Diagnostic agreement between radiologists, dentists and dental students for radiographic detection of approximal caries.

Abstract: Aim: To assess intraobserver and interobserver agreement for radiographic detection of approximal caries among radiologists, dentists and senior dental students. Materials and methods: 75 standardized bitewing radiographs were examined by four maxillofacial radiologists, four dentists and four senior dental students. Interobserver agreement was calculated using a weighted Kappa Cohen test. Two weeks later, the examiners re-evaluated 10% of the sample under the same conditions and intraobserver agreement weighted Kappa Cohen test was calculated. Results: The interobserver Kappa value was 0.68 (good) for the dentist-student pair, 0.51 (moderate) for the student-radiologist and 0.62 (good) for the dentist-radiologist pair. All these differences were significant. The intraobserver agreement Kappa values obtained were 0.56 (moderate) for students (p=0.46), 0.46 (moderate) for dentists (p<0.05) and 0.68 (good) for radiologists. Conclusion: The agreement for proximal caries detection by dental students, dentists and radiologists was moderate to good.

Keywords: Radiography, dental caries, diagnosis, observer variation.

DOI: 10.17126/joralres.2015.034


INTRODUCTION.

Radiographic caries detection allows to evaluate areas that are difficult to assess clinically, particularly approximal surfaces, thus facilitating the decision-making process. Two essential aspects for the evaluation of a diagnostic method are: validity, i.e., demonstrating that the test measures what it intends to measure, and reliability, i.e., that different evaluators obtain the same measures under similar conditions or that the same evaluator obtain the same measure under the same condition. While the validity of results is measured by sensitivity and specificity, reliability is usually expressed in terms of intraobserver and interobserver agreement, generally using Kappa statistic. While the validity of bitewing radiography has been extensively studied, with more than 168 publications available in PubMed to date, there are only a few papers assessing the agreement for radiographic detection of caries. Some publications available from 1997 to date have assessed the agreement on the use of bitewing radiographs and panoramic radiographs to evaluate approximal caries; on the use of high and low resolution radiographs, on direct or indirect digital radiographs, conventional or digital, and on the use of bitewing and oblique lateral radiographs.

The assessment of diagnostic agreement becomes important because it is increasingly common for a patient to be examined by different professionals, either by interconsultation or referral, asking for a second opinion, or due to changes in the patient’s health insurance policy, which may force the patient to go to a different dentist. Although recent publications report on the diagnostic agreement for the clinical evaluation of caries, or between clinical and radiographic evaluation, there are no publications repor-
ting on the agreement with respect to radiographic eva-
ulation between radiologists and different dentists with
varying degrees of professional experience. Therefore,
the aim of this study is to determine the intraobserver
and interobserver agreement for radiographic evaluation
of approximal caries between radiologists, dentists and
dental students.

MATERIALS AND METHODS.
This is a study on diagnostic agreement made ac-
cording to the recommendations given by the GRRAS13.
The study protocol was approved by the Scientific Ethics
Committee of the Faculty of Medicine at Universidad
Austral de Chile.

Sample size
Sample size was determined following the recommen-
dations of Liao 201014, assuming the hypothesis that
all observers agree on the diagnosis, a discordance rate
of 0.20 and probability of tolerance of 90% to 95%. A
sample size of 12 participants was obtained. This sam-
ple size was then divided into three groups comprising:
4 dentists, 4 senior dental students and 4 radiologists.
Bitewing radiographs standardized by a PBR-0546 posi-
tioner holder (Indusbello, Londrina, Brazil) were used.
These radiographs belonged to the clinical records held
by 3rd year dental students at Universidad Austral de
Chile. Radiographs on which at least a healthy surface
with radiolucent shadow reaching into enamel or dentine
were included. Radiographs were digitized and stored in
compressed.8.JPG format. The number of radiographs to
be studied was set at 75 to ensure that all possible pre-
sentations of radiographic caries were examined and to
minimize recall bias of participants.

Enrollment
A convenient sample was used. 55 dentists, 5 radiolo-
gists and 22 students from School of Dentistry at Uni-
versidad Austral de Chile were invited to participate in
the study. The final sample consisted of the first con-
secutive subjects who accepted the invitation. Since there
were few radiologists in the city of Valdivia, radiologists
from neighboring cities were invited to take part in the
study. One radiologist from the city of Osorno accepted
the invitation.

Measurements
Each participant conducted an evaluation session of
75 standardized digital bitewing radiographs in order to
detect approximal caries. Each evaluation was per-
formed in a dim-lighted room. Radiographs were visualized
on a 4th generation iPad (Apple, Cupertino, CA, USA)
in landscape mode (display resolution of 2048x1536 pi-
exels); using Keynote for iOS, screen brightness of 100%.
Images were viewed one at a time and the indications
for the session were given to each participant both ver-
bally and in written. For the assessment of interobserver
agreement, participants were asked to make a diagnosis
for each radiograph for an approximal tooth surface pre-
viously indicated by the same participant. Participants
performed each diagnosis using a scale from 0 to 2 as
seen in Table 1. We decided to use the minimum re-
quired diagnostic criteria to assess the agreement of the
participants so that the result obtained were not affected
by the validity and replicability of the selected criteria.

For the assessment of intraobserver agreement partici-
pants were asked to evaluate 10% of the previously exa-
mined radiographs. This assessment was made two weeks
after the first measurement and under the same initial con-
ditions. Each evaluation was conducted independently.

Statistical analysis
Evaluations made by participants were considered as a
nominal variable (see Table 1). When grouping by cate-
gory student-dentist-radiologist, scores for each diagno-
sis were added. To assess intraobserver and interobserver
agreement a Cohen’s Kappa weighted test was performed.
Kappa values used were <.20 poor, 0.21-0.40 tolerable,
0.41-0.60 moderate, 0.61-0.80 good, and 0.80-1 very

| Table 1. Diagnostic guidelines given to participants. |
| --- | --- |
| **Criterium** | **Code** |
| Sound surface | 0 |
| Enamel lesion | 1 |
| Enamel and dentine lesion | 2 |
good\textsuperscript{15}. The test was chosen because it allows to determine the sources of disagreement among observers on a nominal scale with more than two categories and the effect of the disagreements on the value of Kappa\textsuperscript{16}. Statistical significance was established using the chi-square test comparing the differences between proportions, and the significance level was set at p=0.05.

**RESULTS.**

Characteristics of the participants are shown in Table 2.

The intraobserver agreement showed Kappa values of 0.56 for students (p=0.46), 0.46 for dentists (p<0.05) and 0.68 for radiologists (p<0.01). Details are shown in Table 3.

When assessing the agreement between different groups of participants, the interobserver Kappa calculated was 0.68 for the dentist-student pairing, 0.51 for the student-radiologist, and 0.62 for the dentist-radiologist, as seen in Table 4. All these differences were significant.

**DISCUSSION.**

Moderate to good values of agreement were found among all participants. Kappa values for intraobserver ranged from 0.46 to 0.68, and from 0.51 to 0.68 for interobserver. These values fall within expected ranges considering the high diagnostic variability reported for the clinical evaluations of caries. Recent studies show interobserver agreement using unweighted Kappa values that range from 0.31 to 0.61 in the clinical diagnosis using the criteria ICDAS\textsuperscript{17}, from 0.63 to 0.81 using different types of radiographs\textsuperscript{18}, and from 0.32 to 0.87 using laser fluorescence in approximal lesions\textsuperscript{19}.

The explanation of these results requires an examina-

---

**Table 2. Demographic characteristics of participants.**

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Average age</th>
<th>men : women</th>
<th>Range of experience in years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>4</td>
<td>25</td>
<td>3:1</td>
<td>5-6</td>
</tr>
<tr>
<td>Dentists</td>
<td>4</td>
<td>27</td>
<td>3:1</td>
<td>1-5</td>
</tr>
<tr>
<td>Radiologists</td>
<td>4</td>
<td>35</td>
<td>4:0</td>
<td>3-12</td>
</tr>
</tbody>
</table>

**Table 3. Intraobserver Kappa agreement on radiographic assessment of approximal surfaces.**

<table>
<thead>
<tr>
<th>Second assessment</th>
<th>First assessment</th>
<th>Dentine Kappa</th>
<th>Intraobserver</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sound</td>
<td>Enamel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sound</td>
<td>8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Enamel</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Dentine</td>
<td>1</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Dentists</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sound</td>
<td>9</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Enamel</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Dentine</td>
<td>1</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Radiologists</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sound</td>
<td>15</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Enamel</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Dentine</td>
<td>1</td>
<td>6</td>
<td>10</td>
</tr>
</tbody>
</table>

**Table 4. Interobserver Kappa agreement on radiographic assessment of approximal surfaces between students, dentists and radiologists.**

<table>
<thead>
<tr>
<th></th>
<th>Dentists</th>
<th>Radiologists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>0.68*</td>
<td>0.51*</td>
</tr>
<tr>
<td>Dentists</td>
<td>0.62*</td>
<td></td>
</tr>
</tbody>
</table>

* p<.0001
tion of possible limitations. A potential limitation of the study was the use of an iPad to display the images. The choice of this device was made on the basis that there are no differences between the diagnosis made on a mobile device with a high resolution screen and a traditional negatoscope. Therefore, considering that the aim of the study was to assess the agreement and not diagnostic accuracy, we decided to use the iPad to standardize as much as possible the conditions for reading images.

The sample was a convenience sample of dentists and radiologists from a dental school in particular. The School of Dentistry at Universidad Austral de Chile began operations in 2005, so that most of its members were trained in different universities and colleges. A similar situation occurs in dental public health services. However, the fact that they share the same type of academic work could contribute to a higher diagnostic agreement. Participating dentists showed the lowest level of agreement (weighted Kappa=0.46) so it can be assumed that each dentist maintains his/her own diagnostic opinion. On the other hand, students showed a higher level of agreement (weighted Kappa=0.56). This value suggests that among students there would exist more similar criteria when evaluating radiographs. This also confirms results obtained in previous independent studies that showed a higher level of agreement among dental students than among dentists.

It was assumed that the use of a simple scale, with three possible diagnoses, should have increased the level of agreement. The former scale was chosen because it characterized the clinical radiographic diagnosis more accurately. However, even with this simplified scale, the level of agreement obtained (considered as good) means that a radiologist agrees on a diagnosis in 7 out of 10 radiographs, a student in 6 out of 10, and a dentist in 5 out of 10.

In the study most similar to this investigation, researchers used a sample of 20 participants, divided into groups of 5 senior students from Brazil and Switzerland, and a similar number of dentists from the same countries. They used a 5-point scale, and the level of agreement was calculated using unweighted Kappa. They used the averages of the groups to detect significant differences. Even with these methodological differences with the present study, Diniz et al. found low levels of intraobserver and interobserver agreement, with a slightly higher level of agreement among students than dentists. The authors of the present study share this finding and also the explanation that this level of agreement could be the reflection of the many factors involved in the detection and treatment of caries, as Bader and Shugars reported in 1992.

The results obtained could be in part explained by Bader and Shugars’s explanations, and partly by other factors such as age, training, skills, preferences and experiences of students and professionals, as described in Lazarchik et al. In this regard, Hellen-Halme and Petersson evaluated how variations in educational level and experience influenced diagnostic accuracy for the detection of approximal carious lesions. To achieve this objective, they asked 7 dentists (D) and 21 dental students in their 6th semester (E-6), who had just finished their dental radiology training, in their 10th semester (E-10), the last one before graduation, and dental hygiene students (H) in their final semester. In measuring the area under the ROC curve, they found D=0.598; D-10= 0.539; D-6=0.577 and H=0.576 values, which suggests that education and experience have an effect on the diagnostic accuracy for detecting approximal caries. The findings by Bader and Shugars as well as by Hellen-Halme and Petersson regarding diagnostic accuracy could also explain the results of this study with respect to diagnostic agreement. Considering that radiologists are the ones who usually examine more radiographs, one might speculate that the fact of showing dental radiographs to 75 students and dentists could affect the measurement. This effect would therefore tend to improve intra and inter agreement, by reducing the variation that might be expected with the observation of a few radiographs.
There is consensus that there should be an emphasis on education and training in the diagnosis of caries in undergraduate dental students\(^2^6\) in order to reduce the lack of reproducibility of caries diagnosis reported both in the present Chilean study as in the one carried out in Brazil and Switzerland by Diniz \textit{et al.}\(^2^2\). The lack of reproducibility of diagnosis negatively affects treatment costs\(^2^6\), so it would be advisable, on the one hand, to strengthen the undergraduate curriculum regarding the diagnosis of caries, and, on the other, to propose regular training on diagnostic agreement among dentists in order to detect, verify and correct this lack of reproducibility in the diagnosis.

**CONCLUSION.**

Agreement for the detection of approximal caries among dentists and among dental students is moderate and among radiologists is good. These results suggest that there are differences in caries detection between the groups studied, both interobserver and intraobserver.

**ACKNOWLEDGEMENTS.**

We would like to thank the Institute and School of Dentistry at Universidad Austral de Chile, students and academics for the support provided for the completion of this study, as well as the anonymous reviewers for their valuable comments. SEU proposed the idea of the study, wrote the protocol, performed part of the statistical analysis and wrote the manuscript. FLA wrote the protocol, performed the measurements and did part of the statistical analysis. NLR wrote the protocol, performed the measurements and edited the final draft. KBG and DH prepared the samples for the measurements and edited the final draft. All authors approved the final manuscript. This research is part of FLA and NRM’s requirements to earn a Bachelor Degree in Dentistry. School of Dentistry, Faculty of Medicine, Universidad Austral de Chile.

**REFERENCES.**

Diagnostic agreement between radiologists, dentists and dental students for radiographic detection of approximal caries.


