Intra- and inter-examiner reproducibility of manual probing depth

Roberto Andrade[a]
Manuel Espinoza[a]
Elena Maria Gómez[a]
José Rolando Espinoza[a]
Elizabeth Cruz[a]

[a] Private practice. San Salvador, El Salvador.

Abstract: The periodontal probe remains the best clinical diagnostic tool for the collection of information regarding the health status and the attachment level of periodontal tissues. The aim of this study was to evaluate intra- and inter-examiner reproducibility of probing depth (PD) measurements made with a manual probe. With the approval of an Ethics Committee, 20 individuals without periodontal disease were selected if they presented at least 6 teeth per quadrant. Using a Williams periodontal probe, three calibrated thesis-level students ($k > 0.6$) assessed PD at 6 sites per tooth, from the gingival margin to the bottom of the periodontal sulcus (rounded to the next 0.5 mm). Initial and repeated measurements were performed by the same three examiners. The intra-examiner agreement ($\pm$ 1 mm > 90%) was 99.85%, 100%, and 100% for the three examiners, respectively. When the variables vestibular/lingual surfaces, mesial/distal surfaces, or superior/inferior jaws were evaluated, no significant differences in reproducibility were detected at the inter-examiner level ($p < 0.05$). At this level, the only significant differences observed were in the three examiners' measurements of the anterior and posterior sites. While high intra-examiner reproducibility was detected, inter-examiner level proved to be low. We can conclude that measurement of PD with a manual periodontal probe produced high reproducibility in healthy individuals. The operator's position can affect the reproducibility of repeated measures of PD. Calibration and operator training, rather than operator experience, were fundamental for reproducibility. Other factors, such as individual technique and probing depth force, can affect inter-examiner reproducibility.

Descriptors: Dental Instruments; Reproducibility of Results; Diagnostic Errors; Calibration.

The trial is registered in the Australian-New Zealand Clinical Trials Registry (ANZCTR). ACTRN: ACTRN12611000044921

Declaration of Interests: The authors certify that they have no commercial or associative interest that represents a conflict of interest in connection with the manuscript.

Corresponding author:
Roberto Andrade
E-mail:
robertoandrade_periodoncia@yahoo.com.br

Received for publication on Aug 16, 2011
Accepted for publication on Nov 16, 2011

Introduction

Traditionally, investigators have used the periodontal probe to detect the presence and progression of some periodontal diseases, such as chronic periodontitis, by evaluating numerical data and/or clinical signs of inflammation.[1-3] Typically evaluated parameters include probing depth, gingival levels, presence or absence of bleeding, and clinical attachment levels.[2,4-6] The more exact the measuring tools are, the more reproducible the measurements will be; additionally, a higher level of control of the variables that affect probing will be achieved. Furthermore, an accuracy

of the measuring tools will provide more precise information for diagnosis and earlier detection of disease, which, in turn, can foster proper and timely treatment, as well as long-term periodontal control. Several factors affect periodontal parameters recorded through clinical measurements, and these can lead to a misinterpretation of an individual's true periodontal condition, resulting in a lack or an excess of treatment. Some of these factors are: presence of swollen gingival tissue, dental prosthetics, dental calculus, and diameter of and/or variations in the standardization of probe marks, as well as probing force and operator errors. The aim of this study was to compare inter- and intra-examiner reproducibility using a manual probe to obtain a clinical record of probing depth, with different variables considered.

Methodology

Before beginning the study, three examiners were trained in a calibration process. In a calibration group (not part of the experimental group), study participants were probed by each examiner, using a Williams SE manual probe (Hu-Friedy Co., Chicago, USA), once a week over a period of 4 months. The probing process was repeated until each examiner had substantial correlation as measured by Cohen's Kappa ($k \geq 0.6$). ln addition to the Kappa agreement, the measurements had to show a 90% agreement for ± 1 mm, as well as an exact agreement in 75% of the PD repeated measurements. Once the examiners were trained, the study was initiated. Twenty students (mean age, 21 years) from the Dentistry Program at the Evangelica University of El Salvador were selected (Institutional Ethics Committee Certification No. 9). Exclusion criteria for individuals participating in the study included: loss of clinical attachment levels, the presence of fixed or partial prosthetics, individual crowns, or orthodontics, as well as surgical or non-surgical periodontal treatments in the preceding 8 months, and pregnancy. The trained examiners measured PD in the study participants, to determine the intra- and inter-examiner reproducibility. Probing depth was defined as the distance between the gingival margin and the bottom of the sulcus/pocket.[5] The examiners used a Williams SE manual periodontal probe (1, 2, 3, 5, 7, 8, 9, and 10 mm), which was introduced into the interior of the gingival sulcus following the length of the tooth until resistance was felt by the penetrating probe. Each participant had to have a minimum of 6 teeth per quadrant to be included in the study, and 6 sites per tooth were probed (mesiobuccal, buccal, distal buccal, mesiolingual, lingual, and distolingual), excluding third molars. The PD measurements were done according to the following criteria: All measurements were rounded to the closest 0.5 mm (up or down), and when the PD measurement was halfway between 2 marks on the probe, the closest millimeter immediately above the mark was recorded. The initial probing depths (baseline) were taken as follows: On day 1, examiner 1 did an initial PD; two days later (day 3), examiner 2 did an initial PD; and on day 5, examiner 3 did an initial PD (two days' difference between measurements). On day 7, examiner 1 began the second round (repeated measurements) of PD. Examiners 2 and 3 probed on days 9 and 11, respectively. All these procedures were conducted on participant number 1. This same methodology was used on all 20 participants until complete data were recorded for all of them. Each individual was probed a maximum of 6 times (2 probes per examiner: the initial and the repeated probing, with two days' difference between measurements). Therefore, the 20 participants were probed twice in different appointments by the three examiners. The sequence of the examiners was random. The results were analyzed statistically by a *t*-test for independent samples (intra-examiner reproducibility) and an analysis of variance (ANOVA) for inter-examiner reproducibility ($p = 0.01$). By evaluating the results from the first PD (initial) and the repeated PD measurements taken by the three trained examiners, we could calculate the Kappa value and create mathematical measurements for subsequent statistical analyses. The statistical package used was SPSS for Windows 13.0 (IBM Corporation, Chicago, USA).

Results

Each examiner took 2 PD measurements per site (6,564), for a total of 19,692 measurements among the three examiners.

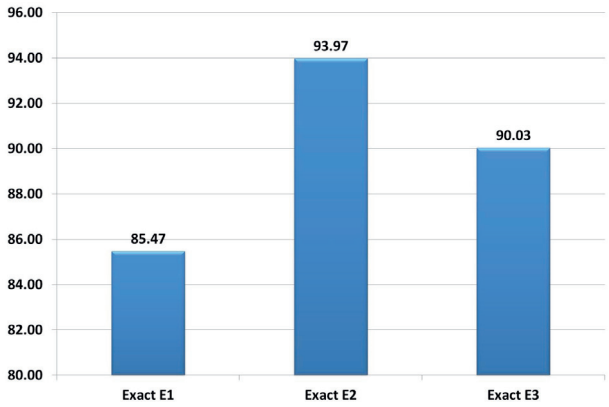

Figure 1 - Percentage of exact intra-examiner correlation > 75% between the initial and repeated probes.

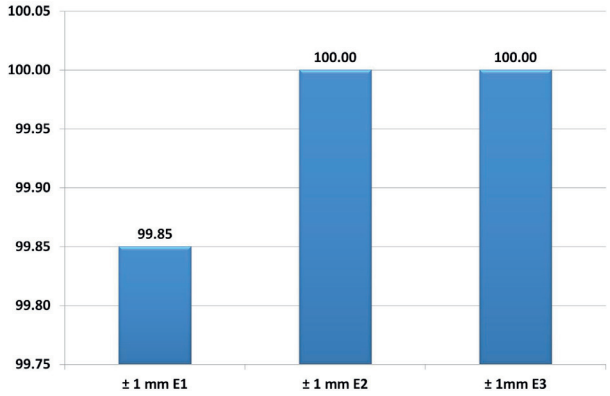

Figure 2 - Percentage of intra-examiner correlation > 90% for ± 1 mm between the initial and repeated probes.

Intra-examiner reproducibility

Based on a criterion of an exact agreement > 75% between the initial and repeated probes per examiner, it was shown that examiner 1 had an exact agreement of 85.47%; examiner 2, 93.97%; and examiner 3, 90.03% (Figure 1).

Based on the criterion of exact agreement > 90% for ± 1 mm between the initial and repeated PD per examiner, it was shown that examiner 1 had an agreement of 99.85%, while examiners 2 and 3 had 100% agreement (Figure 2).

Analysis of intra-examiner variables

A statistical *t*-test for independent samples was used to determine if there was a significant difference between the initial PD measurements and the repeated measurements for each examiner, by analyzing the following variables:

a) PD of mesial surfaces versus distal surfaces;
b) PD of anterior area versus posterior area;
c) PD of buccal surfaces versus lingual surfaces; and
d) PD of superior versus inferior arches.

For examiner 1, the *t*-test for independent samples ($p < 0.01$) results showed no statistically significant differences between each probe for the above-mentioned variables (Table 1).

For examiner 2, the results of the *t*-test for independent samples ($p < 0.01$) showed no statistically significant differences between the initial and repeated probes for the following variables:

- mesial surfaces versus distal surfaces;
- buccal surfaces versus lingual surfaces; and
- superior versus inferior arch PD.

A statistically significant difference did appear between the initial and repeated probes for the PD of anterior versus posterior areas (Table 2).

For examiner 3, the results of the *t*-test for independent samples ($p < 0.01$) showed no statistically significant differences between the initial and repeated probes for the variables:

- mesial surfaces versus distal surfaces;
- buccal surfaces versus lingual surfaces; and
- superior versus inferior arch PD.

Table 1 - Results from the *t*-test for independent samples on the probe depth from examiner 1 for dental surface, area, and arch variables.

	Mean	Range		SEM	*p* value
		Min	Max		
Mesial	0.6	0.19	0.94	0.04	0.413
Distal	0.72	0.27	1.00	0.05	
Anterior	0.72	0.31	1.00	0.04	0.280
Posterior	0.62	0.17	0.91	0.04	
Buccal	0.71	0.40	1.00	0.03	0.400
Lingual	0.66	0.26	0.85	0.04	
Superior	0.68	0.26	0.90	0.03	0.489
Inferior	0.7	0.24	1.00	0.04	

$p < 0.01$. Standard error of the mean.

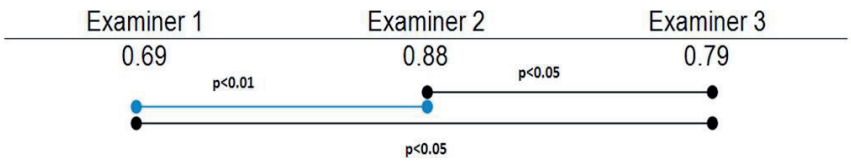

Figure 3 - Scheffé method for the evaluation of inter-examiner probes showing statistically significant differences between and among the examiners.

Table 2 - Results from the *t*-test for independent samples on the probe depth from examiner 2 for dental surface, area, and arch variables.

	Mean	Range		SEM	*p* value
		Min	Max		
Mesial	0.84	0.57	1.00	0.02	0.646
Distal	0.87	0.24	1.00	0.04	
Anterior	0.89	0.71	1.00	0.02	0.162
Posterior	0.86	0.62	1.00	0.02	
Buccal	0.92	0.73	1.00	0.02	0.01*
Lingual	0.83	0.46	1.00	0.03	
Superior	0.88	0.72	1.00	0.02	0.612
Inferior	0.88	0.71	1.00	0.02	

*p < 0.01.

Table 3 - Results from the *t*-test for independent samples on the probe depth from examiner 3 for dental surface, area, and arch variables.

	Mean	Range		SEM	*p* value
		Min	Max		
Mesial	0.77	0.31	0.96	0.04	0.198
Distal	0.85	0.28	1.00	0.04	
Anterior	0.8	0.46	0.97	0.03	0.916
Posterior	0.81	0.51	0.98	0.03	
Buccal	0.87	0.59	1.00	0.03	0.013*
Lingual	0.76	0.39	0.94	0.03	
Superior	0.81	0.58	0.96	0.03	0.419
Inferior	0.79	0.37	0.93	0.03	

*p < 0.01.

A statistically significant difference did appear between the initial and repeated probes for the anterior versus posterior areas (Table 3).

Inter-examiner reproducibility

By evaluating the probe results from the three trained examiners, we could calculate the Kappa value and create mathematical measurements for subsequent application of the *t*-test. An analysis of variance (ANOVA) (p = 0.01, p = 0.05) showed that there were differences between the groups of examiners (Table 4). For detection of the differences (examiner 1 versus examiner 2, examiner 1 versus examiner 3, examiner 2 versus examiner 3), the Scheffé method was applied *a posteriori* (p = 0.01).

The Scheffé method showed that examiners 1 and 2 differed statistically significantly (p < 0.01). Examiner 2 and examiner 3 differed statistically significantly, with p < 0.05, while examiner 3 and examiner 1 also differed at the same level (p < 0.05) (Figure 3).

Table 4 - Analysis of Variance (ANOVA) for inter-examiner reproducibility.

Source of variation	DF	SS	MS	F
Between	2	0.38	0.19	19*
Within	57	0.82	0.01	
Total	59	1.20		

* There are statistically significant differences between the groups. DF: Degree of Freedom; SS: Sum of Squares; MS: Mean Squares; F: Critical value.

Discussion

This study evaluated intra- and inter-examiner reproducibility among three trained and calibrated examiners as they used a manual probe for clinical recording of periodontal probing depth (PD) in healthy individuals. The large number of probing sites by examiner (n = 6,564) and rigorous operator calibration were demonstrated as necessary for obtaining reliable records, which can be used for appropriate decision-making. In our opinion, the high intra-examiner reproducibility observed was the result of the calibration and training program, and was not related to the operators' experience, as have stated other authors,[7-9] who believe that experience is the most important factor in measurement reproducibility. When comparing experienced and

inexperienced examiners, Samuel and co-workers[10] showed that experienced examiners reproduced their probe measurements; however, they also showed that inexperienced examiners had significantly more reproducibility with manual probes (Williams probe), as was also shown in this study. Despite the fact that the examiners in this study were students, their measurements were highly reproducible. As the calibration process progressed (calibration group), the intra-examiner agreement gradually increased as well, a fact that permitted highly reproducible repeated measurements for the experimental group. This speaks positively of the calibration process, as presented by Grossi *et al.*[11] When the results of the intra-examiner measurements were compared with the Kappa test, exact agreement > 75% and > 90% for ± 1 mm was found between the two[12,13] (repeated measures), and the three examiners were positioned above benchmark parameters. These are favorable results in terms of diagnosis and periodontal control, as suggested by Listgarten,[4] who stated that a measurement error of 2 mm or more in PD could lead to problems in the interpretation of results, thus leading to the provision of inappropriate treatment. There are many factors that could cause errors in clinical probing records.[7,12-23] Among the studied variables, our results confirmed the findings presented by Mullally and Linden[12] and Mayfield *et al.*,[24] finding differences in probing measurements between anterior and posterior areas (the only variables with statistically significant differences) (Tables 2 and 3). These differences can be explained by better access, probe position, and visibility of anterior areas or by the unconsciously high force that can be applied when posterior areas are probed. It should be noted that the results of examiners 2 and 3 were different compared with those of examiner 1. This could be due to the fact that the first two examiners changed their working position when appropriate, while examiner 1 did not. Working position directly influences the PD record when a manual probe is being used to perform a periodontal probe, especially when the examiner does not use an angle periodontal probe. This has not been reported in the literature. As other authors have confirmed,[8,9,24-26] our results showed greater reproducibility for measurements with a manual probe, contrary to Listgarten,[4] who stated that reproducibility for manual probes was low. We found that probing a healthy periodontium with a shallow sulcus, which offered greater resistance to probe penetration compared with a diseased sulcus,[27,28] also contributed to high reproducibility.[27,28] This could be considered a limitation of our study, since it was demonstrated that when calculus, inflammation, and bleeding are present, the clinically recorded data can be affected.[29] Despite a high intra-examiner correlation, statistically significant differences were found among the three examiners in terms of inter-examiner reproducibility (Figure 3). Our findings were similar to those presented by other authors, who had shown that differences (inter-examiner reproducibility) were more evident when manual probes were used.[13] It can be estimated that these differences are related to individual differences in probing technique and force, even when high intra-examiner reproducibility is demonstrated. To control the above-mentioned factors, it is recommended that pressure control probes[2] be used to calibrate the operator's probing force, or even an acrylic stent to standardize the operator's technique. We suggest that future research should evaluate other variables that affect periodontal clinical probing[19,21,28,30] and the effects of those variables over the resolution, reproducibility, and accuracy of periodontal probes,[23] not only on healthy tissues but mainly in the presence of periodontal disease. Finally, the methodology was carefully designed to avoid irreversible damage of periodontal tissues as a result of repeated probing. This was avoided by providing sufficient time between the initial and repeated measures and preserving tissue health, thus controlling the likelihood that an operator's memory could induce a bias.[13,25] Greater emphasis should be placed in dental school classrooms on how to take periodontal clinical records, since they are extremely relevant to diagnosis, prognosis, and treatment, and they are highly sensitive to error.

Conclusions

While considering the inherent limitations of this study, we can conclude that probing depth us-

ing a manual periodontal probe produced high reproducibility in healthy individuals. The operator's position can affect the reproducibility of repeated measures. Calibration and operator training, rather than operator experience, were fundamental to reproducibility. However, since this methodology cannot demonstrate the reasons for low inter-examiner reproducibility, we can assume that there are other mitigating factors, such as individual technique and probing depth force.

References

1. Chu SJ, Hochman MN, Fletcher P. A biometric approach to aesthetic crown lengthening: Part II. Pract Proced Aesthet Dent. 2008 Oct;20(9):529-36.
2. Ramachandra SS, Mehta DS, Sandesh N, Baliga V, Amarnath J. Periodontal probing systems: a review of available equipment. Compend Contin Educ Dent. 2011 Mar;32(2):71-7.
3. Hourdin S, Glez D, Sorel O. [Periodontal diagnosis in orthodontics]. Orthod Fr. 2010 Mar;81(1):9-17. French.
4. Listgarten MA. Periodontal probing: what does it mean? J Clin Periodontol. 1980 Jun;7(3):165-76.
5. Mombelli A. Clinical parameters: biological validity and clinical utility. Periodontol 2000. 2005;39(1):30-9.
6. Armitage GC. The complete periodontal examination. Periodontol 2000. 2004;34(1):22-33.
7. Seabra RC, Costa FO, Costa JE, Van Dyke T, Soares RV. Impact of clinical experience on the accuracy of probing depth measurements. Quintessence Int. 2008 Jul-Aug;39(7):559-65.
8. Niederman R. Manual and electronic probes have similar reliability in the measurement of untreated periodontitis. Evid Based Dent. 2009 Jun;10(2):34-61.
9. Silva-Boghossian CM, Amaral CS, Maia LC, Luiz RR, Colombo AP. Manual and electronic probing of the periodontal attachment level in untreated periodontitis: a systematic review. J Dent. 2008 Aug;36(8):651-7. Epub 2008 Jun 5.
10. Samuel ED, Griffiths GS, Petrie A. *In vitro* accuracy and reproducibility of automated and conventional periodontal probes. J Clin Periodontol. 1997 May;24(5):340-5.
11. Grossi SG, Dunford RG, Ho A, Koch G, Machtei EE, Genco RJ. Sources of error for periodontal probing measurements. J Periodontal Res. 1996 Jul;31(5):330-6.
12. Mullally BH, Linden GJ. Comparative reproducibility of proximal probing depth using electronic pressure-controlled and hand probing. J Clin Periodontol. 1994 Apr;21(4):284-8.
13. Buduneli E, Aksoy O, Köse T, Atilla G. Accuracy and reproducibility of two manual periodontal probes; an in-vitro study. J Clin Periodontol. 2004 Oct;31(10):815-9.
14. Barendregt DS, Van der Velden U, Timmerman MF, Van Der Weijden GA. Comparison of two automated periodontal probes and two probes with a conventional readout in periodontal maintenance. J Clin Periodontol. 2006 Apr;33(4):276-82.
15. Bulthuis HM, Barendregt DS, Timmerman MF, Loos BG, van der Velden U. Probe penetration in relation to the connective tissue attachment level: influence of tine shape and probing force. J Clin Periodontol. 1998 May;25(5):417-23.
16. Araujo MW, Hovey KM, Benedek JR, Grossi SG, Dorn J, Wactawski-Wende J, *et al*. Reproducibility of probing depth measurements using a constant-force electronic probe: analysis of inter- and intra-examiner variability. J Periodontol. 2003 Dec;74(12):1736-40.
17. Garnick JJ, Silverstein L. Periodontal probing: probe tip diameter. J Periodontol. 2000 Jan;71(1):96-103.
18. Van Weringh M, Barendregt DS, Rosema NAM, Timmerman MF, Van Der Weijden GA. A thin or thick probe handle: does it make the difference? Int J Dent Hyg. 2006 Aug;4(3):140-4.
19. Hassan MA, Bogle G, Quishenbery M, Stephens D, Riggs M, Egelberg J. Pain experienced by patients during periodontal recall examination using thinner *versus* thicker probes. J Periodontol. 2005 Jun;76(6):980-4.
20. Karpinia K, Magnusson I, Gibbs C, Yang MC. Accuracy of probing attachment levels using a CEJ probe *versus* traditional probes. J Clin Periodontol. 2004 Mar;31(3):173-6.
21. Neto JB, Filho GR, Tramontina VA, Sallum EA, Nociti FH Jr, Sallum AW. Millimeter marks and probe tip diameter standardisation from commercially available periodontal probes. A comparative study. J Int Acad Periodontol. 2001 Jul;3(3):57-60.
22. Eickholz P, Grotkamp FL, Steveling H, Mühling J, Staehle HJ. Reproducibility of peri-implant probing using a force-controlled probe. Clin Oral Implants Res. 2001 Apr;12(2):153-8.
23. Bergenholtz A, al-Harbi N, al-Hummayani FM, Anton P, al-Kahtani S. The accuracy of the Vivacare true pressure-sensitive periodontal probe system in terms of probing force. J Clin Periodontol. 2000 Feb;27(2):93-8.
24. Mayfield L, Bratthall G, Attström R. Periodontal probe precision using 4 different periodontal probes. J Clin Periodontol. 1996 Feb;23(2):76-82.
25. Alves R de V, Machion L, Andia DC, Casati MZ, Sallum AW, Sallum EA. Reproducibility of clinical attachment level and probing depth of a manual probe and a computerized electronic probe. J Int Acad Periodontol. 2005 Jan;7(1):27-30.
26. Ishihata K, Wakabayashi N, Wadachi J, Akizuki T, Izumi Y, Takakuda K, *et al*. Reproducibility of pocket depth measurement by experimental periodontal probe incorporating optical fiber sensor. J Periodontol. 2011 May 16. [Epub ahead of print]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21574830

27. Khan S, Cabanilla LL. Periodontal probing depth measurement: a review. Compend Contin Educ Dent. 2009 Jan-Feb;30(1):12-21.
28. Hatakeyama Y, Uzel MI, Santana RB, Ruben MP. Relationship between position of probe tip and periodontal tissues after periodontal surgery in dogs. Int J Periodontics Restorative Dent. 2005 Jun;25(3):247-55.
29. Oringer RJ, Fiorellini JP, Koch GG, Sharp TJ, Nevins ML, Davis GH, *et al*. Comparison of manual and automated probing in an untreated periodontitis population. J Periodontol. 1997 Dec;68(12):1156-62.
30. Holt LA, Williams KB, Cobb CM, Keselyak NT, Jamison CL, Brand VS. Comparison of probes for microbial contamination following use in periodontal pockets of various depths. J Periodontol. 2004 Mar;75(3):353-9.