Imaging and Radiation Research (2021) Volume 4 Issue 1
doi:10.24294/irr.v4i1.1729

ORIGINAL RESEARCH ARTICLE

Inter-examiner concordance of periodontal findings using conventional periapical radiography

Miguel Simancas-Pallares*, Luisa Arévalo-Tovar, Antonio Díaz-Caballero

Universidad de Cartagena, Cartagena, Colombia. E-mail: msimancasp@unicartagena.edu.co

ABSTRACT

Introduction: Periodontal disease affects more than half of the population in Colombia and is estimated to be one of the leading causes of oral morbidity. Diagnostic aids that allow the evaluation of its extension and severity are of importance since this will provide reliable tools to quantify the severity of the problem. **Objective:** To determine the inter-examiner agreement for the detection of radiographic findings in patients with localized chronic periodontitis using conventional periapical radiography. **Methods:** Study of diagnostic tests including patients with localized chronic periodontitis, the tooth with the worst clinical insertion level and a single conventional radiograph per dental organ using parallelism technique. The radiographic evaluations were performed by two independent and blinded evaluators for the findings: lamina dura, bone defects and type of defect. The agreement obtained was estimated through Cohen's Kappa. **Results:** A total of 125 radiographs were taken. The mean age was 38.8 ± 9.9, and 61.6% were women. Concordance for lamina dura was 0.08 (95% CI: -0.04–0.21), bone defects 1.00 (95% CI: 1.00–1.00); type of defect present 0.31 (95% CI: 0.29–0.38). **Conclusions:** Concordance was evaluated as null, almost perfect and acceptable for the findings lamina dura, presence of bone defects and type of defect respectively. For some findings and given the importance of the diagnostic and therapeutic processes, more accurate evaluations are needed which would result in a higher degree of agreement.

Keywords: Reproducibility of Results; Diagnosis; Periodontics; Dental Radiography

ARTICLE INFO

Received: 29 November 2020
Accepted: 7 January 2021
Available online: 18 January 2021

1. Introduction

Periodontal disease is recognized by gingival inflammation in sites where migration of the junctional epithelium across the root surface has occurred, coupled with loss of connective tissue and alveolar bone due to bacterial invasion[1]. The most prevalent form is chronic periodontitis, with characteristic clinical and radiographic findings that differentiate it from other forms of periodontitis[2]. In Colombia, it is estimated that ~75% of the population over the age of 35 suffers from some form of attachment loss[3], whose diagnosis is based on clinical and radiographic assessment of periodontal tissues[4].

Traditionally, the way to detect bone changes is to determine bone height through radiographic examinations, being able to detect qualitative and quantitative changes and thus representing one of the most important diagnostic aids in dentistry. Periapical radiography, available in film (conventional) and computerized form (digital), is the most widely used due to its multiple benefits.

Clinically, the parameters evaluated for periodontal diagnosis include probing depth, bleeding on probing and clinical attachment levels. The more accurate the measurement tools are, the more reproducible

the assessments will be; additionally, greater control of the variables that affect the identification of periodontal findings at the radiographic level will be achieved. Furthermore, the accuracy of these tools will allow more reliable information for the diagnosis and early detection of the disease, which in turn will promote adequate and timely treatment, as well as long-term periodontal control[5].

Clinicians within dental schools show variations in the interpretation of diagnostic tests. Lewis *et al*. reported low agreement among dentists in study models for occlusal stability and occlusal loss in cases of malocclusion[6]. Likewise, Marbach *et al*. reported considerable variation among clinical instructors when evaluating models for the assessment of bruxism severity[7]. Recent research revealed lack of accuracy and wide variability among periodontists and oral hygienists in assessing radiographic bone loss. Periodontal findings at the radiographic level are of increasing importance because, together with clinical evidence, they lead to the establishment of adequate diagnoses, treatment plans and prognoses as well as long-term evaluation of the periodontium[8].

In this regard, the aim of the present study was to estimate the inter-examiner agreement of periodontal findings using conventional periapical radiography in patients with untreated localized chronic periodontitis.

2. Materials and methods

A study of concordance diagnostic tests was carried out in adult patients; it was approved by the research committee of the Faculty of Dentistry of the University of Cartagena. The ethical aspects expressed in the declaration of Helsinki in 2008, resolution 008430 of 1993 were taken into account and written informed consent was requested from all patients.

The sample consisted of patients over the age of 18, male or female, with a diagnosis of localized chronic periodontitis (mild, moderate and/or severe) and who agreed to enter the study through consecutive sampling until the sample size was reached. The following patients were excluded: pregnant women due to the impossibility of receiving ionizing radiation; patients with thyroid diseases, diabetes mellitus, with periodontal treatment six months before or who reported the intake of antibiotics and/or non-steroidal anti-inflammatory drugs (NSAIDs) six weeks before due to the possible alteration of bone metabolism, therefore, generate a high degree of variability in the measurements when identifying the anatomical reference points for the measurements.

The sample size was calculated taking as parameters: expected Kappa coefficient of 0.85 assuming 95% confidence and power of 80% in the statistical package GenStat V12.1.0.3278 (VSN International Ltd., U.K). Fifty-seven replicates per method were necessary (114 measurements in total). Anticipating 10% follow-up loss and 10% measurement error, a final sample of 136 measurements was required.

The evaluation of the selection criteria, periodontal clinical examination and data collection was performed by a general dentist calibrated in periodontal diagnostic systems. The periodontal diagnosis was carried out taking into account the criteria suggested by the American Association of Periodontology[9]. For the periodontal examination and diagnosis, a periodontal probe type Marquis Hu-Friedy® (N. Rockwell, Chicago, IL. USA) was used. The study protocol followed the following order: clinical examination, radiography in all selected patients.

For conventional periapical radiography, the parallelism technique was used. Only one tooth per subject was selected and one radiograph per patient was taken at the site with the worst clinical insertion level (CIN), because the most severe cases are those that reflect the real state of tissue destruction[10] in posterior teeth (premolars or molars) upper and lower, right and left, and with the presence of chronic periodontitis. If there were two or more teeth with the same CIN, the most posterior tooth was chosen. These criteria were structured by the research team.

The radiographs were obtained by a single oral and maxillofacial radiology professional, with proven experience (>10 years) in taking and processing radiographs, who prepared the radiographic

equipment according to the technical specifications.

Tooth type specific cone and film positioners (XCP Rinn Film Holder, Dentsply®, Dentsply International, Philadelphia, PA, USA) were used; a bite registration was obtained with the JET BLUE bite registration silicone and positioner (Coltène/Whaledent AG, Altstatten, Switzerland), to ensure the patient's occlusion and standardize the geometric projection of the image. Periapical imaging equipment was used (RAIOS X TIMEX 70C PAREDE GELO 127V +4%, Rod Abrao Assed. Km53 +450m—Ribeirao Preto—Sao Paulo—Brazil) presetting the milliamperage and periapical radiographic films type E (Kodak Dental Intraoral E-Speed Film. Carestream Health INC., Rochester, NY, U.S.A.). Once the radiographs were projected, they were developed taking into account the specifications according to the protocol recommended by the manufacturer for the type of film.

For the evaluation of the radiographic conditions, two evaluators were selected, who were widely experienced (≥15 years of periodontal practice and clinical teaching practice) and previously trained in periapical radiographic measurement techniques. In order to sensitize the evaluators with the information collection format, a pilot test was performed with 10% of the sample size, in which the evaluators measured the variables under study and recorded the information in the format designed for this purpose. This test did not yield results that indicated the need to make adjustments to the data collection and recording formats. Likewise, these subjects did not belong to the sample finally studied.

Once the radiographs were obtained and consecutively coded, a simple random statistical method was applied using Microsoft Excel® to determine the order of presentation of the radiographs in a blinded and independent manner for each evaluator. For the presentation of the radiographs, a patient evaluation room was selected with a tabletop X-ray viewer with a 40 cm base, 30 cm high, and white light under controlled artificial light conditions.

The radiographs were arranged on cardboard skeletons, without covering the entire film. Subsequently, the continuity of the hard film (continuous or discontinuous), bone defects (present or absent) and type of bone defects were found (vertical or horizontal). Then, the evaluators proceeded to record the information obtained from the measurement in the instrument designed. Once the data were obtained, the information was tabulated in a matrix table designed by the researchers in Microsoft Excel 2010 for Windows. The data entry was verified periodically and blinded in order to minimize errors.

Data analysis was initially performed using descriptive statistics: mean, median, standard deviation for quantitative data and proportions and confidence intervals for qualitative data. Additionally, normality assumptions were verified with the Shapiro Wilk test for continuous variables. Inter-examiner agreement was estimated with Cohen's Kappa statistic (κ) and 95% confidence intervals. The statistics obtained were evaluated taking into account the criteria proposed by Landis and Koch[11]. The statistical analysis was performed in the IBM SPSS Statistics v.20 statistical package for Microsoft Windows (IBM Corporation—Armond, NY, USA).

3. Results

A total of 136 radiographs were taken. However, due to post-processing quality loss, 125 radiographs were included for a total of 250 observations. The mean age was 38.8 years (SD: 9.9). Regarding sex, 61.6% of the patients were found to be female. On the other hand, 47.2% of the subjects suffered from the severe form; while the moderate and mild form was represented with 39.2% and 13.6% respectively.

Tables 1–**3** show the observed and expected agreements for each of the periodontal findings evaluated; the highest agreement was observed for the bone defect finding (presence/absence) followed by lamina dura.

Finally, the concordance results obtained show for lamina dura 0.08 (95% CI: –0.04 – 0.21 – P = 0.08), bone defects 1.00 (95% CI: 1.00 – 1.00 – P < 0.01) and type of bone defects 0.31 (95% CI: 0.29 – 0.38 – P < 0.01). The results of the degree of

agreement for these findings were 61.6%, 100% and 54.4% respectively.

Table 1. Observed and expected agreement for hard film

		Evaluator 2		
		Continuous n (%)	**Discontinued n (%)**	**Total n**
Evaluator 1	Continuous	7 (4.8)	43 (45.2)	50
	Discontinuous	5 (7.2)	70 (67.8)	75
	Total	12	113	125

Table 2. Observed and expected agreement for presence of bone defects

		Evaluator 2		
		Absent n (%)	**Present n (%)**	**Total n**
Evaluator 1	Absent	51 (20.8)	0 (30.2)	51
	Present	0 (30.2)	74 (43.8)	74
	Total	51	74	125

Table 3. Observed and expected agreement for type of bone defects

		Evaluator 2			
		None n (%)	**Horizontal n (%)**	**Vertical n (%)**	**Total n**
Evaluator 1	None	34 (20.8)	14 (17.5)	3 (12.6)	51
	Horizontal	7 (11)	13 (9.3)	7 (6.7)	27
	Vertical	10 (19.2)	16 (16.2)	21 (11.7)	47
	Total	51	43	31	125

4. Discussion

The results of this investigation, taking into account the findings of presence of bone defects and type of defects, show significant differences in the concordance between the two evaluators. However, for hard lamina, the concordance was null and without statistical significance, which in turn questions its application in the clinical context.

The protocol of the present study did not contemplate the performance of calibration procedures, assuming good prior agreement between the observers according to their experience in periodontal practice (>15 years) and their professional profile. Wolf *et al.* evidenced that the concordance presents statistically significant differences when there is more than one evaluator[12], being the results of the present study consistent with these findings since differences were found in the interpretation of the concordance reported by the evaluators.

In diagnostic test studies, a wide inter-observer variation is to be expected, which is consistent with the results of this study. One explanation for these findings is that when there is more than one evaluator, the definition of the anatomical reference points for measurement is divergent even when calibration procedures exist[13,14]. Tewary, Luzzo and Hartwell in 2011, through a study of diagnostic tests to evaluate intra- and inter-examiner agreement in periapical lesions, compared experienced specialists (5 to 40 years), specialists in training and specialists trained in oral radiology. Each evaluator examined the images on the same computer and under the same lighting conditions. The results revealed that the interpretation by the evaluators lacked the accuracy, precision and consistency that would be expected from trained and highly experienced evaluators, which is consistent with the results of the present study. However, it is important to note that this research did not evaluate accuracy and precision of the methods employed[15]. They also reported that the difference in inter-rater agreement may be attributed to the fact that the interpretation or measurements made on the oral radiographs are not inherent to the technique employed; there are factors that modify agreement such as the experience of the rater and familiarity with the technique employed[15,16].

In general, the greater the experience of the examiner, the higher the expected concordance[17]; however, this did not happen due to some factors such as lack of calibration or previous standardiza-

tion, even when there was a high level of professional experience. On the other hand, additional aspects may explain the difference between evaluators, such as lack of specialized training in radiographic measurement systems. In this regard, Tewary *et al.* in 2011 reported that a minimum experience of at least one year is necessary to obtain a moderate agreement between examiners[15]. In the present study, the examiners received only instructions on the technical procedures of how to perform the measurements but not the calibration procedures.

Pecoraro *et al.* suggested the single evaluation of measurements in order to minimize differences between observers[18]. Other authors suggest consensus measurements in order to minimize inherent observer effects which would then generate differences in agreement[19,20].

Lanning *et al.* in 2006 determined that the design and execution of specific training programs can improve the accuracy and consistency of measurements made by different examiners and thus the agreement between them. They reported that the overall agreement for estimating bone loss in one category (no bone loss) changed from 64.5% to 85.2% after 3 months. Thus the agreement improved from moderate (0.52) to almost perfect (0.80)[20].

In view of all these findings, it is important that the introduction of diagnostic technologies in periodontology should take into account additional procedures to reduce the difference between observers, which would also be reflected in decision making (diagnostic, therapeutic and/or prognostic) in daily clinical practice. It would also be reflected in the training received by dental students, since the mission is to train professionals with adequate clinical skills.

Several authors suggest additional studies to obtain learning curves, which in the context of the present study become important since the objective will be to estimate the number of cases that each evaluator must perform a bone height measurement with the method under study[21,22] in the so-called learning curve. These will constitute a methodological tool which aims to: 1) improve inter-observer performance in studies (of diagnostic tests) and 2) facilitate the transition between diagnostic technologies so that diagnostic, therapeutic and prognostic criteria will not be used subjectively[23].

With the results of the present study, it is possible to infer that there are some factors that could have a possible impact on the concordance such as: experience of the evaluator with the radiographic system and calibration of examiners to perform the measurements. Achieving this, it is expected that the diagnostic ability of clinical instructors is acceptable; this then translates into adequate diagnostic competencies for professionals in training.

Conflict of interest

The authors declare no conflict of interest.

References

1. Teles R, Teles F, Frias Lopez J, *et al*. Lessons learned and unlearned in periodontal microbiology. Periodontol 2000 2013; 62(1): 95–162. doi: 10.1111/prd.12010.
2. Contreras A, Moreno SM, Jaramillo A, *et al*. Periodontal microbiology in Latin America. Periodontol 2000 2015; 67(1): 58–86. doi: 10.1111/prd.12074.
3. 4th National Study on Oral Health. Communication. Bogotá DC (Colombia): Health Ministery, 2015 Contract No. 7.
4. Armitage GC. The complete periodontal examination. Periodontol 2000 2004; 34: 22–33.
5. Andrade R, Espinoza M, Gomez EM, *et al*. Intra- and inter-examiner reproducibility of manual probing depth. Brazilian Oral Research 2012; 26(1): 57–63.
6. Lewis EA, Albino JE, Cunat JJ, *et al*. Reliability and validity of clinical assessments of malocclusion. American Journal of Orthodontics and Dentofacial Orthopedics 1982; 81(6): 473–477.
7. Marbach JJ, Raphael KG, Janal MN, *et al*. Reliability of clinician judgements of bruxism. Journal of Oral Rehabilitation 2003; 30(2): 113–118.
8. Lanning SK, Pelok SD, Williams BC, *et al*. Variation in periodontal diagnosis and treatment planning among clinical instructors. Journal of Dental Education 2005; 69(3): 325–337.
9. Armitage GC. Development of a classification system for periodontal diseases and conditions. Annals of Periodontology 1999; 4(1): 1–6.
10. Page RC, Eke PI. Case definitions for use in population-based surveillance of periodontitis. Journal of Periodontology 2007; 78(7 Suppl): 1387–1399.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33(1): 159–174.

12. Wolf B, von Bethlenfalvy E, Hassfeld S, *et al*. Reliability of assessing interproximal bone loss by digital radiography: Intrabony defects. Journal of Clinical Periodontology 2001; 28(9): 869–878.
13. Hildebolt CF, Pilgram TK, Yokoyama Crothers N, *et al*. Reliability of linear alveolar bone loss measurements of mandibular posterior teeth from digitized bitewing radiographs. Journal of Clinical Periodontology 1998; 25(11 Pt 1): 850–856.
14. Hildebolt CF, Bartlett TQ, Brunsden BS, *et al*. Bitewing-based alveolar bone densitometry: Digital imaging resolution requirements. Dentomaxillofacial Radiology 1994; 23(3): 129–134.
15. Tewary S, Luzzo J, Hartwell G. Endodontic radiography: Who is reading the digital radiograph? Journal of Endodontics 2011; 37(7): 919–921. doi: 10.1016/j.joen.2011.02.027.
16. Rechmann P, Featherstone JD. Quality assurance study of caries risk assessment performance by clinical faculty members in a school of dentistry. Journal of Dental Education 2014; 78(9): 1331–1338.
17. Patel A, Tee BC, Fields H, *et al*. Evaluation of cone-beam computed tomography in the diagnosis of simulated small osseous defects in the mandibular condyle. American Journal of Orthodontics and Dentofacial Orthopedics 2014; 145(2): 143–156.
18. Pecoraro M, Azadivatan le N, Janal M, *et al*. Comparison of observer reliability in assessing alveolar bone height on direct digital and conventional radiographs. Dentomaxillofacial Radiology 2005; 34(5): 279–284.
19. Pepelassi EA, Tsiklakis K, Diamanti-Kipioti A. Radiographic detection and assessment of the periodontal endosseous defects. Journal of Clinical Periodontology 2000; 27(4): 224–230.
20. Lanning SK, Best AM, Temple HJ, *et al*. Accuracy and consistency of radiographic interpretation among clinical instructors using two viewing systems. Journal of Dental Education 2006; 70(2): 149–159.
21. Delamare EL, Liedke GS, Vizzotto MB, *et al*. Influence of a programme of professional calibration in the variability of landmark identification using cone beam computed tomography-synthesized and conventional radiographic cephalograms. Dentomaxillofacial Radiology 2010; 39(7): 414–423.
22. Franceschetti G, Trombelli L, Minenna L, *et al*. Learning curve of a minimally invasive technique for transcrestal sinus floor elevation: A split-group analysis in a prospective case series with multiple clinicians. Implant Dentistry 2015; 24(5): 517–526. doi: 10.1097/ID.0000000000000270.
23. Luz PB, Stringhini CH, Otto BR, *et al*. Performance of undergraduate dental students on ICDAS clinical caries detection after different learning strategies. European Journal of Dental Education 2015; 19(4): 235–241. doi: 10.1111/ejeje.12131.