RESEARCH ARTICLE

A novel classification system for assessment of approximal caries lesion progression in bitewing radiographs

¹Anna Senneby, ²Margareta Elfvin, ³Christina Stebring-Franzon and ¹Madeleine Rohlin

¹Faculty of Odontology, Malmö University, Malmö, Sweden; ²Färjestaden General Dental Clinic, Public Dental Health Service, Sweden; ³Alvesta General Dental Clinic, Public Dental Health Service, Sweden

Objectives: To design and pilot a novel classification system for the assessment of caries lesion progression in bitewing radiography and to report rater agreement of the system.

Methods: A classification system with drawings and text was designed to assess caries lesion progression. Guidelines for Reporting Reliability and Agreement Studies were used to study and report rater agreement. Pairs of posterior bitewing radiographs (baseline and 1-year follow-up) with different status concerning caries lesion progression were selected from files from public dental health clinics. 10 raters, 5 general dental practitioners and 5 specialists in oral and maxillofacial radiology were asked to assess the radiographs with the aid of the classification system. Seven raters repeated their assessments. Rater agreement was expressed as percentage of agreement and kappa.

Results: Kappa for the interrater agreement of 10 raters assessing progression was 0.61, indicating substantial agreement. Agreement was moderate for progression in the outer half of the dentine (kappa 0.55) and within enamel (kappa 0.44). Pairwise interrater agreement varied (range 69–92%; kappa 0.42–0.84). For about half of the pairs of raters, kappa was substantial (≥ 0.61). Intrarater agreement assessing progression was substantial (kappa 0.66–0.82).

Conclusions: We demonstrated the applicability of the proposed classification system on caries lesion progression with respect to rater agreement. This system can provide a common framework for clinical decision-making on caries interventional methods and patient visiting intervals. Scientifically, this system allows for a comparative analysis of different methods of prevention and treatment of caries as well as of different caries risk assessment methods.

Dentomaxillofacial Radiology (2016) **45**, 20160039. doi: [10.1259/dmfr.20160039](https://doi.org/10.1259/dmfr.20160039)

Cite this article as: Senneby A, Elfvin M, Stebring-Franzon C, Rohlin M. A novel classification system for assessment of approximal caries lesion progression in bitewing radiographs. *Dentomaxillofac Radiol* 2016; **45**: 20160039.

Keywords: dental caries; disease progression; visual perception; observer variation; radiography bitewing; classification

Introduction

Alongside visual/tactile examination, radiography is well established as a diagnostic tool to detect caries. According to a recent meta-analysis, the sensitivity for detecting caries

in clinical and *in vitro* studies is 0.24 and 0.42, respectively, for approximal lesions and corresponding specificity is, 0.97 and 0.89.¹ Based on the results of clinical studies, the mean sensitivity for detecting dentinal lesions in radiographs is 0.36 and based on *in vitro* studies, it is 0.45. The mean specificity is 0.94 and 0.95, respectively.¹ Reported rater agreement varies among radiographic studies of caries detection;^{2–5} this may be owing to differences in underlying sample characteristics, such as lesion depth, dentition, surface location and caries prevalence, and to

Correspondence to: Anna Senneby. E-mail: anna.senneby@mah.se

The study was funded by a grant received from Malmö University “Grant for research students” and from the Borrow Foundation. The Borrow Foundation had no role in the study design, data collection, analysis and interpretation of data, in the writing of the report or in the decision to submit the article for publication.

Received 24 January 2016; revised 23 March 2016; accepted 31 March 2016

methodological heterogeneity, such as the number of surfaces, number of raters and scoring categories.

In general dental healthcare, clinicians often have access to a series of radiographs of their patients from different years. The decision on intervention of approximal caries can then be based on not only the clinical examination and the detection of caries lesions and lesion depth in the radiographs but also the caries lesion progression rate. In addition, when evaluating different types of interventions, accurate assessment of caries lesion progression is important in everyday clinical practice and in the context of scientific studies. Assessment of caries lesion progression is also crucial in studies of caries risk assessment, as caries increment, *i.e.* the development of new or more severe disease, comprises the reference standard of choice. A recent systematic review of methods for caries risk assessment demonstrated a lack of clear definition of diagnostic criteria and data on rater variability for the assessment of caries lesion progression. This resulted in a methodological heterogeneity among the included studies.⁶ Although caries lesion progression is assessed radiographically daily in clinical practice and in several scientific studies, the level of agreement among raters is largely unknown. To our knowledge, there is only one previous publication on the reproducibility of radiographic assessment of caries lesion progression.⁷

In view of the importance of the radiographic assessment of caries lesion progression, an established classification system for this assessment could prove valuable. It would have the potential to facilitate clear communication among clinicians and researchers. Preferably, a novel classification system should be evaluated concerning its diagnostic accuracy and correspond to changes in disease status. However, accuracy of radiographic assessment of caries lesion progression in terms of sensitivity and specificity is difficult to achieve owing to difficulties in obtaining a reference standard. Wenzel *et al*⁸ stated, “when no valid expression of the true state of the disease can be obtained a study of precision ... is appropriate”. Swets and Pickett⁹ also highlighted that observer performance could be considered as a measure of the extent to which the inaccuracy of a system is the result of decision-making errors. Based on the reported variation among dentists in the detection and depth estimation of caries lesions, it can be expected that this variation may also exist for the assessment of caries lesion progression.

Hence, the objective of this study was to design a classification system for the assessment of approximal caries lesion progression in bitewing radiography and to pilot its applicability in terms of rater agreement among a group of general dental practitioners (GDPs) and oral and maxillofacial radiologists.

Material and methods

In order to achieve an approach that enables the readership to assess potential bias, this study implemented the guidelines proposed in the Guidelines for Reporting

Reliability and Agreement Studies (GRRAS).¹⁰ We used a GRRAS checklist with modifications (Supplementary Table A).¹¹

Sample of adolescents

Adolescents, aged 12 years, from two public dental health clinics (Kronoberg and Blekinge County, Sweden) were consecutively enrolled when visiting the clinic for regular check-up appointments. The enrolment took place during a recruitment period for a prospective clinical trial in 2012–2014. The adolescents were examined by visual and tactile examination using a mirror and probe. Then, one bitewing radiograph was exposed in each side of the mouth. The examinations were repeated yearly. Ethical approval was given by the Regional Ethical Review Board, Lund, Sweden (diary number: OD62-2012/122).

Radiographic equipment and quality control

Dental X-ray units (Planmeca Intra, Helsinki, Finland) operating at 8 mA and 63 kV with a rectangular positioning device and a focus-to-skin distance of 0.35 m were used. The X-ray units were equipped with electronic timers. Exposure time was set to 0.25 s, but in applicable cases, this was adjusted to 0.32 s. The two imaging plate systems used were Soredex Digora® (Soredex, Finland) with a 40 × 30 mm² effective area and Dürr VistaScan® (Dürr Dental AG, Germany) with a 28 × 36 mm² effective area. The imaging plates offered a pixel size of 39 × 39 and 45 × 45 µm², respectively. The imaging plates were enveloped in a plastic protective sheet and fixed to a Kwik-Bite® film holder (Kerr Corporation, USA).

Prior to the radiographical examination, the first author (AS) checked the radiological equipment for the following parameters: tube voltage, exposure time (reproducibility and linearity), low contrast and spatial resolution. The doses of radiation (microgray) for different exposure times were recorded. In addition, dose ratios (milligray per second) were recorded for applicable exposure times (0.2, 0.25 and 0.32 s). Regarding tube voltage, thresholds for kilovoltage were adjusted (60, 63, 66 and 70 kV) and the actual tube voltage was recorded with RaySafe ThinX Intra equipment (Unfors RaySafe, Fluke Biomedical, WA). When exposure times were registered, a scheme of five repetitions per time setting based on the clinically applicable exposure times was adopted for the reproducibility control. Further, a series ranging from 0.01 to 0.5 s of exposure time was used for the linearity control. Doses of radiation were measured throughout and dose ratios calculated on the obtained values. Spatial resolution was measured by exposing a spatial resolution phantom three times with different exposure times intended to produce optimal visualization of the object in the computer software used. Two plexiglass plates served as a phantom meant to represent the soft tissues of patients, thus being placed in between the spatial resolution phantom and the imaging plate. A value, visible

lines per millimetre, was obtained by reviewing the radiographic image (expressed as a number of 1–10). In order to obtain the best possible image of low contrast reproducibility, a plexiglass rectangle with depressions of different depths was used.

The images were saved in digital software programs: Arion (Pro Curis AB, Sweden) or Effica Viewer (Tieto Healthcare and Software AB, Finland). All images were exported as digital imaging and communication in medicine files¹² to portable storage media from the digital viewing software programs. The images were then imported to Romexis® software (Planmeca, Finland) and were anonymized.

Classification system and selection of bitewing radiographs and approximal surfaces

A classification system with drawings was designed to visualize disease status in terms of no caries lesion progression or caries lesion progression (Figure 1). Parallel to the drawings, text and indices were included to clarify different status of the approximal surfaces.

Two of the authors (AS, MR) selected the bitewing radiographs of approximal surfaces to be included in the study. During the selection process, image quality regarding contrast, sharpness, brightness and projection as well as lesion progression were considered. Images in each patient were paired: one image from baseline and

one image from the 1-year follow-up. Radiographs were selected to represent five surfaces for all categories presented in Figure 1, except for surfaces with caries lesion in the inner half of the dentine (Categories 4, 8, 11, 13 and 14), which were not represented in the sample and not assessed by the raters. 20 sound surfaces without progression were included. Altogether, 65 approximal surfaces with and without caries lesion progression were included, 39 surfaces in first molars and 26 surfaces in premolars from any quadrant with 1 surface reoccurring in the sample. All selected surfaces were non-restored, five surfaces were adjacent to a restored occlusal surface and one surface was adjacent to a restored approximal surface. Fillings in other surfaces were composite alike in radiolucency. Moreover, of the 65 surfaces with 2 radiographs of each (resulting in 130 images), 17 surfaces overlapped with the neighbouring surface, but the overlap did not extend beyond the outer half of the enamel.

Raters and ratings

10 raters, 5 GDPs and 5 specialists in oral and maxillofacial radiology were asked to assess the approximal surfaces. The GDPs worked at five different general dental clinics in three different public dental health counties. The specialists were currently employed at or had recently been employed at the Department of Oral

Figure 1 Classification system for the radiographic assessment of caries lesion progression in approximal surfaces. Caries lesions were categorized as follows: “Outer enamel” represents lesions in the outer half of enamel; “Inner enamel” represents lesions in the inner half of enamel; “Outer dentine” represents lesions in the outer half of dentine; “Inner dentine” represents lesions in the inner half of dentine.

and Maxillofacial Radiology, Malmö University. The selection of raters was made in such a way as to ensure a diverse level of professional experience. The professional experience of the GDPs ranged between 1 and 36 years (mean 13.8 years). The experience of the specialists in their field of speciality ranged between 1 and 33 years (mean 19.2 years).

All assessments took place in one viewing session at the Department of Oral and Maxillofacial Radiology, Malmö University, using Romexis software and a computer (HP Compaq Elite 8100 SFF; Hewlett-Packard, CA). Before the assessments, contrast and brightness of the monitor were visually adjusted with test images TG18-QC and TG18-CT from the American Association of Physicists in Medicine Task Group 18 Assessment of Display Performance for Medical Imaging Systems.¹³

Images were viewed on an 18.1-inch greyscale liquid crystal display monitor (Medical Flat Greyscale Display, 1318, Barco, Belgium) set at a screen resolution of 1280 × 1024 pixels, 8-bit colour depth, with an original luminance of 400 cd m⁻² and 0.2805 pixel pitch. The room had dimmed light (ambient light below 20 lux), and the viewing distance to the screen was 30–40 cm. No image enhancement or altering of image characteristics was allowed. The images were displayed side by side. The software, which was applied, used bilinear interpolation for scaling of images on screen. During the viewing sessions, which were performed independently by each observer, one of the authors (AS) was available to answer questions and to fill out the protocols. For each included approximal surface, the images from baseline were read simultaneously with those from the follow-up. Prior to the ratings, the raters were allowed to view pairs of bitewing images serving as examples of upcoming assessments. The raters were informed that the bitewing radiographs were from adolescents and that the pair of radiographs represented one from baseline and one from a follow-up examination. They did not receive any additional clinical information. The raters were aware that their ratings were to be compared with those of other raters.

The raters independently assessed the surfaces according to the classification system (Figure 1), which was available to the raters during their assessments. The raters were told which tooth and surface to assess. Seven

raters (three GDPs and four specialists) reassessed all surfaces in the same order and same setting after 2–4 weeks. During the second viewing, raters were blinded towards their results from their previous assessments.

Statistical analysis

Statistical analysis was performed with the aid of a statistician at Malmö University. The analysis was performed using STATA 10 (StataCorp, LP, TX). Kappa values with 95% confidence intervals (CI) were generated for intrarater and interrater agreement. For the pairwise interrater and intrarater agreement, the percentage of agreement was also calculated. Interrater agreement calculations were based on the first reading. Kappa values, varying from –1 to 1, were categorized according to the scale suggested by Landis and Koch.¹⁴ According to this scale, values <0 are considered as poor agreement, 0.00–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial and values >0.81 as almost perfect agreement.

Results

All selected 65 surfaces and all rater assessments were included for final analysis. Substantial interrater agreement was achieved for 10 raters for the assessment of caries lesion progression. Intrarater agreement was generally higher than interrater agreement.

Interrater agreement

Agreement in rating progression was substantial among 10 raters (kappa 0.61), being 0.57 among 5 GDPs and 0.65 among 5 specialists (Table 1). Agreement of 10 raters was moderate for the assessment of progression in the outer half of the dentine (kappa 0.55) and within the enamel (kappa 0.44) (Table 1). For surfaces assessed as sound without progression, agreement was substantial (kappa 0.78).

Pairwise interrater agreement regarding progression ranged between 69% and 92% with corresponding kappa values from 0.42 to 0.84 (Table 2). About half of the kappa values indicated at least substantial agreement (≥0.61). Two raters (GDP 2 and oral radiologist 4) presented moderate pairwise agreement with all other raters (Table 2).

Table 1 Interrater agreement expressed as kappa with confidence intervals (CI) for the assessment of caries lesion progression of approximal surfaces (n = 65) in bitewing radiography of adolescents

Raters	Agreement		
	All surfaces assessed to present progression (5, 6, 7, 9, 10, 12) ^a Kappa (CI)	Surfaces assessed to present progression within the enamel (5, 6, 9) ^a Kappa (CI)	Surfaces assessed to present progression in the outer half of the dentine (7, 10, 12) ^a Kappa (CI)
All raters (n = 10)	0.61 (0.54–0.70)	0.44 (0.34–0.54)	0.55 (0.37–0.73)
General dental practitioners (n = 5)	0.57 (0.45–0.69)	0.53 (0.38–0.67)	0.61 (0.36–0.80)
Oral radiologists (n = 5)	0.65 (0.52–0.77)	0.43 (0.30–0.57)	0.64 (0.49–0.79)

^aNumbers refer to categories presented in Figure 1.

Table 2 Pairwise interrater agreement expressed as percentage of agreement (%) and kappa with confidence interval (within brackets) for the assessment of caries lesion progression of approximal surfaces ($n = 65$) in bitewing radiography of adolescents

Raters	Agreement									
	General dental practitioners					Oral radiologists				
	1	2	3	4	5	1	2	3	4	5
General dental practitioners	1	0.63 (0.44–0.82)	0.63 (0.43–0.79)	0.65 (0.46–0.84)	0.69 (0.51–0.86)	0.63 (0.45–0.82)	0.61 (0.42–0.79)	0.62 (0.43–0.81)	0.69 (0.51–0.87)	0.63 (0.45–0.82)
	2	0.60 (0.41–0.80)	0.61 (0.43–0.79)	0.65 (0.46–0.84)	0.69 (0.51–0.86)	0.63 (0.45–0.82)	0.61 (0.42–0.79)	0.62 (0.43–0.81)	0.69 (0.51–0.87)	0.63 (0.45–0.82)
	3	0.42 (0.24–0.60)	0.61 (0.43–0.79)	0.65 (0.46–0.84)	0.69 (0.51–0.86)	0.63 (0.45–0.82)	0.61 (0.42–0.79)	0.62 (0.43–0.81)	0.69 (0.51–0.87)	0.63 (0.45–0.82)
	4	0.52 (0.31–0.74)	0.71 (0.53–0.89)	0.55 (0.35–0.76)	0.52 (0.33–0.71)	0.53 (0.34–0.74)	0.54 (0.35–0.75)	0.54 (0.35–0.75)	0.54 (0.35–0.75)	0.54 (0.35–0.75)
	5	0.46 (0.25–0.67)	0.69 (0.51–0.86)	0.53 (0.33–0.74)	0.60 (0.42–0.79)	0.60 (0.42–0.79)	0.60 (0.42–0.79)	0.60 (0.42–0.79)	0.60 (0.42–0.79)	0.60 (0.42–0.79)
Oral radiologists	1	0.49 (0.30–0.67)	0.76 (0.60–0.91)	0.54 (0.34–0.74)	0.53 (0.34–0.71)	0.54 (0.35–0.75)	0.54 (0.35–0.75)	0.54 (0.35–0.75)	0.54 (0.35–0.75)	0.54 (0.35–0.75)
	2	0.51 (0.29–0.73)	0.77 (0.61–0.93)	0.61 (0.42–0.81)	0.53 (0.34–0.71)	0.54 (0.35–0.75)	0.54 (0.35–0.75)	0.54 (0.35–0.75)	0.54 (0.35–0.75)	0.54 (0.35–0.75)
	3	0.61 (0.43–0.80)	0.84 (0.71–0.98)	0.62 (0.43–0.82)	0.70 (0.53–0.87)	0.68 (0.51–0.86)	0.66 (0.48–0.84)	0.72 (0.56–0.89)	0.62 (0.43–0.81)	0.62 (0.43–0.81)
	4	0.55 (0.35–0.74)	0.65 (0.47–0.84)	0.69 (0.51–0.87)	0.58 (0.38–0.77)	0.68 (0.51–0.86)	0.60 (0.40–0.79)	0.60 (0.41–0.79)	0.69 (0.51–0.87)	0.69 (0.51–0.87)
	5	0.55 (0.35–0.74)	0.65 (0.47–0.84)	0.69 (0.51–0.87)	0.58 (0.38–0.77)	0.68 (0.51–0.86)	0.60 (0.40–0.79)	0.60 (0.41–0.79)	0.69 (0.51–0.87)	0.69 (0.51–0.87)

Intrarater agreement

Intrarater agreement was above 80% for the assessment of caries progression (Table 3). Expressed as kappa, the agreement was substantial for all observers (mean 0.76; range 0.66–0.82). Agreement varied among raters for the assessment of progression in the outer half of the dentine (86–97%; kappa 0.52–0.90) (Table 3) and for progression within the enamel (82–91%; kappa 0.52–0.73) (Table 3). For surfaces assessed as sound without progression, agreement ranged between 94% and 100% (kappa 0.86–1.00). No rater consistently presented the highest or lowest intrarater agreement, but the CI were somewhat wider for some raters.

Discussion

The present study proposes a novel classification system for the assessment of caries lesion progression in bite-wing radiography. The results on rater agreement demonstrate that it is possible to achieve substantial agreement between several raters as well as within raters.

During the last decades, numerous systems have been proposed for the radiographic detection of approximal caries lesions. These systems are based on the interpretation of one radiograph at one point in time.^{15–20} Since a single radiograph with radiolucency should not be misinterpreted as indicating active disease,²¹ an approach with a series of images may be more beneficial to clinical decision-making. Even though it has long been called for in the scientific literature,^{16,22,23} to our knowledge, no classification system based on the radiographic assessment of change of approximal surfaces over time has ever been presented. The classification system presented in this study builds on existing systems of caries lesion detection such as the International Caries Classification and Management System (ICCMSTM),²⁴ which in turn is based on the widely used International Caries Detection and Assessment System (ICDASTM).²⁰ It might be beneficial to follow these criteria, as the staging of caries lesion is already known to clinicians and researchers. However, the fundamental difference is that the classification system presented in this study is designed for the comparison of two images performed with a time interval, in order to detect changes of the dental tissues. The drawings in the classification system are designed to show the caries lesion depths and lesion progression. To facilitate the interpretation of the drawings, the caries lesions within the enamel are presented as triangles. Other appearances of enamel caries lesions, such as a band or a dot, were also represented in the sample. Minor differences of contrast, *e.g.* in intraoral radiography of caries lesion progression, are more evident when comparing a series of images. This is particularly the case when the small changes in density within the same object are imaged against a homogeneous background.²⁵

Table 3 Intrarater agreement expressed as percentage of agreement (%) and kappa with confidence interval (CI) for the assessment of caries lesion progression of approximal surfaces ($n = 65$) in bitewing radiography of adolescents

Raters	Agreement					
	All surfaces assessed to present progression (5, 6, 7, 9, 10, 12) ^a			Surfaces assessed to present progression within enamel (5, 6, 9) ^a		Surfaces assessed to present progression in outer half of dentine (7, 10, 12) ^a
	%	Kappa (CI)		%	Kappa (CI)	% Kappa (CI)
General dental practitioners	1 89	0.75 (0.58–0.92)		89	0.67 (0.45–0.90)	91 0.52 (0.19–0.86)
	2 88	0.74 (0.57–0.91)		88	0.70 (0.52–0.89)	97 0.82 (0.57–1.00)
	3 89	0.78 (0.62–0.93)		83	0.55 (0.32–0.79)	94 0.76 (0.54–0.98)
Oral radiologists	1 83	0.66 (0.47–0.84)		82	0.52 (0.29–0.76)	92 0.74 (0.52–0.96)
	2 91	0.82 (0.68–0.96)		83	0.57 (0.34–0.80)	86 0.60 (0.37–0.83)
	3 89	0.78 (0.63–0.93)		89	0.71 (0.51–0.91)	88 0.56 (0.29–0.83)
	4 91	0.81 (0.67–0.95)		91	0.73 (0.53–0.93)	97 0.90 (0.76–1.00)

^aNumbers refer to categories presented in Figure 1.

Psychologically, the comparison of two images side by side in order to detect signs of tissue changes over time involves a different mindset than the study of a single image. In the diagnostic process, raters make a series of inferences derived from existing observations, one example being historical data.²⁶ In our context, historical data refer to the possibility of comparing two images exposed at different times. With access to serial radiographs over time, any small change of the hard tissues should be assessed. With respect to dental caries, not only lesion progression but also lesion regression, *i.e.* surface remineralization of enamel lesions, is important to assess, in particular when evaluating the efficacy of different non-invasive methods. Caries lesion regression was not included in the present classification system.

Research on observer performance views diagnosis as both a perceptual and a cognitive phenomenon that consists of learning the cues.^{27,28} Various aids have been applied to present cues characteristic of different categories, in order to decrease errors in perception and diagnostic reasoning. The present classification system designed and then piloted in our study has the advantage of presenting cues both as verbal descriptions and as visual patterns. There is ample confirmation from studies in visual diagnosis, primarily in radiology, that when data are elicited as discrete features, there is a large perceptual component that rapidly recognizes patterns rather than a cognitive component that seeks data for further discrimination.²⁹ In line with this concept of pattern recognition, the present classification system was designed to describe caries lesion progression with illustrations. The side-by-side illustrations are drawn to resemble radiographic images with features illustrating approximal surfaces with and without caries lesion progression. Seeing features within an appropriate context induces familiarity effects and will underpin categorization.²⁶

As different raters may have different visual concepts and prior experience, a study of rater agreement anticipates a design with several raters, in particular when the number of surfaces is modest.³⁰ In this study, low caries prevalence and few dentinal lesions limited the number of surfaces. Including more than six or seven

raters has little consequence on the results when a reasonably large sample is examined.⁹ Raters included in this study represented different professional experiences working in different institutions as well as different time lengths of expertise (long, intermediate and short time of experience). Yet, interrater agreement of 10 raters for caries lesion progression indicated substantial agreement (kappa 0.61), which is comparable with the results for 7 raters (kappa 0.70) in the study of caries lesion development by Wenzel *et al.*⁷ Pairwise interrater agreement varied, but substantial agreement was reached for many pairs of raters. The average intrarater agreement was 0.76, identical to that presented by Wenzel *et al.*⁷ for the assessment of caries lesion behaviour. Although difficult to compare, these results indicate higher interrater agreement than for caries detection by bitewing radiography reported in studies of high and medium quality.⁴

The magnitude of agreement is not solely a statistical decision, but also depends on the purpose and consequences of the test results. Even if the guidelines suggested by Landis and Koch¹⁴ are widely adopted when reporting rater agreement, there have been problems when interpreting the kappa coefficients. Interpretation against arbitrary standards does not distinguish between different applications of the guidelines; for example, the reliability may be acceptable for use in a research setting but inadequate for clinical decision-making on individual patients. Clinically, there are different courses of action available in caries lesion progression: (i) non-invasive (early) intervention, the aim of which is to prevent further progression, or (ii) operative intervention, *i.e.* to remove the carious tooth substance and restore it with a filling. Although it is recommended that operative intervention should not be performed unless there is cavitation^{31,32} or if radiographs show definite lesion progression over time,⁸ stated treatment thresholds to intervene surgically differ among clinicians.^{33,34} For example, some clinicians would restore lesions restricted within the enamel in high-risk individuals.³⁵ It may, however, be dubious to base a decision of operative intervention on caries risk assessment, as evidence concerning the validity of existing methods for identifying individuals at high risk is limited.^{36,6}

We applied GRRAS,¹⁰ which outlines important issues to be addressed when studies on rater agreement are designed and reported. To the best of our knowledge, these guidelines have not previously been applied in studies published in dental journals. There is some overlap between GRRAS and the Standards for Reporting of Diagnostic Accuracy,³⁷ which contains a checklist of essential elements to be included in diagnostic accuracy studies. In line with GRRAS, we reported agreement both as a percentage of agreement and as kappa statistics with CI to “allow the reader to get a detailed impression of the degree of the reliability and agreement”.¹⁰

The study has limitations. The study sample was selected by two of the authors in consensus. Although their assessment of caries lesion progression was not used as the reference standard, the selection of images affected the sample characteristics and thereby the results.³⁸ We did not find any report on the frequency of caries lesion progression as assessed in bitewing radiographs in any selected sample or population. Nevertheless, the frequency of surfaces with caries lesion progression may have been higher in the present study than can be expected in a given patient population. On the other hand, the percentage agreement could be higher, when many sound surfaces are included.³⁹ As people seem to be more influenced by representativeness

than probability when judging under uncertainty,⁴⁰ the frequency of caries lesion progression may have had a modest effect on the rater assessments and thereby on the results of their agreement.

The raters were prompted to look for caries lesion progression in specific approximal surfaces. Under such a directed search condition, the rater is explicitly asked to rate the likelihood of a specific category at one high-information region. This did not mimic the clinical situation where “free search” is the case. Furthermore, for the intrarater assessments, the raters were presented with the image pairs in the same order at both occasions. This might have led to them remembering parts of the sample.

Conclusions

We propose a novel classification system for the assessment of caries lesion progression in bitewing radiography. This system can be used to facilitate reporting and communication of changes of caries lesions and is applicable to clinical practice as well as to research of caries risk assessment and of outcomes of different caries interventions. It demonstrated substantial-to-moderate rater agreement, although further studies of different samples and under different conditions are warranted.

References

- Schwendicke F, Tzschoppe M, Paris S. Radiographic caries detection: a systematic review and meta-analysis. *J Dent* 2015; **43**: 924–33. doi: <http://dx.doi.org/10.1016/j.jdent.2015.02.009>
- Pliskin JS, Shwartz M, Gröndahl HG, Boffa J. Reliability of coding depth of approximal carious lesions from non-independent interpretation of serial bitewing radiographs. *Community Dent Oral Epidemiol* 1984; **12**: 366–70. doi: <http://dx.doi.org/10.1111/j.1600-0528.1984.tb01473.x>
- Wenzel A. Computer-automated caries detection in digital bitewings: consistency of a program and its influence on observer agreement. *Caries Res* 2001; **35**: 12–20. doi: <http://dx.doi.org/10.1159/000047425>
- Mejäre I, Axelsson S, Dahlén G, Espelid I, Norlund A, Svensson Å, et al. *Caries—diagnosis, risk assessment and non-invasive treatment. A systematic review. SBU-report no. 188. Summary and Conclusions in English*. Stockholm, Sweden: Swedish Agency for health Technology assessment (SBU); 2007. [Updated 2007; cited April 2016.] Available from: <http://www.sbu.se/en/publications/sbu-assesses/caries-diagnosis-risk-assessment-and-non-invasive-treatment/>
- Senel B, Kamburoglu K, Uçok O, Yüksel SP, Ozen T, Avsever H. Diagnostic accuracy of different imaging modalities in detection of proximal caries. *Dentomaxillofac Radiol* 2010; **39**: 501–11. doi: <http://dx.doi.org/10.1259/dmfr/28628723>
- Senneby A, Mejäre I, Sahlin N-E, Svensäter G, Rohlin M. Diagnostic accuracy of different caries risk assessment methods. A systematic review. *J Dent* 2015; **43**: 1385–93. doi: <http://dx.doi.org/10.1016/j.jdent.2015.10.011>
- Wenzel A, Anthonisen PN, Juul MB. Reproducibility in the assessment of caries lesion behaviour: a comparison between conventional film and subtraction radiography. *Caries Res* 2000; **34**: 214–8. doi: <http://dx.doi.org/10.1159/000016593>
- Wenzel A. Digital radiography and caries diagnosis. *Dentomaxillofac Radiol* 1998; **27**: 3–11. doi: <http://dx.doi.org/10.1038/sj.dmfr.4600321>
- Swets JA, Pickett RM. *Evaluation of diagnostic systems: methods from signal detection theory*. New York, USA: Academic Press; 1982.
- Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud* 2011; **48**: 661–71. doi: <http://dx.doi.org/10.1016/j.ijnurstu.2011.01.016>
- 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; **64**: 96–106. doi: <http://dx.doi.org/10.1016/j.jclinepi.2010.03.002>
- Dicom digital imaging and communication in Medicine*. Arlington, USA: National Electrical Manufacturers Association (NEMA). [Updated 15 October 2015; cited December 2015.] Available from: <http://dicom.nema.org/>
- Assessment of Display Performance for Medical Imaging Systems*. Durham, USA: American Association of Physics in Medicine AAPM Task group 18; [Updated 2006; Cited December 2015]. Available from: <http://deckard.mc.duke.edu/~samei/tg18> (test images: TG18-CT-1k-01 and TG18-QC-1k-01).
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74. doi: <http://dx.doi.org/10.2307/2529310>
- Koch G. Effect of sodium fluoride in dentifrice and mouth wash on incidence of dental caries in schoolchildren. *Odontol Revy* 1967; **18**(Suppl. 12): 1–125.
- Gröndahl HG, Hollender L, Malmcrona E, Sundquist B. Dental caries and restorations in teenagers. III. Social factors and attitudes in relation to dental status and caries increment among urban teenagers in Sweden. *Swed Dent J* 1977; **1**: 107–14.
- Espelid I, Tveit AB, Mjör IA, Eriksen HM, Fjellveit A, Oiestad V. Systems for grading occlusal and approximal carious lesions. [In Norwegian.] *Nor Tannlaegeforen Tid* 1990; **100**: 658–63.
- Rimmer PA, Pitts NB. Effects of diagnostic threshold and overlapped approximal surfaces on reported caries status. *Community*

- Dent Oral Epidemiol* 1991; **19**: 205–12. doi: <http://dx.doi.org/10.1111/j.1600-0528.1991.tb00147.x>
19. Amarante E, Raadal M, Espelid I. Impact of diagnostic criteria on the prevalence of dental caries in Norwegian children aged 5, 12 and 18 years. *Community Dent Oral Epidemiol* 1998; **26**: 87–94. doi: <http://dx.doi.org/10.1111/j.1600-0528.1998.tb01933.x>
 20. Ismail AI, Sohn W, Tellez M, Amaya A, Sen A, Hasson H, et al. The International Caries Detection and Assessment System (ICDAS): an integrated system for measuring dental caries. *Community Dent Oral Epidemiol* 2007; **35**: 170–8. doi: <http://dx.doi.org/10.1111/j.1600-0528.2007.00347.x>
 21. Benn DK. Diagnosis of caries progression from serial bitewings: a comparison between teachers and practitioners. *Br Dent J* 1993; **175**: 26–32. doi: <http://dx.doi.org/10.1038/sj.bdj.4808211>
 22. Pitts NB. Monitoring of caries progression in permanent and primary posterior approximal enamel by bitewing radiography. *Community Dent Oral Epidemiol* 1983; **11**: 228–35. doi: <http://dx.doi.org/10.1111/j.1600-0528.1983.tb01883.x>
 23. Benn DK. Radiographic caries diagnosis and monitoring. *Dentomaxillofac Radiol* 1994; **23**: 69–72. doi: <http://dx.doi.org/10.1259/dmfr.23.2.7835505>
 24. Ismail AI, Pitts NB, Tellez M; Authors of the International Caries Classification and Management System (ICCMS). The International Caries Classification and Management System (ICCMS™). An example of a caries management pathway. *BMC Oral Health* 2015; **15** (Suppl. 1): S9. doi: <http://dx.doi.org/10.1186/1472-6831-15-S1-S9>
 25. Omnell KA. Quantitative roentgenologic studies on changes in mineral content of bone in vivo. *Acta Radiol* 1957; **48**(Suppl. 148): 1–86.
 26. Dore KL, Brooks LR, Weaver B, Norman GR. Influence of familiar features on diagnosis: instantiated features in an applied setting. *J Exp Psychol Appl* 2012; **18**: 109–25. doi: <http://dx.doi.org/10.1037/a0026539>
 27. Blesser B, Ozonoff D. A model for the radiologic process. *Radiology* 1972; **103**: 515–21. doi: <http://dx.doi.org/10.1148/103.3.515>
 28. Brehmer A, Brehmer B. What have we learned about human judgment from thirty years of policy capturing. In: Brehmer B, Joyce CRB, eds. *Human judgment: the SJT view*. Amsterdam, the Netherlands: Elsevier Science Publishers B.V.; 1988. pp. 75–114.
 29. Norman GR, Coblenz CL, Brooks LR, Babcock CJ. Expertise in visual diagnosis: a review of the literature. *Acad Med* 1992; **67** (Suppl. 10): S78–83.
 30. Hintze H, Frydenberg M, Wenzel A. Influence of number of surfaces and observers on statistical power in a multiobserver ROC radiographic caries detection study. *Caries Res* 2003; **37**: 200–5. doi: <http://dx.doi.org/10.1159/000070445>
 31. Mejäre I, Stenlund H, Zelezny-Holmlund C. Caries incidence and lesion progression from adolescence to young adulthood: a prospective 15-year cohort study in Sweden. *Caries Res* 2004; **38**: 130–41.
 32. Baelum V, Hintze H, Wenzel A, Danielsen B, Nyvad B. Implications of caries diagnostic strategies for clinical management decisions. *Community Dent Oral Epidemiol* 2012; **40**: 257–66. doi: <http://dx.doi.org/10.1111/j.1600-0528.2011.00655.x>
 33. Baraba A, Domejean-Orliaguet S, Espelid I, Tveit AB, Miletic I. Survey of Croatian dentists' restorative treatment decisions on approximal caries lesions. *Croat Med J* 2010; **51**: 509–14. doi: <http://dx.doi.org/10.3325/cmj.2010.51.509>
 34. Vidnes-Kopperud S, Tveit AB, Espelid I. Changes in the treatment concept for approximal caries from 1983 to 2009 in Norway. *Caries Res* 2011; **45**: 113–20. doi: <http://dx.doi.org/10.1159/000324810>
 35. Kakudate N, Sumida F, Matsumoto Y, Manabe K, Yokoyama Y, Gilbert GH, et al. Restorative treatment thresholds for proximal caries in dental PBRN. *J Dent Res* 2012; **91**: 1202–8. doi: <http://dx.doi.org/10.1177/0022034512464778>
 36. Tellez M, Gomez J, Pretty I, Ellwood R, Ismail AI. Evidence on existing caries risk assessment systems: are they predictive of future caries? *Community Dent Oral Epidemiol* 2013; **41**: 67–78. doi: <http://dx.doi.org/10.1111/cdoe.12003>
 37. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin Radiol* 2003; **58**: 575–80. doi: [http://dx.doi.org/10.1016/S0009-9260\(03\)00258-7](http://dx.doi.org/10.1016/S0009-9260(03)00258-7)
 38. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005; **37**: 360–3.
 39. de Paola PF, Alman J. Assessment of the reliability of radiographic diagnosis in a clinical caries trial. *J Dent Res* 1972; **51**: 1431–7. doi: <http://dx.doi.org/10.1177/00220345720510053001>
 40. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* 1974; **185**: 1124–31. doi: <http://dx.doi.org/10.1126/science.185.4157.1124>