 ± 19.0 (99)	75.0 (1)	0

[a]Numbers in brackets are percent of providers who responded in that category. For example, 98% of dentists reported providing oral cancer screening always/often, yet only 16% of those dentists' patients received an oral cancer exam.
[b]Discussion about tobacco use was assessed only among patients who self-reported as current users or recent quitters (n = 651).

with dentist visits when both have comparable sensitivities and specificities. Similarly, the lower specificities of CAs for oral examinations and oral hygiene instruction among the hygiene visits reduced the NPV for those procedures.

Table 3 shows the percent of observed patients receiving preventive-related services by category of provider self-report. After placing providers into three categories based on their self-report of frequency of preventive service delivery (always/often, sometimes, rarely/never), we calculated the mean percent of patients who actually received the specific service from those providers. For example, for discussions about diet or nutrition, 28% of dentists reported always or often talking about this subject to patients, yet on average, only 3.9% of these dentists' patients received nutrition counseling in the observed visits. For some services (oral cancer screening, nutrition discussion and oral hygiene instruction), a gradient of service delivery paralleled self-reported general frequencies, i.e. providers who self-reported more frequent delivery of preventive services did provide those services to a greater proportion of their patients compared with providers who estimated lower frequencies by self-report. For example, directly observed nutrition discussions were most frequent (13.7%) among DHs who were in the always/often category, compared with the sometimes category (6.2%) and the rarely/never category (4.2%). Overall, however, self-reported frequency categories overestimated the actual service delivery. Such overestimation occurred even for at-risk populations such as smokers, where comparisons between self-report and direct observation about tobacco use were considered only among current and recent smokers (n = 651) rather than all observed patients.

Finally, we compared the three data collection methods for six preventive services. Fig. 1 shows the percent of patients who received any of the six services from either the dentist or hygienist as documented from direct observation (observed) and chart review (charted) compared with the percent of providers who indicated they provided the procedures almost always or often to their patients (self-report). Thus, if 80% of the observed

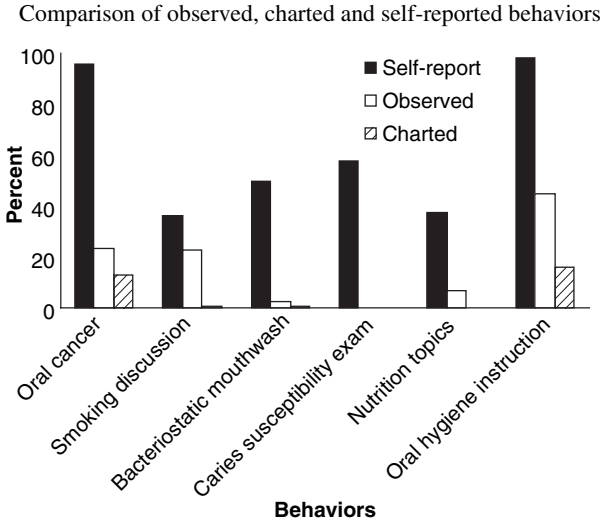

Fig. 1. Percent of patients receiving preventive services from observed and charted data. Self-report represents the percent of providers who indicated that they 'almost always' or 'often' provided that service. For smoking discussion, eligible patients were restricted to current smokers or those who had quit within the past year.

providers (who did in fact account for approximately 80% of all observed patients) indicated that they always/often render a particular service, then we would have expected to see close to 80% of the total observed patients receiving that service. As shown, strikingly fewer patients were observed or charted as receiving these services. For all categories, self-report frequencies exceeded observed and charted behaviors.

Discussion

We directly observed 3751 patient visits in 120 dental practices to assess the validity of commonly used methods such as dental chart review, billing data, or self-reports of practitioners compared with the 'gold standard' of direct observation. The dentists observed were similar in age, gender, and practice years to those on the State of Ohio registration list and the presenting case mix included a wide range of procedures. We standardized the observation and abstraction methods to obtain optimal measures from each of the multiple data sources. Prior to this study, multiple sources of routinely collected dental data had not been compared with direct observation for reliability or validity. We found that compared with direct observation, charted information and BRs provided a reasonable but generally lower estimate of services that are single visit and reimbursable, but substantially underestimate other nonprocedure routine services, particularly those related to prevention. Provider self-reported estimates of performing six selected preventive-related services overestimated the delivery of those services, similar to reports in the medical literature suggesting that self-report practitioner data consistently overstates services provided (7, 8).

Validity measures are best determined in samples that reflect the occurrence of the targeted condition (disease status or behavior) in populations where the screening test will be used (26). This study was conducted in community-based general dental practices, providing realistic estimates of these measures from practitioners' observed behaviors which included services with a range of delivery frequencies. We made no attempt to affect the data sources (i.e. the chart information or BRs), but rather optimized the observation methodology to provide the best possible, although not perfect, gold standard for comparison. Because the kappa values for inter-rater reliability demonstrated excellent or very good agreement among the four observers, the multiple observer approach did not negatively affect the multi-method concordance determination.

For procedures, some categories of the PVC were necessarily general in nature. For instance, oral examinations included any activity that was used to determine the oral health or status of a patient from simple mouth mirror examinations to Diagnodent evaluation, while endodontic treatment included procedures at any stage of treatment. Where there was specific interest (as in prevention) components were broken out and recorded separately as well, i.e. fluoride treatments, sealants, tobacco discussions, nutrition discussions, and oral cancer screenings.

Several findings of this study may be important to researchers investigating dental practice. The sensitivity of the CA and billing data varied based on the type of service when compared with direct observation, while specificity was high for all procedures. Hygiene visit CAs were generally more sensitive and had greater PPV than the dentist's charts. Possible explanations may be either longer visits allowing for additional charting time or hygienist training that emphasizes more consistent charting. Researchers can confidently measure procedures using methods where the sensitivity for those procedures was high, but must appreciate the influence of its prevalence and specificity when considering the predictive value of the data source for that procedure. Thus, for services that are delivered infrequently or not reliably charted (i.e. lack of documentation does not mean the service was not performed, such as counseling services), existing data sources may not be useful for accurately estimating service delivery. For these services, future investigation will require defining and standardizing the recording of those services.

While absolute estimates from self-report data were generally high, self-report and direct observation was often *relatively* similar, that is providers who estimated they were performing services at a higher level were observed doing so compared with their colleagues. However, tobacco discussions appeared to occur in approximately the same proportion of smokers regardless of the provider's self-reported behavior. High self-report may indicate that many providers are aware that provision of these preventive services is desirable and have the intention to perform them or believe they are currently doing so. Although this first report

comparing direct observation to self-report is limited by the global frequency estimates given by the provider, it suggests that further research on survey methodology used to collect self-report information from dental providers is warranted. Further, it will be important to determine where providers are along the pathway from awareness to intentions to action, if efforts to improve preventive service delivery are undertaken.

These concordance findings suggest that strengths and weaknesses of data collection methods must be considered when investigating delivery of dental services especially when using practitioner survey data. When chart abstraction is used as a baseline for practice improvement, for tracking trends in services, or for self-assessment, researchers and practitioners need to be certain that measured changes are real and not an artifact of improved charting. Charting practices may need to be enhanced and standardized among providers to obtain the most accurate estimate of any service when using chart abstraction, whether for research or practice improvement. The validity of survey data would be improved through the development of questions about dental services for specific patient populations with quantitative response variables or through the use of weekly return studies which use daily logs of provided services (27). Finally, questioning patients about receipt of services that are seldom charted may be useful (10, 12).

There is likely to be concern about selection bias in any Practice-Based Research Network based on nonrandom sampling of eligible practices. While this limits generalizability of findings until investigations are repeated in other geographic areas with other samples, studies of sufficient size can provide important data concerning what is actually happening in practice. Although self-selected, these practices appear to represent a wide range of dental offices based on their demographic profiles. The effect of direct observation on providers and patients is difficult to quantify and cannot be completely removed. Observed dentists and hygienists were unaware of the specific goals of the study, and while providers may have attempted to be on their 'best behavior', delivery of many services was not optimal for the recorded behaviors (e.g. prevention and comfort) (28) and visit time limitations did not permit providers to substantially alter their time with patients. Further, previous studies in family medicine showed little influence of observers' presence during observed visits (i.e. Hawthorne effect) (29). Finally, where chart and observed frequencies differed, other possible explanations include observers missing or misclassifying the procedure, misclassification of chart abstraction or charting of a task that the provider did not perform.

In conclusion, this comparison of data collected using multiple methods can be used to determine which data sources are best suited to collect different types of practice information depending on their purpose. Researchers using these data can decide which methods are best suited to study specific services. Primary and secondary preventive services appear to be more challenging to accurately assess through standard data collection methods than many other services. Because self-care knowledge is important to oral health (30) and is transferred to patients through preventive counseling in the dental care setting, appropriate methods to measure preventive counseling are critical to the evaluation of interventions designed to enhance the effectiveness of counseling on patient's self-care regimens.

Acknowledgments

We thank the following for their contributions to this study: James Lalumandier DDS, MPH, Joseph Sudano PhD, and Kristin Williams DDS, MPH for advice and counsel, data clerks Tom Parma and Nida Degyses, observers Marian Kofford, Kate Mingus RDH, Gail Perry RDH, Marlene Rodriguez RDH, Marianne Scherry RDH, Becky Slivka, and Joy Wiedemann; the National Advisory Committee consisting of Howard Bailit DDS, PhD, Duncan Neuhauser PhD, Edward Callahan PhD, Randall Cebul MD, PhD, Benjamin Crabtree PhD, Donna Homenko RDH, PhD, Kurt Stange MD, PhD, and Stephen Zyzanski PhD, and finally the dentists, hygienists, personnel and patients of the 120 practices who graciously allowed us to observe them in action. Supported by NIH/NIDCR grant no. R01DE015171.

References

1. American Dental Association SC. The 2001 survey of dental practice: income from the private practice of dentistry. American Dental Association, Chicago, IL; 2002.
2. Solomon E, Murray J, Dodge W, Redding S, Valenza J, Flaitz C et al. Scope of practice comparison: a tool for curriculum decision making. J Dent Educ 2006; 70:231–45.
3. Manning W, Bailit H, Benjamin B, Newhouse J. The demand for dental care: evidence from a randomized trial in health insurance. J Am Dent Assoc 1985; 110:895–902.

4. Bailit H. Health services research. Adv Dent Res 2003;17:82–5.
5. Parker W, Williams D, Mayotte R, James J, Mangelsdorff A. A model for dental workload measurement. Am J Public Health 1982;72:1022–7.
6. Wan T, Vanostenberg P, Salley J, Singley D, West J. A cost and production analysis of hospital dental care programs. Public Health Rep 1987;102:512–22.
7. Kronstrom M, Palmqvist S, Soderfeldt B, Eriksson T, Carlsson G. Congruence between self-reported and actually provided prosthodontic services among Swedish dentists. Acta Odontol Scand 1999; 57:9–15.
8. Conroy M, Majchrzak N, Silverman C, Chang Y, Regan S, Schneider L et al. Measuring provider adherence to tobacco treatment guidelines: a comparison of electronic medical record review, patient survey, and provider survey. Nicotine Tob Res 2005;7(Suppl. 1):S35–43.
9. Callahan E, Bertakis K. Development and validation of the Davis Observation Code. Fam Med 1991;23: 19–24.
10. Gilbert G, Rose J, Shelton B. A prospective study of the validity of self-reported use of specific types of dental services. Public Health Rep 2003;118:18–26.
11. Swedberg Y, Malmqvist J, Johnsson T. A method for dental care activities time study using observer monitored counting of frequencies. Swed Dent J 1993;17:155–63.
12. Stange K, Zyzanski S, Smith T, Kelly R, Langa D, Flocke S et al. How valid are medical records and patient questionnaires for physician profiling and health services research? A comparison with direct observation of patients visits Med Care 1998;36: 851–67.
13. McDermott M, Lenhardt R, Catrambone C, Walter J, Weiss K. Adequacy of medical chart review to characterize emergency care for asthma: findings from the Illinois Emergency Department Asthma Collaborative. Acad Emerg Med 2006;13:345–8.
14. Bryant M, Essomba R. Measuring time utilization in rural health centres. Health Policy Plan 1995;10: 415–22.
15. Hermida J, Nicholas D, Blumenfeld S. Comparative validity of three methods for assessment of the quality of primary health care. Int J Qual Health Care 1999;11:429–33.
16. Wotman S, Lalumandier J, Nelson S, Stange K. Implications for dental education of a dental school-initiated practice research network. J Dent Educ 2001;65:751–9.
17. US Census Bureau. Summary file 3 (SF 3). Available at: http://factfinder.census.gov/servlet/DTGeoSearchByListServlet?ds_name=DEC_2000_SF3_U&_lang=en&_ts=204408197025 (3 August 2007); generated by Joy Wiedemann; using American Factfinder.
18. Sudano JJ Jr, Kofford B, Wotman S. Using tablet PC's in dental practice research: technology, cost savings, and direct data entry ''on the go''. J Public Health Dent 2005;65:244–5.
19. Parker W, Schopper A, Mangelsdorff A, Cheatham J. Dental treatment time requirements of military personnel. Mil Med 1979;144:376–80.
20. STATA Corporation. STATA, release 8. College Station, TX: STATA Corporation; 2003.
21. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edn. New York: Academic Press; 1988.
22. Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33: 159–74.
23. SPSS Inc. SPSS version 13.0, Chicago, IL: SPSS Inc.; 2004.
24. US Census Bureau. Available at: http://www.census.gov/popest/national/asrh/NC-EST2005/NC-EST2005-02.xls (3 August 2007); generated by C. Demko; using American Factfinder.
25. Wotman S, Adelson R, Douglass C. Disaggregation of oral health care. A model for defining public and private responsibility. J Am Coll Dent 1993;60:24–9.
26. Ransohoff D, Feinstein A. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299:926–30.
27. Green L. The weekly return as a practical instrument for data collection in office based research. Fam Med 1988;20:182–4.
28. Demko CA, Victoroff KZ, Sudano J, Lalumandier J, Williams K, Abbe M et al. Direct observation of prevention counseling and procedures in dental offices. J Dent Res 2006;85(Special Issue A):0803.
29. Goodwin M. Using direct observation in primary care research – the Hawthorne effect: defining the nature and impact of the presence of research observers on patients and physicians in community family practice. Dissertation, Cleveland, OH: Western Reserve University, 2002.
30. Oral Health in America. A report of the surgeon general; 2000. Available at: http://www2.nidcr.nih.gov/sgr/sgrohweb/welcome.htm (3 August 2007).