J Clin Periodontol 2004; 31: 318–323 doi: 10.1111/j.1600-051X.2004.00448.x

Journal of Clinical Periodontology

Reproducibility and accuracy of automated probe measurements of gingiva and bone levels on stone casts following guided bone regeneration treatment

Erwin Van der Zee, Marie-Faustine Vogels, Paul Oosterveld and Marinus Van Waas

Clinic for Oral Implantology, Department of Oral Function, Academic Centre for Dentistry (ACTA), The Netherlands

Van der Zee E, Vogels M-F, Oosterveld P, Van Waas M: Reproducibility and accuracy of automated probe measurements of gingiva and bone levels on stone casts following guided bone regeneration treatment. J Clin Periodontol 2004; 31: 318–323. doi: 10.1111/j.1600-051X.2004.00448.x © Blackwell Munksgaard, 2004.

Abstract

Objectives: For evidence-based evaluation of guided bone regeneration (GBR), accurate registration of changes in gingiva and bone levels is needed. A new method is introduced and evaluated.

Methods: In a clinical trial with 30 patients, alginate impressions of the surgical area including the interproximal gingiva and alveolar bone at the adjacent teeth were made in duplicate prior to and during GBR surgery, fixture installation and abutment connection. Poured in hard stone, the casts were used for repeated measurements of the level of the free gingival margin and the alveolar bone with an automated probe (Florida disc-probe®), using the incisal edge as a fixed reference point. The reproducibility and accuracy of these measurements were evaluated by means of the Intraclass Correlation Coefficients and Generalizability Theory. The effect of treatment was evaluated by multivariate analysis of variance.

Results: Generalizability Theory indicated a high accuracy of the gingiva- and bone-level measurements: the Intraclass Correlation Coefficients for gingiva and bone levels were 0.99 and 0.98, respectively. The intra-cast reproducibility was 0.09 ± 0.07 mm (mean ± SD) and the inter-cast reproducibility was 0.10 ± 0.09 and 0.20 ± 0.07 mm for gingiva and bone levels, respectively. Clinical applicability is demonstrated by the fact that MANOVA revealed on average a small but highly significant ($p = 0.001$) effect of the staged surgical intervention on the gingiva and bone levels at the adjacent teeth.

Conclusion: It is concluded that the presented method makes it possible to evaluate reproducibly and accurately changes in gingiva and bone levels for GBR studies.

Key words: bone resorption; Florida probe; generalizability theory; gingiva recession; guided bone regeneration; oral implantology; probing

Accepted for publication 31 March 2003

Guided bone regeneration (GBR) techniques are used to improve the volume and contour of alveolar bone in order to optimize the position of oral implants and the aesthetics of the soft tissue (Nyman & Lang 1994). Relatively few quantitative data are published in prospective longitudinal clinical trials as needed for evidence-based dentistry; the results are mostly presented in case reports. Accurate and reproducible evaluation methods are lacking. When specified, bone measurements are often clinically recorded with a vernier calliper or a conventional periodontal probe. The location of the gingival margin is mostly recorded with a conventional periodontal probe relative to the cemento-enamel junction (CEJ). These methods have disadvantages: the measurements can only be made at one specific moment – during surgery – and thus cannot be repeated afterwards, and the CEJ is not always easily detectable. Moreover, these recordings are read by the surgeon from the calibration markings placed with a 1–2 mm interval at the tine of the probe, and therefore are relatively crude estimations in 0.5–1 mm steps. A long time interval between different measurements may compromise the precision and comparability: in case of three surgical sessions, 12 months or more are between the measurements at the same sites. This means that these measurements are biased by factors depending on both the surgeon and the measuring device, including the location and angle of insertion, angle and accuracy of visual observation and accuracy of the calibration device (Badersten et al. 1984, Van der Zee et al. 1991).

In the present study, stone casts of the gingiva and alveolar bone were made during GBR treatment for longitudinal evaluation of a group of 30 patients. The levels of the free gingival margin and the alveolar bone at the adjacent teeth were measured on these casts using an automated probe (Florida disc-probe®, Florida Probe Corporation, Gainesville, FL, USA; Marks et al. 1991). The aim was to determine whether these measurements could be used for reproducible and accurate evaluation of changes in these parameters following treatment.

Material and Methods

The study sample consisted of 30 patients with an edentulous space in the maxillary front and/or pre-molar region and received oral implants at the clinic of oral implantology of the Academic Centre for Dentistry (ACTA). The patients met the inclusion criteria of a longitudinal Guided Bone Regeneration study approved by the Medical Ethical Committee of the faculty. After obtaining informed consent, a GBR treatment was applied as described previously (Jovanovic & Nevins 1995, Van der Zee 1999). The surgical protocol involved three sessions: GBR, fixture installation (Fixt) and abutment connection (Abut), each with a 6-month healing interval.

The dimensions of the bone of the alveolar process at the edentulous space as well as of the proximal gingiva and alveolar bone at the adjacent teeth were copied in hard stone casts.

Two alginate impressions (Cavex® CA 37 fast setting sachets, Cavex Holland BV, Haarlem, The Netherlands) were taken of the maxilla before each surgical session using impression trays. Subsequently, while the soft tissues were reflected during a standard mucoperiosteal flap with vestibular vertical releasing incisions, two impressions of the bony defect were made during each surgical session with individualized partial impression trays and immediately poured in hard stone (Fig. 1c, d, f). This resulted in 12 casts per patient: two prior to surgery and two during surgery at the three surgical sessions.

At each session, the levels of the gingiva and the alveolar bone at the adjacent teeth were measured with a conventional William's periodontal probe at the two interproximal sites (S1+S2) facing the edentulous space. After completion of the treatment of all patients, the casts were measured in random order at the same sites with a Florida-probe® (Magnusson et al. 1988). This is a standardized pressure probe equipped with a long probe tine and a disc in order to use the incisal edge as a fixed reference point for measurements (Marks et al. 1991). It is equipped with calibration steps of

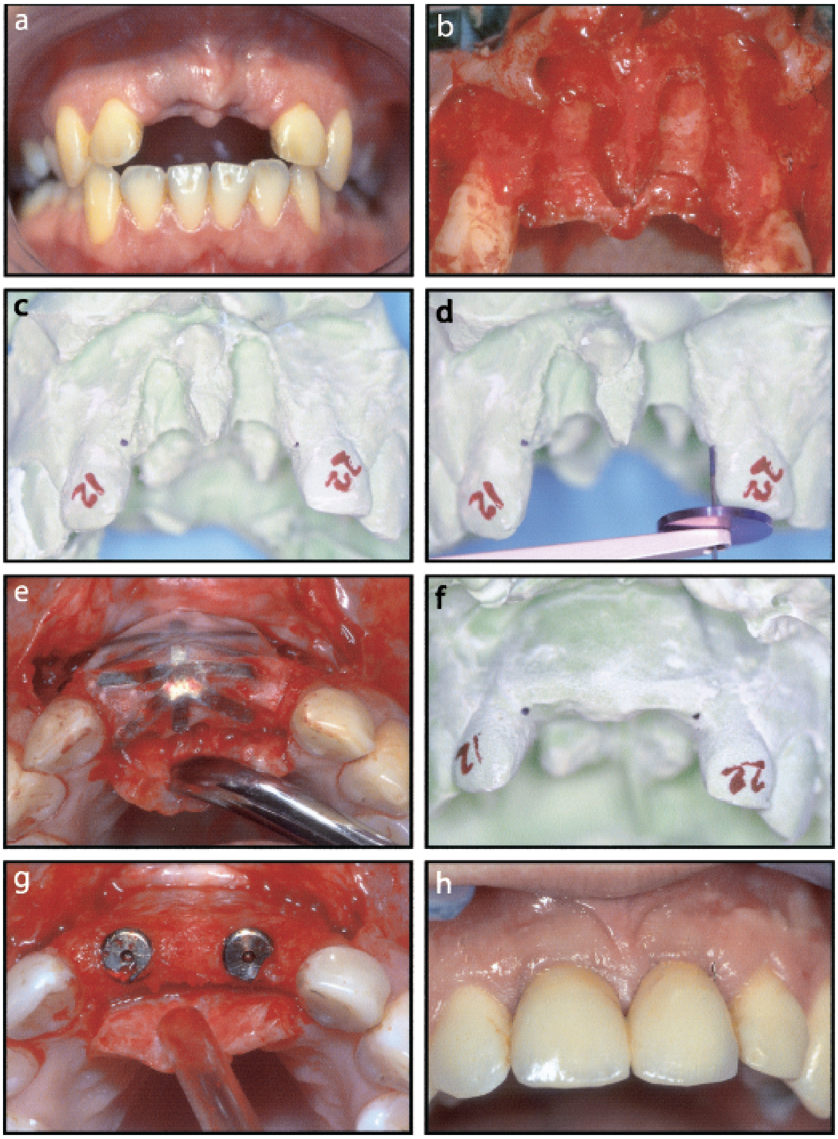

Fig. 1. Example of clinical situation and stone cast during treatment. (a) Clinical situation prior to GBR treatment (female 27 years of age; missing 11 in 21 since 8 weeks due to endodontic failure; primary trauma occurred 2 years ago). (b) Clinical situation of defect area at first surgery. (c) Stone cast of defect area at first surgery with marked sites (tiny ink spot) at adjacent teeth. (d) Stone cast of defect area at first surgery with Florida probe measurement of bone level at marked sites. (e) Clinical situation at second surgery 6 months after GBR treatment and just prior to membrane removal and fixture installation. (f) Stone cast at second surgery just after membrane removal and prior to fixture installation with marked sites at adjacent teeth. (g) Clinical situation at third surgery 6 months after fixture installation and just prior to abutment connection. (h) Clinical result after completion of the treatment with porcelain-metal crowns in situ.

0.2 mm, and has digital readout and storing facilities.

The measurements at the duplicate casts allow the estimation of the inter-cast reproducibility. For the estimation of the intra-cast reproducibility, the bone-level measurements were made in triplicate per cast.

The reproducibility of the measurements was estimated by means of the Generalizability Theory (Cronbach et al. 1972, Cardinet et al. 1981). It is expressed in the Generalizability Coefficient (GC, a measure similar to the Intraclass Correlation Coefficient, ICC; Fleiss & Kingman 1990) and the standard error of measurement. The GC is the ICC for cases where several observed factors (in this context called ''facets'') influence a measurement. The influence of these facets is investigated by calculating the amount of variance that they take into account. For the bone-level measurements five facets were evaluated: patient, site, surgical session, cast and replication. For the gingiva-level measurement there is no replication facet. In the context of the Generalizability Theory, a distinction is made between ''differentiation facets'', i.e., aspects of the units of analysis that we want to investigate, and ''instrumentation facets'', i.e., aspects of the measurement procedure (Cardinet et al. 1981). In the current case ''patient'', ''site'' and ''surgical session'' are differentiation facets; ''cast'' and ''replication'' are instrumentation facets. If there is one differentiation and one instrumentation facet, the GC is identical to the ICC; if there are more facets, the ICC cannot be calculated. In other words, the GC is a generalization of the ICC. The interpretation of the GC and the ICC is the same (Swanson et al. 1999). An attractive feature of Generalizability Theory is that it can be used not only to estimate the reproducibility of the actual measurement procedure (a so-called Generalizability- or G-study) but also to estimate the reproducibility for alternative designs of the measurement, e.g., more or less replications and casts (therefore called a Design- or D-study). A D-study does not require additional data; it utilizes the variance estimates from the G-study.

To investigate the effects of ''casts'' and ''replication'', the data were also analysed with multivariate analysis of variance (MANOVA). In the analyses, the gingiva and bone measurements are the dependent variables, ''site'' and ''surgical session'' are within-subject factors. The effect of different GBR treatments was investigated by running the MANOVA with a between-subjects factor. A significance level of 5% was used.

The validity of the method was investigated by comparing the results of the indirect measurements on the casts with the results of the direct clinical measurement, using the correlation coefficient. This was calculated per ''surgical session'' and ''site'' and summarized in a mean correlation coefficient.

The data of two patients were incomplete: one patient did not report in time for abutment surgery and of another patient, the stone casts of the abutment stage were lost in the dental lab.

Results

Reproducibility of the measurements

The mean difference and standard deviation between repeated bone-level measurements on the same stone cast are 0.09 ± 0.07 mm. The 95% confidence intervals of intra-cast reproducibility are similar for the three surgical sessions (GBR, fixture installation, abutment connection), for both sites per cast and for both casts (Fig. 2). MANOVA reveals no statistically significant differences ($F_{1,11} = 0.73$; $p = 0.69$).

The mean difference and standard deviation of bone-level measurements between duplicate casts are 0.20 mm $\pm$ 0.17. The 95% confidence intervals of inter-cast reproducibility of the bone-level measurements for both sites and for the three surgical sessions are displayed in Fig. 3. MANOVA shows no statistically significant differences ($F_{1,5} = 1.91$; $p = 0.13$).

Comparing the gingiva level between duplicate casts, there is a mean difference and standard deviation of 0.10 ± 0.09 mm. The 95% confidence intervals of inter-cast reproducibility for gingiva levels are similar for the three surgical sessions for each site (Fig. 4). Again, MANOVA reveals no statistically significant differences ($F_{1,5} = 0.64$; $p = 0.67$).

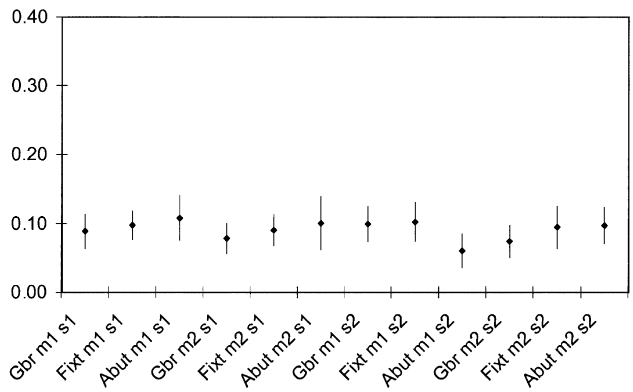

Fig. 2. Intra-cast reproducibility. 95% confidence interval of the mean deviation for triple repeated bone-level measurements of 30 patients with the automated probe at GBR surgery (Gbr), fixture installation (Fixt) and abutment connection (Abut) as registered on two stone casts (m1 and m2) at both sites (S1 and S2). Data represent mean values in mm $\pm$ SEM of 30 patients.

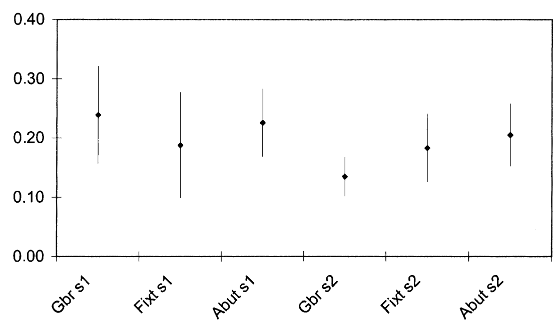

Fig. 3. Inter-cast reproducibility of bone-level measurements. 95% confidence interval of the mean deviation for bone-level measurements on duplicate stone casts with the automated probe at GBR surgery (Gbr), fixture installation (Fixt) and abutment connection (Abut) at both sites (S1 and S2). Data represent mean values in mm $\pm$ SEM of 30 patients.

The accuracy of the measurements

The results of the application of the Generalizability Theory for the estimation of the accuracy are reported in Table 1. In this table, the results of the G- and D-study are given. In the D-study, the effect of replicate measurements and casts is investigated by varying the levels of these factors. For bone-level measurements, there is a larger increase in the standard error of measurement for the reduction of the number of casts than for the gingiva-level measurement.

Validity of gingiva and bone measurements

The distance from the incisal edge to the gingiva and alveolar bone at adjacent teeth during GBR, fixture and abutment surgery is shown in Fig. 5. On average, there is an overall decrease of 0.82 mm in the gingiva level and 0.13 mm in the bone level measured on the stone casts compared with 0.75 and 0.34 mm, respectively, as assessed clinically with a conventional manual probe. A mean correlation of 0.62 was found between the conventional direct clinical measurements and the indirect cast measurements.

Effect of surgical session and site

MANOVA revealed highly statistically significant differences in the gingiva and bone levels between the surgical sessions ($F_{2,25} = 9.71$; $p = 0.001$) and between the left and right sites of the edentulous space ($F_{1,26} = 12.65$; $p = 0.001$).

Discussion

The Florida probe is a clinically well-tested device introduced to make periodontal attachment-level measurements more accurate and reproducible by eliminating or substantially reducing instrumental and operator variables and errors (Magnusson et al. 1988, Marks et al. 1991). To accomplish this, the automated probe is equipped with a standardized pressure and has the possibility of using the incisal edge as a fixed reference point for a disc-tine (Marks et al. 1991, Reddy et al. 1997). The probe, originally equipped with a 0.1 mm resolution, nowadays has calibration steps of 0.2 mm, which is considered to be very precise for a clinical probing measurement (Reddy et al. 1997). Moreover, it has digital readout and storing facilities eliminating ''observation angle'' and transcription errors (Reddy et al. 1997). In this study, we used this device to optimize data collection from stone casts.

The clinical applicability and relevance of the indirect measurement procedure is demonstrated by the fact

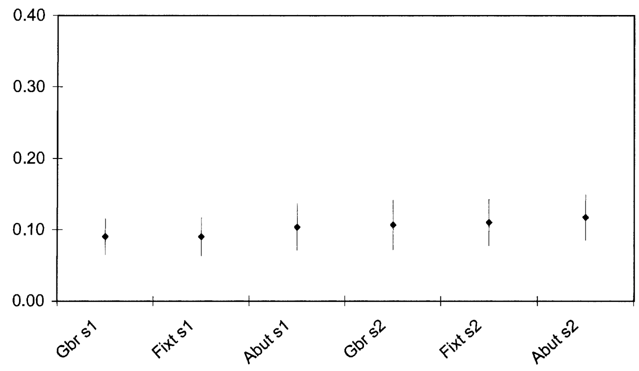

Fig. 4. Inter-cast reproducibility of gingiva-level measurements. 95% confidence interval of the mean deviation for gingiva-level measurements on duplicate stone casts with the automated probe at GBR surgery (Gbr), fixture installation (Fixt) and abutment connection (Abut) at both sites (S1 and S2). Data represent mean values in mm ± SEM of 30 patients.

Table 1. Accuracy estimates for gingiva- and bone-level measurement in G- and D-study

Accuracy estimates	Gingiva level	Bone level
G-study		
generalizability coefficient	0.99	0.98
standard error of measurement	0.13	0.27
D-study		
one cast per session, one replication		
generalizability coefficient	0.99	0.98
standard error of measurement	0.19	0.40
two casts per session, one replication		
generalizability coefficient	0.99*	0.98
standard error of measurement	0.13*	0.29
one cast per session, three replications		
generalizability coefficient		0.97
standard error of measurement		0.38

Note: the coefficients reported are for absolute measurement.
*Design identical to the G-study.

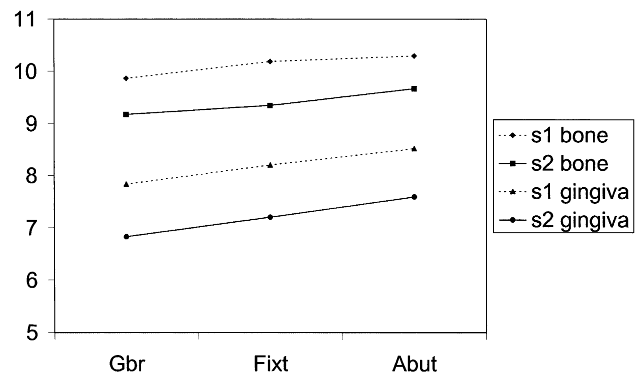

Fig. 5. Gingiva and bone levels. Levels of the gingiva and alveolar bone relative to the incisal edge of the two teeth adjacent to the GBR-treated defect at the two interproximal sites facing the edentulous space (S1+S2) as measured at the GBR surgery (Gbr), fixture installation (Fixt) and abutment connection (Abut) surgery at both sites (S1 and S2). Data represent mean values in mm of 30 patients.

that MANOVA revealed a small but highly significant effect of the staged surgical intervention on the interproximal gingiva (0.82 mm recession) and bone levels (0.13 mm resorption) at adjacent teeth. The differences between the left and right sites of the edentulous space are due to the differences in crown length between the two adjacent teeth in asymmetrical edentulous space and therefore trivial. The mean gingiva recession and bone resorption as assessed by the indirect and direct clinical measurements are similar (within 0.2 mm). This demonstrates the validity of the indirect measurements.

The mean correlation between the measurements on stone casts and clinical measurements shows a reasonable level of agreement. It should be noted that this correlation is dependent on the accuracy of both types of measurements. It seems fair to suggest that the accuracy of the clinical measurements is compromised by the factors mentioned above. Since it is not possible to replicate the clinical measurements blindly at the moment of surgery with conventional probing, the reproducibility of these measurements could not be calculated.

The GCs of the indirect measurement procedure are high, indicating that there is high consistency over casts and replicate measurements. The standard errors are quite low; for both gingiva and bone-level measurement, the 95% confidence interval for a score is less than 1 mm, which makes the measurement procedure accurate.

The high reproducibility of the indirect method can also be concluded from the size of the effects of the replicate measurements and casts. The observed mean differences for both casts and replications are low and not significantly different. For the bone level, this indicates that it is possible to perform the measurements equally accurately prior to GBR, when the bone contour is still irregular, as after completion of this procedure.

The Generalizability Theory results, particularly the standard error of measurements, indicate that the gingiva-level measurements are more accurate than the bone-level measurements; the D-study shows that the duplication of the cast has more impact on the SEM for bone-level measurement than for gingiva-level measurement. Also, intercast reproducibility is lower for the bone level than for the gingiva level, as can be seen in the larger mean difference. Flap reflection and blood are compromising factors when making a bone impression during surgery. These may influence reproducibility.

Using the Generalizability Theory, we can estimate both the GC and the standard error for other study designs. It is found that replications only marginally increase the accuracy of the presented method. The use of duplicate casts in case of bone-level measurements gives some additional information as it results in a small increase in standard error when omitted. Nevertheless, even a minimal measurement procedure involving only one cast per surgical session and no replications still results in a high accuracy for both gingiva- and bone-level measurements.

The observed reproducibility of the indirect measuring technique seems at least in the same order of magnitude as the Florida periodontal probe calibration steps of 0.2 mm. The observed accuracy and reproducibility of gingiva- and bone-level measurements make the evaluation method with this device reliable and suitable for clinical purposes, where probing measurement errors less than 1 mm are advocated (Reddy et al. 1997). In studies using periodontal attachment levels, mean differences recorded with the Florida disc-probe were higher: from 0.33 mm (Grossi et al. 1996) to 0.55 mm (Reddy et al. 1997). With respect to the standard deviation, the presently found mean probing error of 0.09 mm is low compared with that found in attachment-level reproducibility studies (0.3–0.6 mm) using similar automated Florida disc-probes (Gibbs et al. 1988, Osborn et al. 1990, Marks et al. 1991, Yang et al. 1992). An obvious advantage of gingiva- and bone-level measurements as in the present study is the absence of pocket penetration that contributes to inconsistencies in the attachment-level and pocket depth measurements even when standardised pressure is used.

There are more advantages of the indirect measuring technique as presented in this study. Instead of using a surgical stent during clinical measurements (Watts 1987), all sites are marked identically relatively easily with a tiny ink spot at the different stone casts before probe recordings. Subsequently, the data of different surgeries can be registered at one and the same recording session, which facilitates standardization of the evaluation process per site. Owing to the digital readout and storing facilities, this can be carried out in a randomized manner without knowing which patient at what stage was measured. In this way, the use of different sets of measuring instruments is also avoided between the different recording sessions, and inter-operator and intra-operator variations over time are minimized. Another advantage of the indirect technique is that data collection can be repeated as often as required at any convenient moment other than during surgery, and the results can be refined and re-interpreted after restudying the casts.

It is concluded that indirect gingiva- and bone-level measurements on stone casts provide a valuable registration method for the evaluation of hard- and soft-tissue dimensions in longitudinal studies on tissue regeneration techniques.

Acknowledgments

The authors are indebted to Dr. Frans Oort for his valuable comments on the application of Generalizability Theory. All members of the staff of the Clinic of Oral Implantology of ACTA are gratefully acknowledged for their interest in the present study. We especially thank Wieneke Middeldorp, Neeltje Visser and Jacqueline de Cleen for their assistance during clinical treatment sessions and Hanneke Comis and Francis Niessen and Tineke Fortuin for their support in secretarial work. We thank Klaas-Jan van Egmond of the Graphic Design Department for his support in presentation of the photographic material.

References

Badersten, A., Nilveus, R. & Egelberg, J. (1984) Reproducibility of probing attachment level measurements. *Journal of Clinical Periodontology* **11**, 475–485.

Cardinet, J., Tourneur, Y. & Allal, L. (1981) Extension of generalizability theory and its applications in educational measurement. *Journal of Educational Measurement* **18**, 183–204.

Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972) *The dependability of behavioral measurements: theory of generalizability for scores of profiles*. New York: Wiley.

Fleiss, J. L. & Kingman, A. (1990) Statistical management of data in clinical research. *Critical Reviews Oral Biology and Medicine* **1**, 55–66.

Gibbs, C. H., Hirschfeld, J. W., Lee, J. G., Low, S. B., Magnusson, I., Thousand, R. R., Yerneni, P. & Clark, W. B. (1988) Description and clinical evaluation of a new computerized periodontal probe-the Florida probe. *Journal of Clinical Periodontology* **15**, 137–144.

Grossi, S. G., Dunford, R. G., Ho, A., Koch, G., Machtei, E. E. & Genco, R. J. (1996) Sources of error for periodontal probing measurements. *Journal of Periodontal Research* **31**, 330–336.

Jovanovic, S. A. & Nevins, M. (1995) Bone formation utilizing titanium-reinforced barrier membranes. *International Journal of Periodontics and Restorative Dentistry* **1**, 56–69.

Magnusson, I., Fuller, W. W., Heins, P. J., Rau, C. F., Gibbs, C. H., Marks, R. G. & Clark, W. B. (1988) Correlation between electronic and visual readings of pocket depths with a newly developed constant force probe. *Journal of Clinical Periodontology* **15**, 180–184.

Marks, R. G., Low, S. B., Taylor, M., Baggs, R., Magnusson, I. & Clark, W. B. (1991) Reproducibility of attachment level measurements with two models of the Florida Probe. *Journal of Clinical Periodontology* **18**, 780–784.

Nyman, S. R. & Lang, N. P. (1994) Guided tissue regeneration and dental implants. *Periodontology 2000* **4**, 109–118.

Osborn, J., Stoltenberg, J., Huso, B., Aeppli, D. & Pihlstrom, B. (1990) Comparison of measurement variability using a standard and constant force periodontal probe. *Journal of Periodontology* **61**, 497–503.

Reddy, M. S., Palcanis, K. G. & Geurs, N. C. (1997) A comparison of manual and controlled-force attachment-level measurements. *Journal of Clinical Periodontology* **24**, 920–926.

Swanson, D. B., Clauser, B. E. & Case, S. M. (1999) Clinical skills assessment with standardized patients in high-stake tests: a framework for thinking about score precision, equating, and security. *Advances in Health Sciences Education* **4**, 67–106.

Van der Zee, E. (1999) Geleide botregeneratie (GBR) ten behoeve van solitaire tandvervangingen met implantaten. *Nederlands Tijdschrift voor Tandheelkunde* **106**, 195–198.

Van der Zee, E., Davies, E. H. & Newman, H. N. (1991) Marking width, calibration from tip and tine diameter of periodontal probes. *Journal of Clinical Periodontology* **18**, 516–520.

Watts, T. (1987) Constant force probing with and without a stent in untreated periodontal disease: the clinical reproducibility problem and possible sources of error. *Journal of Clinical Periodontology* **14**, 407–411.

Yang, M. C., Marks, R. G., Magnusson, I., Clouser, B. & Clark, W. B. (1992) Reproducibility of an electronic probe in relative attachment level measurements. *Journal of Clinical Periodontology* **19**, 541–548.

Address:
Erwin Van der Zee
Clinic for Oral Implantology
Department of Oral Function
Academic Centre for Dentistry (ACTA)
Louwesweg 1
NL-1068 EA Amsterdam
The Netherlands
Tel.: +31 20 51 88 412
Fax: +31 20 51 88 414
E-mail: E.vd.Zee@acta.nl