Naval Dental Research Institute
NDRI-PR 88-08
July 1988

FALSE POSITIVE RATES IN THE DETERMINATION OF CHANGES IN PROBING DEPTH-RELATED PERIODONTAL MEASUREMENTS

M. E. COHEN

S. A. RALLS

Naval Medical Research and Development Command
Bethesda, Maryland

NAVAL DENTAL RESEARCH INSTITUTE
NAVAL TRAINING CENTER, BUILDING 1-H
GREAT LAKES, ILLINOIS 60088-5259

FALSE POSITIVE RATES IN THE DETERMINATION OF CHANGES IN PROBING DEPTH-RELATED PERIODONTAL MEASUREMENTS

M. E. COHEN

S. A. RALLS

Research Progress Report NDRI-PR-88-08
Work Unit 61152N MR0000101 0053
Naval Medical Research and Development Command
Naval Medical Command, National Capital Region
Bethesda, Maryland 20814-5044

Approved and released by:

R. G. WALTER
Captain, Dental Corps
United States Navy
Commanding Officer

False positive rates in the determination of changes in probing depth-related periodontal measurements

Mark E. Cohen and Stephen A. Ralls
Naval Dental Research Institute, Great Lakes, Illinois, U.S.A.

Cohen ME, Ralls SA: False positive rates in the determination of changes in probing depth related periodontal measurements. J Periodont Res 1988: 23: 161–165.

False positive rates associated with changes in periodontal probing measurements (changes which are of such magnitude as to be construed as due to disease or healing when the observed changes are actually due to measurement error) were estimated by computerized simulation. In the first phase of the simulation study, various distributions of error variances among sites were evaluated for their ability to produce matches to an empirical distribution of differences between replicate measurements. In the second phase of the study, distributions of variances identified in Phase I were used to estimate the false positive rate, under conditions of no actual change, for detection methods based on critical differences between averaged pairs of measurements. This rate was found to be substantially greater than that predicted using normal distribution probabilities and, for a difference of ≥2.5 mm, approached one false detection per examination of 168 sites. In the third phase of the study, simulation procedures were extended to the tolerance detection methodology and the false positive rate, in the absence of real change, was almost one detection per two examinations. This simulation suggested that perhaps one third of tolerance detected 'bursts' of periodontal attachment change may be false positives attributable to measurement error.

Accepted for publication November 12, 1987

Oral Microbiology

Introduction

Evidence for rapid changes in periodontal attachment level and pocket depth is based almost exclusively on differences between sequential periodontal probing measurements. Real attachment losses are postulated to have occurred if changes at or beyond a specified magnitude are present at frequencies substantially above those expected by chance. Should such "excessive" events occur, depending on the time intervals involved, the burst theory of periodontal attachment loss (1, 2) may be supported and implications for clinical intervention drawn. The general problems of periodontal burst detection and the clinical use of this information have been discussed elsewhere (3). The present research is directed toward the evaluation of false positive rates in burst detection under conditions of no actual change (alpha error).

Previous investigators (4) have tended to use detection criteria (e.g., "regression", "tolerance" and running medians methods) for changes in attachment levels which are not easily analyzed with respect to false positive rates, and it has therefore been necessary to depend on simulation to provide these estimates. Although simulation can be useful, this approach requires very careful consideration of method. The conditions under which previous simulations have been run (5), however, are not completely explicit and may not generalize to actual conditions. The present report considers burst detection methods based on attachment level differences both between pairs of replicated (and averaged) scores and in excess of tolerance thresholds.

Method and Results

This simulation study is organized into three phases. Each phase considers distributions of attachment level measurements that would be collected when no real changes have occurred, variability being due solely to measurement error. In the first phase, procedures are identified which can produce distributions of differences in replicate measurements that approximate empirical data. In the second phase, these methods are applied to the estimation of false positive rates when differences between averaged pairs of measurements are used to identify bursts of periodontal attachment loss. In the third phase these methods are extended to the situation where differences between averaged replicate measurements must exceed the three thresholds of the tolerance detection methodology (4).

Phase I

Comprehensive data on the distribution of differences of 48 064 replicate measurements at periodontal probing sites (when there has been no opportunity for real change) are available (5) and have been summarized in the first column of data in Table 1. Based on the

Table 1. Percent of replicate measurements exhibiting differences at specified absolute magnitudes, from empirical data and from 14 simulations of 10 000 sites each

Diff	Goodson[a]	0/0[b]	.1/.1	.2/.2	.3/.3	.4/.4	.5/.5	.3/.4	.3/.5	.3/.6	.3/.7	.3/.8	.3/.9	.3/1.0	.3/1.1
0	63.382	47.75	52.05	55.45	59.74	63.87	66.65	60.73	61.08	62.28	62.69	63.12	63.00	62.93	64.19
1	32.157	44.62	42.60	39.83	34.69	29.61	26.84	33.86	33.63	32.10	31.76	31.54	31.51	31.92	31.04
2	3.722	7.44	5.24	4.52	5.07	5.82	5.42	4.77	4.52	4.79	4.73	4.44	4.61	4.28	3.70
3	0.514	0.18	0.11	0.20	0.46	0.63	0.92	0.53	0.62	0.69	0.62	0.63	0.68	0.58	0.71
4	0.114	0.01	0	0	0.04	0.06	0.15	0.08	0.15	0.13	0.16	0.22	0.15	0.22	0.22
5	0.056	0	0	0	0	0.01	0.02	0.03	0	0.01	0.04	0.05	0.04	0.05	0.09
6	0.029	0	0	0	0	0	0	0	0	0	0	0	0.01	0.02	0.03
7	0.010	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02
8	0.015	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SD[c]	.5464	.6171	.5681	.5464	.5465	.5467	.5464	.5465	.5463	.5468	.5469	.54.65	.5469	.5466	.5480
TRLS with $\sigma < .5464$		0	9975	8931	7565	7404	7113	7944	8423	8452	8837	8935	8890	9168	9244
TRLS with $\sigma > .5464$		0	23	1067	2432	2594	2883	2054	1571	1546	1161	1063	1108	830	751
Chi-square[d]		1281.6	656.9	340.2	105.8	153.3	224.3	56.4	47.1	48.5	37.8	32.3	33.2	22.3	24.6

[a] Based on Table 1 in Goodson, J.M. 1986. *J Clin Perio* 13: 446–455.

[b] The σs used in the simulations of 10 000 sites were within the range .5464 minus the value to the left of the slash to .5464 plus the value to the right of the slash.

[c] This is the SD of individual scores which is estimated from the observed SDdiff.

[d] Chi-square on observed frequencies versus those expected from Goodson's empirical data. The categories were 0, 1, 2, 3, 4, and 5 or more mm. The critical Chi-square ($p < .05$, $df = 5$) is 11.07.

computed standard deviation of differences (SDdiff) of 0.7727, the estimated standard deviation of individual scores (SD) is 0.5464 ($SD = SDdiff/\sqrt{2}$; see Appendix).

This distribution of measurement error is not normal, however, as evidenced in part by kurtosis of 9.714 (6), while for a normal distribution kurtosis should be 3.0 (7). High kurtosis stems from either concentration of probability mass near the population mean (tendency towards a peaked unimodal distribution) or probability mass in the tails (tendency towards a bimodal distribution) (8). Inspection of the data indicates that the increase in kurtosis, from that of a normal distribution, stemmed from both of these sources.

The 14 simulations in Table 1 describe attempts to mirror this empirical distribution of differences. An initial problem in approaching this task is that the SD of 0.5464 incorporates disturbances due to score rounding and variance heterogeneity. It is well established that variation associated with probing depth measurements increases with depth (1, 9), for example. Thus, if one were to eliminate the effect of rounding the actual SD would probably not be 0.5464, and because of variance heterogeneity, probabilities associated with SD values could not be determined using the normal distribution.

The first simulation (0/0) in Table 1 (the methodology of which will be described later) naively uses σ of 0.5464 at all sites. As suggested by the discussion of kurtosis, the resultant distribution of differences exhibits too few differences of both 0 mm and greater than or equal to 3 mm. The sample SD is also too large, which is the result of distributional distortions caused by rounding. Selecting a smaller σ for the simulation would increase the percentage of zero differences but would further reduce the percentage of larger differences. It is therefore appropriate to investigate variance heterogeneity among sites as a means to achieve correspondence to the empirical distribution.

Fourteen simulations, of 10 000 sites (trials) each, were undertaken where score rounding to the nearest whole millimeter and heterogeneity of error variance were incorporated into the methodology. The objective was to define simulation conditions that produced distributions of differences that resembled the empirical data.

On each simulation trial, two random normal deviates (site measurements) were selected from a distribution with specified σ and constant mean. Each measurement was rounded to the nearest whole millimeter and a difference was computed. The SDdiff and SD were computed on the accumulated data on every trial after there were at least two trials simulated and at least one trial had a difference score other than zero. If the SD was greater than 0.5464 then a σ was randomly chosen for the next site from the uniform interval 0.5464 minus a specified value. If the SD was less than 0.5464 then a σ was chosen for the next site from the uniform interval 0.5464 plus a specified value. If the SD was equal to 0.5464 then σ for the next trial was 0.5464. In this way the terminal SD would be very close to 0.5464. The 14 pairs of specified values that determined the lower and upper bounds on σ are described in Table 1 where the value to the left of the slash affected the lower bound on σ, and the value to the right of the slash affected the upper bound.

The decision to use these uniform intervals was not theoretically grounded. However, to the extent that simulations so based are successful in matching the empirical data, some pragmatic appeal accrues to this approach.

A constant value of σ of 0.5464 (0/0) is ineffective in modeling the empirical data. Although error is greater in matching the percentage of zero differences between replicate measurements, the absence of large differences will result in an under-estimation of false positive rates. Other simulations which use the 0.5464 SD value and assume a normal error distribution therefore have limited validity.

The remaining simulations that assume that the upper and lower σ bounds are symmetric (0.1/0.1; 0.2/0.2; 0.3/0.3; 0.4/0.4; and 0.5/0.5) are somewhat more successful, although it does not appear that larger differences can be adequately represented without an over-representation of zero differences.

Simulations where the lower bound is 0.2464 and the upper bound ranges

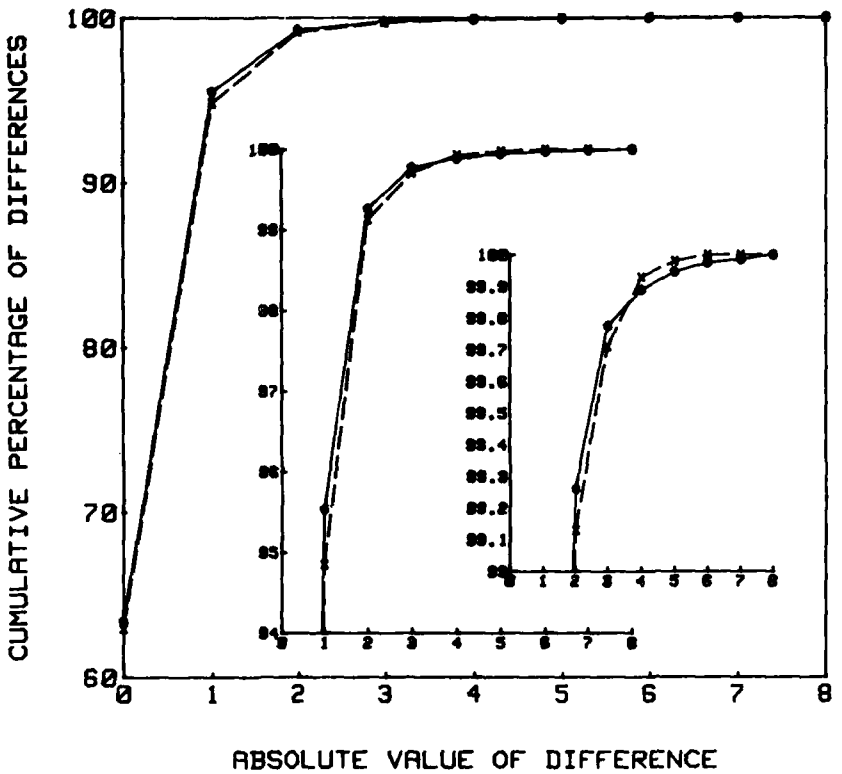

Fig. 1. Cumulative percentages of sites exhibiting differences between replicate measurements as a function of the absolute value of difference magnitude. The inserts provide more detailed analyses of the top portions of the cumulative distribution. The solid lines correspond to Goodson's empirical data while the dashed lines were generated by simulation 0.3/1.0.

from 0.9464 to 1.6464 generally appear to be more successful. Particular attention is directed to simulations 0.3/0.7 through 0.3/1.0 which compare reasonably well with the empirical data.

These conclusions are supported by chi-square tests, described in Table 1, which indicate that simulation 0.3/1.0 exhibited the closest fit to the empirical data. The quality of this fit in terms of cumulative percentages is presented graphically in Fig. 1. All the simulations differed at statistically significant levels from the empirical distribution. However, this appears to result in some cases from minor distributional differences which become influential with sample sizes of 10 000. In simulation 0.3/1.0, the largest discrepancy from the empirical data is at 2 mm; where the simulation exhibits 4.28% of sites and the clinical study 3.722%. The potential effect of this discrepancy on false positive rates may be offset, however, by lower percentages of differences at 5 mm or more. Nevertheless, small distributional differences may have important effects in this type of simulation and continued efforts toward generating a more perfect fit are warranted.

It is of interest to note the frequency of trials where σ was selected below or above 0.5464. As the upper bound on σ increases there is greater opportunity for larger differences between replicate measurements. When large differences occur the computed SD will increase and a substantial number of compensatory trials will follow where σ is selected from the interval below 0.5464. For simulation 0.3/1.0, 9168 sites were generated with σ less than 0.5464 and only 830 sites with σ above this value. If probing depth is related to probing error, then this conceptually corresponds to the situation where there are many shallower pockets that can be measured with accuracy and a few deep pockets that are subject to substantial measurement error.

Phase II

The simulation methodology described was modified in order to investigate the distribution of differences of averaged pairs of measurements. Four random normal deviates were generated for each site, rather than two. Each score was rounded to the nearest whole millimeter, the first two and the second two scores were averaged and the difference between these was the primary datum. The procedure to determine σ for the next trial was identical to Phase I except that the SD was estimated as SDdiff, rather than SDdiff/$\sqrt{2}$, for the case of differences between averaged paired scores (Appendix).

Table 2 provides results for simulations of 50 000 trials each for conditions 0.3/0.4, 0.3/0.7, and .3/1.0. Also reported are empirical results reported in the literature (10) which were collected in a manner to preclude actual disease or healing-related changes. The simulated data tend to predict somewhat fewer large differences than empirically determined. Simulation 0.3/1.0, however, matches the empirical data very well when attention is restricted to differences greater than or equal to 2.0 mm. Using the *de facto* standard critical difference of 2.5 mm between paired averaged scores, 48 sites per 10 000 were detected as changed in the simulation, an estimate close to the reported empirical value of 50.

Phase III

In a clinical study investigating burst detection, site measurements were made approximately bimonthly for 1 yr and burst rates of 393 per 10 000 sites were reported (134 of 3414 sites in 22 subjects changed in attachment level as indicated by the tolerance method) (4). In addition to effects associated with the three detection thresholds, the study also incorporates six opportunities for change at each site (month 0 versus month 2, 2 vs 4, and so forth). The simulation procedures were therefore modified to investigate false positive rates for the tolerance methodology applied over sequences of seven observations at each site.

Fourteen random normal deviates were generated for each of 168 sites for each of 100 simulated subjects. Sequential pairs of scores were rounded and averaged to produce seven site measurements. Although each pair of replicated measurements contributed to the calculated SD value for the patient, σ remained constant for all 14 scores generated within a site. This was done in recognition of variance heterogeneity between sites. The σ value was preset to 0.5464 for the first site in each patient and re-computed on the basis of SD at the start of the simulation of every site

Table 2. Rate per 10000 of replicate averaged pairs of measurements exhibiting differences equal to or greater than the indicated absolute magnitude from, empirical data and from 3 simulations of 50000 each

Diff	Aeppli[a]	.3/.4[b]	.3/.7	.3/1.0
0.0	10000	10000	10000	10000
0.5		5353	5256	5135
1.0	2000	1443	1371	1254
1.5	200	328	319	300
2.0	100	65	91	101
2.5	50	14	25	48
3.0	30	3	7	21
3.5	20	0[c]	2	8
4.0		–[c]	1	3
4.5		–	0	1
5.0		–	–	0
SD		.5464	.5464	.5464
Trials with	$\sigma < .5464$	40036	43337	45807
Trials with	$\sigma > .5464$	9961	6661	4191

[a] Based on Table 2A in Aeppli D.M., Boen, J. R., and Bandt, C. L. 1985. *J Periodontol* **56**: 262–264. The original data were reported in terms of differences not exceeding specified values but have been converted here to probabilities of differences being "equal to or greater than". The Datum for 0.5 was not reported so that the value used here for 1.0 (2000 per 10000) corresponds to the reported probability of 0.8 that the difference does not exceed zero. These values are for probing depth measurements. Attachment loss measurements that were reported exhibited greater numbers of larger differences.

[b] See Table 1 for description of nomenclature

[c] In this table "0" represents the situation where the rate per 10000 is less than 0.5, while "–" represents the absence of cases.

Table 3. Actual and simulated detected changes using tolerance methodology applied to seven observations (six comparisons) per site

	Haffajee[a]	Simulation[b]
Number of sites	3414	16800
Number of sites with changes	134	216
Sites with changes per 10000 sites	393	129
Total number of changes		256
Changes per 10000 observations		25
Average SD	.5798	.5513
Sites with $\sigma < .5464$		15293
Sites with $\sigma > .5464$		1507
Total changes $> = 2.5$ mm		423
Changes per 10000 ($168 \times 6 \times 100/10000$)		42

[a] Detections are based on Table 3 and the SD is based on Table 1 of Haffajee, A. D., Socransky, S. S. and Goodson, J. M. 1983. *J Clin Periodontol* **10**: 298–310.

[b] The tolerance detection thresholds described in the 1983 report were followed except that the subject threshold was not computed on all data for the subject but only on the data for the observation numbers (visits) involved in the comparison.

thereafter using the 0.3/1.0 bounds and the decision rules described for previous simulations.

Six sequential comparisons were conducted on the seven averaged measurements of each site. According to the tolerance methodology a change was detected if: (a) the change exceeded 2 population SDdiff (based on every pair of replicate measurements in the study; a 2 mm minimum change was used); (b) the change exceeded 3 "subject" SDdiff (this was computed on the particular 336 replicated measurements of the 168 sites and two observations defined by the particular comparison; and (c) the change exceeded 3 "pooled standard deviations of the two pairs of measurements at that site." (4). This was interpreted to mean that standard deviations, based on each pair of measurements, were to be computed separately and then pooled.

Table 3 summarizes the results of the simulation and indicates that, under the assumption that there were no actual changes in attachment level in the clinical study (4), 129 changes per 10000 sites (versus the empirical 393) would be detected using the tolerance methodology. When more than a single detected change at a particular site is considered, 25 changes per 10000 observations were detected. A fixed millimeter criterion ($> = 2.5$ mm) applied over the six comparisons yielded 252 detections per 10000 sites, which is consistent with Phase II data when the six observations are considered ($252/6 = 42$ which approximates 48; .3/1.0 condition in Phase II).

Discussion

The three thresholds of the tolerance method are not easily modeled and it would appear that in some simulations a critical value of 2.5 mm between averaged scores has been used in lieu of at least some portions of the complete method (5; pp 448–449). With apparent use of normal distribution assumptions and σ of 0.55, the false positive rate for the 2.5 mm criterion had been estimated at 1.2 per 10000 (5). This compares with the simple normal theory expectation (no threshold aspects of methodology being simulated) of $\Pr(|ZI| \geq 2.5/.5464 = 4.5754) = 0.05$ per 10000. The simulated rate found here of 48 per 10000 is approximately 40 times larger than had been reported when variance heterogeneity was not considered. Under conditions where there has been no real change in attachment level, an examiner can therefore expect to identify, by chance, a single change greater than or equal to 2.5 mm between averaged measurements, per examination of 168 sites (Phase II, $48/10{,}000 \times 168 = 0.81$; Phase III, $42/10{,}000 \times 168 = 0.71$).

The reported change rate (losses and gains) of 393 per 10000 sites over six observations, using the tolerance methodology (4), can be compared to 129 found in the Phase III simulation. The simulation, however, may be an oversimplification of the actual sampling environment. Non-random differences between subjects in the values of SDdiff, which have been reported to range from 0.52 to 1.30 (4), and non-random effects associated with observation number were not considered in the simulation. Effects of site were incorporated (SD values being recomputed between but not within sites) but the degree to which this manipulation reproduces empirical data is unknown. Nevertheless, 33% (129/393) may represent a reasonable estimate of the false positive rate associ-

ated with the complete tolerance detection method, under conditions of no actual change in attachment level. Over the course of six examinations of a single patient, an examiner can therefore expect to detect two ($216/16{,}800 \times 168 = 2.16$) changed sites by chance. On a single examination the probability is almost one in two ($256/100{,}800 \times 168 = 0.43$) of identifying a change.

False positive rates simulated here under conditions of no real change are of important magnitudes. It is also apparent that changes in attachment level of substantially less magnitude than suggested by fixed millimeter or tolerance criteria may be detected because of their superimposition on what has been shown to be influential levels of measurement error. A generalized loss of attachment of relatively small magnitude could account for the asymmetry in the empirical data (70% of tolerance detected changes were losses; 4).

The problem of false positives may or may not be of practical importance, depending upon the purposes and decisions that are involved and the actual incidence of true attachment level changes of clinically important magnitudes. Nevertheless, a Type I (alpha) error rate estimate of 33% for the tolerance methodology necessitates cautious use in the clinical setting and suggests that experimental studies directed towards the identification of disease correlates may have less than expected power due to case misclassification. Construction of more stringent criteria is probably not a viable solution because of the effects this would have on false negative rates.

The burst theory of periodontal attachment loss has been supported by evidence that frequencies of large losses exceed those predicted by chance. The present investigation suggests that previous estimates of alpha error based on normal assumptions may be underestimates and that detection of periodontal sites undergoing rapid change in attachment level may not be as accurate as previously assumed. Although the potential for rapid change is clear, particularly with respect to loss, the present findings re-emphasize the significance and implications of the measurement problem.

Acknowledgments

The opinions expressed herein are those of the authors and cannot be construed as reflecting the views of the Navy Department or the Naval Service at large. The use of commercially available products does not imply endorsement of these products or preference to other similar products on the market. Supported by Naval Medical Research and Development Command Project Number 61152N MR0000101 0053.

The authors wish to express their appreciation to the reviewers of this manuscript for many valuable suggestions.

Appendix

Where X is a variable and a is a constant, it can be shown that:

$$\text{var}[aX] = a^2\text{var}[X].$$

If scores (X_1, X_2, ..., X_n) are randomly (independently) selected from the same distribution the following relationships hold:

$$\text{var}[X_n] = \text{var}[X].$$
$$\text{var}[X_1 \pm X_2] = \text{var}[X_1] + \text{var}[X_2].$$

Using these facts, the standard deviation of the difference between two scores selected from a normal distribution with σ equal to 0.5464 can be shown to equal 0.7727 as follows:

$$\begin{aligned}\sigma^2\text{diff} &= \text{var}[X_1 - X_2] \\ &= \text{var}[X_1] + \text{var}[X_2] \\ &= 2\ \text{var}[X] \\ \sigma\text{diff} &= \sqrt{2}(\sigma[X]) \\ &= \sqrt{2}(0.5464) = 0.7727\end{aligned}$$

The standard deviation of the difference between the mean of two pairs of scores can be derived in a similar fashion.

$$\begin{aligned}\sigma^2\text{diff} &= \text{var}[(X_1 + X_2)/2 - (X_3 + X_4)/2] \\ &= \text{var}[X_1 + X_2)/2] + \text{var}[(X_3 + X_4)/2] \\ &= (1/4)\ \ \text{var}[X_1 + X_2)] + (1/4)\ \text{var}[(X_3 + X_4)] \\ &= (1/4)\ \ (\text{var}[(X_1 + X_2)] + \text{var}[(X_3 + X_4)]) \\ &= (1/4)\ \ (\text{var}[X_1] + \text{var}[X_2] + \text{var}[X_3] + \text{var}[X_4]) \\ &= (1/4)\ (4)\ \text{var}[X] \\ &= \text{var}[X] \\ \text{rdiff} &= \sigma[X] = 0.5464\end{aligned}$$

References

1. Goodson JM, Tanner CR, Haffajee AD, Sornberger GC, Socransky SS. Patterns of progression and regression of advanced destructive periodontal disease. *J Clin Periodontol* 1982; **9**: 472–481.
2. Socransky SS, Haffajee AD, Goodson JM, Lindhe J. New Concepts of destructive periodontal disease. *J Clin Periodontol* 1984; **11**: 21–32.
3. Ralls SA, Cohen ME. Problems in identifying "bursts" of periodontal attachment loss. *J Periodontol* 1986; **57**: 746–752.
4. Haffajee AD, Socransky SS, Goodson JM. Comparison of different data analyses for detecting changes in attachment level. *J Clin Periodontol* 1983; **10**: 298–310.
5. Goodson JM. Clinical measurements of periodontitis. *J Clin Periodontol* 1986; **13**: 446–455.
6. Kent RL, Goodson JM. Statistical analysis of probeable attachment level measurements: Distributional characteristics. *J Dent Res* 1986; **65**: Special Issue: Abstract #523, 227.
7. Snedecor GW, Cochran WG. *Statistical Methods*. 7th ed. Ames: The Iowa State University Press, 1980: 79.
8. Moors JJA. The meaning of kurtosis: Darlington reexamined. *Am Statist* 1986; **40**: 283–284.
9. Badersten A, Nilveus R, Egelberg J. Reproducibility of probing attachment level measurements. *J Clin Periodontol* 1984; **11**: 475–485.
10. Aeppli DM, Boen JR, Bandt CL. Measuring and interpreting increases in probing depth and attachment loss. *J Periodontol* 1985; **56**: 262–264.

Address:

Mark E. Cohen
Naval Dental Research Institute
Naval Training Center, Bldg 1-H
Great Lakes, IL 60088-5259
U.S.A.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ADA200687

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) FALSE POSITIVE RATES IN THE DETERMINATION OF CHANGES IN PROBING DEPTH-RELATED PERIODONTAL MEASUREMENTS		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER NDRI-PR 88-08
7. AUTHOR(s) M. E. COHEN and S. A. RALLS		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Dental Research Institute Naval Training Center, Building 1-H Great Lakes, IL 60088-5259		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61152N MR0000101 0053
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research and Development Command, Naval Medical Command, National Capital Region, Bethesda, MD 20814-5044		12. REPORT DATE July 1988 13. NUMBER OF PAGES 5
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Commander, Naval Medical Command Navy Department Washington, D.C. 20372-5120		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

This document has been approved for public release; distribution unlimited.

18. SUPPLEMENTARY NOTES

Journal of Periodontal Research 23:161-165, 1988

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Periodontics
False Positives
Disease Detection

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

False positive rates associated with changes in periodontal probing measurements (changes which are of such magnitude as to be construed as due to disease or healing when the observed changes are actually due to measurement error) were estimated by computerized simulation. In the first phase of the simulation study, various distributions of error variances among sites were evaluated for their ability to produce matches to an empirical distribution of differences between replicate measurements. In

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

the second phase of the study, distributions of variances identified in Phase I were used to estimate the false positive rate, under conditions of no actual change, for detection methods based on critical differences between averaged pairs of measurements. This rate was found to be substantially greater than that predicted using normal distribution probabilities and, for a difference of ≥2.5 mm, approached one false detection per examination of 168 sites. In the third phase of the study, simulation procedures were extended to the tolerance detection methodology and the false positive rate, in the absence of real change, was almost one detection per two examinations. This simulation suggested that perhaps one third of tolerance detected "bursts" of periodontal attachment change may be false positives attributable to measurement error.