Community Dent Oral Epidemiol 2006; 34: 1–10

Community Dentistry and Oral Epidemiology

Methods

Reducing the bias of probing depth and attachment level estimates using random partial-mouth recording

James D. Beck[1], Daniel J. Caplan[1], John S. Preisser[2] and Kevin Moss[1]

Departments of [1]Dental Ecology and [2]Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Beck JD, Caplan DJ, Preisser JS, Moss K. Reducing the bias of probing depth and attachment level estimates using random partial-mouth recording. Community Dent Oral Epidemiol 2006; 34: 1–10. © Blackwell Munksgaard, 2006

Abstract – *Objectives:* To evaluate the bias and precision of probing depth (PD) and clinical attachment level (CAL) estimates of random and fixed partial examination methods compared with full-mouth examinations. *Methods:* PD and CAL were calculated on six sites for up to 28 teeth (considered to be the gold standard with no bias) and three fixed-site selection methods (FSSMs) that resulted in a partial examination of sites: the Ramfjord method, and the NIDCR methods used in NHANES I, and NHANES 2000. Finally, seven random-site selection methods (RSSMs) were created by sampling the following number of sites: 84, 42, 36, 28, 20, 15, 10 and 6. To compare bias and precision of the methods we calculated percent relative bias and relative error. *Results:* Estimates of means, standard deviations (SD), relative bias and relative error for RSSMs were almost identical to the full-mouth examination, but SDs increase slightly when fewer than 28 sites were sampled and relative bias and error increase for methods sampling fewer than 20 sites. The FSSMs had very low relative error, but much higher relative bias indicating underestimation. The FSSM with the smallest bias and error was the Ramfjord method, but the Random 36 method had less bias and less relative error. The NHANES 2000 method was the FSSM with the lowest bias and relative error for estimates of Extent Scores (percent of sites ≥3, 4, 5, or 5 mm PD or CAL) but random methods sampling fewer sites performed just as well. Both FSSMs and RSSMs underestimated prevalence, especially prevalence of less frequently occurring conditions, but most RSSMs were less likely to underestimate prevalence than the FSSMs. *Conclusion:* The promise of reducing bias and increasing precision of the estimates support the continued development and examination of RSSMs.

Key words: attachment loss; bias; pocket depth; precision; random sampling

James D. Beck, Department of Dental Ecology, CB# 7450, University of North Carolina, Chapel Hill, NC 27516, USA
Tel: +1 919 966 2787
Fax: +1 919 966 6761
e-mail: james_beck@unc.edu

Submitted 29 November 2004;
accepted 3 February 2005

As periodontitis appears to be a site-specific disease, researchers interested in treating and studying this disease agree that it is desirable to monitor as many sites as possible in order to increase the probability of finding disease activity. The sites usually monitored are the mesiobuccal, buccal, distobuccal, mesiolingual, lingual, and distolingual sites. When evaluating a patient's periodontal status, clinicians tend to probe these six sites on all teeth present. Most clinical research studies use the same standards and probe six sites on 28 (or 32) teeth using trained and standardized examiners. These 'full-mouth' examinations require probing a maximum of 192 sites (if third molars are included) and examination time may range from 25 to 45 min, which may be too time-consuming for certain studies. When dental examinations are part of a larger study or when subject characteristics or study budgets cannot support an examination of all teeth, shortened examinations have been proposed (1–3). It is well known that use of a subset of sites tends to underestimate the full-mouth prevalence in a population. However, it is less clear whether so-called partial sampling methods bias estimates of summary measures of the mouth that employ averages, such as mean attachment level or pocket depth or Extent and Severity (4–6).

Two of the most well-known partial sampling methods that reduce the time of periodontal examinations are the Periodontal Disease Index (PDI) system devised by Ramfjord (7, 8) which makes use of a specific set of teeth (index teeth) to represent the mouth, and the National Institute of Dental Research's (NIDCR) method (9, 10), which employs a method that subsamples both teeth and sites. The PDI system devised by Ramfjord makes measurements on six specific teeth: the maxillary right first molar, the maxillary left central incisor, the maxillary left first bicuspid, the mandibular left first molar, the mandibular right central incisor, and the mandibular right first bicuspid. Ramfjord considered these teeth to be representative of the mouth (8). Six sites are probed on each of these teeth resulting in a maximum of 36 sites.

Another well-known partial-mouth recording system was developed by the NIDCR to conduct national surveys in situations where time and financial resources were limited. The NIDCR method employed a strategy of randomly selecting an upper and a contralateral lower quadrant in which to conduct the periodontal measures. In addition, they examined only the mesiobuccal and buccal sites on all teeth in those quadrants. This system has been used to conduct several large studies, such as a study of employed and retired adults (9) and the National Health and Nutrition Examination Survey (NHANES III) (11). In addition, the method has been used by others conducting smaller studies (12, 13).

In order to minimize measurement error, most epidemiology studies measure sites for which there is good examiner reliability. Consequently when choosing a subset of sites, buccal sites, which are directly visible, are usually chosen because of better examiner reliability (14–16). Because certain sites are more likely to exhibit periodontal pocketing (mesial and distal) and recession (buccal) than others, sampling only the directly visible sites may increase underestimation of full-mouth prevalence of the condition relative to alternative selection schemes holding the number of sites fixed. Even though fixed-site partial-mouth recording may result in biased estimates of probing depth (PD) and clinical attachment level (CAL), alternative less biased methods with reduced examination time have not been available. However, over the last 20 years direct data entry of clinical examination data into computers has become more common. Usually the first clinical pass through the mouth involves counting which teeth are present or absent. With this information it is not difficult to write an algorithm to select sites to be sampled. As the selection of sites is a sampling issue, the most straightforward approach would be to select a simple random sample of sites present so that each site present has an equal chance of being included in the examination. Moreover, fixed selection of directly visible sites, such as in NHANES III, may result in biased estimates of average PD or CAL estimates because selected sites are not representative of all sites in the mouth. As simple random samples should allow unbiased estimation of PD or CAL, the question is how many sites should be sampled in a random-site selection method (RSSM) to provide estimates with good precision while having an acceptably small degree of underestimation of full-mouth prevalence.

Our first hypothesis is that RSSM will provide, when jointly considered, less biased and more precise estimates of PD and CAL than fixed-site selection method (FSSM). Our second hypothesis is that RSSMs will result in less underestimation of full-mouth prevalence than existing FSSM methods. While the FSSMs that predominantly target directly visible sites (NHANES I, NHANES III) were designed to produce less measurement error, they are more likely to underestimate prevalence because substantial attachment loss and PDs occur in other sites. Validation of both the hypotheses may lead to sampling fewer sites with RSSMs in future studies. The purpose of this paper is to compare estimates generated from RSSMs of various sizes with those obtained using three well-known FSSMs (the Ramfjord method and two NIDCR methods), using the full-mouth examination as the gold standard, which is considered to be free of bias.

Methods

The Dental ARIC Study provided the data for the present analyses (17). The Dental ARIC study is an ancillary study of the Atherosclerosis Risk in Communities Study (ARIC), a prospective investigation of the etiology and natural history of atherosclerosis and clinical cardiovascular disease among residents of four US communities (Jackson, Mississippi; Washington County, Maryland; suburban Minneapolis, Minnesota; and Forsyth County, North Carolina) (18). A representative sample of 15 792 community-dwelling residents aged 45–64 years at baseline took part in an evaluation of cardiovascular risk factors and their sequelae. The Dental ARIC Study was conducted at ARIC visit 4 in 1996–98. Its aims were to determine the prevalence, extent, and severity of periodontal conditions in the dentate ARIC population, and to describe the associations between those conditions and prevalent coronary heart disease, carotid artery intima-media thickness, presence of carotid artery lesions, and atherosclerosis risk factors. Edentulous persons and those requiring antibiotic prophylaxis for periodontal probing were ineligible resulting in examinations on 6793 participants. Human subjects participated in the study after providing informed consent to a protocol that had been reviewed and approved by the University of North Carolina at Chapel Hill (UNC) School of Dentistry Committee on Research Involving Human Subjects.

Periodontal measures

The Dental ARIC Study contained a periodontal examination that consisted of PDs and cemento-enamel junction (CEJ) measures relative to the gingival margin at six sites per tooth for all teeth in the mouth using a UNC 15 manual probe. Probing depths and CEJ measures were rounded down to the nearest millimeter. CAL was calculated as the sum of PD and CEJ measures.

Periodontal examiners at the ARIC centers were calibrated to a standard examiner and the percent agreement for CAL within 1 mm between each examiner and the standard examiner ranged from 83.2% to 90.2%. Weighted Kappa statistics ranged from 0.76 to 0.86, indicating excellent agreement with the standard examiner. Intraclass correlation coefficients ranged from 0.76 to 0.90, indicating excellent to outstanding agreement. Examiner statistics for PD were similar to those for CAL, and tended to be even better.

'Gold standard' estimates

Our main objective was to evaluate the bias and precision of estimates of PD and CAL derived from partial sampling methods relative to the 'gold standard' estimate. This 'gold standard' estimate was assumed to be the estimate obtained by measuring all six sites for all teeth present, except third molars.

Periodontal estimates using the Ramfjord method

We used Dental ARIC data to estimate PD and CEJ measures at six sites for the maxillary right first molar, the maxillary left central incisor, the maxillary left first bicuspid, the mandibular left first molar, the mandibular right central incisor, and the mandibular right first bicuspid. As indicated by Ramfjord, missing teeth were not substituted (8). This process resulted in PD and CAL scores for up to 36 sites.

Periodontal estimates using NIDCR periodontal methods

We created two separate sets of estimates using the NIDCR protocol. First we randomly selected one upper and the contralateral lower quadrant and recorded PDs and CEJ measures for the mesiobuccal and buccal sites for all teeth in those quadrants except third molars. This process resulted in a maximum of 28 sites examined, depending on the number of missing teeth. As this method was used in the third Health and Nutrition Examination Survey (NHANES III), we labeled these estimates as the 'NHANES III' method.

Beginning with the 1999–2000 NHANES studies, the method was changed by adding an additional site, the distobuccal, to those already measured. To create these estimates, we added PD and CEJ measures for the distobuccal site on each tooth in the quadrants selected. This process resulted in up to 42 sites being examined and estimates from this process were labeled, 'NHANES 2000' method.

Periodontal estimates using simple random sampling of sites

We generated these estimates by designating the number of sites to be sampled and then using a random number generator to select the sites to be assessed for each individual. Third molars were not eligible for selection. If the individual had fewer sites available than were needed for that particular estimate, then all available sites were recorded. We generated eight sets of random estimates based on the number of sites to be sampled. First, we created estimates based on sampling 84 sites in order to have a comparison of bias and precision for sampling half of the sites used in the full-mouth examination. We also generated estimates based on 42, 36, and 28 sites in order to have direct comparisons with the NHANES 2000, Ramfjord, and NHANES III-site selection protocols, respectively. We also generated estimates based on 20, 15, 10 and 6 sites sampled in order to determine whether sampling fewer sites in a random manner would allow greater time savings without substantially decreasing the precision of the estimates. There was no particular rationale for our *a priori* selection of 20, 15, 10 and 6 sites sampled. Our only goal was to produce a range of estimates based on a smaller number of sites sampled than used in the current methods.

Statistical analysis

The analysis is based upon summary statistics that are averages of PD or CAL for each subject. Evaluation of bias and precision of the various partial sampling methods relative to the full-mouth method is based upon means and standard deviations (SD) of these summary measures across all 6793 subjects in the sample. Bias is defined as the mean score of the summary statistics based upon the partial sampling method minus the mean score based upon the full-mouth method. To evaluate bias we calculated the percent relative bias as:

$$\%\ \text{Relative bias} = 100 \times (M_p - M_f)/M_f$$

where M_p is the mean score based on partial method and M_f is the mean score based on full-mouth method.

The precision of a partial method is defined as the mean squared error (MSE) of the measure. The MSE quantifies the variability of the measure around M_f:

$$\text{MSE} = \sum(p - M_f)^2/N = (M_p - M_f)^2 + \sum(p - M_p)^2/N = \text{bias}^2 + [(N-1)/N]\text{SD}^2$$

where p is the score for an individual based on the partial method and summation ($\sum$) is over the $N = 6793$ individuals. Our choice of MSE instead of the usual sample variance (SD^2) to measure precision is based upon the goal of measuring variability about the correct target of the full-mouth mean instead of the (incorrect) estimate of the partial-mouth mean. As shown above, the two are related; the MSE decomposes into the sum of $[(N-1)/N]$ times the sample variance and the square of the bias. As $N = 6793$, the multiplier $[(N-1)/N]$ is effectively equal to 1. Once precision (MSE) is calculated for a partial method, we determine the relative precision, hereafter referred to as relative error, of the full-mouth method with respect to that of the partial-mouth method:

$$\text{Relative error} = \text{MSE}_f/\text{MSE}_p$$

where MSE_f is the mean squared error of full-mouth method and MSE_p is the mean squared error of partial method.

As the full-mouth method is the gold standard its bias is considered to be zero, implying that MSEf equals its usual sample variance. A partial method is said to have good precision if the relative error has a value near 1 or above 1, and poor precision if the relative error is considerably <1 (i.e. the partial method has large variability relative to the full-mouth method). Our strategy is to first compare methods bases on percent relative bias and then consider relative error. While we are concerned about any amount of bias, we are considering a relative bias of 5% or more to be unacceptable. A relative error score below 0.90 is considered poor.

Because there is variability in the random selection process, relative bias and relative error estimates may differ slightly from sample to sample. In order to present more representative estimates in this manuscript, sampling was repeated five times for each RSSM and the average mean and SD across the five replicates was determined. From these averages, percent relative bias and relative error were calculated based upon the formulae above. A similar analysis of percent relative bias and relative error is presented for subject level summary statistics that are percentages of sites with PD of 4, 5, or 6 mm or CAL of 3, 4, 5, or 6 mm.

Comparisons of prevalence estimates are presented as the percent of individuals with one or more sites at or above 3, 4, 5, or 6 mm of CAL and 4, 5, or 6 mm of PD. Underestimation of prevalence is calculated as the percent of individuals with the conditions (as assessed by the full-mouth method) who would not be identified as having the condition by partial-mouth or random methods (false negatives). We assume there is no measurement error in the full-mouth measure (the gold standard). This means that the false-positive rate of all FSSMs and RSSMs is zero, because a partial method cannot identify a positive site that is not identified by the full-mouth method.

Results

Table 1 presents characteristics of the study participants. Ages ranged from 52 to 74 years with mean age of 62 and the number of teeth ranged from 1 to

Table 1. Characteristics of the study participants ($n = 6793$)

Variable	Mean	SD	Minimum–maximum
Age (years)	62.4	5.63	52–74
No. of teeth	21.2	6.86	1–28
	Category	N[a]	%
Gender	Female	3686	54.3
	Male	3107	45.7
Race	African-American	1300	19.2
	White	5468	80.8
Education	Less than High School	918	13.5
	High School Grad	2920	43.0
	Advanced	2946	43.5
Income	<$25 000	1643	25.1
	$25 000 to <$50 000	2410	36.9
	$50 000 or more	2481	38.0
Smoking	Never	3176	48.6
	Former light (<200)[b]	1464	22.4
	Former heavy (200+)	1109	17.0
	Current light	166	2.5
	Current heavy	617	9.4
Type 2 diabetes	No	5586	82.8
	Yes	1158	17.2
Dental visits	Regular	4945	73.2
	Episodic	1814	26.8

[a]The categories for some variables may total to <6793 due to missing information.
[b]Defined as having smoked <200 cigarettes per year.

28 with a mean of 21. As might be expected in this age group, females slightly outnumbered males. Study participants were predominantly white, somewhat skewed toward having advanced educational levels and annual incomes of $50 000 or more. Almost half had never smoked cigarettes and about 12% were current smokers. About 17% had type 2 diabetes and approximately 27% were episodic users of dental services.

In all four panels comprising Fig. 1, the y-axis is the extent of 3+ mm CAL according to the gold standard method, and the x-axis is that extent as generated by the RSSM or FSSM shown. Figure 1a–d shows the NHANES III, NHANES 2000, the RSSM 28 and the RSSM 20 methods, respectively. There is a reasonably wide dispersion of points for every increment of extent with the NHANES III and NHANES 2000 estimates with the latter having a somewhat narrower dispersion of points, likely due to the addition of the distobuccal site, which increases the likelihood that the FSSM will identify sites with attachment loss. Figure 1c is the RSSM that is directly comparable with the NHANES III estimates. The dispersion of points appears narrower than for either Fig. 1a,b. Figure 1d indicates that the RSSM 20 estimates appear to be slightly more dispersed than for the 28-site estimates, but narrower than those for in Fig. 1a,b, likely an effect of sampling fewer sites.

Table 2 presents mean scores, standard deviations, percent relative bias (scores further from zero indicate more relative bias), and relative error (higher scores indicate less relative error) for whole-mouth PD and CAL when compared with estimates from FSSMs and RSSMs. Mean scores for both PD and CAL for FSSMs appear to vary slightly around the full-mouth means with estimates for both NHANES methods showing the most underestimation. This tendency for underestimation is reflected in the negative relative bias scores. Estimates of means produced by RSSMs appear to mirror the whole-mouth scores, although the Relative Bias scores indicate slight over- or underestimations occurred beyond the two decimal points presented. Standard deviations for the NHANES methods were smaller than for the gold standard, likely due to the facial sites that were selected. Standard deviations for the RSSMs increased as the number of sites sampled decreased. While some of the RSSMs produced almost no bias, the bias for both NHANES methods was more substantial for CAL and unacceptable for PD. Relative Error scores above 1.0 for the NHANES methods suggest greater precision than the gold standard but this favorable consideration is outweighed by the presence of the bias previously noted. Conversely, the Ramfjord method showed some loss of precision with a relative error <1.0. Its overall performance was similar to that of the RSSM method that is also based upon 36 sites. The relative error for the RSSMs was slight for PD and CAL for those RSSMs measuring at least 20 sites, but declined to the 0.7–0.8 range when fewer sites were measured.

Table 3a presents the percent relative bias and relative error for estimates of mean percent of sites with PDs of 4, 5, or 6 mm according to various sampling methods. Generally, the percent relative bias becomes greater as the PD levels increase; likely because the prevalence of the condition decreases. Among the FSSMs, it appears that there is less relative bias for the NHANES 2000 method, which is consistent irrespective of PD level. The NHANES III method results in by far the most relative bias and is classified as unacceptable. In general, the relative error is acceptable for NHANES 2000, Ramfjord, and for random methods measuring at least 28 sites. Table 3b presents

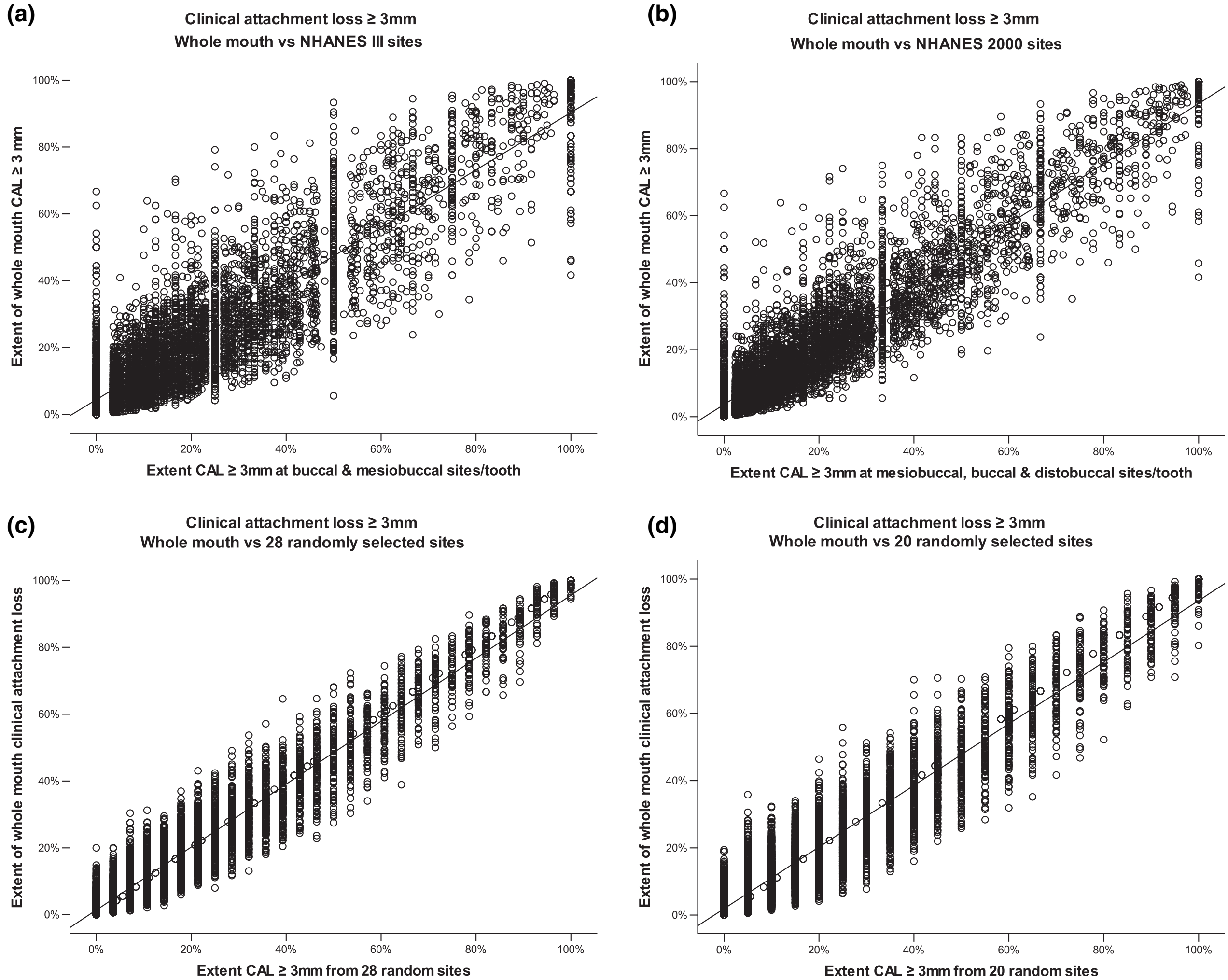

Fig. 1. Estimates of extent of 3+ mm attachment level for four sampling methods.

Table 2. Comparison of means, standard deviations, percent relative bias, and relative error for various sampling methods ($n = 6793$)

	Probing depth (mm)			Attachment level (mm)		
Sampling method	Mean (SD)	% Relative bias	Relative error	Mean (SD)	% Relative bias	Relative error
Whole mouth (168)	1.89 (0.58)	0	Referent	1.77 (1.04)	0	Referent
Fixed partial methods						
NHANES 2000 (42)	1.79 (0.56)	−5.3554	1.0373	1.73 (1.02)	−2.3684	1.0311
Ramfjord teeth (36)	1.87 (0.60)	−0.8843	0.9334	1.78 (1.05)	0.2760	0.9828
NHANES III (28)	1.66 (0.52)	−11.6491	1.0338	1.72 (1.01)	−3.2081	1.0591
Random partial methods						
Random 84	1.89 (0.58)	−0.0101[a]	0.9911	1.77 (1.04)	0.0110	0.9953
Random 42	1.89 (0.59)	−0.0217	0.9662	1.77 (1.05)	−0.0950	0.9859
Random 36	1.89 (0.59)	0.0217	0.9547	1.77 (1.05)	0.0107	0.9771
Random 28	1.89 (0.60)	−0.0112	0.9380	1.77 (1.06)	0.0581	0.9662
Random 20	1.89 (0.61)	−0.0458	0.9088	1.77 (1.06)	−0.0792	0.9535
Random 15	1.89 (0.63)	0.0476	0.8591	1.77 (1.08)	0.0252	0.7043
Random 10	1.89 (0.65)	0.2125	0.7968	1.77 (1.10)	0.1428	0.8922
Random 6	1.89 (0.69)	−0.0738	0.7044	1.77 (1.14)	0.0199	0.8253

[a]The negative percent relative bias reflects the actual underestimation of the mean that is not obvious at two decimal places.

Table 3a. Percent of relative bias and relative error for estimates of mean percent of sites with probing depths of 4, 5 or 6 mm for various sampling methods

	Probing depth (mm)					
	4 mm		5 mm		6 mm	
Sampling method	% Relative bias	Relative error	% Relative bias	Relative error	% Relative bias	Relative error
Whole mouth (168)	0	Referent	0	Referent	0	Referent
Fixed partial methods						
NHANES 2000 (42)	0.3755	0.9405	0.2790	0.9513	−0.4134	0.9472
Ramfjord teeth (36)	0.5519	0.9100	−0.9942	0.9141	−1.5835	0.9218
NHANES III (28)	−40.1062	1.2956	−40.4530	1.3888	−41.5670	1.3587
Random partial methods						
Random 84	−0.1486	0.9849	−0.1687	0.9869	−0.2926	0.9905
Random 42	−0.0759	0.9474	0.0611	0.9493	−0.5186	0.9418
Random 36	0.2648	0.9341	−0.0018	0.9312	0.1831	0.9267
Random 28	0.4173	0.9064	−0.1277	0.9170	−0.7460	0.8977
Random 20	−0.1860	0.8624	−0.3812	0.8638	−1.1669	0.8646
Random 15	−0.3451	0.8157	0.1198	0.8056	1.2158	0.7734
Random 10	0.4117	0.7263	1.8912	0.7129	2.6384	0.6789
Random 6	−0.1642	0.6180	−0.3615	0.6201	−0.8993	0.5736

Table 3b. Percent of relative bias and relative error for estimates of mean percent of sites with attachment levels of 3, 4, 5, or 6 mm for various sampling methods

	Attachment level (mm)							
	3 mm		4 mm		5 mm		6 mm	
Sampling method	% Relative bias	Relative error	% Relative bias	Relative error	% Relative bias	Relative error	% Relative bias	Relative error
Whole mouth (168)	0	Referent	0	Referent	0	Referent	0	Referent
Fixed partial methods								
NHANES 2000 (42)	−0.0814	0.9736	−0.3076	0.9745	−0.2729	0.9729	−0.2289	0.9664
Ramfjord teeth (36)	0.3274	0.9382	0.2542	0.9516	0.1034	0.9534	−0.3097	0.9501
NHANES III (28)	−6.5817	0.9260	−8.7135	0.9895	−14.8444	1.0681	−19.5778	1.1269
Random partial methods								
Random 84	−0.0039	0.9922	−0.0957	0.9943	−0.0143	0.9928	0.0934	0.9932
Random 42	−0.1664	0.9744	−0.2892	0.9810	−0.6565	0.9826	−0.7046	0.9749
Random 36	−0.0227	0.9618	−0.0753	0.9675	0.3506	0.9631	0.2138	0.9680
Random 28	0.2025	0.9407	0.2276	0.9549	0.1633	0.9554	0.3015	0.9461
Random 20	−0.2108	0.9168	−0.3084	0.9317	−0.1644	0.9296	0.1417	0.9185
Random 15	0.1555	0.8828	−0.4422	0.9063	−0.5412	0.9087	−0.6834	0.9089
Random 10	0.2592	0.8223	−0.0265	0.8455	0.5277	0.8454	0.2621	0.8373
Random 6	−0.2385	0.7353	−0.1575	0.7647	0.0875	0.7644	0.4052	0.7401

the same analyses for CAL. Again, the NHANES III methods are considered unacceptable due to having a large amount of relative bias. Both the NHANES 2000 and the Ramfjord methods exhibit low relative bias with acceptably low precision loss relative to the full-mouth measures. RSSMs appear to produce low relative bias scores, even when sampling as few as six sites. However, relative error does increase when 15 or fewer sites are sampled.

The first row of Table 4 presents the prevalence of people with one or more 4, 5, and 6 mm PDs and 3, 4, 5, and 6 mm CALs for the full-mouth examination. We compared the prevalence estimates obtained from the various sampling methods to the full-mouth prevalences and expressed the differences as percent of underestimation. Underestimation of prevalence generally is greater for PD than for CAL and increases as PDs and CALs increase. In addition, for RSSMs the number of sites sampled is inversely related to the amount of underestimation. RSSMs result in less underestimation than comparable FSSMs. For example, NHANES III underestimates prevalence of 4+ mm PDs by 51.3% and the Random 28 method underestimates by 24.7%. For the most severe

Table 4. Comparisons of underestimation[a] of whole-mouth prevalence[b] of probing depth and attachment level for various sampling methods

	Probing depth			Attachment level (mm)			
Sampling method	4+	5+	6+	3+	4+	5+	6+
Whole mouth (168)	69.5[b]	46.7[b]	23.7[b]	97.1[b]	77.2[b]	55.1[b]	35.8[b]
Percent underestimation compared with six sites for 28 teeth							
NHANES 2000 (42)	40.3	52.6	58.3	13.7	27.0	37.4	43.1
Ramfjord teeth (36)	30.3	45.2	54.9	14.1	29.8	41.7	49.8
NHANES III (28)	51.3	64.5	70.8	17.7	34.2	45.8	53.9
Random 84	5.2	11.3	15.9	1.5	6.2	9.9	11.3
Random 42	16.7	27.1	24.3	5.6	17.5	24.0	28.2
Random 36	19.8	32.3	41.2	6.6	19.1	28.4	32.0
Random 28	24.7	39.0	44.8	9.6	24.7	32.9	38.5
Random 20	34.4	48.4	54.6	14.5	32.2	40.6	46.5
Random 15	40.8	54.4	60.1	19.0	37.1	47.1	52.6
Random 10	50.6	63.2	68.8	26.8	46.5	55.5	61.6
Random 6	62.0	73.5	78.0	37.8	56.7	65.2	70.8

[a]Percent of individuals with the condition that would not be identified.
[b]Percent of individuals with one or more sites with the condition.

PDs (6+ mm), the Random 84 method has the smallest underestimation (15.9%) compared with 24% for the Random 42 method. In sum, all partial methods underestimate prevalence and it would appear that to produce more valid estimates, especially of less prevalent conditions, a large number of sites would need to be sampled using RSSM.

Finally, we present Fig. 2 as a summary of estimated prevalence of CALs of 3, 4, 5, and 6+ mm according to the number of sites randomly sampled. Generally, the gold standard prevalence (168 sites) declines by approximately 20% for each millimeter increase in CAL threshold. The rate of change of the estimated prevalence of 3+ mm CAL appears to be greater than for the other levels up to about 28 sites sampled, indicating that sampling more rather than fewer sites produces the greatest gain in the prevalence estimate at this minimal CAL. It appears that there is little additional gain in validity of estimates by sampling more than 42 sites. For example, by doubling the number of sites sampled from 42 to 84, prevalence estimates for 3+ mm of CAL increased from 91.6% to 95.7%.

Discussion

This study supports some aspects of previous studies, which concluded that mean scores from FSSMs, such as NHANES and Ramfjord methods are reasonable estimates of full-mouth means, but that they severely underestimate the prevalence of the conditions. This study further indicates (Table 4) that the severe underestimation of prevalence rates is greater for FSSMs than for RSSMs. Furthermore, RSSMs do appear to produce less biased estimates of mean PD or CAL than the FSSMs and in some instances do so by sampling fewer sites. It is not surprising that the NHANES methods result in little relative error as the sites sampled using those methods (buccal

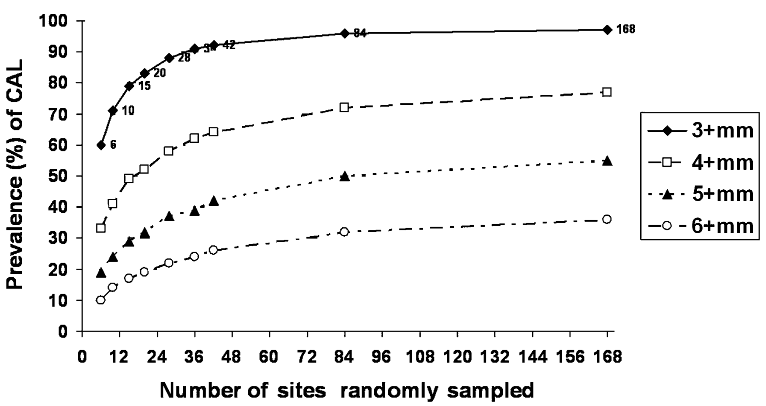

Fig. 2. Estimated prevalence of attachment level according to number of sites randomly sampled.

and mesiobuccal) were selected because they are easier to see and can be examined more reliably; however they may also be consistently shallower sites resulting in less variability and therefore less relative error. If the aims of a study focus on evaluation of summary measures of continuous distributions, such as means (Table 2), and a FSSM must be used, we recommend use of the Ramfjord method because it has the least relative bias and error. However, the Random 28 and 20 methods also perform well and require fewer sites to be examined.

If estimating the extent of PD and CAL at various thresholds is an important aim (Table 3a,b) and an FSSM must be employed, then the NHANES 2000 method appears to perform well in terms of relative bias and error. However, the Random 36 and 28 methods also perform well and require fewer sites to be sampled. The Random 20 method also has little relative bias and error and permits measurement of 22 fewer sites per person than the NHANES 2000 method.

If estimating prevalence is an important aim (Table 4), random methods (28, 36, and 42 sites) provide more accurate estimates than fixed sampling methods in almost all situations. However, they also suffer from substantial underestimation with less prevalent conditions, meaning that if a study objective is to measure prevalence in younger populations that presumably have lower prevalence of periodontal conditions, there is likely to be underestimation with both RSSMs and FSSMs. The Random 84 method requires half of the sites to be examined (assuming 168 sites) and performs better than the other methods. However, it is unlikely that any study requiring partial methods could afford to examine 84 sites per person.

One potential objection to using random methods is that clinical examiners prefer to examine a set pattern of sites on each tooth as it allows them to 'develop a rhythm' and conduct an examination more quickly. This is an important concern, but most examiners have had no experience with random site examination methods. While sites are identified in a random manner by the computer program, that program can sort the sites so that they are listed in the same sequence as a traditional examination and presented to the examiner in that sequence by the recorder. In addition, we showed that random methods using fewer than 28 sites also produce similar prevalence estimates to the full-mouth method. Thus, it may be possible to shorten the examination while decreasing the error of estimation. Missing teeth also would shorten the examination. Individuals may have lost enough teeth that they do not have enough sites for the RSSM being used. In that instance, all sites in the mouth are examined. For example, if the Random 42 method was being used, 7% of the individuals in this study sample would have all of their sites examined. This would result in fewer sites examined and a traditional examination (the recorder would not call each site to be examined) could be used. Similarly, the percent of individuals having all sites examined using Random 36, 28, 20 and 15 methods was 5.1, 2.1, 1.4, and 0.8%, respectively.

A second potential objection to using random methods is that some traditional partial methods only examine facial sites because they are more easily accessed, easier to see and associated with greater reliability (14–16). However, the inclusion of lingual sites by random methods appears to be a major reason why those methods produce more accurate prevalence estimates. Still, simply adding more sites to be examined on each tooth when using partial methods is not the entire solution as the Ramfjord method examines all sites on each tooth, but on a very limited number of teeth. This situation likely results in missing rare events that may not occur on those specific teeth.

Our findings also show that random methods underestimate prevalence and do so severely for less prevalent conditions, such a 6+ mm PDs or CAL (although not as severely as FSSMs). However, Fig. 2 indicates that substantially increasing the number of sites examined beyond 42 sites will result in only a small increase in accuracy of prevalence estimates. Consequently, if the prevalence of the condition is known to be high in the population to be studied, then random partial-mouth measures (possibly assessing 20 sites) could provide estimates with low relative bias and error. If the prevalence is known to be low (likely <50% as seen in Table 4), then any partial-mouth measures will underestimate the condition. Finally, the results of this study may vary with the dental status of the population being studied. In situations where the population has no missing teeth, the FSSMs may perform better, as all sites to be sampled will be present. Conversely, in populations with more missing teeth, the FSSMs may perform more poorly.

In summary, it appears that if the estimation of prevalence is one of the objectives of the study, then substantial underestimation will result using any of the partial sampling measures other than the Random 84 method. If that method still samples too many sites, then the Random 42 method would be the next choice. If estimation of prevalence is not an objective of the study, then the Ramfjord method results in less relative bias and error than the NHANES methods in estimating overall mean scores, but the Random methods also performed well with the possibility of measuring fewer sites. Although this study was based on prevalence data, one might expect incidence estimates to exhibit similar patterns, perhaps with more resembling the patterns seen for less prevalent events.

In conclusion, all partial-mouth measures tend to underestimate the prevalence of periodontal conditions, but the problem is more severe for three commonly used fixed sampling partial-mouth measures than for RSSMs. Additionally, for summary measures such as mean PD or CALs, or percent of sites with those conditions, RSSMs performed favorably compared with FSSMs in terms of bias and precision. The promise of potentially decreasing costs and time spent in data collection while reducing bias and increasing precision of the estimates support the continued development and examination of RSSMs.

Acknowledgement

This work was supported by NIH grant DE-11551.

References

1. Alexander A. Partial mouth recordings of gingivitis, plaque and calculus in epidemiologic surveys. J Periodontol Res 1970;5:141–5.
2. Grossman F. The navy periodontal screening examination. J Am Soc Prev Dent 1973;3:41–3.
3. Mills W, Thompson G, Beagries G. Partial-mouth recording of plaque and periodontal pockets. J Periodontol Res 1975;10:36–41.
4. Hunt R, Fann S. Effect of examining half the teeth in a partial periodontal recording of older adults. J Dent Res 1991;70:1380–5.
5. Hunt R. The efficiency of half-mouth examinations in estimating the prevalence of periodontal disease. J Dent Res 1987;66:1044–8.
6. Kingman A, Loe H, Anerud A, Boysen H. Errors in measuring parameters associated with periodontal health and disease. J Periodontol 1991;62:477–86.
7. Ramfjord S. Devices for prevalence and incidence of periodontal disease. J Periodontol 1959;30:51–9.
8. Ramfjord SP. Indices for prevalence and incidence of periodontal disease. J Periodontol 1959;30:51–9.
9. Miller A, Brunelle J, Carlos J, Brown L, Loe H, editors. Oral health of United States adults: the national survey of oral health in U.S. employed adults. Bethesda, MD: National Institute of Dental Research, NIH Publication No. 87-2868; 1987.
10. Brown LJ, Oliver RC, Loe H. Evaluating periodontal status of US employed adults. J Am Dent Assoc 1990;121:226–32.
11. Albandar JM, Brunelle JA, Kingman A. Destructive periodontal disease in adults 30 years of age and older in the United states, 1988–1994. J Periodontol 1999;70:13–29.
12. Beck JD, Koch GG, Offenbacher S. Incidence of attachment loss over 3 years in older adults – new and progressing lesions. Community Dent Oral Epidemiol 1995;23:291–6.
13. Slade GD, Spencer AJ, Gorkic E, Andrews G. Oral health status and treatment needs of non-institutionalized persons aged 60+ in Adelaide, South Australia. Aust Dent J 1993;38:373–80.
14. Badersten A, Nilveus R, Egelbert J. Reproducibility of probing attachment level measurements. J Clin Periodontol 1984;11:475–9.
15. Glavind L, Loe H. Errors in the clinical assessment of periodontal destruction. J Periodontol Res 1967; 66:1044–8.
16. Watts T. Constant force probing with and without a stent in untreated periodontal disease. J Clin Periodontol 1987;14:407–13.
17. Beck J, Elter J, Heiss G, Couper D, Mauriello S, Offenbacher S. Relationship of periodontal disease to carotid artery intimal-media wall thickness: The Atherosclerosis Risk In Communities (ARIC) Study. Arterioscler Thromb Vasc Biol 2001;21:1816–22.
18. Heiss G. Atherosclerosis Risk in Communities: a follow-up study of early arterial lesions in the general population. In: Glagov S, editor. Pathology of human atherosclerotic plaque. New York: Springer-Verlag; 1989. p. 877–88.