Impact of Scoring Single or Multiple Occlusal Lesions on Estimates of Diagnostic Accuracy of the Visual ICDAS-II System

Anahita Jablonski-Momeni,1 David N. J. Ricketts,2 Monika Heinzl-Gutenbrunner,1 Richard Stoll,3 Vitus Stachniss,3 and Klaus Pieper1

1 Department of Paediatric and Community Dentistry, Dental School, Philipps University of Marburg, Georg-Voigt-Straße 3, 35033 Marburg, Germany
2 Section of Operative Dentistry, Fixed Prosthodontics and Endodontology, Dundee Dental School, University of Dundee, Dundee DD1 4HN, UK
3 Department of Operative Dentistry, Dental School, Philipps-University of Marburg, Georg-Voigt-Straße 3, 35033 Marburg, Germany

Correspondence should be addressed to Anahita Jablonski-Momeni, momeni@staff.uni-marburg.de

Received 9 September 2009; Accepted 18 December 2009

Recommended by Roland Frankenberger

Caries lesions can occur at different sites on the occlusal surfaces of teeth and may differ in appearance and severity. This study aimed to evaluate how estimates of reproducibility and accuracy of ICDAS-II were affected when all lesions on occlusal surfaces, or only a representative lesion, were scored. 100 permanent teeth with 1–4 investigation sites on the occlusal surface were examined visually by four examiners. Serial sections of the teeth were assessed for lesion depth. Intra- and interexaminer reproducibility (weighted kappa values), sensitivity, and specificity were calculated for all investigation sites and for a randomly selected site per tooth. Comparing the kappa values for the whole sample and the independent sites, no effect only as a slight effect was found. Comparing the areas under the ROC-curves no effect could be shown. Examining multiple sites on teeth leads to results comparable to when a single independent site is chosen per tooth.

Copyright © 2009 Anahita Jablonski-Momeni et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The turn of the twentieth century has seen the need for improved visual caries detection. In 2001, a systematic review of the literature reported that the strength of the evidence for the diagnostic performance of visual/tactile method for coronal caries was poor due to the small numbers of studies available [1]. For those studies that were included, it was reported that whilst the specificity for the visual/tactile method was high, sensitivity was low.

In the year 2002, an International Consensus Workshop on Caries Clinical Trials (ICW_CCT) was held in Loch Lommond, Scotland and at this meeting a similar systematic review of visual/tactile caries diagnosis was presented [2]. Only twenty nine papers met the inclusion criteria, and the main conclusions were that there was a great variation in the disease process being measured and the examination conditions. In relation to the former conclusion, some studies only recorded cavitated lesions and some noncavitated lesions. Some made an attempt at determining lesion activity, whilst the majority did not. Few gave explicit information about the lesions being recorded and whether they were differentiated from noncarious changes. In relation to the second conclusion, there was no consistency on the use of blunt probes, whether the teeth were cleaned or whether they were dried.

Due to such variation it is clear that comparisons between studies are problematic and the criteria used are not related to the histopathology of the disease. It is these factors that led an International Group to develop the International Caries Detection and Assessment System (ICDAS-II), in the hope that a standardized approach to recording and characterising carious lesions that relate to the histopathology of the disease could be developed for use by researchers, epidemiologists, clinicians, and teachers [3].
Whilst the ICDAS-II can be used on all tooth surfaces, it is noted that carious lesions on the occlusal surface of posterior teeth can occur at different discrete sites and be of different appearance and severity. This is important for the clinician in treating caries, especially when operative intervention is concerned as not all sites may require operative intervention. Clinicians therefore have to make site specific judgements on a single tooth surface. As such, when diagnostic methods or tools are used in laboratory or clinical studies to assess their diagnostic accuracy, the presence of multiple discrete investigation sites on a single occlusal has often been used to increase the sample size. Sites have usually been chosen on the premise that they are discrete and easily relocated to allow assessment of intra- and interexaminer reproducibility [4–6].

However, such studies could and have frequently been criticised at peer review for not having statistically independent data. It can be argued that the position and appearance of a lesion in one part of the fissure system could bias the judgement of the examiner about the appearance of a separate lesion elsewhere in the fissure system and hence skew results.

Bader et al. [7] stated in their systematic review of the performance of methods for identifying carious lesions that in many studies on caries diagnosis the choice of sites rather than surfaces may pose a threat to external validity because most occlusal surfaces will present multiple sites for assessment. The results of site assessment do not summarize the status of the entire surface, which would be important for an epidemiological survey for example. This clearly demonstrates the problem that lesions can vary in severity across the occlusal surface and the dentist’s need to evaluate all sites to determine the worse status against a background of differing visual appearances. There are obvious differing needs between the epidemiologist who requires information on the entire surface and the clinician planning operative intervention; both need to be able to determine site specific information against a background of noise.

This study therefore aimed to evaluate how estimates of reproducibility, sensitivity, and specificity of ICDAS-II were affected by whether all lesions on occlusal surfaces, or only a representative lesion, were scored.

2. Materials and Methods

One hundred unrestored permanent molar \((n = 85)\) and premolar \((n = 15)\) teeth were selected from a group of extracted teeth collected at Dundee and Marburg Dental Schools. These were cleaned thoroughly and stored in water. The teeth were collected and used in accordance with the appropriate legislation and regulations in place in the UK and Germany predating September 2006.

The teeth used in this study were those used in a previous study [8]. Initially one hundred teeth were selected which had between one to four discrete sites within the pit and fissure system suitable for investigation. In total there were 181 discrete investigation sites which were marked on digital photographs of the occlusal surfaces. Black and white copies, printed in draft quality on plain paper, were used by the examiners during this study and were only suitable for lesion location. The distribution of the investigation sites was: 33 teeth had 1 investigation site, 54 teeth had 2, 12 teeth had 3, and 1 tooth had 4 investigation sites. Only those teeth with two or more discrete investigation sites were used in this study \((n = 67)\) giving a total of 148 investigation sites.

Prior to the visual examination, the reference examiner (D. N. J. R.) trained 3 other examiners (K. P., V. S., and A. I.-M.) in the ICDAS-II classification system in a 2-hour session. For details see Jablonski-Momeni et al. [8]. Each investigation site was visually examined by the four investigators blind to each other using the International Caries Detection and Assessment System (ICDAS-II) (Table 1, http://www.icdas.org/) [9].

After three weeks, 3 of the trained examiners re-examined all of the teeth and investigation sites in order to determine intraexaminer reproducibility. The reference examiner was not available to repeat the examinations, being abroad.

2.1. Histological Preparation. Following the visual examination the roots of the teeth were resected using a horizontal cut just apical to the cement-enamel junction as described previously [8, 10]. In brief, a photograph of the occlusal surface of the tooth was taken at the same magnification as a right-angled coordinate system which surrounded the tooth in order that the exact investigation site (IS) could be determined by its \(x\)- and \(y\)-coordinates (Figure 1). The mesial surface of each tooth was placed face down in contact with a mounting plate at the base of an embedding cylinder, such that the cut root face was vertical and the distal surface of the tooth was upper most. A right-angled triangle of blue foil with a base to height length of 1:2 (\(\beta = 63.5^\circ\)) was mounted in relation to the cut root surface, parallel with the occlusal surface of the tooth in a vertical direction and with its base aligned horizontally with the mounting plate.

The teeth and foil were embedded in acrylic and sectioned in a bucco-lingual direction, starting from the upper most distal surface. Each cut therefore contained a section of the tooth and the blue foil in the form of a distinct line in relation to the cut root face. If the length of the blue line at the base of the section is ML the formulae determined in Figure 1 \((H = 2x (a - ML))\) will give the height \((H)\), or \(y\)-coordinate of the section. As such the correct section can be determined for each investigation site. As to where this is along each section, the \(x\)-axis coordinate is measured and marked on an image of the section.

Eleven to fifteen sections were produced per crown \((\text{width} = 200 \mu m \pm 30 \mu m)\) and 1–4 sections per investigation site were available.

2.2. Histological Examination. For each investigation site the selected sections were examined by all 4 examiners using a binocular microscope (Wild Heerbrugg AG, Gais, Switzerland) using 16x magnification and reflected light. The Downer histological classification system [11] was used to record caries severity at each investigation site and this was carried out blindly to the other examiners (Table 2).
Table 1: The ICDAS-II criteria.

<table>
<thead>
<tr>
<th>ICDAS-II code</th>
<th>Criteria [9]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Sound tooth surface: no evidence of caries after prolonged air drying (5 seconds)</td>
</tr>
<tr>
<td>1</td>
<td>First visual change in enamel: opacity or discoloration (white or brown) is visible at the entrance to the pit or fissure after prolonged air drying, which is not or hardly seen on a wet surface</td>
</tr>
<tr>
<td>2</td>
<td>Distinct visual change in enamel: opacity or discoloration distinctly visible at the entrance to the pit and fissure when wet, lesion must still be visible when dry</td>
</tr>
<tr>
<td>3</td>
<td>Localized enamel breakdown due to caries with no visible dentine or underlying shadow: opacity or discoloration wider than the natural fissure/fossa when wet and after prolonged air drying</td>
</tr>
<tr>
<td>4</td>
<td>Underlying dark shadow from dentine +/- localised enamel breakdown</td>
</tr>
<tr>
<td>5</td>
<td>Distinct cavity with visible dentine: visual evidence of demineralisation and dentine exposed</td>
</tr>
<tr>
<td>6</td>
<td>Extensive distinct cavity with visible dentine and more than half of the surface involved</td>
</tr>
</tbody>
</table>

Figure 1: Occlusal view of a molar tooth with three investigation sites and the corresponding histological sections. The length of the embedded coloured foil allows accurate location of the section in the y-axis, using the formula $H = 2 (a - ML)$. The position of the lesion along each section can be then determined by the x-axis coordinate.

Up to 4 sections were available which could be assigned to each investigation site. A histological score was given to each section and the worst/deepest score was taken as the definitive score for further analysis. Caries extent was based upon colour and structural changes in enamel and dentine, with emphasis being placed on differentiating carious changes from protective changes of the pulp-dentine complex, such as tubular sclerosis and reactionary dentine formation.

For each investigation site the results from all 4 examiners were then compared to achieve a consensus histological score to be used in the subsequent analysis. Where 3 or more examiners agreed on the histological score, this was taken as consensus. Where there was greater disagreement,


2.3. Data Management and Statistical Evaluation. Both the ICDAS-II and histology scores were recorded on data collection forms and later transferred to an Excel table. For each tooth one investigation site was randomly chosen (SPSS 15.0) in order to avoid dependencies resulting from several measurements on the same tooth. Only those with more than one investigation site were included. Thus the subsequent analyses were carried out for the whole sample (n = 148 investigation sites) and the independent sites (independent data n = 67).

For the ICDAS-II scores, interexaminer reproducibility was calculated for all pairs of examiners and intraexaminer reproducibility was calculated for the trained examiners using weighted Cohen’s kappa (linear weights, ComKappa version 1.0).

The consensus Downer histology was used to calculate sensitivity and specificity at the $D_1$ and $D_3$ diagnostic threshold. At the $D_1$ diagnostic threshold all histological scores 1–4 were classed as caries and each ICDAS-II cut-off was used to calculate sensitivity and specificity for each examiner. Similarly for the $D_3$ diagnostic threshold histological scores 3 and 4 were classed as caries only and sensitivity and specificity calculated at each ICDAS-II cut-off. Using these sensitivity and specificity values Receiver Operating Characteristic (ROC) analyses were carried out at the $D_1$ and $D_3$ thresholds for each examiner.

To test whether kappa values, and the areas under the ROC-curve differed systematically between the two sets of data (whole sample and independent sites), the effect size (Cohen’s d) [12] was calculated between the results for all the data and the results for the randomized independent group (SPSS 15.0). According to Cohen [12] when the effect size is between 0.0 to 0.2 it indicates no effect, 0.2-0.5 indicates a small effect, 0.5–0.8 a medium effect, and >0.8 a large effect.

### Table 2: Criteria used in the histological examinations.

<table>
<thead>
<tr>
<th>Score</th>
<th>Criteria used in the Downer histological examination [11]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No enamel demineralisation or a narrow surface zone of opacity (edge phenomenon)</td>
</tr>
<tr>
<td>1</td>
<td>Enamel demineralisation limited to the outer 50% of the enamel layer</td>
</tr>
<tr>
<td>2</td>
<td>Demineralisation involving the inner 50% of the enamel, up to the enamel-dentine junction</td>
</tr>
<tr>
<td>3</td>
<td>Demineralisation involving the outer 50% of the dentine</td>
</tr>
<tr>
<td>4</td>
<td>Demineralisation involving the inner 50% of the dentine</td>
</tr>
</tbody>
</table>

3. Results

Initially, 148 investigation sites were planned for statistical evaluation, but, owing to section damage on some teeth only 146 investigation sites and 291 corresponding sections were available for analysis (34 investigations sites with 1 section, 84 with 2 sections, 23 with 3 sections, and 5 with 4 sections).

Three or more examiners agreed on the histological assessment in 82% of investigation sites when using the Downer classification. The interexaminer weighted kappa values were 0.69–0.78 for the histology. Thus the agreement between examiners in the histological assessment of the sections was good to substantial, but where disagreement occurred, a consensus decision was made following discussion and this was used in subsequent analyses.

For examiners 1, 2, 3, and 4 the ICDAS-II scores differed from one site to another in 73.1%, 70.1%, 56.7%, and 77.6% (mean = 69.4%) of teeth, respectively. For the consensus Downer score, in 71.6% of teeth the histological scores also differed from one site to another.

Figure 2 shows the distribution in the differences between the scores at each investigation site in each surface for each examiner when using ICDAS-II and for the consensus histological scores.

The weighted kappa values for inter- and intraexaminer reproducibility are shown in Table 3. These demonstrate good to substantial agreement between examiners and for each examiner when examinations are repeated. When comparing the kappa values for the whole sample and the independent sites, the effect size was mainly between 0.0 and 0.2 indicating no effect or difference between the two samples.

Table 4 shows for the $D_1$ and $D_3$ diagnostic thresholds optimum sensitivity, specificity, the corresponding ICDAS-II cut-off used to achieve this and the area under the curves (AUC) for each examiner when the whole sample and independent sites were analysed. The effect size between...
the AUCs for the whole data and the randomized independent data shows no effect at the \( D_1 \) threshold (effect size between 0.01 and 0.08) or \( D_3 \) threshold (effect size between 0.05 and 0.15).

### 4. Discussion

It has been recognised that detection of caries from a visual examination alone is problematical [1, 2]. The occlusal surface poses a particular problem as this surface is the most commonly affected by caries in children, adolescents, and young adults [13, 14]. The invaginated anatomy and histopathology of the disease process in this surface also complicates early caries detection; initial lesions occur on the walls or at the entrance to the fissure, spread laterally through the enamel and widely and often deeply into dentine before frank cavitation occurs. Subtle changes at the surface of the tooth in previous visual classification systems have not been related to the histopathology of the disease and as such many lesions have been missed or simply not included in the criteria [2]. One of the purposes of the ICDAS-II system is to overcome this short fall, to characterise and describe the earliest visible changes due to caries on all coronal surfaces (and specifically in this study in occlusal surfaces) through to frank cavitation and how these stages relate to the histopathology of the disease.

Even when using a detailed system such as the ICDAS-II there is a degree of subjective interpretation due to perhaps visual perception, lighting, and potential bias. Such bias may arise from other surfaces on the same teeth and in clinical studies of other teeth, or within other areas of the same surface. It might be argued that, for example, if a smooth buccal surface is carious, the likelihood of an approximal surface on the same tooth being carious is much higher. Similarly if the distal fossa of an upper molar tooth is carious it could be argued that the mesial fossa also has a higher risk of being carious and the operator might be inclined to change the visual score to reflect this. However, on the occlusal surface the anatomical sites are often discrete, have individual susceptibilities and risk factors, and it can also be argued that if any technique or detection tool has a place in the dental market they should be able to overcome such biases. One might argue that a blinded design for a study is possible by masking the other sites in the occlusal surfaces with multiple sites so eliminating bias. This could be done on photos (visual diagnosis) or direct on the surfaces. But this would not reflect the in vivo clinical situation when the all teeth are subject to visual examination and all sites within one tooth and other teeth would be visible anyway.

Multiple examination sites per occlusal surface have been used in a number of studies. For example, Ferreira Zandonà et al. [15] used premolars with three investigation sites on each occlusal surface to compare the performance of visual examination and laser fluorescence for detection of demineralization in occlusal pits and fissures. They stated that the use of multiple sites on a single tooth provided an opportunity to correlate the methods precisely, and was justified since the sites were well separated and contained distinct demineralization as judged either by histology or by confocal laser microscopy. Lussi et al. [16] used 1–3 sites of 26 occlusal surfaces (total 41 sites) to study the accuracy of an electrical resistance monitor in diagnosing occlusal caries. Subsequent histological examinations revealed no carious connections between multiple test sites on a tooth. The sites were selected according to the following criteria: (1) easily located topographic position, and where there was more than one site on a tooth, the sites were clearly separated. The authors stated that the use of multiple sites in a single tooth was fully justified. It reflects the clinical situation where several occlusal sites on a single tooth require evaluation. In this study no distinction was made between the results whether one site or all of the investigation sites were analysed.

This study has clearly shown that the histological severity of caries between investigation sites on the same occlusal surface can vary considerably (Figure 2). However, dentists can also be site specific in making ICDAS-II decisions as the codes given to individual discrete investigation sites on the same tooth surface also differ (Figure 2). The results also
Table 4: The area under the ROC curve, optimum sensitivity and specificity and corresponding ICDAS-II threshold used, for each examiner at $D_1$ and $D_3$ diagnostic threshold.

<table>
<thead>
<tr>
<th>Examiner 1</th>
<th>Examiner 2</th>
<th>Examiner 3</th>
<th>Examiner 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>All data</td>
<td>Randomised data</td>
<td>All data</td>
<td>Randomised data</td>
</tr>
<tr>
<td>$D_1$ diagnostic threshold</td>
<td>$D_1$ diagnostic threshold</td>
<td>$D_1$ diagnostic threshold</td>
<td>$D_1$ diagnostic threshold</td>
</tr>
<tr>
<td>Opt Sens</td>
<td>0.70</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>Opt Spec</td>
<td>0.83</td>
<td>0.90</td>
<td>0.81</td>
</tr>
<tr>
<td>ICDAS cut-off</td>
<td>1-2</td>
<td>1-2</td>
<td>1-2</td>
</tr>
<tr>
<td>AUC (SE)</td>
<td>0.81 (0.04)</td>
<td>0.82 (0.06)</td>
<td>0.80 (0.04)</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.72-0.89</td>
<td>0.69-0.94</td>
<td>0.72-0.88</td>
</tr>
<tr>
<td>ES for AUC</td>
<td>0.03</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>$D_3$ diagnostic threshold</td>
<td>$D_3$ diagnostic threshold</td>
<td>$D_3$ diagnostic threshold</td>
<td>$D_3$ diagnostic threshold</td>
</tr>
<tr>
<td>Opt Sens</td>
<td>0.71</td>
<td>0.63</td>
<td>0.60</td>
</tr>
<tr>
<td>Opt Spec</td>
<td>0.88</td>
<td>0.85</td>
<td>0.87</td>
</tr>
<tr>
<td>ICDAS cut-off</td>
<td>2-3</td>
<td>2-3</td>
<td>2-3</td>
</tr>
<tr>
<td>AUC (SE)</td>
<td>0.85 (0.03)</td>
<td>0.81 (0.06)</td>
<td>0.84 (0.03)</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.79-0.92</td>
<td>0.69-0.92</td>
<td>0.78-0.91</td>
</tr>
<tr>
<td>ES for AUC</td>
<td>0.11</td>
<td>0.08</td>
<td>0.15</td>
</tr>
</tbody>
</table>

AUC: Area under the curve; SE: Standard error; CI: Confidence interval, ES: Effect size.

show good inter- and intraexaminer reproducibility for the ICDAS-II system and acceptable diagnostic accuracy at both the $D_1$ and $D_3$ diagnostic threshold, irrespective of whether the whole sample is analysed or whether one investigation site is randomly chosen to represent independent data for each tooth (effect sizes from 0.00 to 0.25, Tables 3 and 4).

The kappa values for ICDAS-II found in this study are similar to those kappa values published by Ismail et al. [17] who reported that the reliability of six examiners to classify tooth surfaces by their ICDAS carious status ranged between good to excellent (kappa coefficients ranged between 0.59 and 0.91, using two different weighting schemes). In a study where ICDAS-II codes were used in both primary and permanent teeth [18] intra- and interexaminer reproducibility were found to be excellent (kappa values > 0.82). Rodrigues et al. [19] obtained unweighted kappa values of 0.61 for intraexaminer reproducibility (permanent teeth) and 0.51 for interexaminer reproducibility when ICDAS-II was used in permanent teeth.

In many publications it is not always clear how investigation sites were selected. Some authors selected sites which have an easily located topographical position [4, 6]. Or an easily identifiable interlobal groove on the occlusal surface where caries would be likely to occur would be chosen for investigation [5]. In our study we used teeth which had between one to four discrete sites within the pit and fissure system suitable for investigation.

Regarding the Downer histological classification [11], the area under the ROC curve showed good performance of the ICDAS-II in detecting occlusal caries lesions at both $D_1$ and $D_3$ diagnostic thresholds (AUC from 0.67 to 0.86, Table 4). Other studies had shown that the ICDAS-II produced areas under the ROC curves of 0.75 [19] and 0.73 [20] which are within the range of our results.

This study supports the view that dentists can be site specific in applying ICDAS-II visual criteria to multiple discrete sites within the same occlusal surface and not necessarily be biased by the appearance at an adjacent site. Dentists' education and clinical experience tell them that lesions can be variable across one tooth surface and the appearance at one site should not dictate the treatment for all sites, each should be taken on their own merit. These results have important implications in relation to the use of human tissues for research, regulations over which, now severely limit the availability and use of extracted teeth for research in many countries.

### Conflict of Interest

The authors declare that they have no conflict of interest.

### References


