Reproducibility between conventional and digital periapical radiography for bone height measurement

Concordancia entre la radiografía periapical convencional y digital para medición de la altura ósea

Miguel Simancas-Pallares • Jorge Andrés Rubio-Romero • Edgar Cortés-Reyes

Received: 29/04/2015   Accepted: 18/06/2015

Background. Several diagnostic aids are available for bone height measurement. Digital and conventional radiographs are the two ones most used in Dentistry. Few studies accounting for accuracy and precision have been conducted to compare these methods.

Objective. The aim of this study was to estimate reproducibility between conventional and digital periapical radiography in bone height measurement in patients with chronic periodontitis.

Methods. A consistency diagnostic test study was performed. 136 patients with chronic periodontitis were included, selecting the worst prognosis teeth and two radiographs —conventional and digital— were taken for each one. Two experienced and blinded examiners performed radiographic measurements. Reproducibility was obtained through Lin’s concordance correlation coefficient by using the statistical package STATA™ for Windows.

Results. Average age was 38.8 (SD: 9.9) and 61.6 % were female patients. 125 pairs of matched radiographs for 1000 measurements were evaluated. Overall reproducibility between the methods for mesial and distal measurements were 0.62 (95% CI: 0.55-0.70) and 0.64 (95% CI: 0.57-0.71) respectively.

Conclusions. Reproducibility between methods was considered poor, including subgroup analysis, therefore, reproducibility between methods is minimal. Usage of these methods in periodontics should be made implementing the whole knowledge of the technical features and the advantages of these systems.

Keywords: Reproducibility of Results; Periodontics; Digital Dental Radiography; Epidemiology (MeSH).

Resumen

Antecedentes. Diversas ayudas diagnósticas están disponibles para la medición de la altura ósea; las dos más empleadas en odontología son la radiografía periapical convencional y la digital. A la fecha se cuenta con pocos estudios que den cuenta de la precisión y exactitud al compararlos.

Objetivo. Estimar la concordancia entre la radiografía periapical convencional y la digital para la medición de la altura ósea en pacientes con periodontitis crónica.

Métodos. Se realizó un estudio de pruebas diagnósticas de consistencia en 136 pacientes con periodontitis crónica seleccionando el diente con peor pronóstico. Se obtuvieron dos radiografías —convencional y digital— para cada diente.
y dos examinadores cegados realizaron las mediciones radiográficas. La concordancia se obtuvo con el coeficiente de correlación y concordancia de Lin empleando el paquete Stata para Windows.

**Resultados.** La edad promedio fue 38.8 años (DE: 9.9) y 61.6% de los pacientes fueron mujeres. Se evaluaron 125 pares de radiografías para 1000 mediciones en total. La concordancia global fue 0.62 (IC 95%: 0.55-0.70) y 0.64 (IC 95%: 0.57-0.71) para las mediciones mesiales y distales respectivamente.

**Conclusiones.** La concordancia entre los métodos se consideró pobre incluso en el análisis por subgrupos por tanto la reproducibilidad es mínima. El uso de estos métodos en periodoncia debe hacerse empleando el conocimiento completo de las características técnicas y ventajas de estos sistemas.

**Palabras clave:** Reproducibilidad de resultados; Periodoncia; Radiografía digital dental; Epidemiología (DeCS).


**Background**

Periodontal diseases are recognized by gingival inflammation in sites where junction epithelium migrates through radicular surfaces, which leads to bone and connective tissue loss due to bacterial invasion (1). The most prevalent form is chronic periodontitis with clinical and radiographic findings that allow differential diagnosis of other forms of the disease (2). In Colombia, ~50.2% of the adult population suffers loss of attachment (3) and the disease diagnosis is based on the clinical and radiographic examination of the periodontal tissue.

Usually the way to detect bone changes is by means of measuring bone height through a radiographic examination. So radiographies are one of the most relevant diagnostic tools as they permit to detect qualitative and quantitative changes. The most used radiography is the periapical one due to its several advantages as it is available film-based or digitally.

Additionally, unlike conventional radiography, the digital one allows low time consumption for each patient, low radiation doses, low rate of shooting mistakes, easily storage and environment preservation.

Khocht et al. compared digital and conventional radiographies in 25 subjects having periodontitis, obtaining better results in conventional radiographies for maxillary bone height measurements (P<0.02) (5); however, digital radiographies were better than conventional ones in mandibular anterior bone height measurement (P=0.00). Digital x-rays showed more sites with bone loss than conventional radiography. Nevertheless, statistical methods used for reproducibility evaluation did not reflect accurate measurements.

The aim of this study was to determine the reproducibility between conventional and digital periapical radiography for bone height measurement in patients with untreated chronic periodontitis.

**Materials and Methods**

A diagnostic test study was performed. Patients were selected from the Dental Clinics of the Faculty of Dentistry of the University of Cartagena. The study has ethical approval issued by the board of the Faculty of Medicine of the National University of Colombia (Record CE-035/2011) and the Research Committee of the Faculty of Dentistry of the University of Cartagena (Record 03/2011). Moreover, this study was conducted taking into account the Helsinki Statement and the Decree 008430, issued by the Colombian Ministry of Health. All the patients signed a consent form to authorize their participation in the study.

The sample consisted of adult subjects that were diagnosed with untreated chronic periodontitis. Patients were recruited until the sample size completion. Sample size was determined using the following parameters: expected Lin’s concordance correlation coefficient: 0.95 (6); mean shift value: 0.20 mm, and Pearson’s expected correlation: 0.97 and using the GenStat statistical package (V.12.1.0.3278–VSN International Ltd., U.K). 57 replications per method —114 measurements in overall— were necessary. By anticipating 10% of follow-up loss and 10% of measurement errors, the final calculated sample was 136 measurements (teeth).

For data collection a calibrated General Dentist applied the selection criteria, performed the periodontal exam and gathered the data in a questionnaire designed by the research staff. Periodontal diagnosis was performed according to the criteria suggested by the American Association of Periodontics.
A Hu-Friedy —N. Rockwell, Chicago, IL. USA— periodontal probe was used for periodontal examination. Once the clinical exam was completed, radiographs were taken —conventionally and digitally—.

For x-rays shooting, the parallelism technique was used. X rays were taken with both techniques at the same time before beginning the periodontal treatment following the conventional-digital sequence for each tooth selected. Just one tooth per patient at the site with worst clinical attachment level (CAL) in posterior, upper or lower, left or right jaw was included. In the cases where there were two sites with the same CAL, the most posterior tooth was chosen. These criteria were determined by the research staff in order to avoid selection and measurement bias.

The same oral and maxillofacial radiographic technician that had adequate experience performed all the X-rays. Additionally, the X-ray device was prepared, according to its technical specifications, by the same technician.

Conventional x-rays were taken using film holders —XCP Rinn Film Holder, Dentsply®, Dentsply International, Philadelphia, PA, USA—. To ensure the patient’s bite was reproducible for each technique, the X-ray technician placed an impression material on the plastic bite block —JET BLUE Coltène/Whaledent AG, Altstätten, Switzerland—.

X-rays were taken using a wall-mounted device with 7mA and a 70kV intensity —RAIOS X TIMEX 70C PAREDE GELO 127V +4%, Rod Abrao Assed. Km 53 +450m – Ribeirão Preto – Sao Paulo – Brasil— and using E films —Kodak Dental Intraoral E-Speed Film. Carestream Health INC., Rochester., N.Y., U.S.A.—. The same technician, following the manufacturer technical specifications, developed the films.

Digital x-rays were obtained by the same technician using the Dr Suni Plus® software —Suni Medical Imaging. 6840 Via Del Oro. Suite Nº 160. San Jose, CA 95119. USA— with the same intensity degree than the conventional x-rays.

Two blinded assessors examined each pair of x-rays. To avoid measurement bias they reviewed them with a 15-day interval between each type of radiography, in an independent process. Simple randomization was used to determine the sequence of x-ray assessment. Prior to the beginning of the study, the assessors were calibrated for radiographic measurement.

Conventional x-rays were disposed in cardboard holders in a room with controlled light conditions. Bone height measurements —osseous crest to the most apical point— were realized using a plastic ruler and then assessors gathered this information on the clinical records. After 15 days (8-10), digital x-rays were presented using the radiographic software without brightness or contrast manipulation in a 15.6” screen laptop PC —DELL® Latitude E6510-Dell INC, USA—.

Data analysis was performed through descriptive statistics —mean, median, standard deviations— taking into account the normality assumptions with the Shapiro-Wilk test for continuous data. Reproducibility was determined by employing the Lin’s concordance correlation coefficient even in non-normal distributed data (11). In addition, Bland & Altman (12) plots were obtained to determine agreement limits between the methods. Obtained reproducibility was assessed with the McBride criteria (6).

Analysis was shown initially in overall terms and then according to subgroups —teeth and assessor—. Statistical analysis was done using the Stata software v.12.0 for Windows —4905 Lakeway Drive College Station, Texas, USA—.

### Results

136 pairs (272) periapical x-rays were obtained. Nevertheless, due to processing errors only 125 pairs for measurement and further analysis were available.

Average age was 38.8 (SD: 9.9) years old and 61.6% of the population were females. Regarding periodontal disease severity, moderate form was the most frequent finding in 39.2% of the cases. Descriptive statistics of the bone height measurement are shown in Table 1.

<table>
<thead>
<tr>
<th>Surface</th>
<th>Median (mm)</th>
<th>Interquartile Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conventional</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mesial</td>
<td>14</td>
<td>3.5</td>
</tr>
<tr>
<td>Distal</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td><strong>Digital</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mesial</td>
<td>13.9</td>
<td>3.05</td>
</tr>
<tr>
<td>Distal</td>
<td>12.3</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Source: authors’ development through the research data.
The mean and the mean difference between the methods are shown through the Bland & Altman plots in Figure 1 —mesial: 0.65±2.1mm / distal: 0.76±2.1mm—. Additionally, the agreement limits for the obtained reproducibility —mesial: -3.4 – 4.7mm / distal: -3.3 – 4.9mm— are shown. Subgroup analysis revealed a better reproducibility for assessor 1 in premolar teeth as depicted in Table 3.

### Table 3. Observed agreement by assessor and type of tooth.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Lin’s conc ordance coefficient</th>
<th>95% CI Lin’s concordance coefficient</th>
<th>Pearson’s $\rho$</th>
<th>$C_b$</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assessor 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessor 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.71-0.84</td>
<td>0.79</td>
<td>0.97</td>
<td>1.08</td>
<td>0.67</td>
<td>0.95</td>
<td>1.07</td>
</tr>
<tr>
<td>0.71-0.85</td>
<td>0.79</td>
<td>0.98</td>
<td>1.09</td>
<td>0.65</td>
<td>0.96</td>
<td>1.08</td>
</tr>
<tr>
<td><strong>Assessor 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessor 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.18-0.48</td>
<td>0.35</td>
<td>0.93</td>
<td>1.05</td>
<td>0.65</td>
<td>0.96</td>
<td>1.08</td>
</tr>
<tr>
<td>0.30-0.56</td>
<td>0.48</td>
<td>0.88</td>
<td>0.95</td>
<td>0.67</td>
<td>0.95</td>
<td>1.07</td>
</tr>
<tr>
<td><strong>Premolar</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premolar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.44-0.68</td>
<td>0.59</td>
<td>0.94</td>
<td>1.26</td>
<td>0.67</td>
<td>0.95</td>
<td>1.07</td>
</tr>
<tr>
<td>0.36-0.62</td>
<td>0.51</td>
<td>0.95</td>
<td>1.17</td>
<td>0.65</td>
<td>0.96</td>
<td>1.08</td>
</tr>
<tr>
<td><strong>Molar</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.54-0.73</td>
<td>0.66</td>
<td>0.96</td>
<td>1.02</td>
<td>0.65</td>
<td>0.96</td>
<td>1.08</td>
</tr>
<tr>
<td>0.54-0.73</td>
<td>0.67</td>
<td>0.93</td>
<td>1.1</td>
<td>0.67</td>
<td>0.95</td>
<td>1.07</td>
</tr>
</tbody>
</table>

$\rho$: Lin’s correlation concordance coefficient. 95% CI $\rho$: 95% confidence interval for the Lin’s concordance correlation coefficient. Pearson’s $\rho$: Pearson correlation coefficient. $C_b$: bias correction factor. Source: authors’ development through the research data.

**Figure 1.** Bland & Altman plots for the agreement limits between conventional and digital periapical radiographies. The upper squares show the mesial agreement limits, while the lower ones show the distal agreement limits. Source: authors’ development through the research data.
Discussion

Even when the most important finding for radiographic diagnosis of periodontal disease is the discontinuity of the lamina dura, bone loss is a main characteristic in periodontitis diagnosis and treatment. Nevertheless, these radiographic findings are not enough for a diagnosis establishment (13).

Bone height radiographic assessment tends to underestimate the amount of bone loss (14-16). Furthermore, it only provides a 2D image of a 3D structure that can change the bone level geometry. Having this in mind, digital processing and manipulation can modify the diagnostic interpretation of the x-rays (17).

When overall reproducibility between the methods was assessed, poor reproducibility but moderate correlation in mesial and distal surfaces was obtained. Khocht et al. reported correlation values between 0.57 and 0.83 (P=0.01) (5). Aside this work, Kim et al. reported Pearson’s correlation between two digital methods of radiographic measurement between 0.76 and 0.79 (P<0.05) (18), which is consistent with the overall reproducibility values reported in this study. Nevertheless, it is not suggested to use this correlation coefficient due to its several statistical limitations. Thus, Lin’s coefficient was obtained since it assesses precision and accuracy of the obtained results.

Bland & Altman plots for overall reproducibility analysis showed differences between the methods. The results also showed underestimation of the radiographic measurements performed with the digital method (0.65-0.76mm); these results are not coincident with the reported results of Khocht et al. (5) and Kim et al. (18), whose results demonstrated overestimation of the bone height measurements with the digital method (0.3-0.5 mm). In addition, this underestimation is clinically significant when clinicians have to establish the diagnosis, the therapy or prognosis decisions (7,19,20).

Other factors that can explain the differences found between the methods could be attributed to film size variations and film/sensor flexibility. Even when the sensor is smaller than the film, it is difficult to place it in the oral cavity due to its stiffness. These conditions could influence positions and angulations. Additionally, having an attached USB cord to the digital system might interfere with the patient’s bite, thus altering the images (5). Furthermore, film holder’s usage could standardize the geometric projection and generate reliable images. Besides bite blocks, the holder can be standardized with higher accuracy. Eickholz et al. demonstrated that the obtained tooth-holder difference for a three-month period was lower that the measurement error from necessary angulations to obtain film positioning (21). Taking this into account, stabilized —bite blocks— film/sensor holders were used in order to minimize x-rays error probabilities, thus increasing geometry projection.

Poor reproducibility was found in the tooth subgroup analysis. Reproducibility could be influenced by the following parameters: defect dimensions, bone walls number and jaw positioning. Regarding this a better reproducibility for molar teeth was achieved when compared with premolars. However, these results are not coincident with the suggested evidence by Pepelassi et al. that shows better results for premolars and anterior teeth. Thus, in this study we did not include anterior teeth in order to standardize the x-ray process (22).

Even when the results found suggest that reproducibility was poor according to tooth subgroup; Khocht et al., in a quadrant analysis, revealed correlation coefficients between 0.57 and 0.83 (P<0.01) suggesting that these two methods do maintain a reproducibility pattern without regarding quadrant/tooth differences. In this study Pearson’s coefficients according to tooth type between 0.51 and 0.67 (P<0.05) were found. These results reveal consistent findings between the studies, even when correlation coefficients are not good enough (5).

Another factor that influences the measurement quality and therefore reproducibility is the lesion status. Tonetti et al. concluded underestimation in untreated lesions (23). Pepelassi et al. stated that bone height measurements could be underestimated in mild cases; have relative accuracy in moderate cases, and are underestimated in severe cases (22). this situation can explain the reproducibility obtained in this study: 47.2 % of the cases were severe. In this study Bland & Altman plots presented underestimation of the radiographic measurements carried out through the digital method.

Although the degree of comparison among the digital methods of image visualization is adequate, it is important to consider that the measurement is not as accurate as the image resolution suggests. This way, the image model and the assessor skills could affect measurements. Thereby, digital x-rays are not more accurate than conventional x-rays (24).

Statistically significant differences in reproducibility between the methods were found when performing assessor subgroup analysis. Wolf et al. showed no statistically significant differences in this subgroup, which is not coincident with the reported results in this study due to the poor reproducibility obtained. Moreover, they concluded that poor reproducibility can be explained when there is more than one assessor (17), which is consistent with the results obtained in the present study.

Hildebolt et al. concluded that in diagnostic studies it is possible to find higher rates of inter-examiner differences

Rev. Fac. Med. 2015 Vol. 63 No. 4: 625-31
than intra-examiner, as our results suggest. Having more than one examiner could generate additional variation in the anatomic landmarks criteria even when calibration trials can be performed (25).

Eickholz et al. (26) showed that multiple examiners with adequate training and experience is not a factor that affects the validity of the computerized measurements. In this regard, they concluded that patient and x-ray related factors are the ones who could affect the image validity and then reproducibility between methods. The poor reproducibility achieved could be partially explained by the examiners’ lack of experience in digital measurement processes. Other factors like bone density and amount of exposure were not measured. Tewary et al. attributed these examiners’ differences to the fact that x-ray measurement readings are not related to the technique being used. They showed that the examiner experience and the technique familiarity are high related to the reproducibility to be achieved (27).

Pecoraro et al. suggested single image measurement, thus minimizing observer’s differences (28). Another study suggests to obtain the mean between both measurements, generating higher reproducibility differences (5). Considering this evidence, single measurements were the only ones to be performed.

It is important to consider additional procedures in order to decrease the examiners’ differences, which could be reflected in the clinical decision making process. Delamare and Chambers recommended additional studies to obtain learning curves and thus determine the number of measurements needed to get a reliable measurement (29, 30). Learning curves allow to improve the inter-examiner performance and to encourage the diagnostic technology transition, which produces an objective diagnosis, therapy and prognostic criteria.

Conclusions

The results of this study suggest that reproducibility between methods is minimal. In addition, some factors like the assessors’ lack of experience and previous calibration processes might have a negative impact on the obtained agreement. Likewise, learning curves might have a potential role for the standardization of clinicians.

Conflicts of interest

None stated by the authors.

Financing

No external or institutional financing declared by the authors.

Acknowledgements

None stated by the authors.

Disclaimer: This manuscript was performed according to the results of an original research performed as a partial requisite to obtain the first author’s Masters in Clinical Epidemiology degree.

References


