ORIGINAL ARTICLE

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

Sergio E Uribe.

Felipe Lara-Abarzúa.

Nicolás Ruiz-Miranda.

Katherine Bravo-Gajardo.

Daniela Huichaman.

School of Dentistry, Universidad Austral de Chile, Valdivia, Chile.

Corresponding author: Sergio E Uribe. Associate Professor, Radiology Service, School of Dentistry, Faculty of Medicine, Universidad Austral de Chile, Valdivia, Chile. Rudloff # 1640, Valdivia. Chile. Phone:(+56 63) 2293743. E-mail: sergiouribe@uach.cl

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. Iinterobserver 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.

INTRODUCTION.

Radiographic caries detection allows to evaluate areas that are difficult to assess clinically, particularly approximal surfaces, thus facilitating the decision-making process[1]. 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[2]. 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[3]. While the validity of bitewing radiography has been extensively studied, with more than 168 publications available in PubMed to date[4], 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[5]; on the use of high and low resolution radiographs[6], on direct or indirect digital radiographs[7], conventional or digital[8], and on the use of bitewing and oblique lateral radiographs[9].

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[10]. Although recent publications report on the diagnostic agreement for the clinical evaluation of caries[11], or between clinical and radiographic evaluation[12], there are no publications repor-

ting on the agreement with respect to radiographic evaluation 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 according to the recommendations given by the GRRAS[13]. 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 recommendations of Liao 2010[14], 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 sample size was then divided into three groups comprising: 4 dentists, 4 senior dental students and 4 radiologists. Bitewing radiographs standardized by a PBR-0546 positioner 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 presentations of radiographic caries were examined and to minimize recall bias of participants.

Enrollment

A convenient sample was used. 55 dentists, 5 radiologists and 22 students from School of Dentistry at Universidad Austral de Chile were invited to participate in the study. The final sample consisted of the first consecutive 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 performed in a dim-lighted room. Radiographs were visualized on a 4th generation iPad (Apple, Cupertino, CA, USA) in landscape mode (display resolution of 2048x1536 pixels); 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 verbally and in written. For the assessment of interobserver agreement, participants were asked to make a diagnosis for each radiograph for an approximal tooth surface previously 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 required 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 participants were asked to evaluate 10% of the previously examined radiographs. This assessment was made two weeks after the first measurement and under the same initial conditions. Each evaluation was conducted independently.

Statistical analysis

Evaluations made by participants were considered as a nominal variable (see Table 1). When grouping by category student-dentist-radiologist, scores for each diagnosis 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

Table 2. Demographic characteristics of participants.

Group	n	Average age	men : women	Range of experience in years
Students	4	25	3:1	5-6
Dentists	4	27	3:1	1-5
Radiologists	4	35	4:0	3-12

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

| | | First assesment | | | Intraobserver | p |
	Second assesment	Sound	Enamel	Dentine	weighted Kappa	
Students	Sound	8	2	2		
	Enamel	3	5	2	0.56	0.4624
	Dentine	1	3	14		
Dentists	Sound	9	4	3		
	Enamel	4	4	2	0.46	0.0005
	Dentine	1	2	11		
Radiologists	Sound	15	2	0		
	Enamel	1	4	1	0.68	<.0001
	Dentine	1	6	10		

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

	Dentists	Radiologists
Students	0.68*	0.51*
Dentists	--	0.62*

* p<.0001

good[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[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[17], from 0.63 to 0.81 using different types of radiographs[18], and from 0.32 to 0.87 using laser fluorescence in approximal lesions[19].

The explanation of these results requires an examina-

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[20]. 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[21].

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[22]. 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[23].

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.*[24]. In this regard, Héllen-Halme and Petersson[25] 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[26] 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 *et al.*[22]. The lack of reproducibility of diagnosis negatively affects treatment costs[26], 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.

Acuerdo diagnóstico entre radiólogos, odontólogos y estudiantes de odontología para la detección radiográfica de caries proximales.

Resumen: Objetivo: Determinar el acuerdo intra e interobservador para la evaluación radiográfica de caries proximales entre radiólogos, dentistas y estudiantes de odontología de último año. Método: 75 radiografías bitewing estandarizadas convencionales digitalizadas fueron utilizadas para determinar la presencia de caries proximales por 12 examinadores, (4 radiólogos maxilofaciales, 4 odontólogos y 4 estudiantes de odontología de último año). Con los diagnósticos obtenidos se calculó el acuerdo diagnóstico interobservador mediante un test de Kappa de Cohen ponderado. Los mismos evaluadores volvieron a diagnosticar un 10% de la muestra 2 semanas después bajo las mismas condiciones para obtener de esta segunda evaluación el acuerdo diagnóstico intraobservador mediante un test de Kappa de Cohen. Resultados: El Kappa interobservador fue 0.68 (bueno) para el binomio dentista-estudiante, 0.51 (moderado) para estudiante-radiólogo y 0.62 (bueno) para dentista-radiólogo. Todas estas diferencias fueron significativas. El acuerdo intraobservador mostró valores de Kappa de 0.56 (moderado) para estudiantes (p=0.46), de 0.46 (moderado) para dentistas (p<0.05) y 0.68 (bueno) para radiólogos. Todos los acuerdos se consideran buenos.Conclusión: Existe un acuerdo moderado a bueno entre estudiantes de odontología, dentistas y radiológos, tanto intra como interobservador para la detección radiográfica de caries proximales.

Palabras clave: *radiografía, caries dental, diagnóstico, variaciones dependientes del observador.*

REFERENCES.

1. Baelum V. What is an appropriate caries diagnosis? Acta Odontol Scand. 2010;68(2):65–79.

2. De Smet E, Jacobs R, Gijbels F, Naert I. The accuracy and reliability of radiographic methods for the assessment of marginal bone level around oral implants. Dentomaxillofacial Radiol. 2002;31(3):176–81.

3. Sim J, Wright CC. The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005;85(3):257–68.

4. National Center for Biotechnology Information. Resultado búsqueda [Internet]. 2014 [cited 2014 Oct 22]. Available from: http://j.mp/1s9aeZd

5. Kamburoglu K, Kolsuz E, Murat S, Yüksel S, Ozen T. Proximal caries detection accuracy using intraoral bitewing radiography, extraoral bitewing radiography and panoramic radiography. Dentomaxillofacial Radiol. 2012;41(6):450–9.

6. Berkhout WER, Verheij JGC, Syriopoulos K, Li G, Sanderink GCH, van der Stelt PF. Detection of proximal caries with high-resolution and standard resolution digital radiographic systems. Dento Maxillo Facial Radiol. 2007;36(4):204–10.

7. Naitoh M, Yuasa H, Toyama M, Shiojima M, Nakamura M, Ushida M, Iida H, Hayashi M, Ariji ENakamura M, Ushida M, et al. Observer agreement in the detection of proximal caries with direct digital intraoral radiography. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 1998;85(1):107–12.

8. Lawrence HP, Benn DK, Sheiham A. Digital radiographic measurement of approximal caries progression in fluoridated and non-fluoridated areas of Rio de Janeiro, Brasil. Community Dent Oral Epidemiol. 1997;25(6):412–8.

9. Townsend D. Detection of dentine caries using the oblique lateral radiograph. Int J Paediatr Dent Br Paedodontic Soc Int Assoc Dent Child. 2000 Jun;10(2):145–9.

10. Guay AH. The evolution of dental group practices. J Calif Dent Assoc. 2013;41(12):899–904.

11. Iranzo-Cortés JE, Montiel-Company JM, Almerich-Silla JM. Caries diagnosis: agreement between WHO and ICDAS II criteria in epidemiological surveys. Community Dent Health. 2013;30(2):108–11.

12. Ritter AV, Ramos MD, Astorga F, Shugars DA, Bader JD. Visual-tactile versus radiographic caries detection agreement in caries-active adults. J Public Health Dent. 2013;73(3):252–60.

13. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011 Jan;64(1):96–106.

14. Liao JJZ. Sample size calculation for an agreement study. Pharm Stat. 2010;9(2):125–32.

15. Fleiss JL. Statistical methods for rates and proportions. 3rd ed. Hoboken, N.J: J. Wiley; 2003. 760 p.

16. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ. 1992 Jun 6;304(6840):1491–4.

17. Tikhonova SM, Feine JS, Pustavoitava NN, Allison PJ. Reproducibility and diagnostic outcomes of two visual-tactile criteria used by dentists to assess caries lesion activity: a cross-over study. Caries Res. 2014;48(2):126–36.

18. Senel B, Kamburoglu K, Uçok O, Yüksel SP, Ozen T, Avsever H. Diagnostic accuracy of different imaging modalities in detection of proximal caries. Dento Maxillo Facial Radiol. 2010;39(8):501–11.

19. Neuhaus KW, Ciucchi P, Rodrigues JA, Hug I, Emerich M, Lussi A. Diagnostic performance of a new red light LED device for approximal caries detection. Lasers Med Sci. 2014;In press.

20. Uribe S, Gomez E. Handheld mobile devices for radiographic diagnosis of approximal caries. [abstract 50]. Caries Res. 2012;46:289.

21. Brocklehurst P, Ashley J, Walsh T, Tickle M. Relative performance of different dental professional groups in screening for occlusal caries. Community Dent Oral Epidemiol. 2012;40(3):239–46.

22. Diniz MB, Rodrigues JA, Neuhaus KW, Cordeiro RCL, Lussi A. Influence of examiner's clinical experience on the reproducibility and accuracy of radiographic examination in detecting occlusal caries. Clin Oral Investig. 2010;14(5):515–23.

23. Bader JD, Shugars DA. Understanding dentists' restorative treatment decisions. J Public Health Dent. 1992;52(2):102–10.

24. Lazarchik DA, Firestone AR, Heaven TJ, Filler SJ, Lussi A. Radiographic evaluation of occlusal caries: effect of training and experience. Caries Res. 1995;29(5):355–8.

25. Hellén-Halme K, Petersson GH. Influence of education level and experience on detection of approximal caries in digital dental radiographs. An in vitro study. Swed Dent J. 2010;34(2):63–9.

26. Schulte AG, Pitts NB, Huysmans MCDNJM, Splieth C, Buchalla W. European Core Curriculum in Cariology for undergraduate dental students. Eur J Dent Educ Off J Assoc Dent Educ Eur. 2011;15(Suppl 1):9–17.

27. Pitts N, Melo P, Martignon S, Ekstrand K, Ismail A. Caries risk assessment, diagnosis and synthesis in the context of a European Core Curriculum in Cariology. Eur J Dent Educ Off J Assoc Dent Educ Eur. 2011;15(Suppl 1):23–31.

28. Da Silva RP, Meneghim MC, Correr AB, Pereira AC, Ambrosano GMB, Mialhe EL. Variations in caries diagnoses and treatment recommendations and their impacts on the costs of oral health care. Community Dent Health. 2012;29(1):25–8.