

Reliability of plaque and periodontal measurements estimated by the internal consistency method

Jostein Rise and Tore Tollefsen

Institute of Community Dentistry, and Department of Periodontology,
Dental Faculty, University of Oslo, Oslo, Norway

Rise J, Tollefsen T. Reliability of plaque and periodontal measurements estimated by the internal consistency method. *Acta Odontol Scand* 1984;42: 293–296. Oslo. ISSN 0001–6357.

The purpose of this paper was to assess whether the internal consistency method could be used to estimate reliability of periodontal and plaque measurements. This method is based on the assumption of comparability of data from the left and right halves of the mouth. The material comprised data from 61 subjects (mean age, 43 years). Percentage of tooth surfaces with plaque index scores of at least 1, percentage of bleeding units (GI = 2), and the mean pocket depth were calculated for each half of the mouth. At the group level the mean prevalence of the two sides of the mouth did not differ significantly with regard to any of the variables, and there was a substantial degree of bilateral symmetry for all variables as measured by Pearson's *r*. Reliability coefficients were computed in accordance with the Spearman–Brown formula. Highest reliability coefficient was observed for pocket depth measurements (0.97), followed by plaque measurements (0.95), and, finally, gingival bleeding measurements (0.90). The internal consistency method seems promising in estimating reliability in epidemiologic surveys, but further studies should be initiated to substantiate this contention. □ *Clinical methods; gingivitis; periodontology; pocket depth*

Jostein Rise, Institute of Community Dentistry, University of Oslo, P.O.B. 1052, Blindern, Oslo 3, Norway

Repeatability has been defined as 'the extent to which a test provides the same results on the same subjects on two or more occasions either in the hands of the same or different observers, the subjects of the test being in the same state of health or disease on the two occasions' (1). Reproducibility can be considered to be equivalent to repeatability (2), of which the test-retest approach, which is a measure of the proximity of scores obtained by the same examiner on two occasions, has been extensively used in periodontal research (see, for example, Refs. 3–6). It has been shown empirically that the use of probes as measuring devices inevitably changes the environment, and thereby the repeat examinations of plaque scores introduce systematic errors provided that the replications are performed within a short time interval (5, 6). On the other hand, for gingival scores the results were conflicting (5, 6). Furthermore, the examiners may remember their recordings from the first examination. The problems may to some extent be avoided by performing the dupli-

cate examinations on consecutive days (5), but instead errors may be introduced owing to the reversible nature of gingivitis and plaque (6). Concerning reproducibility of measuring pocket depth and loss of attachment, no systematic errors seem to have been detected (3, 7). Another disadvantage of the test-retest approach, particularly with regard to large-scale epidemiologic surveys, is that it requires reexamination of all or part of the study population.

The inherent weaknesses of the test-retest method do not seem to have been remedied through the use of alternative ways. In addition, the assessment of measurement error in terms of repeatability does not permit the influence of error to be quantified (8). This may be achieved by evaluating measurement error in terms of the concept of reliability. Reliability of a set of measurements can be defined as the proportion of the variance which is true variance, and the total variance is composed of true and error variances (9).

One likely candidate would appear to be

the internal consistency method (9), which has been used to measure reliability of caries prevalence and incremental data (2) and denture plaque (10). Although the method has been mentioned by Alman (11), it has not been used in empirical analysis of periodontal measurements. The rationale of the internal consistency method is based on the assumptions of no systematic differences of scores between the right and left sides of the mouth and of bilateral symmetry (2). The former assumption in terms of plaque and periodontal scores has been substantiated in empirical studies (12, 13). No empirical data seem to have been presented concerning the latter assumption of bilateral symmetry, which may be assessed by correlating the prevalence scores of the two halves for each individual.

The purpose of this study was to ascertain whether the two assumptions are upheld, to estimate the reliability of plaque, gingival, and pocket depth measurements by means of the internal consistency method.

Materials and method

The empirical basis of this paper consisted of data recorded in patients ($n = 61$) drawn from the files of a private dental practice (mean age, 43 years). Their periodontal status was assessed by a periodontist (T.T.). The plaque index (PII) and the gingival index (GI) were scored on four surfaces of all teeth present (14). Finally, the pocket depths of four surfaces of all teeth were measured to the nearest whole millimeter, using a periodontal probe. Pockets were probed at the mid-buccal and the mid-lingual aspects, whereas the approximal pockets were measured buccally to the contact point.

Treatment of the data

The prevalence of plaque, gingivitis, and periodontal disease was calculated for the right and left sides separately and expressed in terms of percentage of tooth surfaces with a plaque score of 1, percentage of bleeding units (GI = 2), and mean pocket depth. On the basis of prevalence scores of each half

of the mouth for the same individuals the hypothesis of no difference between the scores of the two halves was tested by means of a paired t test, which is suitable for correlated data (15).

To test the assumption of bilateral symmetry, Pearson's correlation coefficients were calculated between the scores of the three variables for the two sides of the mouth. Since this coefficient is numerically equal to the reliability coefficient of only one half of the mouth, a correction—the Spearman-Brown formula—was applied to provide an estimate of the reliability coefficient of both sides of the mouth (9):

$$r_{tt} = \frac{2r_{hh}}{1 + r_{hh}}$$

where r_{hh} is the correlation between the two halves (Pearson's r) and r_{tt} is the reliability coefficient of both halves of the mouth.

A basic definition of the reliability of a set of measurements is that it is the proportion of the total variance that is true variance (9). If it is assumed that the error component is independent of the true component, then the variance of the total scores (S_t^2) is the sum of the true variance (S_∞^2) and error variance (S_e^2) (9): $S_t^2 = S_\infty^2 + S_e^2$.

The basic equations for the coefficient of reliability (r_{tt}) would appear to be:

$$r_{tt} = \frac{S_\infty^2}{S_t^2} \quad \text{or} \quad r_{tt} = 1 - \frac{S_e^2}{S_t^2}$$

Then, if e^2 stands for the proportion of error variance in the total variance, $r_{tt} + e^2 = 1.00$.

Results

The mean prevalence of the two sides did not differ significantly with regard to any variable, and thus the null hypothesis of homogeneity of the two halves could not be rejected at the 5% level (Table 1). However, there was a tendency towards a higher percentage of bleeding units of the left than of the right side, and the variance tended to be consistently higher for all variables on the left side.

There was a substantial degree of bilateral symmetry for all variables as measured by Pearson's r (Table 1). The reliability coefficients for all three variables reached the numerical value of 0.90. The highest coefficient was obtained for pocket depth (0.97) and the lowest for gingival scores (0.90), with that for plaque scores in an intermediate position (0.95).

Discussion

According to Guilford & Fruchter (9), the concept of internal consistency comes closest to the basic idea of reliability, providing estimates of 'on-the-spot' reliability. However, comparability of the two halves of the mouth with regard to the phenomenon being studied must be assumed. At the group level, comparability is indicated by prevalence scores in terms of means and variances. To judge from the present study, this assumption is tenable for pocket depth, plaque scores, and gingival scores, and accordingly, no systematic differences seem to exist between the halves. Admittedly, the means and variances of gingival scores tended to be higher on the left than on the right side, but the difference was not statistically significant. These data differ from those of Alexander (13), who consistently observed higher gingival scores for the right side than for the left side. However, since neither of these differences was statistically significant, chance circumstance may presumably account for these divergent results.

Bilateral symmetry may be assessed by

correlating the scores of the two sides for each individual. The bilateral symmetry as measured by Pearson's r turned out to be high for pocket depth and plaque scores but not so high for gingival bleeding. Since these estimates represent the reliability coefficients of only one half of the mouth, the correction formula had to be applied to determine the reliability coefficient of both sides of the mouth.

Since there will always be some true differences between the right and left sides of the mouth, the fraction of the error variance due to this difference cannot be separated from the genuine error variance (2). Thus, the internal consistency method cannot provide a precise estimate of the error variance in order to study the influence of error on the efficiency of a clinical trial. However, in this study the genuine error variances of all variables were less than 1/10th ($r > 0.90$) of the total variance.

The greatest advantage of the internal consistency method as compared with the test-retest method is that study subjects, or a sample of them, do not have to be re-examined. It is thus applicable and seems to be the method of choice, particularly in large-scale epidemiologic surveys.

The present material is derived from a limited number of patients in a private practice, and broad generalizations of the results to other populations may be unwarranted. Therefore, epidemiologic surveys should be initiated to ascertain whether the assumptions of no systematic differences between the two sides of the mouth and of bilateral symmetry are upheld.

Table 1. Reliability coefficients and Pearson's r of plaque scores, gingivitis and pocket depth estimated by the internal consistency method. Means and variances of the right (R) and left (L) sides are also given ($n = 61$)

| Variable | Side | Mean | Variance | t values, d.f. = 60 | Pearson's r | Reliability coefficient |
|---------------------|------|------|----------|--------------------------|------------------|----------------------------|
| Plaque scores (%) | R | 45.9 | 461.65 | $t = 0.234$ | 0.90 | 0.95 |
| | L | 46.2 | 480.03 | | | |
| Gingival scores (%) | R | 22.8 | 367.67 | $t = 1.128$ | 0.82 | 0.90 |
| | L | 24.7 | 439.93 | | | |
| Pocket depth (mm) | R | 2.4 | 0.23 | $t = 0.110$ | 0.94 | 0.97 |
| | L | 2.4 | 0.27 | | | |

References

1. Fletcher CM. Criteria for diagnosis and assessment in clinical trials. In: Bradford Hill A, ed. *Controlled clinical trials*. Oxford: Blackwell, 1960;19.
2. Rugg-Gunn AJ, Holloway PJ. Methods of measuring the reliability of caries prevalence and incremental data. *Community Dent Oral Epidemiol* 1974;2:287-94.
3. Glavind L, Løe H. Errors in the clinical assessment of periodontal destruction. *J Periodontal Res* 1976;2:180-4.
4. Clemmer BA, Barbano JP. Reproducibility of periodontal scores in clinical trials. *J Periodontal Res* 1974;9(suppl 14):118-22.
5. Birkeland JM, Jorkjend L. The influence of examination on the assessment of the intra-examiner error by using the plaque and gingival index systems. *Community Dent Oral Epidemiol* 1975;3:214-6.
6. Shaw L, Murray JJ. Diagnostic reproducibility of periodontal indices. *J Periodontal Res* 1977;12:141-7.
7. Abbas F, Hart AAM, Oasting J, Velden van der U. Effect of training and probing force on the reproducibility of pocket depth measurements. *J Periodontal Res* 1982;17:226-34.
8. Rugg-Gunn AJ, Downer MC, Ashley FP. Reliability of caries data in three clinical trials. *Community Dent Oral Epidemiol* 1976;4:15-8.
9. Guilford JP, Fruchter B. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1973, 407-34.
10. Ambjørnsen E, Rise J, Haugejorden O. A study of examiner errors associated with measurement of denture plaque. *Acta Odontol Scand* 1984;42:183-91.
11. Alman JE. Reproducibility of periodontal scores in clinical trials. Formal discussion. *J Periodontal Res* 1974;9(suppl 14):123-5.
12. Suomi J, Barbano JP. Patterns of gingivitis. *J Periodontol* 1968;39:71-4.
13. Alexander AC. Partial mouth recording of gingivitis, plaque and calculus in epidemiological surveys. *J Periodontal Res* 1970;5:141-7.
14. Løe H. The gingival index, the plaque index and the retention index system. *J Periodontol* 1967;38:610-6.
15. Daniel WW. *Biostatistics: a foundation for analysis in the health sciences*. New York: John Wiley & Sons, Inc. 1974;163-78.

Received for publication 18 July 1983