Room for improvement?

The accuracy of dental practitioners who diagnose bony pathoses with radiographs

Stephen E. Stheeman, PhD, DDS,a Phil A. Mileman, PhD, MScD, DDPHRCS, BDS,b Martin van ‘t Hof, PhD, MSc,c and Paul F. van der Stelt, PhD, DDS,d Amsterdam, The Netherlands

ACADEMIC CENTRE FOR DENTISTRY AMSTERDAM AND UNIVERSITY OF NIJMEGEN

Objectives. The impact of any effort aimed at improving diagnostic accuracy by improving clinical decision making in diagnostic radiology will be limited by the ability of the clinician to correctly recognize the presence of abnormalities on radiographs. We carried out a study designed to examine whether dentists are able to correctly identify various kinds of periapical bone lesions visible on intraoral radiographs and diagnose their pathologic nature.

Study design. General dental practitioners (n = 98) assessed 32 radiographs that showed either normal bone (10) or one abnormality (22) in the periapical bone. The “gold standard” for pathosis was histopathologic analysis. The dentists were asked to judge for the presence of an abnormality and to decide whether an active pathologic process was present.

Results. On average dentists identified 81% of all visible abnormalities correctly. Subsequently, they diagnosed 59% of all the pathologic cases correctly. Dentists, however, incorrectly identified 55% lesions on radiographs when experts had stated that no abnormality was visible.

Conclusion. There is room for improvement of diagnostic accuracy of bony pathology.


Considerable interobserver variation has been shown to exist in the assessment of the presence of pathoses in radiographs.1 In addition, dentists use different treatment criteria that result in differences in treatment planning.2 Ways to reduce this variation have been suggested.3,4 For instance, a checklist to prompt the observer to focus on parts of the radiographic image with known significance for the probable diagnosis has been developed, and it successfullly increased diagnostic accuracy in an experimental setting.5,6 However, in practice efforts aimed at improving clinical decision making for radiographically visible lesions will be limited by the practitioner’s ability to detect an abnormality in the radiographic image and to classify its pathologic nature correctly. Subsequently, the clinician may decide to investigate a case in more detail. Because an abnormality that is either not detected or is classified as being insignificant will not receive further attention, the degree of improvement in diagnostic accuracy attainable will be limited. Only positive assessments from radiographs of the presence of pathoses will trigger the practitioner into further investigations.

The aims of the present study were the following: (1) to describe the ability of general dental practitioners (GDPs) to recognize abnormalities present in the image of the bone on periapical radiographs; (2) to describe GDPs’ diagnostic accuracy for bony pathoses visible on periapical radiographs; (3) to describe the variation in treatment proposals for bony pathoses between GDPs when they use periapical radiographs; (4) to estimate the room for improvement in the clinical diagnosis of bony pathoses.

MATERIAL AND METHOD

The sample

GDPs (n = 107) were subjects in this study. They had registered to take part in continuing education courses on endodontics (n = 79) or prosthodontics (n = 28) in the 2 weeks after the study. These GDPs were considered to be representative of practitioners likely to be receptive to innovations in diagnosis. Each GDP was mailed a questionnaire that contained 32 periapical radiographs. The response was 95% (102 questionnaires were returned). Because a few dentists enrolled for more than one course during the period in which this study was to be performed or failed to fill out the questionnaire completely, the actual number of questionnaires used was 98 (92%).

*Staff Fellow, Department of Oral Radiology, Academic Centre for Dentistry Amsterdam.

Assistant Professor, Department of Oral Radiology, Academic Centre for Dentistry Amsterdam.

Associate Professor, Department of Medical Informatics and Epidemiology, University of Nijmegen.

Professor, Department of Oral Radiology, Academic Centre for Dentistry Amsterdam.

Received for publication Feb. 27, 1995; returned for revision Apr. 12, and July 6, 1995; accepted for publication Sept. 5, 1995. Copyright © 1996 by Mosby-Year Book, Inc. 1079-2104/96/$5.00 + 0 7/16/69323
The questionnaire
A series of 32 duplicated radiographs of the mandibular premolar and molar area were used in the study. The practitioners were provided with a schematic diagram that showed the region of interest (Fig. 1). For each radiograph the GDPs were asked to examine the periapical bone in the region indicated on the questionnaire and to state (1) whether an abnormality could be seen in the periapical bone; (2) whether it had generally a radiolucent or radiopaque appearance; (3) whether it was certainly pathologic; and (4) whether it was in need of treatment.

The duplicated radiographs
The original radiographs had been mounted on transparent film and duplicated on to Kodak X-Omat duplicating film (Eastman Kodak Co., Rochester, N.Y.) with the BXR MKII radiograph duplicating printer (Blu-Ray Company, Essex, Conn.). All duplicating films were developed in an X-Omat M35 processor with RP X-Omat chemicals according to the manufacturer's instructions (Eastman Kodak Co.). Kodak's X-Omat duplicating film is considered to be capable of producing clinically acceptable duplicate radiographs.7

The films were checked after duplication for quality by an independent oral radiologist to ensure acceptable quality levels. Any film that contained artifacts or was of unacceptable density or contrast was discarded. The films were then individually mounted in the Rinn Corporation's Ezeemount dental X-ray film mounts (Rinn Corporation, Elgin, Ill.).

Table I. Accuracy of recognition of bony abnormalities from periapical radiographs

<table>
<thead>
<tr>
<th></th>
<th>Mean ± SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly recognized radiographs with normal bone</td>
<td>45% ± 14</td>
<td>10%-80%</td>
</tr>
<tr>
<td>Correctly recognized radiographs with an abnormality</td>
<td>81% ± 11</td>
<td>50%-100%</td>
</tr>
<tr>
<td>Correctly recognized radiographs with a radiopaque abnormality</td>
<td>40% ± 12</td>
<td>6% ± 63%</td>
</tr>
<tr>
<td>Correctly recognized radiographs with a radiolucent abnormality</td>
<td>54% ± 16</td>
<td>0%-83%</td>
</tr>
</tbody>
</table>

SD, Standard deviation.

Table II. Diagnostic accuracy for bony pathoses

<table>
<thead>
<tr>
<th></th>
<th>Mean ± SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity (overall)</td>
<td>61% ± 12</td>
<td>33%-92%</td>
</tr>
<tr>
<td>Sensitivity (overall)</td>
<td>59% ± 18</td>
<td>10%-100%</td>
</tr>
<tr>
<td>Sensitivity (radiopaque)</td>
<td>65% ± 19</td>
<td>14%-100%</td>
</tr>
<tr>
<td>Sensitivity (radiolucent)</td>
<td>45% ± 28</td>
<td>0%-100%</td>
</tr>
</tbody>
</table>

SD, Standard deviation.

The radiographic and diagnostic norms
A radiographic silver standard (RSS) was established by a Delphi consensus panel for the presence of abnormalities visible in the image of the periapical bone on the 32 radiographs. This panel consisted of four independently chosen, international, expert, oral radiologists described elsewhere.8 The panelists assessed the radiographs during a Delphi consensus procedure,9 stating whether any abnormality was visible in the region of the periapical bone indicated and whether the abnormality was generally radiopaque or radiolucent.

After three rounds of anonymous correspondence the panel agreed on the assessments made for the 32 radiographs. In 91% of all cases all panelists agreed; in the remaining three cases the majority of panelists agreed on the assessment. According to this RSS, the set of 32 radiographs contained 10 radiographs without an abnormality visible in the image of the periapical bone in the region indicated, 16 radiographs with a radiopaque abnormality, and 6 radiographs that showed a radiolucent abnormality.

In order to be able to measure dentists' accuracy in diagnosing pathoses from periapical radiographs, a diagnostic gold standard was established for 12 cases on which a visible abnormality had been detected by the Delphi panel. These had been the subject of histopathologic examination to confirm the diagnosis (Gröndahl H-G, personal communication, 1993). The gold standard comprised two cases of idiopathic osteosclerosis, six cases of condensing osteitis, two
cases of osteosarcoma, and two cases of central ossifying fibroma. The two cases of idiopathic osteosclerosis were not considered to represent a pathologic process or to require treatment; all other cases were considered to have a pathologic nature and to be in need of treatment.\textsuperscript{10,11}

**Statistical analysis**

The data were analyzed with the SPSS-X statistical package (SPSS, Inc., Chicago, Ill.) on a mainframe computer (Hitachi, Japan). The accuracy of the recognition of abnormalities from the radiographs was calculated for the GDPs with the RSS. Accuracy is expressed as the percentage of correctly recognized radiographs with a visible bony abnormality (n = 22) and the percentage of correctly recognized radiographs without visible abnormality (n = 10). Specificity and sensitivity were calculated for the GDPs' diagnosis of pathoses.\textsuperscript{12} Specificity is expressed as the number of correctly diagnosed healthy cases as a percentage of all healthy cases in the test set (n = 12) (TN = true negatives; FP = false positives):

\[
\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{\text{True Negative}}{\text{All Actual Negative}}
\]

Sensitivity is expressed as the number of correctly diagnosed pathologic conditions as a percentage of the cases of pathoses according to the diagnostic gold standard (n = 10) (TP = true positives; FN = false negatives)

\[
\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{True Positives}}{\text{All Actual Positives}}
\]

Overtreatment proposed by the GDPs was calculated as the percentage of cases subjectively diagnosed as healthy for which treatment was suggested. Undertreatment was calculated as the percentage of cases subjectively diagnosed as pathologic for which no treatment was proposed.

Pearson's correlation coefficients were calculated for the relationships between the recognition and the classification of abnormalities and treatment planning.

**RESULTS**

An overview is shown in Table I of the accuracy of recognition of bony abnormalities among GDPs for radiopaque as well as for radiolucent abnormalities. On average, dentists performed better in the recognition of the presence of abnormalities (81\% ± 4\% 11) than in the recognition of normality (45\% ± 14).

GDPs' average overall sensitivity and specificity for all cases as well as for radiopaque and radiolucent pathoses is shown in Table II. Differences in accuracy for diagnosing pathoses (59\% ± 18) compared with the accuracy of the diagnosis of health (61\% ± 12) are smaller than the difference between the recognition of the presence and absence of abnormalities. The interobserver variation remains large. Dentists in our study diagnosed the presence of radiopaque pathoses (65\% ± 19) more accurately than the presence of radiolucent pathoses (45\% ± 28) although they recognized radiolucent abnormalities more accurately (54\% ± 16) than radiopaque abnormalities (40\% ± 12). Over- and undertreatment suggested by the GDPs compared with their subjective diagnoses are shown in Table III. Interestingly, dentists suggested treatment in a mean of 3\% of cases they had earlier diagnosed as healthy. On average, the dentists were carefully suggesting treatment that resulted in 11\% undertreatment of subjectively diagnosed pathoses. Pearson's correlation coefficients were calculated for various relationships between the accuracy of the recognition of abnormalities, diagnostic accuracy, and the accuracy of overtreatment and undertreatment compared to GDPs' subjective diagnoses. An overview is shown in Table IV.

**DISCUSSION**

Interobserver variation in this study was high. The accuracy of recognition of radiographs with an abnormality present ranged between 50\% and 100\% (Table I). On the other hand, the accuracy of dentists in the recognition of normality varied between 10\% and 80\% of radiographs with normal bone according to the RSS (Table I). The dentists viewed the radio-
graphs in their own practice or home environment. Variation in viewing conditions is unlikely to have contributed significantly to the observer variation found in this study.\textsuperscript{13,14} Nor is it likely that the duplication procedure would have adversely affected dentists’ performances.\textsuperscript{7}

Dentists seem to be quite inaccurate in identifying both radiopaque and radiolucent abnormalities when compared with their overall accuracy in recognizing the presence of abnormality (Table I). In calculating these accuracies, however, dentists were first judged for their accuracy in the recognition of visible abnormalities according to a RSS. Their ability to correctly identify the radiographic density of the abnormalities was then assessed. The accuracy of detecting radiopaque and radiolucent abnormalities in the radiographic image is therefore limited by the dentist’s accuracy in detecting the presence of a visible abnormality.

The percentage of radiographs with normal bone according to the RSS on which dentists incorrectly recognized the presence of an abnormality (19%; Table I, 100% minus the percentage of correctly recognized radiographs with an abnormality) is low compared with their false-negative diagnoses of pathoses (41%; Table II, 100% minus sensitivity). In this study the dentists were explicitly asked to diagnose a pathosis only if they were certain of its presence. It is likely that this diagnostic threshold, on average, was more conservative for dentists in our study than the threshold they might use to decide on the presence of a visible abnormality. This may have resulted in fewer false-positive diagnoses at the expense of an increased percentage of false-negative diagnoses.

On average, dentists in our study showed a tendency to undertreat pathoses compared with their subjective diagnoses (Table III). Dentists may be using an even more conservative threshold for their treatment decisions. This might be caused by their judgment of the seriousness of an incorrect decision.\textsuperscript{15} They might fear false-positive diagnoses of pathoses that result in unnecessary treatment and be more prepared to tolerate the presence of chronic disease. The practitioners’ estimates of the chances of successful treatment or retreatment in individual cases especially in relation to the consequences of such treatment might also make them more conservative in suggesting treatment.

Our results show, on average, 39% false-positive diagnoses of bony pathoses (Table II, 100% minus specificity). Dentists in this study were more accurate in diagnosing the presence of radiopaque pathoses than they were in diagnosing the presence of radiolucent pathoses (Table II). Possibly, dental practitioners find radiopaque abnormalities more alarming than radiolucent ones, which results in an increased sensitivity for radiopaque pathoses. It may, on the other hand, be expected that this results in an overestimation of the presence of pathoses among radiopaque abnormalities such as cases of idiopathic osteosclerosis. Idiopathic osteosclerosis was considered by us to be a variant of normal bone.\textsuperscript{10}

We conclude that there is room for improving diagnostic accuracy of bony pathoses by dentists with the use of a decision aid. This might especially apply to radiopaque abnormalities on radiographs.

We thank Dr. Gerard Sanderink for checking the duplicate radiographs used in this study for technical quality.

\section*{References}


\textit{Reprint requests:}

Stephen E. Stheeman, PhD, DDS Department of Oral Radiology Academic Centre for Dentistry Amsterdam Louwesweg 1 1066 EA Amsterdam, The Netherlands