Diagnostic Accuracy of Cone-beam Computed Tomography and Conventional Radiography on Apical Periodontitis: A Systematic Review and Meta-analysis

Kamile Leonardi Dutra, DDS, MSc, Leticia Haas, DDS, MSc, André Luis Porporatti, DDS, MSc, Carlos Flores-Mir, DDS, MSc, PbD, Juliana Nascimento Santos, DDS, MSc, PbD, Luis Andrè Mezzomo, DDS, MSc, PbD, Márcio Corrêa, DDS, MSc, PbD, and Graziela De Luca Canto, DDS, MSc, PbD

Abstract

Introduction: Endodontic diagnosis depends on accurate radiographic examination. Assessment of the location and extent of apical periodontitis (AP) can influence treatment planning and subsequent treatment outcomes. Therefore, this systematic review and meta-analysis assessed the diagnostic accuracy of conventional radiography and cone-beam computed tomographic (CBCT) imaging on the discrimination of AP from no lesion. Methods: Eight electronic databases with no language or time limitations were searched. Articles in which the primary objective was to evaluate the accuracy (sensitivity and specificity) of any type of radiographic technique to assess AP in humans were selected. The gold standard was the histologic examination for actual AP (in vivo) or in situ visualization of bone defects for induced artificial AP (in vitro). Accuracy measurements described in the studies were transformed to construct receiver operating characteristic curves and forest plots with the aid of Review Manager v.5.2 (The Nordic Cochrane Centre, Copenhagen, Denmark) and MetaDisc v.1.4. software (Unit of Clinical Biostatistics Team of the Ramón y Cajal Hospital, Madrid, Spain). The methodology of the selected studies was evaluated using the Quality Assessment Tool for Diagnostic Accuracy Studies-2. Results: Only 9 studies met the inclusion criteria and were subjected to a qualitative analysis. A meta-analysis was conducted on 6 of these articles. All of these articles studied artificial AP with induced bone defects. The accuracy values (area under the curve) were 0.96 for CBCT imaging, 0.73 for conventional periapical radiography, and 0.72 for digital periapical radiography. No evidence was found for panoramic radiography. Conclusions: Periapical radiographs (digital and conventional) reported good diagnostic accuracy on the discrimination of artificial AP from no lesions, whereas CBCT imaging showed excellent accuracy values. (J Endod 2016;42:356–364)

Key Words
dental radiography, diagnosis, meta-analysis, periapical periodontitis, review, sensitivity, specificity, X-ray computed tomographic imaging

Apical periodontitis (AP) is an inflammatory/immune response in the apical periodontium that often results from intracanal microorganisms (1). The resulting apical bone resorption is a defense mechanism that prevents the spread of infection and appears radiolucent on radiographs (2, 3).

Because AP is usually asymptomatic, it is frequently only detected during routine radiographic examination (4). In this sense, radiography is essential for the successful and timely diagnosis of AP and historically has been limited to periapical and panoramic radiographs (5). However, periapical radiographs and panoramic imaging have inherent limitations such as superimposition and distortion of important structures that commonly mask lesions (6, 7). In addition, lesions in cancellous bone cannot be consistently detected with these radiographic techniques (8). Therefore, in some cases, extensive bone resorption may be present even when there is no radiographic evidence of it (8,9).

Currently, the use of cone-beam computed tomographic (CBCT) imaging has made it possible to visualize the related anatomic structures in 3 dimensions with higher resolution. This has improved the overall diagnostic efficacy and made early diagnosis possible for some specific clinical situations (10, 11). In endodontic practice, CBCT imaging with limited field of view (FOV) has been suggested for diagnosis in patients with contradictory or nonspecific clinical signs and symptoms (12).

Clinicians must be aware of how accurate each radiographic method is and which image can provide the most reliable information regarding bone resorption around the AP. To the best of the authors’ knowledge, there are no studies correlating the diagnostic accuracy (sensitivity and specificity) of conventional radiography and CBCT imaging by means of a meta-analysis.

A previous systematic review of radiologic diagnosis of AP in endodontics found that the diagnosis made through conventional periapical radiography (CPR) presented similar accuracy outcomes compared with digital periapical radiography (DPR) and CBCT imaging (13). Another systematic review that evaluated the diagnostic efficacy
of CBCT imaging for AP focused on a 6-tiered hierarchical model and concluded that the standard use of CBCT imaging for diagnosing AP was not justified (14). However, these systematic reviews did not provide a specific quantitative synthesis (meta-analysis) of the available evidence regarding the diagnostic capability of the radiographic methods.

Therefore, the present systematic review and meta-analysis were performed to answer the following focused question: "What is the diagnostic accuracy of conventional radiography and CBCT imaging on the discrimination of AP versus no lesion?"

Methods

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist (15) and the appropriate Cochrane guidelines (16).

Protocol and Registration

The systematic review protocol was registered at the international prospective register of systematic reviews (PROSPERO) (17) under number CRD42015024704.

Study Design

A systematic review of human studies was undertaken to evaluate the diagnostic accuracy of conventional radiography and CBCT imaging, including conventional and digital panoramic and periapical radiographs, to assess AP. The accuracy of the images was used to discriminate the diagnosis of AP from no lesion.

Eligibility Criteria

Inclusion Criteria. Studies in which the primary objective was to evaluate the diagnostic accuracy of conventional radiography and CBCT imaging, including conventional and digital panoramic and periapical radiographs, to assess AP in humans were included. The gold standard was histologic examination of periapical biopsy specimens for in vivo and postmortem studies or in situ visualization of bone defects for induced artificial AP for in vitro studies. Studies performing a diagnosis of actual or artificial AP compared with a control group with no lesions were selected. No language or time restrictions were applied.

Exclusion Criteria. Exclusion criteria included studies that did not assess actual or artificial AP; reviews, case reports, letters, personal opinions, book chapters, and conference abstracts; studies conducted using animal models; studies that evaluated other diagnostic methods such as micro–computed tomographic imaging, ultrasonography, and radiographic subtraction; studies that did not use the preconized gold standard; noncontrolled studies in which the presence or absence of lesions could not be assessed; and studies that did not report sensitivity or specificity or in which the data reported could not be used to extrapolate diagnostic accuracy.

Information Sources

Detailed individual search strategies for each of the following electronic databases were performed: Embase, Google Scholar, LILACS, MEDLINE using the Ovid search engine, PubMed, Science Direct, Scopus, and Web of Science. ProQuest was searched for additional partial gray literature for any references that may have been missed. All searches were conducted on May 18, 2015. In addition to the electronic search, a hand search was made, and reference lists of the selected articles were screened.

Search Strategy

Appropriate key words and Medical Subject Heading terms were selected and combined with the use of Boolean operators (AND, OR). The search strategy was adapted for each database search with the support of a health sciences librarian (Supplemental Table S1 is available online at www.jendodon.com).

The following search string summarizes the initial search done in PubMed: ("Periapical lesions" OR "periapical periodontitis" OR "periapical radiolucency" OR "dental pulp diseases" OR "periapical diseases" OR "apical pathology" OR "endodontic pathology" OR "periapical healing" OR "periapical repair"] AND ["sensitivity" OR "specificity" OR "accuracy" OR "receiver operating characteristics curve"] AND ["dental radiography" OR "panoramic radiography" OR "periapical radiography" OR "cone beam computed tomography"]).

EndNote basic software (Thompson Reuters, New York, NY) was used to remove any duplicate articles.

Study Selection

A 2-phase selection of articles was conducted. In phase 1, 2 reviewers with expertise in endodontics and oral radiology independently reviewed the titles and abstracts of all the references. Any articles that did not appear to meet the inclusion criteria were excluded. In phase 2, the selected full articles were independently reviewed and screened by the same reviewers. Any disagreement was resolved by means of discussion. When mutual agreement between the 2 reviewers was not reached, a third reviewer with expertise in oral radiology was involved to make a final decision. The final selection was always based on the full text of the publication.

Data Items and Collection Process

For all the included studies, the following descriptive characteristics were recorded: study characteristics (authors, year, and country of publication), sample characteristics (type and size), intervention characteristics (lesions, control group, index test, and reference standard), observer characteristics (number, type, and kappa value), and findings. One reviewer collected the required information from the selected articles. The second reviewer cross-checked the collected information and confirmed its accuracy. Again, any disagreement in either phase was resolved by means of discussion, and the third reviewer made a final decision if consensus was not reached by the first 2 reviewers. If the required data were not complete or the data presented could not be extrapolated, attempts were made to contact the authors to retrieve the missing information. No further information was obtained through these contact attempts.

Risk of Bias in Individual Studies

The methodologic quality of the selected studies was evaluated using the Quality Assessment Tool for Diagnostic Accuracy Studies-2 (QUADAS-2) (18). It consisted of 4 key domains that discussed patient selection, index test, reference standard, flow of patients through the study, and timing of the index tests and reference standard. Two authors independently assessed each domain in terms of the potential risk of bias. Risk of bias was judged as “low,” “high,” or “unclear,” and a third reviewer resolved any disagreement when needed.

Summary Measures

The diagnostic sensitivity and specificity of radiographic methods in the detection of AP against controls (no lesions) were considered as the main outcome.

Synthesis of Results

To decrease the heterogeneity, the studies were analyzed in 3 groups (only ones for which data were available): CPR, DPR, and CBCT imaging. The accuracy of the different imaging methods to
discriminate AP from the no lesion controls was evaluated through a meta-analysis following the appropriate Cochrane guidelines (16). Meta-analysis data were combined using random effect models with restricted maximum likelihood estimation and the DerSimonian pooled method. All statistical analyses were crude and had no adjustment for potential confounders. Accuracy measurements described in the data items were transformed to construct receiver operating characteristic (ROC) curves and forest plots with the aid of Review Manager v.5.2 (The Nordic Cochrane Centre, Copenhagen, Denmark) and MetaDisc v.1.4 (Unit of Clinical Biostatistics Team of the Ramón y Cajal Hospital, Madrid, Spain). Heterogeneity was calculated by inconsistency indexes ($I^2$), and a value greater than 50% was considered an indicator of substantial heterogeneity between studies (19). The significance level was set at 5.0%.

Figure 1. Flowchart of the literature search and selection criteria. Adapted from PRISMA. *References of these 33 excluded articles are listed in Appendix 2.
<table>
<thead>
<tr>
<th>Study characteristics</th>
<th>Sample characteristics</th>
<th>Intervention characteristics</th>
<th>Observers</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author, year</td>
<td>Country</td>
<td>Type</td>
<td>Size (n)</td>
<td>Lesions</td>
</tr>
<tr>
<td>Holtzmann et al,</td>
<td>US</td>
<td>Human jaws</td>
<td>n = 28</td>
<td>True pathosis (n = 19)</td>
</tr>
<tr>
<td>Lennon et al, 2011</td>
<td>UK</td>
<td>Dry mandibles</td>
<td>n = 5</td>
<td>Mechanically induced (n = 10)</td>
</tr>
<tr>
<td>Lie et al, 2012</td>
<td>China</td>
<td>Dry mandibles</td>
<td>n = 15</td>
<td>Mechanically induced (n = 63)</td>
</tr>
<tr>
<td>Patel et al, 2009</td>
<td>UK</td>
<td>Dry mandible</td>
<td>n = 6</td>
<td>Mechanically induced (n = 12)</td>
</tr>
<tr>
<td>Paurazas et al, 2000</td>
<td>US</td>
<td>Dry mandible</td>
<td>n = 10</td>
<td>Mechanically induced (n = 56)</td>
</tr>
<tr>
<td>Sogur et al, 2009</td>
<td>Turkey</td>
<td>Dry mandible</td>
<td>n = 12</td>
<td>Chemically induced (n = 84)</td>
</tr>
<tr>
<td>Sullivan et al, 2000</td>
<td>US</td>
<td>Dry mandible</td>
<td>n = 6</td>
<td>Mechanically induced (n = 64)</td>
</tr>
<tr>
<td>Tsai et al, 2012</td>
<td>US</td>
<td>Dry mandible</td>
<td>n = 6</td>
<td>Mechanically induced (n = 80)</td>
</tr>
<tr>
<td>Wallace et al, 2001</td>
<td>US</td>
<td>Dry mandible</td>
<td>n = 4</td>
<td>Mechanically induced (n = 96)</td>
</tr>
</tbody>
</table>

CBCT, cone-beam computed tomographic; CPR, conventional periapical radiography; DPR, digital periapical radiography; IaO, intraobserver kappa value; IeO, interobserver kappa value.
TABLE 2. Quality Assessment with QUADAS-2

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Risk of bias</th>
<th>Applicability concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Patient selection</td>
<td>Index test</td>
</tr>
<tr>
<td>Holtzmann et al (28)</td>
<td>1998</td>
<td>(2)</td>
<td>?</td>
</tr>
<tr>
<td>Lennon et al (26)</td>
<td>2011</td>
<td>(2)</td>
<td>(2)</td>
</tr>
<tr>
<td>Liang et al (27)</td>
<td>2014</td>
<td>(2)</td>
<td>?</td>
</tr>
<tr>
<td>Patel et al (25)</td>
<td>2009</td>
<td>(2)</td>
<td>(2)</td>
</tr>
<tr>
<td>Paurazas et al (24)</td>
<td>2000</td>
<td>(2)</td>
<td>(2)</td>
</tr>
<tr>
<td>Sogur et al (20)</td>
<td>2009</td>
<td>(2)</td>
<td>(2)</td>
</tr>
<tr>
<td>Sullivan et al (23)</td>
<td>2000</td>
<td>(2)</td>
<td>(2)</td>
</tr>
<tr>
<td>Tsai et al (21)</td>
<td>2012</td>
<td>(2)</td>
<td>?</td>
</tr>
<tr>
<td>Wallace et al (22)</td>
<td>2001</td>
<td>(2)</td>
<td>(2)</td>
</tr>
</tbody>
</table>

(2) low risk/low concerns; (2), high risk/high concerns; ?, unclear; QUADAS-2, Quality Assessment Tool for Diagnostic Accuracy Studies-2.

Risk of Bias across Studies

The heterogeneity of the studies was analyzed through comparison of adequate study participation, methodologic points, and appreciation of the outcomes.

Additional Analysis

Additional analysis was performed using the positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative likelihood ratio (LR−), and diagnostic odds ratio (DOR).

Results

Study Selection

A flowchart of the process of identification, inclusion, and exclusion of studies is shown in Figure 1. In phase 1, 39 of the 924 articles found across the 8 electronic databases were selected for the next phase. A partial gray literature search was performed using ProQuest identifying 129 studies, whereas only 1 study was selected for further assessment. The reference lists of the combined 40 studies were screened, and 2 additional studies were included. Subsequently, 42 articles were retrieved for full-text reading. Thirty-three of these articles were excluded later (Supplemental Table S2 is available online at www.jendodon.com). Therefore, only 9 studies fulfilled the eligibility criteria and were included in the qualitative synthesis. Of those, only 6 were adequate to use for the meta-analysis.

Study Characteristics

From the 9 selected studies for the qualitative synthesis, all of them included radiographic examinations of dry human mandibles (20–27) or fresh human jaws (28). Sample sizes ranged from 4 (22) to 28 (28) mandibles, resulting in apical sites studied ranging from 18 (25) to 120 (22). A summary of the descriptive characteristics of included articles is provided in Table 1.

Risk of Bias within Studies

Although none of the studies fulfilled all the criteria from the risk of bias QUADAS-2 assessment tool, overall the studies’ methods were very homogeneous, and all presented a low risk of bias for applicability concerns (Supplemental Table S3 is available online at www.jendodon.com). The main areas for risk of bias within studies were related to poor reporting for the domain 1 item (ie, “patient selection”), which was scored with a high risk of bias. For every study, question 1 of domain 1 that discussed patient selection (“Was a consecutive or random sample of patients enrolled?”) was scored as “no” because all studies recruited an experimental sample without patients. Moreover, if all studies had recruited a group of no lesion controls and a group of induced lesions, the case-control design was necessary to allow the discrimination of AP from no lesion. For this reason, the second question of domain 1 (“Was a case-control design avoided?”) was also answered “no”. Additionally, the risk of bias of “index test” (domain 2, “Could the conduct or interpretation of the index test have introduced bias?”) were scored “unclear” for 3 studies related to no information about intra- or interobserver agreement. Table 2 summarizes the QUADAS-2 assessments.

Results of Individual Studies

To improve our interpretation of results, the studies were clustered in 3 groups according to the index test (only when data were available) (Supplemental Table S4 is available online at www.jendodon.com). No data were identified for CPR or DPR. Unvalidated cutoff values were selected for sensitivity, specificity, and area under the curve (AUC). No consensus values have been reported in the literature. Sensitivity, specificity, and AUC values above 80% were considered as excellent outcomes, between 70% and 80% as good, between 60% and 69% fair, and below 60% poor outcomes for a diagnostic test. The 3 groups were as follows:

1. CPR: Diagnostic accuracy of CPR for the detection of artificial AP in human jaws was assessed by 5 studies (20, 22–24, 28). All of them compared the results with the diagnostic accuracy of DPR and found no significant difference in the diagnostic performance. Three studies had sufficient information and were included in the meta-analysis (20, 22, 23). The pooled sensitivity and specificity of CPR were 0.58 (poor) and 0.70 (good), respectively.

2. DPR: Diagnostic accuracy of DPR for the detection of artificial AP was assessed by 8 studies (20–25, 27, 28). Five of them had sufficient information and were included in the meta-analysis (20, 22, 23, 25, 27). The pooled sensitivity and specificity of DPR was 0.56 (poor) and 0.78 (good), respectively. No significant difference in the diagnostic performance was found regardless of the sensor type (ie, charge-coupled device or complementary metal oxide semiconductor) (23).

3. CBCT imaging: 5 studies assessed the diagnostic accuracy of CBCT imaging (20, 21, 25–27) and concluded that CBCT imaging had excellent capacity to correctly identify artificial AP and no lesion controls. Four of them were included in the meta-analysis (20, 25–27). The pooled sensitivity and specificity of CBCT imaging were 0.95 and 0.88, respectively (both excellent).

Synthesis of Results

A diagnostic table was constructed using the data extracted from each article (Table 3). In this table, all prevalence and accuracy...
measurements (sensitivity, specificity, PPV, NPV, LR+, LR−, and DOR) are presented.

The diagnostic accuracy (sensitivity, specificity, and 95% confidence interval) of the studies included in a meta-analysis is shown in Figures 2 and 3. The sensitivity and specificity for the different selected studies varied substantially from 24%–100% and from 41%–100%, respectively. Only 1 study reported excellent sensitivity for CPR (84.33%) and DPR (80.33%) (20). In contrast, all the CBCT studies (20, 25–27) reported excellent sensitivity (90.66%–100%). Regarding specificity values, only 1 study that tested CPR reported excellent performance (94%) (22). Two studies reported excellent specificity (100%) for both DPR and CBCT imaging (25, 27).

A meta-analysis was conducted with the 6 selected studies (20, 22, 23, 25–27). We chose to showcase the meta-analysis results in ROC curves and AUC (Fig. 2). Because of differences in the radiographic methods, no cutoff point measures were justified and so no threshold effect was possible; therefore, a symmetric curve was applied. All the information about the meta-analysis of individual studies is described in Figure 3 and Supplemental Table S4. The heterogeneity found among the studies was high, ranging from 70.7%–93.4%; therefore, a random model was chosen (19). Accordingly, sensitivity and specificity values found for CBCT imaging (AUC = 0.96) were considered excellent, and the values found for CPR (AUC = 0.73) and DPR (AUC = 0.72) were considered good.

Risk of Bias across Studies

The main methodologic limitations across the studies were related to the sample selection. All of the identified studies used in vitro methods with induced lesions. Moreover, investigators did not assess intra- and interobserver agreement in some of the studies.

Additional Analysis

Regarding PPV values, the highest PPV values reported for CBCT imaging and DPR (25, 27) showed that these techniques were able to discriminate artificial AP from no lesion data 100% of the time. CBCT imaging was also reported to have the highest NPV, distinguishing control patients from those with artificial AP 100% of the time (25, 27).

Regarding LR values, 4 studies showed LR+ greater than 1.00 for CPR, DPR, and CBCT imaging (20, 22, 23, 26), which means that all radiographic techniques argue for the diagnostic of interest (29). The highest LR+ value was reported for CPR (LR+ = 7.00) (22) followed by CBCT imaging (LR+ = 3.37) (26). LR− values closer to 0 were reported for CBCT imaging (20, 25–27), which means a low probability of disease when it is absent in the examination (29).

Finally, the highest DOR reported was for CBCT imaging (25, 27), indicating better discriminatory test performance (30). In general, these results showed that CBCT imaging satisfied the criteria required for an excellent diagnostic test. CPR and DPR satisfied the criteria for a good diagnostic test (Table 3).

Discussion

This systematic review and meta-analysis investigated the available evidence regarding the diagnostic accuracy of conventional radiography and CBCT imaging to discriminate AP from controls with no lesions in humans. Adequate reference testing may be examination of patient

### TABLE 3. Diagnostic Accuracy of Index Test

<table>
<thead>
<tr>
<th>Index test</th>
<th>Study, authors, year</th>
<th>Sample size (n)</th>
<th>Prevalence %</th>
<th>Sensitivity %</th>
<th>Specificity %</th>
<th>PPV %</th>
<th>NPV %</th>
<th>LR+</th>
<th>LR−</th>
<th>DOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional radiography</td>
<td>Sogur et al, 2009 (20)</td>
<td>84</td>
<td>75.00*</td>
<td>84.33*</td>
<td>52.00*</td>
<td>84.00*</td>
<td>52.00*</td>
<td>1.75*</td>
<td>0.36*</td>
<td>5.83*</td>
</tr>
<tr>
<td></td>
<td>Sullivan et al, 2000 (23)</td>
<td>80</td>
<td>80.00*</td>
<td>60.93</td>
<td>56.25</td>
<td>84.51*</td>
<td>25.93*</td>
<td>1.36*</td>
<td>0.71*</td>
<td>1.87*</td>
</tr>
<tr>
<td></td>
<td>Wallace et al, 2001 (22)</td>
<td>120</td>
<td>80.00*</td>
<td>42.00</td>
<td>94.00</td>
<td>96.00*</td>
<td>28.00*</td>
<td>7.00*</td>
<td>0.61*</td>
<td>16.42*</td>
</tr>
<tr>
<td>Digital radiography</td>
<td>Liang et al, 2014 (27)</td>
<td>100</td>
<td>63.00*</td>
<td>67.00</td>
<td>100.00</td>
<td>100.00</td>
<td>64.00</td>
<td>—</td>
<td>0.33*</td>
<td>148.26*</td>
</tr>
<tr>
<td></td>
<td>Patel et al, 2009 (25)</td>
<td>18</td>
<td>66.00*</td>
<td>24.00</td>
<td>100.00</td>
<td>100.00</td>
<td>38.00</td>
<td>—</td>
<td>0.76*</td>
<td>4.78*</td>
</tr>
<tr>
<td></td>
<td>Sogur et al, 2009 (20)</td>
<td>84</td>
<td>75.00*</td>
<td>80.33*</td>
<td>67.00*</td>
<td>87.00*</td>
<td>52.00*</td>
<td>2.41*</td>
<td>0.22*</td>
<td>7.69*</td>
</tr>
<tr>
<td></td>
<td>Sullivan et al, 2000 (23)</td>
<td>80</td>
<td>80.00*</td>
<td>62.70*</td>
<td>94.00</td>
<td>90.78*</td>
<td>21.24*</td>
<td>1.05*</td>
<td>0.92*</td>
<td>1.29*</td>
</tr>
<tr>
<td></td>
<td>Wallace et al, 2001 (22)</td>
<td>120</td>
<td>80.00*</td>
<td>34.50*</td>
<td>72.50*</td>
<td>82.00*</td>
<td>21.00*</td>
<td>1.21*</td>
<td>0.91*</td>
<td>1.27*</td>
</tr>
<tr>
<td>Cone-beam computed tomographic imaging</td>
<td>Lennon et al, 2011 (26)</td>
<td>20</td>
<td>50.00*</td>
<td>91.00</td>
<td>73.00</td>
<td>89.10</td>
<td>76.00</td>
<td>3.37*</td>
<td>0.12*</td>
<td>21.00*</td>
</tr>
<tr>
<td></td>
<td>Liang et al, 2014 (27)</td>
<td>100</td>
<td>63.00*</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>—</td>
<td>0.00*</td>
<td>9525.00*</td>
</tr>
<tr>
<td></td>
<td>Patel et al, 2009 (25)</td>
<td>18</td>
<td>66.00*</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>—</td>
<td>0.00*</td>
<td>325.00*</td>
</tr>
<tr>
<td></td>
<td>Sogur et al, 2009 (20)</td>
<td>84</td>
<td>75.00*</td>
<td>90.66*</td>
<td>70.00*</td>
<td>90.00*</td>
<td>70.00*</td>
<td>3.00*</td>
<td>0.14*</td>
<td>23.75*</td>
</tr>
</tbody>
</table>

—, data not available in the original article; DOR, diagnostic odds ratio; LR+, positive likelihood ratio; LR−, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

*Data calculated from information available in the article.

---

**Figure 2.** ROC curves.

---
biopsies (13). However, the ethical considerations regarding this practice make it impossible to perform.

We found only 9 eligible studies with data concerning CPR, DPR, and CBCT imaging (20–28). No data were found for panoramic radiography. The sensitivity and specificity index values of tests varied among the included studies. The differences in these scores probably reflect the variation among bone lesion sizes, the anatomic region examined, the observer experience, and so on. Concerning the images’ observer experience, most of the included studies reported poor or no information about inter- or intraobserver agreement (20–24, 26–28). Regarding the anatomic region examined in each study, lesions in cortical bone were detected with greater accuracy than in trabecular bone (25). Thus, higher false-negative rates could be related to induced lesions in the lacunae of the cancellous bone, increasing different results among the studies (31). Moreover, diagnostic accuracy was related to lesion size in some of the included studies. Larger lesions were detected more frequently regardless of the type of radiographic system used (23, 24).

A meta-analysis was conducted over the 6 selected studies (20, 22, 23, 25–27). All of these studies were conducted with artificial AP induced in skeletal material rather than actual AP. The ROC curves and AUC show meta-analysis results based on a combination of both sensitivity and specificity values (Fig. 2). The greater the area under the ROC curve, the higher the accuracy to discriminate artificial AP from control data and the more useful the radiographic method will be (32). Based on this, CBCT imaging provided excellent diagnostic accuracy (AUC value above 80%) enabling the identification of those who have artificial AP, whereas CPR and DPR only performed good accuracy (AUC values between 70% and 80%).

The PPV and NPV values confirmed the ROC curve results. The higher PPV related to CBCT imaging showed a lower risk of false-positive results (ie, finding more lesions than actually exist). In addition, a high NPV noticed in CBCT evaluations indicates that there is also a lower risk of underdiagnosis. This was not noticed in CPR and DPR evaluations, which showed NPV values ranging from 21%–64%. It is important to emphasize that the prevalence of a disease can affect PPV and NPV values. When prevalence is high, true-positive results are more likely to be found in the population instead of false-positives, increasing the NPV and decreasing the PPV, respectively (33). Similarly, the DOR values of index tests reported indicate that CBCT imaging had better discriminatory test performance (25, 27). Finally, LR+ and LR– values expressed better diagnostic accuracy for CBCT imaging (20, 25–27), with the exception of 1 study that reported a better value for CPR (22).

To the best of the authors’ knowledge, this is the first systematic review to validate sensitivity and specificity values of CPR, DPR, and CBCT imaging. These values, added to PPV, NPV, LR+ , LR–, and ROC curve analyses, were used for diagnostic accuracy. In a clinical setting, it is important to know how a diagnostic radiographic method discriminates diseased patients from healthy ones. However, all of these radiographic examinations can only identify the presence or absence of bone resorption (13). These methods are not sensitive enough to provide a pathological state, and to obtain this information, a histologic analysis must be performed (13).

In this study, the best diagnostic accuracy results were reported when using CBCT imaging. Even so, if CBCT imaging is chosen for AP diagnosis, the adjustment of FOV should be considered in order to use the lowest radiation dose possible on the patient (31, 34). As low as reasonably achievable principles must be considered always (12). Accordingly, the latest American Academy of Oral and Maxillofacial Radiology/American Association of Endodontists joint position statement reported that limited FOV CBCT imaging should be considered the modality of choice for diagnosis in patients who present with contradictory or nonspecific clinical signs and symptoms (12). All included studies regarding CBCT imaging in this systematic review used similar exposure parameters with limited FOV and high resolution (Supplemental Table S5 is available online at www.jendodon.com).

Figure 3. Forest plot with the diagnostic accuracy (sensitivity, specificity, and 95% confidence interval [CI]) of each study. FN, false-negative; FP, false-positive; TN, true-negative; TP, true-positive.
Studies that did not compare an index test with the gold standard were not included (35–47) because only an acceptable gold standard can prove the clinical relevance and reduce the risks of both false-positive and false-negative findings (31). Adequate reference testing may be the examination of biopsies from patients and autopsies from deceased individuals (13, 28, 31). However, ethical considerations regarding this practice make it impossible to perform. In this study, we considered the in situ visualization of artificial lesions (in vitro studies) (20–27) or the histologic analysis of the surgical specimen (in vivo and postmortem studies) (28) as the gold standard. Studies using an animal model were excluded (31, 48–50) because of the different apical anatomy that may be correlated to poor accuracy results, leading to no clinical extrapolation (31).

The main methodologic limitation of our systematic review is that all included studies used in vitro methods in which artificial AP were induced in the skeletal material by drilled holes (21–27) or acid applied at the periapical bone tissue (20). Even though mechanically induced lesions can be accurately measured, it does not produce the diffuse borders of natural AP, and chemically induced lesions have a more natural appearance but are difficult to measure (13).

In conclusion, CPR and DPR showed good diagnostic accuracy on the discrimination of artificial AP from no lesion, whereas CBCT imaging reported excellent accuracy values.

Acknowledgments

The authors thank Mrs Maria Gorete Monteguti Savi for her invaluable contribution at the stage of the search strategy construction.

The authors deny any conflicts of interest related to this study.

Supplementary Material

Supplementary material associated with this article can be found in the online version at www.jendodont.com (http://dx.doi.org/10.1016/j.joen.2015.12.015).

References


