

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/25140>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

Letters to the Editor

To the Editor:

Re: Epidemiology of Periodontal Diseases (position paper)(1996;67:935-945)

We would like to comment on the section regarding the threshold dilemma for defining clinical attachment loss (CAL) changes, since it is of utmost importance in longitudinal periodontal research.

It is discussed in the position paper that a change in attachment level needs to be at least 2 mm (i.e., 2 to 3 times the standard deviation [SD] of the differences between duplicate CAL measurements) before the investigators can be confident that they are seeing real change rather than measurement error. The question arises "how confident the investigators can be that they are seeing real changes." Since the authors refer to two particular studies, it seems that they believe that the probability of making a wrong decision about a real change in CAL (type I error) can be calculated by application of parametric statistics. Along this line of thinking, with a SD of differences between duplicate CAL measurements of around 0.8 mm, which is the presented SD in the position paper, a change in CAL of at least 3 mm (the threshold for real change in most studies) exceeds the SD around 3.7 times. This results in a *calculated* probability of $P < 0.0001$ that a ≥ 3 mm CAL change is due to the measurement error. In that case the investigators can be confident that they are seeing real change. However, here we face a problem. The *real* probability of encountering a CAL change of ≥ 3 mm which is caused by the measurement error can be obtained from studies on duplicate measurements.¹⁻⁵ These studies indicate that measurement errors of ≥ 3 mm occur with a frequency of around 1% to 2%. Thus the *real* probability ($P = 0.01$ to 0.02) of making a type I error in case of a ≥ 3 mm CAL change is much larger than the *calculated* probability of $P < 0.0001$. The discrepancy between the real and calculated probability of the type I error is explained by the non-Gaussian distribution of the measurement error.^{4,6} As far as the sign testing of means and values in the interval $mean \pm SD$ is concerned, the use of parametric statistics is robust in case of a small deviation of the normal distribution. However, the distribution of the measurement error exhibits heavier tails than a normal (Gaussian) distribution. Therefore, the application of parametric statistical calculations for the probability of outliers (≥ 3 mm) of the measurement error results in too low P -values.

In some prospective cohort studies, pairs of CAL measurements at each time point were taken in order to increase the reliability of the method.^{2,7} Although the SD

of the method error decreased with $\sqrt{2}$, the occurrence of large measurement errors could not be prevented.⁵

The use of the more recently introduced electronic probes seems to increase the reliability of the CAL measurements. The reported SDs of the measurement error of around 0.3 mm is much smaller than the manual probes with visual recording.^{8,9} Unfortunately, it has not been reported to what extent these electronic probes prevent the occurrence of large measurement errors. Thus the suggestion in the literature, that the threshold for defining real CAL changes could be cut down to ≥ 1 mm, is somewhat premature.

The *real* probability of 1% to 2% to encounter a ≥ 3 mm CAL change due to the manual probing measurement error is in the range of reported proportions of sites with ≥ 3 mm CAL change in several recent prospective cohort studies.^{10,11} This means that the type I error can account for a substantial proportion of false-positive sites. With an observed 3% of sites showing a CAL change of ≥ 3 mm in the cited prospective studies and a type I error of 1.5%, the proportion of false-positive sites will reach 50%. The problem of erroneously assigning sites in the category of real CAL changes (type I error) is inextricably bound up with the problem of erroneously assigning sites in the category of *no* real CAL changes (type II error). The position paper did not address the problem of the type II error. The magnitude of the type I error and the type II error and the problem of how to keep the rate of false-positive *and* false-negative sites as low as possible is of particular interest in prospective studies aiming to evaluate clinical, microbiological, and biochemical site-related diagnostic tests. One can question the high threshold of ≥ 3 mm for a CAL change, since all sites with real changing CAL below that threshold are not recognized. The aforementioned studies on duplicate measurements indicate that the frequency of smaller measurement errors of ≥ 2 mm varies between 2% to 10%. Although the type I error is higher with a lower threshold of ≥ 2 mm for real CAL change, the rate of false-positive sites may be favorably influenced if the frequency of observed ≥ 2 mm CAL changes has increased more than the type I error. On the other hand, the lower threshold for real CAL change has definitely a favorable influence on the rate of false-negative sites.

The choice of threshold as to which the investigators assign sites as having experienced CAL changes might rather be a matter of meticulous consideration when a study is designed. Weighing the importance of type I against type II error might be more suitable than simply applying a rule of thumb. We do not offer a final solution to the discussed problems of type I and type II errors in CAL measurements. We consider this comment a valu-

able addition to the position paper.—*W.H. van Palenstein Helderma and M.F. Timmerman, University of Nijmegen, WHO Collaborating Centre, Nijmegen, The Netherlands.*

REFERENCES

1. Badersten A, Nilvéus R, Egelberg J. Reproducibility of probing attachment level measurements. *J Clin Periodontol* 1984;11:475-485.
2. Goodson JM. Clinical measurements of periodontitis. *J Clin Periodontol* 1986;13:446-455.
3. Loos B, Kiger B, Egelberg J. An evaluation of basic periodontal therapy using sonic and ultrasonic scalers. *J Clin Periodontol* 1987;14:29-33.
4. Janssen PTM, Faber JAJ, van Palenstein Helderma WH. Non-Gaussian distribution of differences between duplicate probing depth measurements. *J Clin Periodontol* 1987;14:345-349.
5. Janssen PTM, Drayer A, Faber JAJ, van Palenstein Helderma WH. Accuracy of repeated single versus averages of repeated duplicates of probing depth measurements. *J Clin Periodontol* 1988;15:569-574.
6. Gunsolley JC, Best AM. Change in attachment level. *J Periodontol* 1988;59:450-456.
7. Haffajee AD, Socransky SS, Goodson JM. Comparison of different data analysis for detecting changes in attachment level. *J Clin Periodontol* 1983;10:298-310.
8. Gibbs CH, Hirschfeld JW, Lee JG, et al. Description and clinical evaluation of a new computerized periodontal probe - the Florida Probe. *J Clin Periodontol* 1988;15:137-144.
9. Yang MCK, Marks RG, Magnusson I, Clouser B, Clark WB. Reproducibility of an electronic probe in relative attachment level measurements. *J Clin Periodontol* 1992;19:541-548.
10. Maechte EE, Norderyd J, Koch G, Dunford R, Grossi S, Genco RJ. The rate of periodontal attachment loss in subjects with established periodontitis. *J Periodontol* 1993;64:713-718.
11. Claffey N, Kelly A, Bergquist J, Egelberg J. Patterns of attachment loss in advanced periodontitis patients monitored following initial periodontal treatment. *J Clin Periodontol* 1996;23:523-531.

Author's Response:

Drs. van Palenstein Helderma and Timmerman raise an issue that is academically interesting but which can raise the frustration level for periodontal researchers. The root of the problem is that with our standard clinical instruments we cannot measure attachment loss with any more precision today than we could 40 years ago. Next to the remarkable advances we are beginning to see at the molecular biology level, our clinical methods of measuring attachment loss appear more and more crude. Drs. van Palenstein Helderma and Timmerman rightly point out that the computerized probes are more precise (i.e., they reduce the degree of random error) than manual probes, but they still do not remove all our clinical measurement problems. Clinical measurement of attachment loss is, and always will be, a relatively inexact procedure.

Measurement is the foundation upon which science is built, and periodontal attachment levels are by no means the only biomedical measurements that have to get by with far less precision than we would like. It makes for less efficiency in research, though science can still prog-

ress so long as this human limitation is recognized. Studies must be carefully planned to ensure sufficient power; study protocols should include replicate measurements to help us judge the reliability of measurements; and alternative measures can be added as appropriate. Dental records, questionnaires, and radiographs are traditional added measures, and molecular biology is promising some new ones. Molecular biology holds the promise that we will have measurements with a degree of validity and precision that clinical measurements never will, but that does not mean that clinical measures will not remain a fundamental aspect of clinical and epidemiological research. The tradeoff between how much type I error (i.e., accepting that attachment loss has occurred when it really has not) and type II error (recording too many sites as unchanged when attachment loss really has occurred) can be accepted should be a standard part of the planning for any longitudinal study, as urged by Drs. van Palenstein Helderma and Timmerman.

Drs. van Palenstein Helderma and Timmerman offer no solution to this measurement problem, and indeed it seems there really isn't one beyond recognizing the imprecision of clinical measures and how to deal with that fact. Good science can still result, even if it takes more people in the study and more clinical measurements than we ideally would like. Whether the threshold for change should be 2 mm or 3 mm is for each researcher to determine and justify according to the aims of the study and the need to ensure adequate power. Our inability to measure clinical attachment loss more precisely is not the sign of an inadequate researcher; it is part of the human condition.—*Brian A. Burt, University of Michigan, Ann Arbor, MI.*