RADIOGRAPHIC DIAGNOSES AND TREATMENT DECISIONS ON APPROXIMAL CARIES

IVAR ESPELID
RADIOGRAPHIC DIAGNOSES AND TREATMENT DECISIONS
ON APPROXIMAL CARIES

by

Ivar Espeland

BERGEN
1987
CONTENTS

ACKNOWLEDGEMENTS ........................................... 4
PREFACE ..................................................... 5

SUMMARY

INTRODUCTION ..................................................... 7
THE AIMS ....................................................... 8

MATERIALS AND METHODS WITH DISCUSSION OF METHODS .............. 9
TEETH .......................................................... 9
RADIOGRAPHIC PROCEDURE ...................................... 9
VALIDATION OF RADIOGRAPHIC DIAGNOSIS ......................... 10
DIAGNOSTIC SESSIONS ........................................ 11
MEASUREMENT OF OBSERVER PERFORMANCE AND DIAGNOSTIC
QUALITY ........................................................ 12

OBSERVER PERFORMANCE INDICES .................................. 17
QUESTIONNAIRE STUDY ........................................... 19

RESULTS AND GENERAL DISCUSSION ................................ 19
HOW EFFECTIVE IS THE RADIOGRAPHIC DIAGNOSIS OF MINERAL
LOSS IN APPROXIMAL ENAMEL? .................................. 19
RADIOGRAPHIC DIAGNOSIS OF APPROXIMAL CARIES ............. 21
RADIOGRAPHIC DIAGNOSIS COMPARED TO CLINICAL
CHARACTERISTICS OF THE LESIONS ............................... 21
ARE RADIOGRAPHIC DIAGNOSIS MADE IN ENAMEL AND DENTIN EQUALLY EFFECTIVE?  

IS QUALITY OF RADIOGRAPHIC CARIES DIAGNOSIS AFFECTED BY FACTORS SUCH AS VIEWING CONDITIONS AND RADIOPACITY OF FILLINGS?  

TREATMENT DECISIONS ON APPROXIMAL CARIES  

EXAMPLE  

OBSERVER CALIBRATION?  

THE CLINICAL RELEVANCE OF THE FINDINGS  

CONCLUSIONS  

REFERENCES
ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to:

My co-authors who in addition to their authorship, have contributed to this work in different ways.

Dr. odont Anne Bjørg Tveit for never failing enthusiasm and inspiring and valuable discussions.

Professor Ola Haugejorden for valuable advices.

Paul Riordan B.D.S. M.D. for linguistic advices.

Cand. odont. Nils Roar Gjerdet for invaluable help during the computer analyses.

The observers and their efforts are greatly appreciated.

Professor Agnar Halse who initiated this work.

Professor Knut A. Selvig, professor Per Rasmussen and professor emeritus Einar Hals for valuable advices.

Cand. odont. & cand. real. Olav Bøe for statistical advice.

Mrs. Randi Hansen for skilful computer work and for typing most of the text.

Mr. Cato Olsen, Mrs. Rita Eriksen, Mrs. Sidsel Aasen, Ms. Borgny Vik and Ms. Siri Fredriksen for skilful assistance.

I want to thank Mr. Kjell Steffner, Mr. Rune Haakonsen and Mr. Torolf Engen for the photographic work.

In part the work was carried out using laboratory facilities in the Departments of Dental Research and of Dental Materials. My gratitudes also include these departments.
PREFACE

This thesis consists of a summary and the papers listed below. The papers are referred to in the summary by their roman numbers.


Paper VII  Espelid I, Tveit AR, Haugejorden O, Riordan PJ.

SUMMARY

INTRODUCTION
Radiography plays an important role in the diagnosis and treatment planning of dental caries. According to the model described by Blesser and Ozonoff (1) the radiologic process is divided into 3 phases: "psychophysical, psychological, and nosological. Psychophysical processing refers to x-ray imaging and reception by the peripheral nervous system; physiological functioning concerns the basis for transforming the image into a meaningful pattern; nosological considerations pertain to the clinical evaluation of the patterns." All three sections contribute to the final decision about presence or absence of disease.

Dentists have traditionally used radiographs for diagnostic purposes to identify caries at an early stage. Approximal dental surfaces are usually not directly accessible to inspection or probing, so that radiography represents a useful, non-destructive method to provide information on these "hidden" surfaces. It is reasonable to believe that a high proportion of the intra-oral radiographs are taken for detection of caries even without any indication that caries is present (2). The number of annual intra-oral radiographs increases (3) despite the reported decrease in caries incidence (4).

Radiographic tests have costs in addition to time and money, i.e. the exposure of the patient to radiation and consequences of the built in errors of the test. Therefore, knowledge is needed about
the validity and the reliability of the information provided during the radiographic examination for caries. Uncertainty will always be associated with diagnostic tests, whether they are based on laboratory data or on radiographic interpretations.

Clinical strategies for treatment of approximal caries should in general be based on, among other factors, the available information about incidence and progression rate of disease. The radiographic examination is the most important source for the collection of such information for groups and for individuals. The clinician has thus a pertinent need for valid information about the quality of radiographic caries diagnosis when he has to interpret the epidemiological data or data on progression rate and decide whether caries is present or not on a specific tooth surface. He needs knowledge about the possibilities of making erroneous decisions when action is taken on operative therapy and when intervals between radiographic examinations are decided. Knowledge about the validity of diagnosis and treatment decisions is important for both clinicians and epidemiologists. High quality diagnosis and treatment decisions contribute to professional quality assurance and are in the interest of patients, society and the dental profession.

THE AIMS

1. To investigate the relationship between approximal mineral loss and the quality of radiographic diagnosis;
2. To assess the validity of radiographic diagnosis of approximal caries;

3. To compare the quality of radiographic caries diagnosis in enamel and in dentin;

4. To assess the influence of viewing conditions and radiopacity of fillings on the observer performance during radiographic caries diagnosis.

5. To investigate treatment decisions on approximal caries among dentists in Norway.

MATERIALS AND METHODS WITH DISCUSSION OF METHODS

TEETH

In Papers I-VI extracted premolars and molars with unknown history were used. The teeth were stored in 2% benzalconium chloride since extraction. In general, teeth with any defect due to other causes than caries, were excluded.

RADIOPHGRAPIC PROCEDURE (PAPERS I-VI)

All radiographs were exposed by a Ritter Explorer II dental X-ray machine operated at 65 kV, 15 mA and 1 mm Al of added filtration. The half-value layer was 2.7 mm Al. A 10 mm wide Plexiglas container filled with water was placed between the teeth and focus to simulate soft tissues. The exposure times and focus-object-film distances varied to some extent in the different studies. In general original radiographs were viewed except for Paper VI in
which copies (Kodak X-Omat duplicating film) were examined by the
dentists. In all studies the radiographs of the same series were
processed in one batch by a developing machine (Refrema XR-Minor T,
processed in Agfa-Gaevert's G 150 and G 334 chemicals).

VALIDATION OF RADIOGRAPHIC DIAGNOSIS

The measurements of diagnostic quality require valid information
about the "true status". This means that information about the
presence of caries should be obtained using a presumptively more
valid method than the one under investigation (5). The confirmed
diagnosis was based on direct inspection and probing of the teeth
under optimal conditions. Neither microscope nor lens was used for
magnification, since it was stressed that the truth should have
"clinical relevance" with respect to the detection of caries. The
relationship between mineral loss and radiographic detection (Paper
I) was studied using sectioned premolars and molars with artifi-
cially created, standardized defects of known size.

To study the quality of radiographic caries diagnosis (Papers II-
VI), extracted premolars and molars with or without approximal
caries were radiographed. The sound teeth served as negative
controls. The validation of the teeth was based on direct
observations of approximal surfaces (Papers II-IV). In Papers II
and III the extension of caries towards the pulp was monitored in a
"clinical" manner by direct inspection and probing in prepared
Class II cavities.
In the study of the relationship between different viewing conditions and radiographic diagnosis (Paper IV) 3 different "signals" were used: approximal caries on extracted teeth, artificially created approximal defects (same material as in Paper I) and Plexiglas phantoms i.e. Plexiglas boxes with holes or cylinders which were randomly placed as signals (phantom lesions).

In the study of secondary caries diagnosis and marginal defects in connection with Class II restorations (Paper V) 2 materials with different radiopacity were inserted in the same cavities, one at a time. The materials under study were amalgam and a zinc-glass containing composite (P30). About 1/3 of the restorations were without failure, 1/3 had an artificial marginal defect created and 1/3 of the teeth had primary caries left in the gingival part of the Class II preparation, to illustrate secondary caries.

**DIAGNOSTIC SESSIONS (PAPERS I-V)**

Due to the observer errors and the fact that one "extreme" observer could affect the results significantly, it is important to use several observers for validating the radiographic method (6,7,8,9,10). The number of observers who examined the different series ranged from 2 to 11. During the examination the observers had no time restrictions and they used a viewing box with a 2x magnifying lens. The light sources and viewing conditions were standardized within each trial.

All diagnoses were made according to a 5-category rating of confidence of caries (Papers II-VI) or caries-like lesions.
(Paper I), marginal defects along Class II restorations (Paper V) or phantom lesions (Paper IV) being present. In principle the rating scale expressed the following confidence levels concerning the presence or absence of a "signal":

1. Almost definitely not present
2. Probably not present
3. Unsure
4. Probably present
5. Almost definitely present

The scores were dichotomized for some purposes as follows:
positive diagnosis = scores 4 + 5; negative diagnosis = scores 1 + 2 + 3.

MEASUREMENT OF OBSERVER PERFORMANCE AND DIAGNOSTIC QUALITY

Presence or absence of disease or binary response categories are usually applied for diagnostic tasks. For descriptive purposes diagnoses may be classified into a 2x2 decision matrix according to the true, verified state of the cases. A perfect test should always be positive in the presence of disease and the negative in absence of disease. Unfortunately, tests are biased by errors, false positives and false negatives. In sum, four possible decision outcomes exist, namely true positive (TP), true negative (TN), false positive (FP) and false negative (FN). A set of diagnoses may be characterized by proportions derived from the decision matrix:
The **true positive ratio** is the proportion of actually positive cases which are correctly classified as positives:

\[
\frac{TP}{TP + FN}
\]

The **true negative ratio** is the proportion of the negative cases which are denoted as negatives:

\[
\frac{TN}{TN + FP}
\]

The **false positive ratio** is the proportion of positive reports in sound cases:

\[
\frac{FP}{FP + TN}
\]

The **false negative ratio** is the proportion of negative diagnosis in cases with disease:

\[
\frac{FN}{FN + TP}
\]

The confidence ratings used in the studies were gradually cumulated (scores 5, 5+4, 5+4+3 and 5+4+3+2) to obtain 4 different decision thresholds of lesion being present and the corresponding true positive ratio and false positive ratio were calculated. The scores were classified according to the true state using a specially written computer program. The data were treated according to the receiver operating characteristic (ROC) method (11,12,13,14,15).

The ROC curve gives a continuous, monotonic representation of the compromises between TP and FP decisions made when the diagnostic criterion or confidence threshold is varied (Fig. 1). As shown in Fig. 1 the ROC curves gives a description of the different trade-
offs by using various degrees of caries towards the pulp as the diagnostic threshold level. The lowermost cut off point on the curve in Fig. 1 was defined by the TP ratio and FP ratio when caries was scored radiographically in inner half of dentin (denoted R6 in Paper II). The next point on the ROC-curve was obtained using cumulative number of diagnosis in the inner half of dentin and the next diagnostic level towards the surface, namely between the inner half and the outer third of dentin (R5+R6) and by a gradual combination of the different degrees of caries monitored radiographically, all cut off points were obtained.

**Fig. 1**

ROC curves can be created in different ways. In this example it was based on radiographic caries diagnosis data from Fig. 2 in Paper II. This provides an illustration of the typical ROC curve obtained by combining true positive diagnosis of different threshold level with the corresponding false positive diagnosis. In this case the validating criterion was a positive finding of dentin caries in the prepared cavity.

The ROC curves used in the present studies were created in another, but similar way, by combining different confidence ratings that caries or another "signal" to be registered, was present.
The shape of ROC curves derived from data in medical imaging settings, usually fit into a binormal model based on the assumption of normal probability distribution of the underlying observations (16,17). The ROC points usually follow a straight line when plotted on binormal coordinate paper with normal deviate axes (17) (Fig. 2). A binormal ROC curve plotted on a normal deviate axes is defined by 2 parameters, namely intercept with the y-axis and the slope, usually denoted A and B and illustrated in Fig. 2.

![Diagram](image)

The empirical ROC curve in Fig. 1 drawn as a straight line in a binormal graph based on maximum likelihood estimates of curve parameters. The normal deviate values of true positive and false positive proportions are scaled linearly on the ordinate and on the abscissa respectively. The ROC curves are not fully described by the $A_z$ value, but the paired parameters A (y-intercept) and B (slope) give a complete description of the binormal ROC. Two iso-$\beta$ curves are drawn to illustrate the trade-off between different diagnostic criteria.

The scoring in Fig. 2 along the ordinate and abscissa is given in normal deviate values and the corresponding probability values are
noted. The closer to the upper left corner the curve is positioned, the better is the quality of diagnosis. Iso-$\beta$ curves for $\beta = 2.0$ and $\beta = 0.5$ are drawn in Fig. 2, thereby comparisons of diagnostic strategy at different points along the ROC curve can be made. The $\beta$ value reflects the observer's weighting of relative costs of mistakes (the cost of additional false positive errors versus the cost of additional false negative errors) in his reading. The respective values are given in normal deviate values. Testing for the significance of differences in the shape of ROC curves needs a simultaneous comparison of both the intercepts and the slopes \((18,19)\). In the present studies single ROC indices for comparison, were used: Area beneath binormal ROC curves \(A_z\), cut-off points \(Z_k\) and the corresponding ROC slopes for curves plotted on linear probability scales \(\beta\). The area index \(A_z\), provides a measure of diagnostic quality which is independent of variations in the diagnostic criterion. \(A_z\) ranges from 0.0 to 1.0 which is perfect performance. \(A_z = 0.5\) represents an ROC along the diagonal between the corners \((0,0)\) and \((1,1)\) indicating diagnosis at pure chance level. \(Z_k\) and \(\beta\) are decision strategy indices. The use and statistical treatment of these single indices have been reviewed and discussed by others \((11,17,20,21)\). Maximum likelihood estimates of the ROC parameters have been recommended \((5,11,17,21)\). The rating data were pooled for each setting in Papers I,II,IV-VI and treated as a single data set. In Paper III the average performance was calculated across the observers, according to the method suggested by Hanley and McNeil \((22)\) and which is based on the general formula for paired data given by Swets and Pickett \((11)\). This method takes into
account three types of variances and the corresponding correlations due to case sample matching and reader matching.

**OBSERVER PERFORMANCE INDICES**

Several types of indices have been used for discrimination between different observers, settings and diagnostic systems. Indices provided by the ROC technique have been compared to other indices used for similar tasks (11,23,24). The types of accuracy indices used in dental radiology varies widely and a few examples will be given. Arnold (10) used a weighting of the true positive diagnosis inversely related to the depth of the prepared approximal lesions. Welander et al. (25) used perceptibility curves which indicate the relationship between the film density and number of perceptible details under different settings. Sewerin and Andersen (26) used a similar system comparing the number of positive registrations made under various conditions. None of these methods consider the occurrence of false positive diagnoses. Mileman (27) used the radiographic diagnoses from specially trained observers as a validating norm. Kappa statistics which provide an agreement rate taking account of the contribution of chance agreements, was used in the statistical testing with the norm as the truth. According to Swets and Pickett (11) and Swets (24) the kappa statistic does not provide a criterion-free index of discrimination because the model is invalid. They stress that criterion-free indices are preferable and none of the examples from dental studies previously mentioned, satisfies this demand. It is essential, but difficult, to establish the true diagnoses for patients. Douglass et al. (29) used the simultaneous interpretation of all three radiographic
methods under study as the consensus radiographic standard. In general, verification bias may arise when a "gold standard" is based on the methods under investigation (29,30,31,32).

A problem which arises during diagnostic trials is the possible shift in diagnostic criteria used by the observers, and this may affect the diagnostic outcome and cause difficulties in statistical testing. To assess the diagnostic quality under different settings a criterion-free index is needed. Usually the observer will be concerned with the benefits and costs of the various outcomes and thus the strictness of the criterion used might be affected. The ROC method was chosen in the present studies since this method provides parameters of diagnostic quality which are independent of the diagnostic criterion adopted and uncontaminated by expectation and motivation (5,11,33,34,35,36).

The ROC analysis has the ability to isolate the effects of response bias and provides criterion-free indices of diagnostic efficacy. The method has been described in detail along with a review of its applications in medical radiology, perception and decision making, weather forecasting, information retrieval, industrial quality control, military monitoring and crime investigation by Swets and Pickett (11). In studies on lesion detectability in general radiology the observer performance analysis provided by the ROC method, has been the most common approach (37).

The ROC technique has to some extent been used in evaluation of observer performance in oral radiology. The method has been
applied to quantify diagnostic quality using different intra-oral films (38,39,40,41,42,43), film-screen combinations (44), the effect of variations in beam-energy (kV) (41), in efficacy studies on digital subtraction technique (45,46) and xeroradiography (39,43,47) In addition, the method has also been used for description of variation in decision strategies in dentistry (49,49,50).

QUESTIONNAIRE STUDY (PAPER VII)

A pre-coded questionnaire was sent to a random sample of dentists in Norway in March 1983. Out of 741 dentists, 616 replied (83%). In addition to demographic variables, the dentists were asked 10 questions about opinions, experience, treatment criteria and routines in connection with the radiographic examinations for approximal caries. For description of data, statistical analyses and testing the statistical packages BMDP (51) and Minitab (52) were used.

RESULTS AND GENERAL DISCUSSION

HOW EFFECTIVE IS THE RADIOGRAPHIC DIAGNOSIS OF MINERAL LOSS IN APPROXIMAL ENAMEL (PAPERS I,II)?

The answer to this question is linked to the information sought when radiographic examinations of approximal caries are carried out. One definition of efficacy in this context might be "the extent to which an observer can distinguish among various states of disease and health by using the imaging procedure" (17), and in the present studies the term diagnostic performance is used synonymously with efficacy. In dental radiography several factors influence the detectability of a lesion: the loss of minerals; the
bucco-lingual extent of the lesion; the pulpal extent of the lesion; the tooth morphology and finally the projection geometry. Traditionally decisions on the presence of caries are dichotomous (53) and all the factors mentioned contribute to the final diagnosis. There is an urgent need for a validated scoring system which monitors the degree of approximal caries lesions for the purpose of more detailed information put into the therapeutic decision making process as well as for studies on caries increment and progression in epidemiologic surveys or in clinical trials (54,55). The relative number of small, incipient lesions has increased during the last decades, due to slower progression rather than decreased number of new lesions (56). This may lead to an increased number of underscorings in epidemiological surveys since the degree of caries has become less severe on average. Different degrees of caries, ranging from demineralization in the outer third of enamel to involvement of the dentino-enamel junction, have been suggested as threshold lesions for radiographic diagnosis (57,58,59). Data for the relationship between the radiographic diagnosis and mineral loss, however, was not found in the literature, and therefore such a comparison was carried out in Paper I.

The diagnostic quality measured as $A_Z$, indicated a substantial weaker radiographic signal strength when the relative mineral loss in the direction of the X-ray beam decreased from 10% to 5%. If a threshold value is to be chosen, the 10% mineral loss may be regarded as such a value. Horizontal angulation $12.5^\circ$ to each side from a tangential projection did not affect the diagnosis except for the estimates of the degree of caries. A higher frequency of
overscoring was observed on radiographs produced using eccentric projections. Sensitivity and specificity estimates for artificially created lesions with 5% mineral loss were at the same level as for caries lesions without cavity formation (Papers I,II).

RADIOGRAPHIC DIAGNOSIS OF APPROXIMAL CARIES (PAPERS II,III,VI)
RADIOGRAPHIC DIAGNOSIS COMPARED TO CLINICAL CHARACTERISTICS OF THE LESIONS
The radiographic diagnoses were compared to the clinical characteristics of the approximal caries lesions. These were assessed simultaneously by the 2 authors of Papers II and III to reach a joint decision. The results showed that the probability of a positive radiographic diagnosis increased as the severity of the lesion as observed on the surface increased (Papers II,III,VI). The smallest caries lesions characterized as discolorations only, without any break in the surface, were on average classified as being sound radiographically in 61.3% of diagnoses (Paper I). The TP percentage was 96.4% for surfaces with cavities (>1.5mm). However, 15.7% of the radiographic diagnosis concerning the actually sound surfaces were FP diagnoses, of which nearly one third were judged to penetrate into dentin. Estimates of lesion depth based on the radiographic outline were compared with the "clinical" appearance in prepared Class II cavities and revealed a fairly close relationship, but a slight tendency towards underestimation of the radiographic depth (Paper II). On average 6 out of 10 assessments of radiographical depth were correct with regard to the extension into enamel or dentin.
Another group of dental practitioners, 243 participants in courses dealing with approximal caries, diagnosed radiographically 7.4% of sound surfaces as having caries in another sample of extracted teeth (Paper VI). The lower ratio of false positives might reflect another, more strict, diagnostic strategy since these dentists should also consider the treatment need. The caries lesions without cavitation were positively diagnosed in about one out of four surfaces, but the true positive rate increased to about four out of five surfaces when a "big" cavity (>1mm) was present. Results from different studies show that 57-90% of lesions which are quantitatively evaluated on radiographs are correctly judged either to be confined to enamel or to penetrate dentin (Table 1).
### Table 1

The degree of approximal caries monitored radiographically validated against findings on microradiographs or in prepared Class II cavities. Compiled data.

<table>
<thead>
<tr>
<th>Authors/Paper</th>
<th>Type of validation</th>
<th>Radiographic diagnosis</th>
<th>Sound Enamel Dentin caries caries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bille &amp; Thylstrup (60)</td>
<td>In cavity (in vivo)</td>
<td>6/0a 39/51 61/107</td>
<td>76.5% 57.0%</td>
</tr>
<tr>
<td>Thylstrup et al. (61)</td>
<td>In cavity (in vivo)</td>
<td>- 84/102 417/558</td>
<td>- 82.4% 74.7%</td>
</tr>
<tr>
<td>Mejäre &amp; Malmgren (62)</td>
<td>In cavity (in vivo)</td>
<td>- 9/10 31/50</td>
<td>- 90.0% 62.0%</td>
</tr>
<tr>
<td>Paper IIB</td>
<td>In cavity (in vitro)</td>
<td>24/28 33/57 39/66</td>
<td>85.7% 57.9% 59.1%</td>
</tr>
<tr>
<td>Furdell-Lewis et al. (59)</td>
<td>Microradiography</td>
<td>18/18 41/64 12/18</td>
<td>100.0% 64.1% 66.7%</td>
</tr>
</tbody>
</table>

The frequency of true negative registrations were recorded in only 2 of these studies (59, Paper II) and a high specificity is indicated. The radiographic diagnoses which do not correspond to the "true" depth mostly represent under-registration and thus indicate the relative weighting the observers put on additional false positive registrations compared to additional true positive registrations. However, in 2 studies only, the teeth were selected randomly (62, Paper II). The in vivo studies of Bille and Thylstrup (60); Thylstrup et al. (61) and Mejäre and Malmgren (62)

---

**a** The notation indicates underscoring

**b** Mean values for 6 observers given without decimals
were all based on teeth which were decided to be restored due to
caries, and therefore small lesions represented most likely a very
small part of this material. This may explain why these papers
report a higher frequency of agreement for recordings in enamel.

Detection of cavities has been regarded as important, since a
break in the outer enamel surface cannot be totally remineralized.
Based on the **in vitro** studies in Table 2, 4 to 6 of every 10
cavities are radiographically confined to enamel, 2 to 6 out of
every 10 are diagnosed as extending into the dentin while up to
17.9% false negative registrations or cavities are not detected
radiographically. In the works of Bille and Thylstrup (60) and
Thylstrup et al. (61) cavities are considerably more often
diagnosed into dentin radiographically (73.3-90.9%) compared to
other studies (21.4-59.5%) (Table 2). This can be explained by the
scoring system used or by methodological problems concerning the
registration of cavities **in vivo** using several dentists and lack of
standardized technique. Mejäre and Malmgren (62) have modified the
scoring system and have also defined that a cavity should extend at
least 1/5 of the enamel thickness to be registered as such.
Table 2
The degree of approximal caries monitored radiographically validated against a break in the outer surface (cavity).

<table>
<thead>
<tr>
<th>Authors/Paper</th>
<th>Type of validation</th>
<th>Radiographic diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sound</td>
</tr>
<tr>
<td>Bille &amp; Thylstrup (60)</td>
<td>Any cavity (in vivo)</td>
<td>0/53</td>
</tr>
<tr>
<td>Thylstrup et al. (61)</td>
<td>Any cavity (in vivo)</td>
<td>4/286</td>
</tr>
<tr>
<td>Rugg-Gunn (63)</td>
<td>Any cavity (in vivo)</td>
<td>2/34</td>
</tr>
<tr>
<td>Mejäré &amp; Malmgren (62)</td>
<td>Cavity &gt;1/5 of enamel (in vivo)</td>
<td>-</td>
</tr>
<tr>
<td>Mejäré et al. (64)</td>
<td>Any cavity (in vitro)</td>
<td>5/28</td>
</tr>
<tr>
<td>Paper IIC</td>
<td>Any cavity (in vitro)</td>
<td>11/100</td>
</tr>
<tr>
<td>Purdell-Lewis et al. (59)</td>
<td>Cavity &gt;1/2 mm (in vitro)</td>
<td>0/17</td>
</tr>
<tr>
<td>Marthaler &amp; Germann (65)</td>
<td>Any cavity (in vitro)</td>
<td>9/83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is interesting to notice the findings reported by Mejäré et al.; in one paper 59.5% of the cavities were diagnosed radiographically as extending into dentin (62) while in another paper only 21.4% of such lesions were diagnosed into dentin (64). An explanation

C Mean values for 6 observers given without decimals.
might be that in one study the teeth were prepared and filled (62) and in the other extracted due to orthodontic reasons (64). Thereby, the selection of cases, type of teeth (molars vs premolars) and the diagnostic strategy may have been different.

ARE RADIOGRAPHIC DIAGNOSIS MADE IN ENAMEL AND DENTIN EQUALLY EFFICIENT (PAPERS II,III)?

When comparing diagnoses made at 5 different levels toward the pulp, small and statistically insignificant (p>0.05) variations were found with respect to diagnostic quality (A_z). The diagnostic strategies differed according to the recording levels of depth. The β values at the operational points for the 2 cumulated scores of highest confidence that caries was present (score 4+5) indicated that the dentists considered cost of additional FP scores relative to additional FN scores to be about 13 times higher in inner dentin compared to outer enamel. The frequency of overestimated depth was more than 3 times higher for outer enamel lesions compared to lesions in the middle of dentin. This indicates that clinical reasoning affects the observer performance in radiographic caries diagnosis. When levels closer to the pulp were diagnosed the inter- and intra-observer variation increased, in accordance with Backer-Dirks, 1964 (66), and the diagnostic criterion became stricter. Monitoring the degree of caries seem to be of increasing importance for individual treatment of caries, for the collection of epidemiological data, and for research purposes regarding measurements of caries progression, and for clinical trials for evaluation of different prophylactic measures (67). The results indicate that the quality of radiographic diagnoses are not
affected whether it is scored in enamel or dentin. On the other hand, the clinician should be aware of the tendency toward a stricter diagnostic behaviour when dentin is examined, compared to examination for lesions in enamel.

In the clinical situation serial radiographs are often available which provide information about caries progression, but it is reported that intra- and inter-observer variations are similar to the reading of radiographs independently (68). However, no report is found about the quality of radiographically based lesion depth estimates comparing independent and non-independent readings.

**IS QUALITY OF RADIOGRAPHIC CARIES DIAGNOSIS AFFECTED BY FACTORS SUCH AS VIEWING CONDITIONS AND RADIOPACITY OF FILLINGS (PAPERS IV, V)?**

It has been claimed that it is important for the quality of radiographic diagnoses that certain requirements with respect to viewing conditions, are met (26, 69, 70, 71, 72, 73, 74, 75). Observer performance were compared for radiographic examinations obtained under two different viewing conditions ("poor" and "ideal" conditions) (Paper IV). The quality of diagnosis showed small variations according to viewing conditions and the results indicated that viewing requirements depend on the subject of interest and density of film. The results indicate that viewing conditions are not critical for the quality of radiographic caries diagnosis.
This conclusion is not in agreement with the findings Welander et al. (25) or Sewerin and Andersen (26) made. The reason may be that results from studies based on subtle and hardly visible phantom lesions (25,26) are not always directly applicable in the clinical situation. Other studies using dental caries or artificially created dental lesions as test pathology, do not confirm that viewing conditions play a crucial role in radiographic caries diagnosis on radiographs (10,76). No report was found about the possible influence of viewing conditions on the diagnostic threshold. In Paper IV it was concluded that diagnostic strategies (cut-off points and $\beta$ values) were not affected substantially by viewing conditions.

The radiographic diagnosis of secondary caries and residual caries may be more difficult and unreliable than the diagnosis of primary approximal caries (77,78). Radiopacity of the filling material is required for radiographic diagnosis of secondary caries and defects. The different composite materials for use in posterior teeth differ widely with respect to radiopacity (79). The degree of radiopacity is usually based on subjective assessment and observer preference (80,81) and in Paper II it is shown that radiopacity influences the quality of radiographic diagnosis of secondary caries and defects. From a caries diagnostic point of view a semi-radiopaque restoration seems preferable to a filling with very high radiopacity (amalgam). Both the TP ratio (sensitivity) and TN ratio (specificity) were higher in connection with the composite restorations compared to amalgam fillings. The higher number of TP registrations may be explained by the masking
effect a very radiopaque filling has on detection of an underlying lesion. The higher degree of FP scores in connection with a very radiopaque amalgam filling, could be explained by the Mach band effect (82). Further perception studies on caries, overhangs, voids and marginal defects are needed to find out the preferable degree of radiopacity.

TREATMENT DECISIONS ON APPROXIMAL CARIES (PAPERS VI, VII)

With reference to the questionnaire study (Paper VII), most dentists in Norway (66%) would fill an approximal lesion before it passes the dentino-enamel junction, judged radiographically. Sixty-nine percent of the dentists were of the opinion that the radiographic outline usually underestimated the true depth revealed during cavity preparation. If any radiolucency in enamel is chosen as indicative of restoration the consequence will be that about 60% of lesions with dentin caries will be restored; on the other hand, 5.0% of sound surfaces and 8.1% of lesions confined to enamel will be filled (Paper II).

Most dentists find it important to restore a caries lesion at the stage of cavitation (Paper VII). Paper VI indicates however, that dentists' radiographically based, restorative criteria, lead to restoration of less than 50% of these lesions. Lesions diagnosed into dentin have a high probability (p=0.908) of receiving restorative therapy according to restorative proposals analyzed from 243 dentists (Paper VI). However, some degree of overtreatment is likely to occur, especially if the philosophy is to intervene at the earliest stage of the caries process. About 2 out of 100
dentists seem to adhere to this philosophy even if the patient is not considered to be highly susceptible to caries (Paper VII). About 22% of lesions radiographically judged not to penetrate more than half way through enamel, were proposed restored (Paper VI). This treatment would most likely "hit" a lesion with a big cavity compared to a lesion with no break in the tooth surface. The dentists who more often tended to classify correctly dentin lesions also had a higher tendency toward proposals of filling in cases actually being sound.

EXAMPLE
The following hypothetical situation is used to illustrate the clinical consequences of a diagnostic test applied to a situation with low prevalence of disease. On average a twelve year old child has about 50 approximal surfaces which are susceptible to caries. The average approximal caries incidence at this age is 2 new lesions per year (83). Only 1 out of 10 is considered to be in need of restorative therapy (84) if we assume that the criterion for restoration is a break in the outer enamel surface; i.e. cavitation >1mm in width. The prevalences of caries and lesions in need of restorative therapy will then be 0.040 (2/50) and 0.004 (0.2/50) respectively. To what extent are lesions with discoloration only or a small cavity diagnosed radiographically? Suppose that 2,000 children aged 12 years from an average population have been subjected to a radiographic bitewing examination. By other means it was measured that out of 100,000 surfaces under examination actually 4% (4,000) had caries and only 0.4% (400) had cavities extending 1mm of the width on the surface. Based on
numbers from Figs. 1-4 in Paper VI the following calculations were made based on true caries prevalence being 0.04:

Diagnostic outcomes:

TP diagnoses: \(100,000 \times 0.04 \times 0.57 = 2,280\)

FN diagnoses: \(100,000 \times 0.04 \times 0.43 = 1,720\)

TN diagnoses: \(100,000 \times 0.96 \times 0.93 = 89,280\)

FP diagnoses: \(100,000 \times 0.96 \times 0.07 = 6,720\)

Total number of surfaces: \(100,000\)

Positive predictive value of radiographic caries diagnosis in outer enamel:

\[
p(s|S) = \frac{TP}{TP + FP} = \frac{2,280}{2,280 + 6,720} = 0.253
\]

Negative predictive value of radiographic caries diagnosis in outer enamel:

\[
p(n|N) = \frac{TN}{TN + FN} = \frac{89,280}{89,280 + 1,720} = 0.981
\]

Treatment decisions:

Cavities >1mm are considered in need of operative treatment in this hypothetical example.
TP decisions: 100,000 x 0.004 x 0.41 = 164
FN decisions: 100,000 x 0.004 x 0.59 = 236
TN decisions: 100,000 x 0.996 x 0.95 = 94,620
FP decisions: 100,000 x 0.996 x 0.05 = 4,980
Total number of surfaces: = 100,000

Positive predictive value (cavity) of a treatment decision:

\[ p(s|s) = \frac{TP}{TP + FP} = \frac{164}{164 + 4,980} = 0.032 \]

Negative predictive value (sound surface or lesion without cavity) of a treatment decision:

\[ p(n|N) = \frac{TN}{TN + FN} = \frac{94,620}{94,620 + 236} = 0.998 \]

These calculations show that the probability that a positive decision on treatment, based on single radiographs only, actually coincides with a caries lesion with cavity formation, is about 3 out of 100 decisions. The reason for this low value is the relatively low number of teeth in need of a restoration. This emphasizes the necessity of serial radiographs to follow the progression rate of caries (68) and the importance of taking into account clinical reasoning on individual caries risk, clinical findings and perhaps the patient's own preferences with respect to recalls (85).

Criteria for treatment of caries should be chosen in a way which maximizes the benefit to the patient with regard to reliability of
diagnosis and minimizes the probabilities of over- and under-treatment. Utility values should be placed on each possible outcome of the diagnostic and therapeutic process. A decision analytical model can provide useful information about clinical judgement, but the basic biostatistical problem lies in the choice of test material (e.g. teeth and observers) that is the source of the data (23). Mileman et al. (86) have assessed utility values for the diagnosis and treatment of caries and put them into a decision analysis on approximal caries. The approach used in that study is useful, but the conclusions reached will never be more reliable than the reliability of the data put into it. Sensitivity and specificity values given in Papers II and VI might offer validated data to be put into such an analysis and thereby highlight the effectiveness offered by the radiographic method and other diagnostic tools to the clinician, under given conditions.

Fig. 3

Positive predictive value of radiographic caries diagnoses as a function of caries prevalence according to Bayes' theorem (87). The combination of TP and FP values for each curve, was derived from Paper II (Fig. 2) and represented the actual values for different radiographic criteria. The validation criteria for a caries lesion was a dentin lesion revealed by preparing a Class II cavity.
Fig. 4

Negative predictive value of radiographic caries diagnoses as a function of caries prevalence. Five pairs of TN and FN were obtained from Fig. 2 in Paper II, representing radiographic caries diagnosis of different stringency. The curves correspond to the curves in Fig. 3 and the true lesion was defined as a dentin lesion verified by direct inspection and probing in a Class II preparation.

Figures 3 and 4 illustrate the predictive value of the radiographic method when caries in dentin is searched for and it is obvious that low prevalence gives low positive prediction value. In a low caries population the FP diagnoses might constitute the majority of positive diagnoses. It has been claimed that radiographs are not needed when surveys for caries are conducted in low caries populations, since this will not result in a substantial loss of information about the caries prevalence (88).

OBSERVER CALIBRATION?

The consequences for each individual treated of applying an extremely restorative philosophy or basing treatment decisions on poor diagnostic quality, should be investigated. From a societal point of view the economic consequences should be taken into consideration as well. Papers VII and VI have indicated an urgent need for calibration and training of the practising dentists. It has been reported that the effect on interexaminer reliability of a calibration program was limited with respect to radiographic caries diagnosis when the diagnostic criteria were well defined.
throughout the study (9). It does not seem likely that the observers in this study include the relatively few extreme individuals who contribute most to inferior treatment decisions. These dentists should be reached in general programs including individual diagnostic exercises and consensus on treatment criteria among large groups of dentists. Another approach for minimizing discrepancies in treatment criteria is to give general guidelines to the dental profession (89), but the possible effect of guidelines needs to be evaluated.

THE CLINICAL RELEVANCE OF THE FINDINGS

The present studies confirm that radiographic caries diagnosis as a diagnostic test is not absolutely valid. It has built-in errors. The clinician has to decide on a trade-off between the two types of inversely related errors, namely overrecording (FP) and underrecordings (FN). Figures 3 and 4 indicate that the prevalence of disease plays an important role, as does the relative cost of the different therapeutic outcomes. The monitored decision behaviour of dentists concerning approximal caries may lead to overprescription of radiographs and fillings as the prevalence of primary and secondary caries decreases (56, 90, 91). More research appears to be needed into methods of postgraduate training in clinical decision making. Computer assisted decision making for clinical purposes in medicine, has even been described, and programmable, hand-held calculators computing Bayesian probabilities have been introduced (87). Research in dentistry on diagnostic quality and probabilities of alternative outcomes is fundamental for the creation of an expert system in dentistry.
Artificial intelligence used in an expert system may help the clinician to reach the most likely correct decision under uncertainty.

**CONCLUSIONS**

Mineral loss which represents a threshold value for radiographic diagnosis, cannot be defined exactly. For clinical use 10% mineral loss in the direction of the X-ray beam may constitute a borderline lesion for radiographic detection, and caries lesions without cavitation seemed to be beyond this diagnostic threshold.

The degree of caries estimated by using radiographs is fairly closely related to the depth of the tissue changes recorded in the prepared cavity.

Radiographic examinations more often lead to underestimation than overestimation of the degree of caries.

Radiographic caries diagnoses made at different degrees of penetration toward the pulp showed insignificant variations with respect to quality, but the observers were more confident of caries being present (used more strict criterion) when they scored caries in inner dentin.

Consensus on diagnostic criteria and improved diagnostic quality are considerably more important to the quality of therapeutic decisions on approximal caries than viewing conditions and film density.
A semi-radiopaque material in Class II fillings seems to offer advantages compared to amalgam in respect of the diagnosis of secondary caries and marginal defects.

In Norway there is a danger that dentists will restore approximal caries lesions too early and before these can be diagnosed in dentin radiographically.

REFERENCES

Presented according to recommendations given by the International Committee of Medical Journal Editors:


61. Thylstrup A, Bille J, Qvist V. Radiographic and observed tissue changes in approximal carious lesions at the time of operative treatment. Caries Res 1986;20:75-84.


89. Socialstyrelsens författningssamling. Socialstyrelsens almänna råd om diagnostik, behandling och registrering av karies på approximalytor samt rutiner för överföring av patient till annan tandläkare. SOSFS 1984;3.


PAPER I
Radiographic Diagnosis of Mineral Loss in Approximal Enamel

I. Espelid, Anne Bjørg Tveit

Department of Oral Radiology, Department of Cariology and Endodontics and Department of Dental Research, School of Dentistry, University of Bergen, Norway

Key Words. Enamel lesions · Observer performance · Radiography, dental · Receiver Operating Characteristic (ROC) analysis

Abstract. The relationship between mineral loss in approximal enamel and the precision of the radiographic diagnosis was investigated using experimental lesions and Receiver Operating Characteristic (ROC) technique. The ROC curves indicated a substantial loss of radiographic information concerning the presence or absence of a lesion when mineral loss was reduced from 10 to 5%, but the diagnoses were better than pure guessing. Variations in horizontal angulation within a range of 25° did not affect detection, but lesion depth was overestimated more often at eccentric projections. The results indicate that radiographic diagnosis is unreliable for small carious lesions with about 5% mineral loss in the X-ray beam direction, and that positive diagnosis is frequently due to observer error.

The extent of carious lesions is generally underestimated from intraoral radiographs compared to clinical or histological findings [Burket, 1941; Darling, 1959; Jung, 1965; Leijon, 1969; Marthaler and Germandt, 1970; Gwinnett, 1971; Rug-Gunn, 1972; Purdell-Lewis et al., 1974; Downer, 1975; Bille and Thylstrup, 1982]. Several studies have indicated that, clinically, a threshold level for the radiographic diagnosis of caries exists. Purdell-Lewis et al. [1974] have suggested that a demineralization which can be detected radiographically involves at least the outer third of enamel. Gwinnett [1971] found that half the enamel is frequently affected before a carious lesion can be diagnosed. Darling [1959] has claimed that carious lesions which can be diagnosed radiographically have already reached the dentino-enamel junction.

The radiographic detectability of a carious lesion depends mainly on the mineral loss in the direction of the X-ray beam. The purpose of this study was to examine the relationship between mineral loss and radiographic diagnosis, and the consequences of variations in horizontal angulation on the extent of the radiographic lesion. A major problem in diagnostic trials involving human observers is to adjust for
variations in diagnostic strategies during the experiment. The so-called ROC (Receiver Operating Characteristic) technique was used here because it is independent of such variations [Metz, 1978].

**Materials and Methods**

The experimental material consisted of 13 extracted permanent teeth, 5 premolars and 8 molars. They were all clinically sound and had been stored in a 10% formalin solution since extraction. Each tooth was sectioned twice so as to produce an 0.5–0.7 mm thick mesio-distal longitudinal section (fig. 1). The teeth with the section in position were mounted in methyl methacrylate tubes with elastic impression material. The tubes were fixed in a device which made it possible to repeat radiographic exposures with identical projections. Radiographs of each tooth were taken in three different projections, one tangential to the proximal surface, one 12.5° mesio-eccentric and one 12.5° disto-eccentric (fig. 2). These variations in horizontal angulation were found to be within the clinical range, based on test radiographs. Exposures were made at 65 kV and 15 mA using a dental X-ray machine with an electronic timer (Ritter® Explorer II). The exposure time was 1.2 s, the focus-object distance 32.0 cm, and the object-film distance 1.5 cm. A 1.0-cm wide Plexiglas® container filled with water was placed between focus and object to simulate soft tissues. Kodak® DF-58 (double-pack) film was used. All radiographs were processed by a standardized procedure.

The sections were removed from the tubes and an air rotter with a 1-mm cylindrical diamond bur was used to prepare lesions in the enamel from the most prominent part of the mesial or distal surface to the dentino-enamel junction (fig. 1). The lesions were about 1 mm wide. Then the sections were replaced between the buccal and lingual parts of the teeth and new radiographs taken. The sections were removed 4 times and each time the whole of the section was reduced about 100 µm in thickness, replaced and radiographed, until the lesions were hardly visible in the radiographs. Exposure parameters and geometry were the same in each exposure.

The teeth were then embedded in methyl methacrylate and sectioned transversely through the lesions. In order to calculate the mineral loss in the X-ray beam direction the thickness of each section, measured with a micrometer caliper (EJ, Sweden), was compared to the width of its tooth. This dimension was measured at an arbitrary depth of 1 mm below the lesion border, approximately corresponding to the enamel thickness in this region, in a direction parallel to the central beam. The measurement error (%) of the tooth width was calculated from 30 repeated measurements according to the formula

$$M_e = \frac{\sqrt{\frac{\Delta^2}{n}}}{D} \times 100,$$

where $M_e$ is measurement error in percent, $\Delta$ is the difference between the first and second measurement, $n$ is the number of dimensions to be measured and $D$ is the mean value of the measured dimensions.

*Fig. 1. A premolar cut longitudinally into three parts. An approximal groove is prepared in the middle section, extending from the prominence to the dentino-enamel junction.*

*Fig. 2. Cross-sectional view of the tooth in figure 1. The three parts are put together and radiographed with different horizontal X-ray beam angles.*
A total of 127 radiographs showing teeth with lesions, and 109 radiographs taken before the artificial lesions were prepared, were examined by 5 experienced dentists. They used a viewing box with a 2 x magnifying lens and had no time restrictions. The light sources available had luminous intensities of 4,500 and 96,000 lx, the latter to optimize the interpretation of the very darkest images. The diagnoses were made according to a 5-point confidence scale: 0 = lesion not present; 1 = lesion probably not present; 2 = equal chance of being present or not; 3 = lesion probably present; 4 = lesion present.

The scorings were classified into a so-called decision matrix and then treated according to the ROC method [Swets and Pickett, 1982; Swets, 1973]. A maximum likelihood estimate of ROC-curve parameters was obtained using a computer program developed by Dorfman and Alf [1968].

The prevalence of lesions was known in this study, and the probability of lesions, given positive diagnosis \( p(S|S) \), could be calculated according to Bayes' theorem [Weinstein et al., 1980]:

\[
p(S|S) = \frac{p(S|S)p(s)}{p(S|S)p(s) + p(S|n)p(n)}
\]

where \( p(S|S) \) is the ratio of true-positive scores, \( p(S|n) \) is the ratio of false-positive scores, \( p(s) \) is the prevalence of lesions and \( p(n) \) is the frequency of sound surfaces.

The probability of a lesion, given score 0, was calculated by replacing the \( p(S|S) \) with the ratio of false-negative scores, \( p(N|S) \), and \( p(S|n) \) with the ratio of the true-negative scores, \( p(N|n) \).

**Table I.** Classification of lesions according to extent in percent of the tooth dimension in X-ray beam direction

<table>
<thead>
<tr>
<th>Lesion grouping (mineral loss)</th>
<th>S,</th>
<th>Range</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>4.94 ± 0.90</td>
<td>3.47-5.97</td>
<td>57</td>
</tr>
<tr>
<td>7%</td>
<td>6.97 ± 0.60</td>
<td>6.07-7.94</td>
<td>43</td>
</tr>
<tr>
<td>10%</td>
<td>10.18 ± 1.94</td>
<td>8.00-14.42</td>
<td>27</td>
</tr>
</tbody>
</table>

\( S, \) is standard deviation and \( n \) denotes number of lesions.

From those radiographs which revealed a lesion, 10 were randomly chosen and scanned microdensitometrically (Joyce & Loebl® 3CS). Several scans with a distance of 0.15 mm were made in the approximal enamel across the lesion in a direction parallel to the dentino-enamel junction. The aperture was adjusted so that the width of each scan was 0.1 mm.

**Results**

The lesions were grouped according to size in percent of the tooth width, measured in the X-ray beam direction. The various groups are characterized by the values given in table I. In the following, the lesions within each group are referred to as 5, 7 and 10% lesions, respectively. The measurement error represented ± 6.6%, indicating acceptable precision of measurements.

The distribution of the scores within each group is shown in figure 3. The cumulative positive score, which is the total of all scores from 1 to 4, increased with le-
The diagnostic efficiency is best for curves which are located in the upper left corner of the coordinate space. The area beneath the ROC curve ($A_z$) is a single parameter representing this diagnostic efficiency. The $A_z$ value is calculated by using the fitted, binormal ROC curve which is a maximum likelihood estimate of the empirical curve. The goodness of fit is checked by using the $\chi^2$-test. The probabilities derived did not reject the assumption of normal distribution because $p(\chi^2) > 0.05$. Table II gives $A_z$ values related to the different sizes of the lesions. This parameter of diagnostic efficiency indicates a proportionality between sensitivity and mineral loss of the lesion. The difference between 5% and 10% loss of mineral was significant at the 5% level under the assumption of normality for the $A_z$ estimates.

At the actual prevalence (0.54) the probability of a positive diagnosis being true decreased from 0.91 when 10% lesions had score 4 ('certainly present'), to 0.79 at 5% lesions. When lesions of the same size were rated score 1 ('probably not present') the probabilities were 0.42 (10% lesions) and 0.44 (5% lesions), respectively. The probability that a negative diagnosis was wrong was calculated from the number receiving score 0 ('certainly not present'). These values varied from 0.23 (10% lesions) to 0.38 (5% lesions). The probabilities, calculated by applying Bayes' theorem, are given in figure 5, according to diagnostic confidence score and lesion group.

A 12.5° variation in horizontal angulation of the X-ray beam did not result in any substantial difference in diagnostic efficiency (fig. 6; table III).
Fig. 5. The probability that different scores hit an actual lesion.

<table>
<thead>
<tr>
<th>Mineral loss</th>
<th>( A_j )</th>
<th>( S_{A_j} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5°</td>
<td>0.674 ± 0.021</td>
<td>NS</td>
</tr>
<tr>
<td>7°</td>
<td>0.737 ± 0.022</td>
<td>NS, ( p &lt; 0.05 )</td>
</tr>
<tr>
<td>10°</td>
<td>0.790 ± 0.024</td>
<td></td>
</tr>
</tbody>
</table>

\( S_{A_j} \) is standard error.

Table III. Area underneath ROC curves (\( A_j \)) representing 12.5° difference in horizontal angulation

<table>
<thead>
<tr>
<th></th>
<th>( A_j )</th>
<th>( S_{A_j} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ 12.5° eccentric</td>
<td>0.725 ± 0.026</td>
<td>NS</td>
</tr>
<tr>
<td>Tangential</td>
<td>0.732 ± 0.026</td>
<td>NS</td>
</tr>
<tr>
<td>- 12.5° eccentric</td>
<td>0.729 ± 0.026</td>
<td></td>
</tr>
</tbody>
</table>

\( S_{A_j} \) is standard error.

All the experimental lesions were prepared to the dentino-enamel junction, but in 17.4% of the true-positive observations, the dentists diagnosed the lesion as including dentin. A higher frequency of over-scoring was observed on eccentric projections. There was no significant difference in size between those lesions which were overscored into dentin and those not overscored (t test). Dentin lesions were diagnosed in 6.8% of the radiographs which did not actually have a lesion. Most of these scores were made with a high degree of certainty, namely scores 3 and 4.

The densitometric tracings across the lesions did not reveal any sharp boundary. The density fluctuations in the border area of the lesions were mostly less than 0.1 D.
Fig. 7. Lesion contrast on radiographs was examined by densitometric tracings across the lesions in longitudinal tooth direction. The upper curve represents the outer approximal surface. One scan is made for each 0.15 mm and the lowest curve shows density variations along the dentino-enamel border.

and the lesions could hardly be identified by reading the curves (fig. 7). By repeated tracings the curves were precisely reproduced.

Discussion

This study demonstrates a relationship between lesion width in the X-ray beam direction and diagnostic quality. The $A_z$ values which reflect this association do not correspond to any of the actual response frequencies within this trial, but represent the proportion of correct scores within a so-called forced choice test [Green and Sweis, 1974]. In a forced choice test a comparison is made of one true-positive and one negative case, and the test observer has to decide which is which. All the experimental lesions gave considerably smaller $A_z$ values than clinical cavities [Espelid, 1982], but all values were greater than 0.5, which corresponds to pure guessing.

The Bayesian probabilities that different scores hit the actual lesions, gives a posteriori information about the validity of the diagnoses. The probabilities characterize various cutoff points of the ROC curve and change according to variation in disease prevalence, and they represent another way of presenting the information content of the diagnoses. In our calculations we used the same prevalence values because all lesions were pooled in the diagnostic trial. The proportional relationship between Bayesian probability of correct positive score and lesion size was restricted to the highest level of observer confidence that a lesion was present. The diagnoses at lower confidence levels showed the inverse tendency.

The overscoring within dentin may reflect either a misinterpretation of radiographic tooth anatomy or be due to variations in observer performance. The frequency of overscoring in dentin was increased by using eccentric projections. This is in accordance with previous observations [Leijon, 1969; Sewerin, 1981] and could be explained by superimposition of the lesion onto the dentin in the two-dimensional radiographic image. The diagnostic consequences of variations of the vertical angulation of the X-ray beam were not considered in this study.

The artificial lesions in the present study were all uniform in shape and, in contrast to carious decay, had sharp boundaries. They were, therefore, expected to be easier to recognize visually than carious lesions with a similar mineral loss. The densitometric tracings across the arti-
ficial lesions indicated that the image of the tooth overshadowed the lesion boundary, which was very difficult to identify. As a consequence it was not possible to compare lesion contrast given by the densitometric tracings and diagnostic sensitivity parameters. The recognition of these very weak, visual signals may be explained by the fact that the search for proximal caries is restricted to a small, well-defined part of the tooth. Kundel et al. [1979] reported that decreased edge gradient (increased blur) of artificial nodules on chest films gave a monotonical decrease in human detection. The mean contrast of the nodules in that study was 0.12 D. This indicates that sharpness of lesion boundaries may also have importance for radiographic diagnosis of caries.

Variations in the diagnostic criteria change the operational point along the same ROC curve, with corresponding new values of under- and overscorings. If an observer wants to diagnose a very small mineral loss, as represented by the 5% lesions, and detect most of the lesions (90%), he will get about 80% overscoring, while the same frequency of true-positive scores will result in less than 60% overscoring of the 10% lesions. Therefore, the observer should be aware of the clinical consequences of his radiographic diagnoses, and take these into account.

The clinician who interprets radiographs showing small caries-like density changes should consider the predictive value of his diagnoses. The diagnostic quality could be checked by a test observation of dental radiographs of proximal surfaces, whose condition is known. The radiographic diagnosis could also be compared to clinical observations of cavity preparation during treatment [Bille and Thylstrup, 1982].

The Bayesian probability that different scores hit a true lesion varies according to lesion prevalence. In this study 54% of the examined surfaces had true-positive lesions. If, for example, only 5% of the surfaces had a lesion, a substantial decrease in $p(s|S)$ from 0.79 to 0.36, and $p(s|N)$ from 0.38 to 0.03, would take place for the 5% lesion group, score 4 and score 0, respectively. It is therefore important to take the lesion prevalence into consideration when the diagnostic confidence threshold is to be chosen.

Although radiographic examination is a sensitive method for registration of mineral loss in enamel, interpretation of the radiographic findings may be difficult. Lesions of different sizes and with varying degrees of mineral loss may have identical outlines in radiographs. The radiographic summation picture of a lesion can represent one of several clinical situations concerning lesion morphology and mineral loss, and is also a function of X-ray beam angulation. Consequently, it is important when making treatment decisions based on radiographic images of carious lesions, to take into account clinical findings such as the size of contact area and the width of the tooth in the X-ray beam direction.

An exact diagnostic threshold value cannot be deduced, but radiographs of the 5% lesions had poor information content. Radiographic diagnosis made on such lesions would seem to be inappropriate for clinical use.

The classical definition of a threshold is 50% probability of perception after correction for chance success. The idea of the so-called high-threshold theory is that sensory threshold is a well-defined cutoff le-
vel which the signal has to exceed to be detected [Green and Swets, 1974]. This gives a straight 'threshold line' in the ROC diagram through the X and Y probability coordinates, respectively (0, 0.5) and (1, 1). The line lies most close to the ROC curve of the 10% lesion group, indicating a possible answer to the question of a threshold value.

The tooth dimension in the X-ray beam direction is affected by changes in horizontal angulation. This did not influence the detection of lesions, but eccentric projections more often gave overscoring in lesion depth.

References


PAPER II
Clinical and radiographic assessment of approximal carious lesions

Ivar Espelid and Anne Bjørg Tveit
Department of Pedodontics, Department of Oral Radiology, and Department of Cariology and Endodontics, School of Dentistry, University of Bergen, Bergen, Norway


The aim of the study was to compare the radiographic diagnosis of approximal carious lesions with visual observations of the approximal surfaces and within drilled Class II cavities (made into the pulp). Sound \((n = 28)\) and carious \((n = 123)\) approximal surfaces of extracted premolars and molars were radiographed. The radiographs were studied by seven observers to diagnose caries. Lesions without cavitation were most often classified as sound \((61.3\%)\). When lesions had cavities, the rate of detection increased to \(89.1\%). Sound surfaces were erroneously classified as carious in \(15.7\%)\) of cases. Statistically, about 6 out of every 10 qualitative assessments of lesion depth on the basis of radiographs correctly recorded lesions as being in enamel or extending into dentin. The interexaminer variation in radiographic caries diagnoses were mostly due to difference in diagnostic criteria, whereas differences in diagnostic capability were less important. □ Dental caries; diagnosis; radiography

Ivar Espelid, Department of Pedodontics, School of Dentistry, Årstadveien 17, N-5000 Bergen, Norway

The radiographic examination of approximal dental surfaces is the most reliable source of information about the presence or absence of caries. Radiographic diagnosis of approximal caries plays a crucial role in the treatment decision \((1)\). The interpretation of the radiographs is, however, subject to error, and the sensitivity and specificity of the diagnosis should therefore be taken into consideration when the treatment decision is made.

Cavitation has usually been considered a criterion for operative treatment of an approximal lesion \((2, 3)\), and it is generally accepted that restorative treatment should be initiated when the radiographic appearance of an approximal lesion indicates discontinuity of the enamel surface.

When the relationship between the radiographic and clinical appearance of approximal lesions has been considered \((2, 4-6)\), the most tenable conclusion seems to be that more than 8 out of the 10 radiolucencies that appear to have reached the dentin represent clinical cavitation. However, in one clinical study, in which the radiological appearance was compared with visible tissue changes in opened, drilled cavities, cavitation was found in only 52\% of the lesions registered radiographically halfway through the dentin \((1)\). These authors claim that 'an adjustment seems reasonable of the hitherto accepted relation between radiographic and clinical caries in epidemiological studies'. For clinicians, these contradictory results may lead to confusion in the diagnosis of caries and have consequences for treatment planning.

Previously, when the clinical and radiographic appearances of lesions have been compared, one \((1, 2, 5)\), two \((6, 7)\), or three \((4)\) dentists scored each lesion. It is well known that great variation exists between different observers \((8)\), and this fact should be taken into account when considering such studies.

Definite knowledge about risks and benefits is not available when a dentist has to decide whether to make a restoration. Dentists may place different values on clinical findings and information about the patient. Any decision is only as good as the reliability of the data put into it. To obtain knowledge about the quality of dental care, such as treatment of approximal caries, it is
essential to validate the information on which the treatment decisions are based. For approximal caries it is important to study the relationship between radiographic and clinical findings during drilling. The most reliable knowledge will be obtained by using an in vitro model with extracted teeth, which excludes many clinically confounding variables. Validation of other factors that might influence the final treatment decision, such as bacteria tests, saliva tests, and patient history, is outside the scope of the present investigation.

The aim of this study was to validate the radiographic method for detecting approximal carious lesions of different severities. The validating criteria were (a) the extent of the lesion as observed on the approximal tooth surface and (b) the penetration of the lesion observed during preparation of a Class II cavity.

Materials and methods

Extracted premolars (n = 105) and molars (n = 46) that had been stored in 2% benzalkonium chloride since extraction were used in this study. The history of the teeth was unknown, but most of them were young teeth that had apparently been extracted recently for orthodontic reasons. The approximal surfaces were thoroughly air-dried, visually inspected, and examined by probing. The surfaces were then allocated to sound (n = 28) or carious (n = 123) groups, after those with hypoplasia or other non-carcious defects had been excluded. The carious surfaces were designated S1 to S5 (left column in Fig. 1) in accordance with the buccolingual extent of the lesions and the maximum diameter of the cavity if present. Any break in the enamel surface detected with a probe was registered as a cavity: if the buccolingual extent was 1.5 mm or less, it was noted as a small cavity, otherwise as a big cavity. The direct classification of the approximal surfaces was done independently by both authors, and, on disagreement, the final scoring was based on a joint decision. Each surface was radiographed, using double-pack film (Kodak DF 58) and a dental X-ray machine (Ritter Explorer II).

Exposures were made at 65 kV and 15 mA with exposure time ½ sec at an object-focus distance of 30 cm and object-film distance of 2 cm. The horizontal projection was tangential to the approximal surface. A 10-mm-wide Plexiglas container filled with water was placed between the focus and the object to simulate soft tissues. All films were processed, fixed, and rinsed in batch by a standardized procedure. Seven experienced dentists interpreted the radiographs under standardized conditions, using an X-ray viewer with a ×2 magnifying lens. They classified the extent of radiographic images of lesions as follows: R2 = not more than halfway into enamel; R3 = up to the dentino-enamel junction; R4 = into the outer third of dentin; R5 = between outer third and half of dentin; R6 = more than halfway into dentin. Negative caries diagnoses were denoted R1. The observers used a 5-point confidence rating scale to express their level of confidence that a lesion was present: 1 = lesion not present; 2 = lesion probably not present; 3 = equal chance of being present or not; 4 = lesion probably present; 5 = lesion present.

To record the extent of lesions in a pulpal direction, modified Class II preparations into the pulp were cut with a water-cooled air-rotor. The size of the lesion was compared with the total enamel/dentin thickness. To ensure that the registrations of maximum depth were correct, drilling continued until the lesions no longer could be detected in the gingival wall. The classification used corresponded to the radiographic scores (R1-R6) and were designated LI-L6 (Fig. 2). Any visible color change of enamel and dentin and/or reduced hardness compared with normal tissue detected by probing, was taken as evidence of caries. The lesions were assessed by two observers (joint decision).

The radiographic diagnoses were processed by the ROC technique. A more detailed description of this method is given in another paper dealing with the identical material (9). ROC curves describe the relationship between true positive (TP) scores, which are the positive caries diagnoses corresponding to actual carious lesions, and false positive (FP) scores, which
are sound surfaces diagnosed on radiographs as having caries. The estimated area ($A_2$) beneath the ROC curves was used as a measure of diagnostic quality.

**Results**

**Diagnostic quality**

The mean area beneath the curves are given for each observer in Table 1. The differences between observers were not statistically significant ($p > 0.05$).

**Radiographic diagnosis**

In the following, the five scores in the confidence rating scale were reduced to two categories. Scores 5 and 4 were interpreted as positive caries diagnoses, and scores 3, 2, and 1 as negative caries diagnoses.

Many surfaces were given different radiographic scores. Owing to this disagreement among the observers, the mean score for each lesion was considered the best estimate in comparing radiographic and direct scoring. To express the variation between the dentists, the 'extreme' scores, those farthest from the mean, are also given. One extreme is named the 'strict' observer, who gives score 4 or 5 only when he is quite sure that a lesion is present, whereas the other is called the 'lax' observer because he gives score 4 or 5 even when a bit uncertain about the diagnosis.

The possible influence of tooth type on the radiographic detection type was examined. Lesion severity, as recorded on the surface and within the drilled cavity, was scored 1-7 and 1-5, respectively. Score 1 denoted no lesion. On an average, the lesions were more severe in the premolars than in the molars. The mean scores on the surface were 4.4 ($SD = 2.0$) for premolars and 3.9 ($SD = 2.3$) for molars, and the corresponding scores in the drilled cavity were 3.2 ($SD = 1.4$) and 2.8 ($SD = 1.5$). A stepwise multiple regression analysis was performed, taking lesion severity into consideration. The test did not show any statistically significant ($p > 0.05$) difference in radiographic detectability due to tooth type.

**Table 1. Area beneath ROC curves ($A_2$) for different observers used as index of diagnostic quality. Maximum area is 1, obtained when all diagnoses are correctly made with highest degree of confidence. The value 0.5 represents pure guessing. The values are averaged across the level of depth under examination. $SE_{A_2}$ denotes standard error. Differences between observers were not statistically significant ($p > 0.05$)**

<table>
<thead>
<tr>
<th>Observer</th>
<th>$A_2$</th>
<th>$SE_{A_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.854</td>
<td>0.096</td>
</tr>
<tr>
<td>2</td>
<td>0.899</td>
<td>0.054</td>
</tr>
<tr>
<td>3</td>
<td>0.869</td>
<td>0.033</td>
</tr>
<tr>
<td>4</td>
<td>0.883</td>
<td>0.087</td>
</tr>
<tr>
<td>5</td>
<td>0.933</td>
<td>0.042</td>
</tr>
<tr>
<td>6</td>
<td>0.864</td>
<td>0.042</td>
</tr>
<tr>
<td>7</td>
<td>0.842</td>
<td>0.052</td>
</tr>
</tbody>
</table>

**Fig. 1. Square table of surface recordings (S1-S5) and radiographic diagnoses (R1-R6) of approximal surfaces. S1 = sound surface; S2 = lesion without cavitation; S3 = small cavitation; S4 = lesion showing both discoloration and a small cavity; S5 = lesion showing both discoloration and a big cavity. R1 = no lesion; R2 = not more than halfway into enamel; R3 = lesion showing both discoloration and a small cavity; and R6 = more than halfway into dentin. The heavy types in each box denote the average number for seven observers. The upper number belongs to the observer who made fewest positive caries diagnoses, and the lower number belongs to the observer who recorded caries most often.**
**Relationship between the radiographic and clinical appearance**

The comparisons between the radiographic scoring and the direct scoring of the lesion as observed on the tooth surface (S1-S5) are given in Fig. 1. The upper numbers belong to the strict observer, the lower to the lax one, whereas those in between are the means based on the scores given by all dentists. The strict observer diagnosed correctly all intact surfaces, whereas the lax one diagnosed about 43% of them as carious, giving a high percentage of false positive scores. On the other hand, the lax dentist diagnosed correctly the actual lesions more often than the strict one, resulting in a higher percentage of true positive scores. The frequency of detection of lesions without cavitation showed great variation between observers, from 91.3% to 47.8%. On the basis of the mean scoring, it could be concluded that 61% of carious lesions without cavitation (S2) were not detected radiographically. The rest were mainly scored R2 or R3—that is, radiolucencies extending to the dentinoenamel junction. About 5% were diagnosed on radiographs as extending into dentin, about the same proportion as for sound surfaces. After correcting for success by chance, the probability of a positive radiographic diagnosis of a lesion without a cavity was calculated to be about 0.2.

About 63% of lesions with small cavitations (S3) were either diagnosed as intact or found on radiographs to be limited to the enamel. The other 37% were given the diagnosis 'lesion in dentin'. White or brown discoloration in addition to a cavitation (S4) did not influence the radiographic diagnoses to any significant extent. Fifty-five percent of the lesions with cavities greater than 1.5 mm were recorded on the radiographs as radiolucencies in dentin (R4-R6).

The relationship between the pulpal extent of the approximal lesions, as observed in opened, prepared cavities, and the radiographic scoring is shown in Fig. 2. Cavitation is not included in this comparison. The variation in the diagnosis and the difference in overscoring/underscoring between the strict and lax observer can be read in the upper and lower rows in each box. Considering the mean values, 60–80% of the lesions were scored on radiographs given the same score as or one score lower than observed directly.

Whether the radiographic outline indicates carious lesions with cavitation is of interest. In Fig. 3 three different radiographic outlines are compared with the corresponding clinical appearances of lesions with and without cavitation. The mean values indicate that when caries is diagnosed on radiographs to extend up to halfway through the enamel, 8% of the diagnoses are false positive diagnoses (FP). 68% of the diagnoses represent lesions confined to

---

![Fig. 2: Square table of recordings in opened, drilled Class II cavities (L1-L6) and radiographic diagnoses (R1-R6) of approximal surfaces. The scoring system of the lesion extent in the drilled cavities: L1 = no lesion; L2 = not more than halfway into enamel; L3 = up to the dentinoenamel junction; L4 = into the outer third of dentin; L5 = between outer third and half of dentin; and L6 = more than halfway into dentin. The radiographic scoring system is explained in the legend to Fig. 1. The heavy types in each box denote the average number for seven observers. The upper number belongs to the observer who made fewest positive caries diagnoses and the lower number belongs to the observer who recorded caries most often.](image-url)
Fig. 3. Comparison of three different radiographic outlines (R2, R3, and R4) and the corresponding recordings in drilled cavities. The carious lesions are separated by degree of caries and cavitation.

enamel clinically, and the other 24% correspond to lesions that clinically can be observed in dentin. Seventy-three per cent of the radiographic diagnoses ‘halfway or less into the enamel’ represented lesions with cavitations.

With regard to the radiographic diagnoses ‘radiolucency to the dentinoenamel junction’, 88% were lesions with visible cavitation, and 49% were directly observed to extend into the outer third of the dentin or deeper.

Most of the surfaces (84%) given score R4 radiographically had been observed to have lesions extending into dentin, and nearly all of them (93%) had cavities.

Discussion

The results confirm that the outcome of the process of caries diagnosis on radiographs cannot be predicted with certainty, not even for the deepest lesions or intact surfaces. Radiographic diagnosis of caries frequently underestimates the true lesion depth (5, 10, 11), and the opposite also occurs regularly (2, 11, 12). In the present study 12.5% of the most deeply penetrating lesions (in the inner half of the dentin) were diagnosed as sound on radiographs, whereas 15.7% of the sound surfaces were diagnosed on radiographs as extending into dentin or enamel. It is not possible to eliminate simultaneously these two different kinds of ‘errors’ (under- and over-scoring). A statistical approach based on probability distributions of the diagnoses may, however, give some general guidelines for understanding the relationship between radiographic and clinical appearance of approximal caries. The relative distribution of radiographic diagnoses for a sample of lesions is affected by the diagnostic threshold chosen—that is, the actual decision criterion based on a weighted choice of the relative value of a true positive (TP) diagnosis versus a false positive (FP) diagnosis. The relationship between TP and FP diagnoses in the present study is illustrated more completely by the ROC curves in another paper (9).

Lesions without cavitation are most likely not to be detected on radiographs: 61.3% were classified as sound in the present study (Fig. 1). The radiographic diagnoses of such lesions showed wide disparity, confirming earlier findings that they seldom are detected radiographically (1, 2, 5).

Statistically, there is a great likelihood of a break in the surface continuity when the lesion is diagnosed on radiographs as extending into the outer third of dentin. Our results indicate that this probability is greater than 0.9. This is in accordance with previous investigations in which cavity formation was found in 84.4% to 100% of lesions with a radiographic outline in the outer part of dentin (2, 4, 5, 7). Bille & Thylstrup (1) however, reported that only 52% of the lesions that on radiographs were diagnosed as extending up to halfway through the dentin had cavitation. These recordings were made clinically by different dentists during the drilling procedure and therefore might be less accurate than direct inspection of extracted teeth under standardized conditions.
It would be of particular interest to the clinician to know the relationship between clinical and radiographic lesion depth. In this study the probability that an actual enamel lesion (L2 and L3) was diagnosed on radiographs as being within enamel (R2 and R3) was 58.5%, whereas directly observed dentin lesions (L4, L5, and L6) were diagnosed on radiographs to extend into dentin (R4, R5, and R6) in 60.3% of the cases. The corresponding values reported by Bille & Thylstrup (1) were about 77% and 57%. The discrepancy between the values for enamel lesions might be due to difference in the severity of the lesions, since in the work of Bille & Thylstrup all surfaces were restored, suggesting enamel lesions of greater severity. Statistically, about 6 out of every 10 qualitative assessments of lesion depth based on radiographs are correct with regard to extent into enamel or dentin. The radiographic diagnoses that did not correspond to the validation criterion were mostly under­ scorings. Histologic and microradiographic validation of the radiographic technique confirms the tendency to underestimation (7, 13).

The frequencies of true positive scores (TP) and false positive (FP) scores were higher in this study than in previous inves­ tigations (1, 2, 5). The explanation could be that different diagnostic criteria were used by the observers. From the signal detection theory it is known that the numbers of TP scores and FP scores are positively correlated in detection experiments (14). If one observer wants to increase the number of TP scores, this is usually accompanied by an increase in FP scores. For radiographic caries diagnosis no specific diagnostic confidence about presence or absence of caries or decision criterion can be used to distinguish clearly between sound and carious surfaces. The choice of an appropriate strategy (lax, moderate, or strict) in radiographic caries diagnoses will be influenced by the consequences of the outcome. To avoid extremely strict or lax diagnostic threshold, pooled diagnoses from seven observers were used in this study. The frequency of false­positive scores (FP) ranged from 0% (strict observer) to 42.9% (lax observer), and it is obvious that an investigation of the reliability of the radiographic technique should not be based on one observer, who could be an extreme one.

On the basis of these pooled scores it could be concluded that the radiographic outline of approximal carious lesions is fairly closely related to the depth of the lesions. The deeper the lesion penetrates, the higher is the detectability. A lesion that involves enamel and dentin and has a cavity on the surface has a consistently high radiographic detectability. The radiographic technique will always be subject to errors. More attention should be given to false positive errors, since caries prevalence in the population has decreased (15).

References

Received for publication 6 June 1985
PAPER III
Diagnostic quality and observer variation in radiographic diagnoses of approximal caries

Ivar Espelid and Anne Bjørg Tveit
Department of Pedodontics, Department of Oral Radiology, and Department of Cariology and Endodontics, School of Dentistry, University of Bergen, Bergen, Norway


The aim of the study was to compare observer variations, observer strategy, and diagnostic quality with regard to radiographic caries diagnoses made at different depth levels. Qualitative assessment of approximal carious lesions on the basis of radiographs were made for sound (n = 28) and carious (n = 123) lesions by seven dentists. The inter- and intra-observer variations were lowest when lesions were diagnosed as being in the outermost parts of the teeth. The frequency of false positive scores was lower when dentin was examined as compared with enamel. The quality of radiographic diagnoses showed small variations (p < 0.05) when different levels of pulpal depths were interpreted. Denial caries: observer performance, receiver operating characteristic (ROC) analysis.

Ivar Espelid, Department of Pedodontics, School of Dentistry, Arstadveien 17, N-5044 Bergen, Norway

In planning treatment of approximal caries the dentist must consider the extent of the lesion and determine the state of the lesion as progressing or arrested. In most instances treatment decisions are based only on the radiographic outlining (1). Although the probability of radiographic detection increases with the depth of the lesion or mineral loss in lesion (2, 3) the quality of diagnoses made at different penetration levels is not known. Radiographic estimates of depth have been assessed against different validating norms such as histological appearance of lesions and the extent of tissue changes as observed in prepared cavities, and the results indicated that the radiographic technique provides a net underestimation of lesion penetration (1, 2, 4–6). However, changes in the relative number of true positive or false positive radiographic recordings do not necessarily reflect variation in diagnostic quality but may indicate differences in diagnostic strategies or criteria (7). The receiver operating characteristic (ROC) method provides the most reliable estimates of diagnostic quality and is free of bias due to changes in the diagnostic criteria of observers (8).

Grøndahl (9) has shown that observers easily change their diagnostic criteria. This can be triggered by giving them new information about the cases to be interpreted. It is reasonable to assume that observer performance might alter if the treatment outcome of a diagnostic decision changes as well. A lesion confined to outer enamel and one penetrating to dentin are usually deemed to require different treatments, such as a fluoride application in the first instance and usually a filling in the second case. The first choice leaves a chance for the lesion to become arrested or even remineralized, whereas restorative treatment is irreversible. False-positive errors may therefore occur more frequently in enamel than in dentin because the consequences of overdose diagnoses are less serious.

The aims of the present investigation were (a) to study the inter- and intra-observer variation in radiographic diagnoses of approximal carious lesions; (b) to examine whether the diagnostic strategy changes when observers diagnose at different depths; and (c) to compare the quality of radiographic caries diagnoses made at different depths.
Materials and methods

The material consisted of 105 extracted premolars and 46 extracted molars. The radiographs were taken under conditions similar to bitewing radiographs, but after extraction. The material is more fully described in another report (10). Seven dentists interpreted the radiographs under standardized conditions, using an X-ray viewer with a \( \times 2 \) magnifying lens. They classified the radiographic images of lesions at five levels (R2–R6): A = not more than halfway into enamel; B = up to the dentinoenamel junction; C = into the outer third of the dentin; D = between the outer third and halfway into the dentin; and E = more than halfway into the dentin. The observers used a five-point confidence rating scale to express their level of confidence that a lesion was present: 1 = lesion not present; 2 = lesion probably not present; 3 = equal chance of being present or not; 4 = lesion probably present; and 5 = lesion present. One of the investigators was present during the diagnostic sessions and noted the scores. At least 3 weeks later the interpretations were repeated by three observers, to estimate intraexaminer variance. By probing and visual examination of the teeth, the approximal surfaces were classified as being sound (\( n = 28 \)) or having caries (\( n = 123 \)) (joint decision). To record the extent of the carious lesions in a pulpal direction, a Class II cavity was prepared. Visible color changes of the dental hard tissues and/or reduced hardness detected by probing were used as evidence of caries. The distribution of lesions by extent of caries is given in Table 1; the classification used (L2–L6) corresponds to the radiographic scores (R2–R6). A description of statistical methods used is included in the Appendix.

Results

Observer variation

The inter- and intra-observer variations are described by indices in Tables 2 and 3, respectively. Inter- and intra-observer agreement was greater when the outer levels (A and B) were diagnosed compared with the inner levels (D and E).

Diagnostic strategies

The operational points on ROC curves in Fig. 1 indicate that a shift in diagnostic decision strategies took place as deeper levels were diagnosed. In Table 4 are given the \( Z_k \) values of the operational points for the cumulated scores of the most confident ratings that a lesion was present, namely scores 4 and 5. The \( \beta \) values in Table 4 reflect that the observers consider the relative cost of additional false positive (FP)
Table 4. Diagnostic strategies at different recording levels of approximal surface, characterized by the cut-off point and the corresponding slope (β) of the ROC curve. The most confident scores that a lesion was present (scores 4+5) were used. The cut-off point is represented by the normal deviate value of the false-positive ratio, \( Z_k \). Standard error, \( SE_{Z_k} \). Pooled data from seven observers.

<table>
<thead>
<tr>
<th>Recording level</th>
<th>( Z_k )</th>
<th>( SE_{Z_k} )</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-1.030_{A}</td>
<td>0.099</td>
<td>0.54</td>
</tr>
<tr>
<td>B</td>
<td>-1.105_{B}</td>
<td>0.070</td>
<td>1.28</td>
</tr>
<tr>
<td>C</td>
<td>-1.437_{C}</td>
<td>0.071</td>
<td>2.47</td>
</tr>
<tr>
<td>D</td>
<td>-1.658_{D}</td>
<td>0.068</td>
<td>3.92</td>
</tr>
<tr>
<td>E</td>
<td>-2.134_{E}</td>
<td>0.095</td>
<td>6.87</td>
</tr>
</tbody>
</table>

Table 5. Radiographic overestimation (one depth level) for different types of lesions as compared with clinical registrations in opened, drilled cavities. Confidence scores 4 and 5 were used as positive caries diagnoses. Pooled data from seven observers.

<table>
<thead>
<tr>
<th>True lesion penetration</th>
<th>Overestimation one level of depth (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2(A)</td>
<td>18.9 (into B)</td>
</tr>
<tr>
<td>L3(B)</td>
<td>13.0 (into C)</td>
</tr>
<tr>
<td>L4(C)</td>
<td>10.2 (into D)</td>
</tr>
<tr>
<td>L5(D)</td>
<td>5.5 (into E)</td>
</tr>
</tbody>
</table>

The diagnostic strategy used by the observers is also reflected by the relative distribution of ratings at different levels of depth. Figs. 2 and 3 show that the observers were most confident about a lesion being present in the outermost parts of the tooth.

Diagnostic quality

The number of lesions clinically recorded within each level (L2-L6) is shown in Table 1. ROC curves for the radiographic diagnoses made at different levels are shown in Fig. 1. The validating criteria were the direct inspection and probing within the corresponding level. The areas beneath ROC
RADIOGRAPHIC DIAGNOSES OF
SOUND SURFACES

Fig. 2. Distribution of radiographic scores for sound surfaces and levels without caries penetrating up to the actual level. The scores from 1 to 5 represent a five-graded confidence rating about the presence or absence of caries judged radiographically. Score 1 denotes the most confident score that the surface was sound, and score 5 represents the contrary opinion.

RADIOGRAPHIC DIAGNOSES OF
CARIOS SURFACES

Fig. 3. The radiographic score distribution for different levels that actually had a carious lesion penetrating to this depth or deeper. See Fig. 2 legend for explanation of the five-graded rating.
Extent of approximal caries

Table 6. Caries predictive values of radiographic diagnoses with different confidence scores that lesion is present. Diagnoses were made at different depths

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>Pulpal depth (radiographic)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>0.473</td>
</tr>
<tr>
<td>2</td>
<td>0.550</td>
</tr>
<tr>
<td>3</td>
<td>0.588</td>
</tr>
<tr>
<td>4</td>
<td>0.802</td>
</tr>
<tr>
<td>5</td>
<td>0.983</td>
</tr>
</tbody>
</table>

Table 7. Area beneath ROC curves (A<sub>z</sub>) indicating the quality of caries diagnoses made at different depths. Mean values from seven observers

<table>
<thead>
<tr>
<th>Radiographic level</th>
<th>A&lt;sub&gt;z&lt;/sub&gt;</th>
<th>SE&lt;sub&gt;A&lt;sub&gt;z&lt;/sub&gt;&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (R2)</td>
<td>0.903&lt;sub&gt;NS&lt;/sub&gt;</td>
<td>0.033</td>
</tr>
<tr>
<td>B (R3)</td>
<td>0.884&lt;sub&gt;NS&lt;/sub&gt;</td>
<td>0.031</td>
</tr>
<tr>
<td>C (R4)</td>
<td>0.867&lt;sub&gt;NS&lt;/sub&gt;</td>
<td>0.035</td>
</tr>
<tr>
<td>D (R5)</td>
<td>0.879&lt;sub&gt;NS&lt;/sub&gt;</td>
<td>0.049</td>
</tr>
<tr>
<td>E (R6)</td>
<td>0.860&lt;sub&gt;NS&lt;/sub&gt;</td>
<td>0.135</td>
</tr>
</tbody>
</table>

curves, which represent the diagnostic quality at different levels, are listed in Table 7. None of the differences between areas were statistically significant (p>0.05). However, the quality of the radiographic caries diagnosis was strongly associated with extent of caries. The area beneath ROC curves increased as deeper lesions were diagnosed, except for level E (Table 8). The sound surfaces served as negative controls for all lesions independent of lesion depth.

Table 8. Diagnostic quality indicated by area beneath ROC curve, A<sub>z</sub>, for lesions of different depths as observed in opened, drilled cavities. Radiographic diagnoses were made at the respective levels. Pooled data from seven observers. Standard errors, SE<sub>A<sub>z</sub></sub>, of the maximum-likelihood estimates are given in parentheses

<table>
<thead>
<tr>
<th>Lesion type according to clinical depth</th>
<th>Radiographic recording level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (L2)</td>
<td>A (R2)</td>
</tr>
<tr>
<td>B (L3)</td>
<td>0.734 (0.026)</td>
</tr>
<tr>
<td>C (L4)</td>
<td>0.936 (0.011)</td>
</tr>
<tr>
<td>D (L5)</td>
<td>0.998*</td>
</tr>
<tr>
<td>E (L6)</td>
<td>0.921 (0.033)</td>
</tr>
</tbody>
</table>

* The SE<sub>A<sub>z</sub></sub> value could not be calculated because the frequency of usage of some of the response categories was too small for maximum-likelihood calculation. The given A<sub>z</sub> value is a least square estimate.

Discussion

Inter- and intra-observer variation differed with examined level, and both were lowest in the outer part of enamel, as indicated by the reliability coefficients and Kendall's tau in Tables 2 and 3. However, Scott's pi showed the inverse tendency and thus indicated that the frequency of equal scores was highest at the inner levels. Score 1 was used very often. This could be explained by the fact that caries prevalence was lower and that observers applied a stricter diagnostic threshold at this level. Scott's pi tells nothing about the quantitative difference between unequal scores. Kendall's tau is a ranking test that takes into account the relative values but ignores the absolute differences between two scores. In the reliability coefficient the absolute differences are squared; thus this coefficient is a measure of the absolute differences between the scores.

Clinical reasoning seemed to affect observer strategy in radiographic caries diagnosis. The cut-off points gradually moved toward a stricter diagnostic threshold as deeper parts of the tooth were examined (Table 4, Fig. 1). A reasonable explanation is that the cost of diagnostic errors is weighted differently in restorative treatment decisions and decisions about preventive therapy. The observers seemed, whether consciously or not, to apply a stricter diagnostic criterion when the outcome of the decision was irreversible treatment. A reduction in caries prevalence will reduce the predictive value of the caries diagnosis when the criterion for
diagnosis is kept unchanged. The decreasing prevalence of caries in deeper levels compared with the outermost part is an underlying factor that probably also influences observer performance. However, the caries predictive values (Table 6) indicate that the change in diagnostic criteria due to decreasing prevalence resulted in an undercompensation. The diagnosing of caries was most efficient in outer enamel and decreased monotonically toward the pulp. From a statistical point of view, the most reliable caries diagnoses in this study were made in the outermost parts of the tooth.

The quality of radiographic interpretations did change with the depths to be diagnosed, but the differences were not statistically significant (p>0.05) (Table 7). However, a total number of about 190 cases are necessary to demonstrate whether the difference between the areas of 0.90 and 0.85 is statistically significant (11). The area beneath an ROC curve may represent the probability that a randomly chosen pair of sound and carious approximal surfaces is correctly separated. This area is intimately connected with the quantity obtained by the nonparametric statistical Wilcoxon or Mann-Whitney tests performed on the scores (11). The differences between ROC curves may be interpreted in various ways. From a clinical point of view calculation of the equivocal response rate might be of interest (the frequency of questionable or borderline cases). Traditionally, dentists have used a binary diagnostic grading for caries detection, namely caries present or not. However, in some cases the diagnostic decision is made under great uncertainty, with an equal chance of a positive or a negative diagnosis. On the assumption that the sum of erroneously and correctly classified surfaces is 100% minus the questionable lesions, the number of questionable lesions can be calculated for both sound and carious surfaces (8). We require the upper limits for the two kinds of decision errors—that is, FP and FN errors. If we choose a 10% upper limit for errors on the ROC curves in Fig. 1, we find the proportions of questionable surfaces (averaged for sound and carious surfaces) to be 28% in level A, 27% in B, 29% in C, 31% in D, and 40% in E. This means that more than 3 of the diagnoses made at levels A to D and about 2 of the diagnoses made at level E were definitive.

The curve that describes observer performance at level A is considerably slacker than the other ROC curves in Fig. 1. This suggests that if the diagnostic criteria are changed from a strict threshold toward a laxer one, the rate of FP errors would increase more in outer enamel than in the other levels. Thus if a very lax criterion is applied, the curve for level A will intersect the other at an FP ratio greater than 0.4-0.5. The TP ratio will now be lower in A than in the other levels.

The probability of lesion detection varied with the true depth of lesions. This means that the probability of a positive radiographic caries diagnosis was greater for a lesion that clinically penetrated the dentin than for a lesion that was confined to enamel only. This is confirmed by others (6), but their areas beneath ROC curves were consistently lower. This may be due to the severity of cases selected, the method for validation (they used histological examination), or the method for calculating ROC area. According to Swets & Pickett (8), a linear ROC estimate may give somewhat larger ROC area than a computation based on empirical values when the actual operational points are not well spread across the ROC space.

It can be stated that (a) both the inter- and intra-observer variations for radiographic caries diagnoses were greatest when the level closest to the pulp was diagnosed; (b) the superficial levels were diagnosed with laxer diagnostic criteria than the deeper levels, and (c) the quality of radiographic diagnoses showed small variations between levels.

References

3. Espelid I, Tveit AB. Radiographic diagnosis of min-

Appendix: Statistical methods

Observer variation

The observer variation is represented by three different indices on the basis of correlation between ratings within the actual levels (intraobserver variation) and correlation between independently made repeated ratings within each level. The reliability coefficient was calculated in accordance with Winer (12):

\[ r_k = 1 - \frac{MS_{wr}}{MS_{br}}. \]

where \( MS_{wr} \) = mean squared value observed when one observer rereads a set of cases independently, and \( MS_{br} \) = mean squared value observed when a set of observers read a set of cases once.

This parametric estimate of observer variation was calculated by means of the statistical software package BMDP (13). The non-parametric correlation coefficients Scott's pi (14) and Kendall's tau (13) were also calculated.

Diagnostic strategies

The choice of diagnostic decision strategy was analyzed with the two indices \( Z_k \) and \( \beta \) (8). \( Z_k \) is the frequency of overscoring at the operating point, given in normal deviate values from the bivariate normally distributed ROC graph. The value of \( \beta \) is the slope of a tangent to the ROC curve at the operational point. The \( \beta \) reflects the observer's evaluation of the cost of additional false positive (FP) scores compared with the cost of additional false negative (FN) scores (15). The computational formula is as follows:

\[ \beta = B \exp\{[(1 - B^2)Z_k^2 + 2ABZ_k] - A^2\}, \]

where \( A \) is the Y intercept of the bivariate normally distributed ROC curve, and \( B \) is the slope of the bivariate normally distributed ROC curve (8). All parameters used in these calculations were obtained from the maximum-likelihood estimates from the output of a computer program developed by Dorfman & Alf (16). The calculations are based on pooled ratings from all observers.

The caries predictive values of the scores were calculated by Bayes' theorem as given by Weinstein et al. (17).
Diagnostic quality

The observer ratings were analyzed by the receiver operating characteristic (ROC) technique. The area beneath the ROC curve was used as a measure of diagnostic quality. The method is described by Swets & Pickett (8). The computer program developed by Dorfman & Alf (16) was used to calculate maximum-likelihood estimates of ROC area and other parameters of performance. Statistical analyses of paired observations were applied in accordance with Hanley & McNeil (18) and McNeil & Hanley (19). This method provides increased statistical power compared with settings with different cases and/or observers for each interpretation. The testing of the difference between diagnostic quality at various levels was conducted by calculating the z value according to the equation:

\[ z = \frac{A_{z(dif)}}{SE_{z(dif)}} \]

where \( A_{z(dif)} \) = the difference between areas under two ROC curves, and \( SE_{z(dif)} \) = the standard error of the difference.

The ratio was referred to a table of area under the normally distributed curve for verification or rejection of the null hypothesis (rejection if alpha>0.05). The standard error for the difference is a summation of the standard errors due to the case-sample (c), between-reader (br), and within-reader (wr) variability. In all settings the readers and the teeth were the same, and the covariance between observations was therefore taken into consideration. The computational formula as given in equation 5 in Swets & Pickett (8) was as follows:

\[ SE_{z(dif)} = 2 \left[ S_{c+wr}^2 (1 - r_{c-wr}) + \frac{S_{br+wr}^2}{1 - r_{br-wr}} \right] \]

where \( S_{c+wr}^2 = S_{c}^2 + S_{wr}^2 \) is the variance observed in area when one observer reads each of a set of different case samples once; \( S_{br+wr}^2 = S_{br}^2 + S_{wr}^2 \) is the variance observed in area when one case sample is read once by each observer; \( S_{wr}^2 \) is the variance observed in area when one observer reads one case sample on two or more independent occasions; \( r_{br-wr} \) is the correlation observed between areas when different observers read the same case sample in the five settings; \( r_{c-wr} \) is the correlation observed between areas when one observer reads a set of subsamples in the five settings; and \( l \) denotes the number of readers.

The calculations were based on the \( S_{c+wr}^2 \) value obtained from the output of the Dorfman & Alf computer program averaged across readers. The estimates of \( S_{br+wr}^2, S_{wr}^2, \) and \( r_{br-wr} \) were calculated by the method of Swets & Pickett (8). The correlation coefficient \( r_{c-wr} \) was estimated as described by Hanley & McNeil (18), referring Kendall's tau correlation coefficient between ratings to the tables in this paper.

---

Received for publication 6 June 1985
The influence of viewing conditions on observer performance in dental radiology

Ivar Espelid
Department of Pedodontics, School of Dentistry, University of Bergen, Bergen, Norway


Observer performance under different viewing conditions was monitored by means of radiographs, showing (a) teeth with approximal caries, (b) teeth with simulated approximal caries, and (c) Plexiglas phantoms. Series a and c were exposed at two different exposure times, thus obtaining a dark and a light set of radiographs. The series a, b, and c were read by 11, 2, and 5 observers, respectively. All radiographs were examined under two viewing conditions on different occasions as follows: ceiling light in a room without a window, and X-ray viewer (×2 magnification) with two alternative light sources. Diagnostic quality, measured as the ROC area, showed small variations in accordance with viewing conditions. For dark radiographs the X-ray viewer improved diagnostic quality compared with the viewing against ceiling illumination; however, room illumination provided the best viewing conditions when light radiographs were examined. Most of these differences were not statistically significant (p > 0.05). Only small differences in diagnostic strategies (cut-off points) were recorded between viewing conditions. □ Dental caries: receiver operating characteristic (ROC) analysis; visual perception

Ivar Espelid, Department of Pedodontics, School of Dentistry, Årstadveien 17, N-5009 Bergen, Norway

Textbooks in dental radiology emphasize the importance of optimal viewing conditions during the examination of dental radiographs (1–3). Many papers that have focused on radiographic film interpretation conclude that viewing conditions may be of significance to visual detection (4–11). However, most frequently phantom lesions radiographed on extraoral (5, 9, 10, 12) or intraoral film format (13, 14) have been used in studies of viewing conditions. Only a few studies have used dental caries or artificially created dental lesions as test pathological material, and the results from these studies do not confirm that viewing conditions play a crucial role in diagnosis of caries on radiographs (15, 16).

According to Leijon (17), the viewing of dental radiographs should be performed under the following conditions: adjustable viewing light; mounted films; masking for extraneous light; film reflexes should be avoided; and a ×2 magnifying lens should be used. However, it is reasonable to assume that these requirements are not met in most dental practices, although most dentists seem to have a viewing box (18, 19).

Welander et al. (14) claim that clinicians who view radiographs under less than optimal viewing conditions ‘are compromising their diagnostic efficiency’. In this study the perceptibility, which was defined as the number of phantom radiolucencies recorded (20), was improved when using masked frames or a viewer. Viewing against the ceiling light or a viewing box with no masking of extraneous light decreased the perceptibility.

Some data indicate that density of films processed in dental practice may vary considerably even within the same practice (21). It has been shown that the density of radiographs affects interpretation and that radiographs of low density seem to require other viewing conditions than high-density ones (14, 16).

There is a need for consistent criteria for diagnosis and therapy among dentists (22). Variation in viewing conditions and film density may affect the application of diagnostic
criteria and the quality of radiographic diagnosis. This study was carried out to measure observer performance and to quantify diagnostic quality when different viewing conditions and film densities were used.

Materials and methods

Observer performance under two viewing conditions was assessed in this study. The first condition consisted of using the ceiling light in a room without a window. The average level of illumination at the plane was about 215 lx (Hagner photometer). The second alternative consisted of using two alternative viewboxes with luminance of about 4800 lambert and 100,000 lambert (Phillips Densoscope). Radiographs were read in a darkened room, and the observers were asked to use the low-intensity viewbox routinely, and Densoscope when more intense light was necessary. A viewer with a ×2 magnification lens (X-Produkter) was used by the observers under the second condition.

Three series of radiographs with different types of images were examined under the two viewing conditions with at least 2 weeks’ interval.

Radiographs of 60 approximal surfaces constituted series 1. On the basis of direct inspection and probing the surfaces were classified as sound \( (n = 23) \) or carious \( (n = 37) \). When a break in the outer enamel was detected, the surface was classified as carious, otherwise sound. The films in series 1 were interpreted by 11 dentists.

Series 2 consisted of 236 radiographs of 13 extracted teeth that originally had no caries or defects. By using a diamond bur (diameter, 1 mm) artificial lesions extending to the dentinoenamel junction were made in the most prominent part of the approximal surfaces. Several exposures were made of each tooth, and the relative mineral loss of the lesions was changed between each exposure by a technique more fully described elsewhere (23). Radiographs were also taken before preparation of lesions, and these images served as negative controls \( (n = 109) \). Two dentists diagnosed series 2.

Series 3 contained radiographs of 25 Plexiglas phantoms. The phantoms were square blocks measuring \( 30 \times 40 \text{ mm}^2 \) and 35 mm high. Holes or cylinders were randomly placed in up to four fixed positions on these blocks. Thus the number of ‘signals’ or phantom lesions varied from 0 to 4 on each film. The radiographs \( (n = 25) \) contained 20 sites with images of holes and 29 of cylinders and 51 negative controls. The diameter of the holes and cylinders was 9 mm and the depth/height ranged from \(-9 \text{ mm} \) to \(+12 \text{ mm} \). Series 3 was interpreted by five observers.

Series 1 and 3 contained two sets of radiographs showing the same images. The only difference between the sets was film density, which varied with exposure time. Thus two series, which are termed ‘light’ or ‘dark’ in the text, were obtained.

Density measurements were performed with a Macbeth TD502 densitometer, and the film density of tooth images was measured just in dentin, just beyond the dentinoenamel junction. A positive D-value denotes a radiolucent phantom lesion, and a negative value a radiopaque phantom lesion. Radiographs in series 1 had a mean density of 0.5 and 1.1 for the light and dark images, respectively. In series 2 the corresponding density was 1.3. The phantoms in series 3 had mean background densities of 0.7 (light pictures) and 1.9 (dark pictures). The density values for the relative object contrast ranged from 0.00 to \(-0.05 \) (median, \(-0.02 \)) and from 0.00 to 0.04 (median, 0.01) for the light pictures. For the dark radiographs the corresponding values ranged from \(-0.01 \) to \(-0.23 \) (median, \(-0.07 \)) and from 0.01 to 0.13 (median, 0.03).

Analytical methods

The confidence ratings were sorted into decision matrices in accordance with the true state of the lesions to be diagnosed. The findings were then analyzed in accordance with the receiver operating characteristic (ROC) technique as described by Swets & Pickett (24), using a computer program developed by Dorfman & Alf (25), which fitted binormal ROC curves to the pooled observer scores by maximum likelihood estimation. The diagnostic quality was measured
by using the area beneath ROC curves, $A_z$, and diagnostic strategies were analyzed by using the parameters $Z_k$ and $\beta$. $Z_k$ is the frequency of overscoring at the operating point, given in normal deviate values from the bivariate normally distributed ROC graph. $\beta$ denotes the slope of a tangent to the curvilinear ROC curve at the operating point. The $\beta$ value indicates the observers' relative weighting of the cost of additional false-positive (FP) scores versus additional false-negative (FN) scores (26). The computational formula applied to draw the curves of the constant $\beta$ was:

$$\beta = B \exp\{\frac{1}{2}(1 - B^2)Z_k^2 + 2ABZ_k - A^2\}.$$
Table 1. Quality of radiographic caries diagnosis obtained under different viewing conditions given by the area beneath ROC curve, $A_r$, with standard error in parentheses. The calculations were based on pooled scores from 11 dentists. NS denotes no statistically significant difference between viewing conditions ($p > 0.05$)

<table>
<thead>
<tr>
<th>Viewing condition</th>
<th>Room lighting $A_r$</th>
<th>Viewer $A_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dark radiographs ($n = 60$)</td>
<td>0.919 (0.014)</td>
<td>NS 0.923 (0.011)</td>
</tr>
<tr>
<td>Light radiographs ($n = 60$)</td>
<td>0.836 (0.027)</td>
<td>NS 0.813 (0.024)</td>
</tr>
</tbody>
</table>

Table 2. Diagnostic quality in identification of simulated caries lesions (236 radiographs). $A_r$ denotes area beneath ROC curve, and standard error is given in parentheses. Pooled data from two dentists. NS = difference not statistically significant ($p > 0.05$)

<table>
<thead>
<tr>
<th>Viewing condition</th>
<th>Room lighting $A_r$</th>
<th>Viewer $A_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dark radiographs ($n = 25$)</td>
<td>0.909 (0.018)</td>
<td>S 0.994 (0.012)</td>
</tr>
<tr>
<td>Light radiographs ($n = 25$)</td>
<td>0.862 (0.024)</td>
<td>NS 0.791 (0.029)</td>
</tr>
</tbody>
</table>

Table 3. Efficiency of observer detection of phantom lesions in a Plexiglas phantom. $A_r$ denotes area beneath ROC curve, and standard error is given in parentheses. Pooled scores from five observers were used. S = difference statistically significant ($p < 0.001$); NS = difference not statistically significant ($p > 0.05$)

<table>
<thead>
<tr>
<th>Viewing condition</th>
<th>Room lighting $A_r$</th>
<th>Viewer $A_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dark radiographs ($n = 25$)</td>
<td>0.909 (0.018)</td>
<td>S 0.994 (0.012)</td>
</tr>
<tr>
<td>Light radiographs ($n = 25$)</td>
<td>0.862 (0.024)</td>
<td>NS 0.791 (0.029)</td>
</tr>
</tbody>
</table>

of diagnostic quality and the values are given in Tables 1–3. For the dark radiographs the X-ray viewer and a viewing box improved diagnostic quality compared with viewing against ceiling illumination (Tables 1–3). The differences were only statistically significant when Plexiglas phantoms were detected ($p < 0.001$) (Table 3). In contrast, when light radiographs were examined, room illumination provided the best viewing conditions, but the differences were not statistically significant ($p > 0.05$).

Figs. 2–6 show pairwise ROC plots representing observer performance under the viewing conditions that were compared. Figs. 2 and 3 indicate only small, statistically non-significant differences in observer performance between the viewing conditions when dental caries was diagnosed. The corresponding cut-off points did not differ significantly between viewing conditions ($p > 0.05$). When different film densities were compared, the position of the cut-off points indicated that a stricter diagnostic threshold was applied when the light radiographs were interpreted (Figs. 2 and 3). This tendency was statistically significant in most comparisons between corresponding curve points ($p < 0.05$). Fig. 2 shows that the cost of additional FP scores is valued more than five times the cost of additional FN scores when light radiographs are interpreted by using an X-ray viewer. This relative weighting between additional FP and FN scores decreased to about three when ceiling illumination was used. When dark radiographs were interpreted, this value was about three under both viewing conditions (Fig. 3). Scores 4 and 5 were pooled and used as the actual cut-off point of most clinical relevance in these comparisons.

Discussion

This study indicates that viewing conditions are not critical for the quality of radiographic caries diagnosis, although the importance of masking and/or magnification is usually stressed in the literature (1–4, 7, 8, 13, 27). The present results did not confirm that an X-ray viewer with an X2 magnifying lens always provides optimal viewing conditions. However, when the film is relatively dark, it was shown that masking of extraneous light might be important. An appropriate viewing condition for clinical use seems to be adjustable viewing light in combination with a
Fig. 2. Binormally fitted ROC curves for comparison of radiographic caries diagnosis under two different viewing conditions using light radiographs (series I). Fig. 1 legend gives more detailed information about the ROC method. The $A_t$ values are given in Table 1.

Fig. 3. Observer performance when caries is diagnosed radiographically, using fairly well blackened films ('clinical' judgement), is indicated by one binormal ROC curve for each setting. Fig. 1 legend gives additional information; see also Table 1.
viewing box for masking and magnification purposes. The light transmitted through the film should be brought to a level that is best perceived by the eye (3). This may explain the fact that overhead room light of low intensity turned out to be at least as effective as an X-ray viewer and high-intensity illumination when low-density radiographs were interpreted. These results are partly confirmed by other investigators (12, 16).

Merrild-Hansen & Ratjen (12) used simulated defects, which were radiographed on 24 × 30 cm² films. The films were examined at a distance of 50 cm by four observers. Within the film densities examined (relatively light) and luminance used (540 to 10,800 lambert), they concluded that glare was of little importance.

Arnold (16) examined the influence of illumination and the use of accessories on radiographic detection of drilled approximal defects in tooth-like phantoms. He used 10 observers. He concluded that it might be an advantage to interpret light radiographs against the light of a window, which is in accordance with the findings of the present study. His general conclusion was that the best viewing conditions were provided in a semi-darkened room with a viewbox with high illumination and good masking against scattered light and that variation in illumination conditions had only small influence on detection. On the other hand, he found that a magnifying glass (×5) or an X-viewer (×2) produced a significant improvement in detection.

Mileman (15) confirmed that the effect of different viewing illumination levels was of minor clinical importance. His study was based on bitewings that were interpreted by 12 dental teachers. Inter- and intra-observer variation seem to be more important sources of diagnostic variation than different viewing conditions (15, 16).

An observer’s thinking about the diagnosis and therapy might be affected by his feelings about diagnostic accuracy. For instance, if he subjectively judges the viewing conditions and/or film density not to be optimal, it is reasonable to expect that he will take precautions by consciously or unconsciously adjusting the diagnostic threshold. The results indicated that the diagnostic quality did not differ substantially and that viewing conditions did affect the relative costs that were associated with incorrect decision outcomes (FP and FN scores). The viewing against ceiling light using dark radiographs seemed to increase the relative cost of additional FN scores compared with additional FP scores whereas the opposite observer performance took place during the examination of light radiographs, as indicated in Figs. 2-6.

Detection of low-contrast lesions is affected by viewing conditions. This is demonstrated by using the test image provided by Ravindra et al. (28). Their conclusion was that extraneous light peripheral to an image decreases visual sensitivity. By creating subtle phantom lesions (series 3) a situation was created in which diagnostic quality depended on viewing conditions. However, when clinically relevant radiographs showing dental caries were examined, the effect of viewing conditions was overshadowed by film density effect. It is therefore reasonable to raise the question of how relevant it is to extrapolate results from general perception studies and make general recommendations for examination of dental radiographs, without taking into account variables like film density and type of pathologic condition.

Incipient caries in approximal enamel has a radiographic outline that is relatively often recognized by a trained observer (29). It has been shown by means of digitized radiographic displays showing approximal caries that the amount of quantum noise in the image could be increased substantially while maintaining a comparable diagnostic performance (30).

Most studies that assign a significant influence to the viewing conditions are based on phantoms with no similarity to the clinical situation (5, 9, 10, 12-14). Some of these studies deal with borderline lesions that are hardly visible. The relevance of focusing on, for instance, viewing conditions and not on the observer error, which in general plays an important part in diagnosis (31), might be questioned.

Quantitative assessment of diagnostic quality is necessary to compare observer performance under different conditions. Some
Fig. 4. Binormal ROC plots indicating observer performance for radiographic examination of simulated carious lesions under different viewing conditions. See Fig. 1 legend for further explanation: Table 2 shows an index of diagnostic quality.

Fig. 5. Phantom lesions diagnosed on light radiographs under two viewing conditions gave these binormal ROC curves. See Fig. 1 legend for further information. Table 3 gives a parameter of diagnostic quality which is based on the actual ROC curves.
studies use test patterns and monitor the number of positive registrations (5, 9, 13, 14, 16) or fraction of true-positive registrations (10). In one study consistency was measured (15). These criteria cannot separate differences between two systems, since they do not take into account changes in decision criterion and diagnostic performance. Usually, diagnosticians do not want to identify pathologic condition at any price. The price may be a lot of FP registrations. Observers might desire to minimize the errors by weighting the pay-off and balance the over- and under-scoring (24). The expected value of a decision may change in accordance with the knowledge and experience about the importance of film density and viewing conditions. By using the ROC technique, it is possible to obtain an estimate of diagnostic accuracy which is not biased by changes in diagnostic threshold (24). The ROC analysis of observer performance has been the commonest approach to studies of lesion detectability (32).

The present findings suggest that the viewing requirements needed in dental radiography depend on the subject of interest and the density of the film. Studies based on subtle and hardly visible phantom lesions do not always give reliable results that are relevant to the clinical situation.

References
PAPER V
Radiographic diagnosis of caries and marginal defects in connection with radiopaque composite fillings


Abstract The aim of this study was to find out if carious lesions and marginal defects were as easy to diagnose radiographically in connection with the radiopaque composite P-30 as with amalgam. Amalgam and P-30 Class II restorations with and without secondary caries and simulated marginal defects were made in extracted premolars. Radiographs of the teeth were examined by 11 dentists to diagnose lesions and defects. The results showed that a higher percentage of secondary carious lesions and marginal defects was detected near the radiopaque composite than near the amalgam fillings, and that the frequency of overlooking was lower in connection with P-30. This indicates that this composite compared to amalgam, has a degree of radiopacity which increases the detection rate of caries and defects adjacent to the restorations.

Material and methods

Some five extracted premolars were used in the study. The teeth had been stored in 2% sodium hypochlorite since extraction. The teeth were divided into 3 groups:

Experimental Group 1 (EG1): 25 teeth
Experimental Group 2 (EG2): 30 teeth
Control Group (CG1): 20 teeth

EG1

The teeth in this group had approximal caries lesions. One standardized Class II preparation was made in each tooth, but half of the lesion was left in the gingival wall. The preparation was made without undercuts and facilitated the removal of the fillings. The lesions appeared as secondary caries when filling material was inserted in the preparations. To secure that no experimental filling material penetrated the caries lesions, they were covered with a thin layer of the radiopaque material, with low radiopacity (Silux). Then a Class II matrix was adapted and amalgam (Amalcap) inserted and condensed by hand instruments. The teeth were mounted in standardized plaster blocks, fixed in a device which made it possible to repeat radiographic exposures with identical projections. Radiographs of each tooth were taken in a projection tangential to the approximal surface. Exposures were made at 65 kV and 15 mA using a dental X-ray machine with an electronic timer (Ritter, Explorer II). The exposure time was 1.2 s, and a focus-object distance of 32 cm and an object-film distance of 15 cm were used. Kodak (R) DF-57 (double-pack) film was used. All radio-
graphs were processed by a standardized procedure.

The amalgam fillings were then removed, without disturbing the gingival wall. Pilot studies had shown that this could be achieved by drilling a groove in the proximal part of the filling and then levering the rest of the amalgam out of the cavity.

P-30 was inserted in the same cavities, following the manufacturer’s instructions, but no acid etching was performed prior to inserting the composite. Radiographs were made, using the procedure described above.

EG2

This group contained teeth with sound approximal surfaces. One standardized Class II preparation was made in each tooth as for the teeth in EG1. Before inserting the amalgam, about 0.5 mm thick layer of composite (Silux) was applied in the gingival wall. This layer was to illustrate a marginal defect on the radiographs. In these cases, acid etching was performed before placement of the layer of composite with low radiopacity. Following insertion of the amalgam the teeth were mounted and radiographed as described above. The amalgam fillings were removed, following the same procedure as described before, and without touching the composite material in the gingival wall. The P-30 Class II fillings were placed and radiographed made as for the teeth in EG1. In this way, the secondary caries lesions in EG1 and the “marginal defects” in EG2 were identical whether the teeth were radiographed with amalgam or composite material.

CG

Twenty teeth with sound approximal surfaces constituted the control group. One Class II preparation was made in each tooth. The insertion of amalgam and the P-30 composite material and the radiographic procedure were the same as for the other 2 groups, but the control teeth had no lesion or marginal defect in connection with the fillings.

Diagnostic procedure

The radiographs were examined by 10 experienced dentists, using a standardized illumination source and a viewing box with 2x magnifying lens. The observers had no time restrictions. The dentists diagnosed “lesions” and “defects” according to a 5-point confidence scale:

Caries diagnosis:

1=caries definitely or almost definitely not present
2=caries probably not present
3=caries possibly present
4=caries probably present
5=caries definitely or almost definitely present

Diagnosis of simulated marginal defects:

1=marginal defect definitely or almost definitely not present
2=marginal defect probably not present
3=marginal defect possibly present
4=marginal defect probably present
5=marginal defect definitely or almost definitely present

The scoring was classified into a decision matrix and treated according to the ROC method (p. 7) and to Bayesian statistics (16). Dichotomized scores were used for calculations of sensitivity (fraction of true positive cases), specificity (fraction of true negative cases) and Bayesian probabilities (predictive value). Scores 4 and 5 were condensed and read as positive caries diagnoses. Scores 1, 2 and 1 were read as negative caries diagnoses. For further details about the procedure, see Espeland & Bent, (9). The statistical testing of paired samples (sensitivity and specificity values) was performed using t-test.

Results

The radiopacities of the composite fillings and the corresponding amalgam fillings are illustrated in Fig. 1. All composite fillings appeared homogenous on the radiographs. Visually, the radiopacity of the composite fillings exceeded that of enameled.

A higher sensitivity, i.e. a higher proportion of true positive scores (TP), was found for all but 2 observers diagnosing secondary caries lesions in connection with P-30 (Fig. 2). The specificity, i.e. the proportion of true negative scores (TN) was also higher in the composite group compared to the amalgam group for all but 2 observers (Fig. 2).

When pooling the scores from all observers, the results showed that a higher percentage of the secondary caries lesions was detected with high degree of confidence (Scores 4 & 5) near the radiopaque composite than near the amalgam fillings (Table 1). In addition, the fillings without secondary caries or simulated defects were more often diagnosed correctly in the composite group (Tables 1 & 2), which means that the frequency of false positive scores (FP) was lower in connection with P-30.

The diagnoses of simulated marginal defects showed the same tendency as for the diagnoses of secondary caries: the observers showed a higher sensitivity in the composite group compared to the amalgam group (Fig 3). In other words, the frequency of FP scores was higher and rate of FN scores was lower with regard to the composite. Pooled scores for simulated marginal defects are given in Table 2.

The quality of the diagnoses was measured by means of the respective areas under the ROC curves. The areas which are listed in Table 3, indicate that the diagnoses of secondary caries in the composite group was better than in the amalgam group. The same tendency was true for diagnoses of simulated marginal defects (Table 3). Figs. 4 & 5 show how the prevalence of secondary caries and marginal defects affects the positive prediction value. The curves which are based on Bayes‘ theorem illustrate a theoretical model based on fixed TP and FP values. The
Fig. 2 Radiographic diagnosis of secondary caries associated with P-30 and amalgam restorations. The points represent the values for sensitivity and specificity for the observers. The filled circle represents 3 observers. The differences between P-30 and amalgam were statistically significant with respect to sensitivity (p < 0.05) but not for specificity (p > 0.05).

Fig. 3 Radiographic diagnosis of marginal defects associated with P-30 and amalgam restorations. The points represent the values for sensitivity and specificity for the observers. The filled circle represents 3 observers, the X, 2 observers. The differences between P-30 and amalgam were statistically significant with respect to sensitivity (p < 0.05) but not for specificity (p > 0.05).

Table 1. Diagnosis of secondary caries. Distribution of confidence ratings according to filling material. Increasing confidence of caries being present from score 1 to 5. Ten observers.

<table>
<thead>
<tr>
<th>Filler</th>
<th>Confidence rating</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-30</td>
<td>(n)</td>
<td>42</td>
<td>35</td>
<td>41</td>
<td>73</td>
<td>50</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>(percent)</td>
<td>16.8</td>
<td>14.0</td>
<td>16.4</td>
<td>29.2</td>
<td>23.6</td>
<td>100.0</td>
</tr>
<tr>
<td>Amalgam</td>
<td>(n)</td>
<td>41</td>
<td>49</td>
<td>29</td>
<td>51</td>
<td>60</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>(percent)</td>
<td>16.4</td>
<td>19.6</td>
<td>11.6</td>
<td>20.4</td>
<td>24.0</td>
<td>100.0</td>
</tr>
<tr>
<td>No caries</td>
<td>(n)</td>
<td>150</td>
<td>19</td>
<td>17</td>
<td>11</td>
<td>3</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>(percent)</td>
<td>75.0</td>
<td>9.5</td>
<td>8.5</td>
<td>5.5</td>
<td>1.5</td>
<td>100.0</td>
</tr>
<tr>
<td>Amalgam</td>
<td>(n)</td>
<td>141</td>
<td>30</td>
<td>12</td>
<td>19</td>
<td>2</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>(percent)</td>
<td>65.5</td>
<td>15.0</td>
<td>6.0</td>
<td>9.5</td>
<td>1.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Discussion

The main reason for replacing Class II fillings is secondary caries (10, 11). The majority of defects and caries in connection with such fillings is found on the proximal surfaces especially the gingival part (12-14).

Intraoral radiographs are of help in detecting caries and in evaluating the success for failure of any radiopaque filling placed on proximal surfaces. High radiopacity of a filling material has been considered as a prerequisite for adequate diagnosis of caries, defects and overhangs in association with Class II fillings. A general opinion among practitioners is that the higher the contrast between a filling material and the adjacent dental hard tissue, the more efficient is the radiographic diagnosis. This assumption is not supported by the present results.

A higher frequency of both under and over-scoring of secondary caries and simulated marginal defects was found in connection with amalgam compared to the radiopaque composite. This indicates that the composite compared to amalgam, has a degree of radiopacity which increases the detection rate of caries and defects adjacent to the restorations.

The fact that a moderate radiopacity of a restorative material might be favorable, could be explained by the Mach effect (15), an increase in the perceived density near the boundary of areas which differ widely in density. As a consequence caries lesions or marginal defects may be over-diagnosed near fillings with high radiopacity. Another explanation might be that caries near the lingual and buccal margin of a Class II restoration will be masked by a highly radiopaque filling material, but be detected when a material of moderate radiopacity is used. Theoretically, a moderate radiopacity might be preferable to a high radiopacity of a material due to the masking effect of the latter. However, further investigations are needed to confirm this hypothesis.

In a study of a two-component radiopaque composite and diagnosis of simulated defects filled with wax, a high proportion of over-diagnoses was found (16). This was in contrast to the present results and may be due to the heterogeneity of the two-component
Table 2. Diagnosis of marginal defects. Distribution of confidence ratings according to filling material. Increasing confidence of defects being present from score 1 to 5. Ten observers.

<table>
<thead>
<tr>
<th>Marginal defect</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(percent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amalgam</td>
<td>2.5</td>
<td>9.0</td>
<td>4.5</td>
<td>11.0</td>
<td>16.0</td>
<td>100.0</td>
</tr>
<tr>
<td>(n)</td>
<td>13</td>
<td>32</td>
<td>10</td>
<td>54</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>(percent)</td>
<td>6.5</td>
<td>16.0</td>
<td>5.0</td>
<td>22.5</td>
<td>50.0</td>
<td>100.0</td>
</tr>
<tr>
<td>No marginal defect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(percent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amalgam</td>
<td>163</td>
<td>18</td>
<td>12</td>
<td>6</td>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>(n)</td>
<td>81.5</td>
<td>9.0</td>
<td>6.0</td>
<td>3.0</td>
<td>0.5</td>
<td>100.0</td>
</tr>
<tr>
<td>(percent)</td>
<td>1.18</td>
<td>32</td>
<td>1.0</td>
<td>7.0</td>
<td>0.4</td>
<td>100.0</td>
</tr>
</tbody>
</table>

composite fillings. Such radiolucencies in the radiographic image of the restorations could be mistaken for caries or marginal defects.

The ability to detect marginal defects was considerably higher than for detecting caries and the reason may be that most of the defects were actually big and clearly depicted on radiographs. The absolute detection rate of marginal defects in this study cannot apply to the clinical situation, since the defects were artificial. However, the defects seemed adequate for a standardized comparison of the present materials.

The prevalences of secondary caries (0.38%) and marginal defects (0.31) were extremely high in the present study. The theoretical considerations illustrated in Figs. 4 and 5 indicate that the differences in positive prediction values will be greater with a lower prevalence of secondary caries or marginal defects. This means that the difference in radiopacity between the present materials may be of greater importance in the clinical situation where the actual prevalence of secondary caries is lower.

In conclusion it could be stated that, the radiopaque composite P-31 permits diagnosis of secondary caries and marginal defects at least as well as amalgam.

References

1. Revised guidelines for submission of composite resin materials for occlusal Class I and Class II restorations. Council on Dental Materials, Instruments, and Equipment American Dental Association August 1, 1984


Radiographic diagnoses and treatment decisions on approximal caries

Ivar Espelid
Departments of Oral Radiology and of Pedodontics, School of Dentistry, University of Bergen, Bergen, Norway

Accepted for publication 16 February 1986

Key words: dental caries, diagnosis, dental fillings, radiography, dental

Material and methods

Six molars and 30 premolars were selected from extracted teeth stored in 2% benzyolperoxide solution. The approximal surfaces were inspected and examined by probing and classified according to the findings: sound surfaces (0), discolored lesions without cavities (1), lesions with cavities ≤ 1 mm (2), small cavities (3), and lesions with cavities ≥ 1 mm (4). Surfaces with hypoplastic pits or hypomineralizations were excluded (0).

The teeth were mounted in blocks of wax, four teeth in each block, radiographed using projectional geometry similar to that used in bite-wing exposure. Kodak DI 57 double-pack film and a Ritter Explorer II dental X-ray machine with an electronic timer were used, operated at 15 mA and 65 kVp. The film-object distance was 1.5 cm, the film-focus distance 30.0 cm and a 1.0 cm wide water-filled Plexiglas container was placed between focus and object to simulate soft tissues. The radiographs, after standardized processing, were mounted on a 24.0 x 15.0 cm sheet and masked. Each radiograph was numbered and duplicates were made on Kodak X-Omat duplicating film. The duplicating film may increase the contrast.

The mean density value for the 68 approximal surfaces measured in dentin was 0.86 (SD 0.28, range 0.50 - 1.45). The variation in density between different duplicated films was insignificant and did not exceed the measurement error.

The some 350 dentists who were asked to examine the radiographs, were to participate in several courses dealing with approximal caries, 243 (69%) responded to the request. They were asked to diagnose the surfaces according to a 5-graded confidence rating: 1 - lesion not present, 2 - lesion probably present, 3 - equal chance of being present or not, 4 - lesion probably present, 5 - lesion present. They were required to diagnose caries in each halves of enamel and in dentin separately, and were asked whether or not they would restore the surfaces given the following patient information: first visit of a 15-year-old patient with contact gingivitis. These variables were assumed not to change until the next recall, at the earliest after 1 yr. The dentists were asked to return the diagnoses and treatment proposals before the course started.

The resultant scores were classified...
into decision matrices using a computer program. Each observer's performance was evaluated by ROC technique applying the RSCORE II computer program (8) which gave likelihood estimates for several parameters of diagnostic accuracy (9). In this study the area beneath the ROC curve was used. For analysis of data, programs in the BMDP statistical package were used (10). The means of two samples were compared using the Mann-Whitney test (11). Equality of variances between scores in different levels in the pulp were compared using Levene's test (10). Dichotomized caries scores were coded as follows: caries present scores 1, 2 or 3; caries present scores 4 or 5.

Results

On average the 243 dentists proposed to restore 15.6 (SD = 6.5 range 0-66) of 66 surfaces. Figs. 1 to 4 show the distribution of diagnoses in enamel and dentin and the corresponding treatment proposals for sound surfaces and three groups of different lesion severities. Restorative therapy was proposed in 46.9% of the treatment proposals for surfaces which actually were sound (Fig 1). The probability that a dentist would suggest operative therapy for a true lesion increased with lesion severity. In only 5.0% of lesions without a cavity filling was therapy proposed (Fig 2). The proportion increased to 27.6% for small cavities (Fig 3) and 51.2% for big cavities (Fig 4). The dentists proposed fillings in 91.0% of the lesions which were diagnosed radiographically as extending into dentin and 75.7% of the lesions in the inner half of enamel but only 21.9% of lesions confined to the outer half of enamel.

Proposals to fill sound surfaces or surfaces with discolored lesions were invariably made by fewer than 50% of the dentists. For cavities, however, a positive filling decision was usually supported by a majority of the dentists, on average 94.2% of the dentists agreed to fill small cavities and 91.0% big cavities. Of all decisions to restore, 76.1% were supported by more than 50% of dentists. A majority of the dentists would have filled 13 of the surfaces, whereas 37 surfaces would have been filled by fewer than 10% of the dentists, and 7 of these by only 1 or 2 dentists.

Table 1. Mean confidence of caries being present in outer and inner half of enamel and dentin, based on radiographic interpretation. Five response categories were used. Score 1 denoted lesion not present and score 5 lesion present. In parentheses are given standard deviation values

<table>
<thead>
<tr>
<th>Radiographic decision</th>
<th>No lesion</th>
<th>Very slight</th>
<th>Small cavity</th>
<th>Big cavity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer enamel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No filling</td>
<td>1.29 (0.32)</td>
<td>2.09 (0.42)</td>
<td>3.17 (0.51)</td>
<td>3.57 (0.63)</td>
</tr>
<tr>
<td>Filling</td>
<td>1.56 (0.62)</td>
<td>4.58 (0.83)</td>
<td>4.74 (0.78)</td>
<td>4.92 (0.77)</td>
</tr>
<tr>
<td>Inner enamel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No filling</td>
<td>1.14 (0.66)</td>
<td>2.14 (0.92)</td>
<td>3.18 (1.12)</td>
<td>2.20 (1.04)</td>
</tr>
<tr>
<td>Filling</td>
<td>1.28 (0.51)</td>
<td>2.46 (0.75)</td>
<td>4.58 (0.98)</td>
<td>4.78 (0.67)</td>
</tr>
<tr>
<td>Dentin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No filling</td>
<td>1.14 (0.66)</td>
<td>1.14 (0.47)</td>
<td>1.49 (0.71)</td>
<td>1.90 (0.73)</td>
</tr>
<tr>
<td>Filling</td>
<td>1.37 (0.53)</td>
<td>2.51 (0.72)</td>
<td>4.21 (1.25)</td>
<td>4.78 (1.44)</td>
</tr>
</tbody>
</table>
In Table 1 the observers' confidence of caries being present in enamel or dentin was compared according to the therapy suggested and type of lesion. The variances as a measure of dispersion of confidence ratings, were compared using Levene's test of equality. All differences between variances were statistically significant \( (P < 0.01) \). Great variation in scores was noted for the enamel diagnoses of sound surfaces which were suggested restored. For lesions to be filled the variance of enamel scores decreased as lesion severity increased (only 2410\(^{th} \)) of the sound surfaces would not have been filled by any of the dentists. Four surfaces, all in molars, accounted for more than 60\(^{th} \) of the proposals to fill sound surfaces. When the dentists suggested filling sound surfaces the scoring revealed that they were more confident of caries being present in dentin than enamel. This tendency was reversed when surfaces having “true” caries were suggested filled.

Caries confidence scores in dentin were better correlated with restorative therapy than caries scores in enamel (Table 2). Very few dentists \((n \sim 10)\) would consistently restore all surfaces they judged to have dentin caries and only one dentist would fill every lesion with a positive caries diagnosis (scores 4 and 5). The number of dentists in agreement about restorative therapy according to type and number of lesions to be filled is shown in Fig. 5. The number of dentists who agreed upon restoration was strongly correlated to lesion severity.

The quality of radiographic diagnoses showed great variation among the dentists. The area beneath the ROC curves was used as a measure of diagnostic quality and for radiographic diagnosis of caries in enamel the mean area was 0.828 SD 0.023, range 0.747 0.999 in 1211. The computational method for calculating the area requires estimates for both “true” lesions and sound surfaces.\(^{17} \) Dentists scores did not fulfill these requirements and were not included in this comparison.

Table 3 gives the correlation coefficients between confidence ratings of caries being present in enamel and dentin therapy and the extent of the lesions measured on the tooth surface. The radiographic scores in enamel were the parameter which was best correlated with lesion severity monitored by direct inspection.

The dentists were separated into two groups according to frequency of proposals to treat sound surfaces. One group constituted dentists who suggested that two or more of these surfaces should be filled or \( 62 \), 24\(^{th} \), these dentists accounted for more than 80\(^{th} \) of the de-

| Table 2 | Average correlation coefficients (Pearson r) for 243 dentists between proposals for therapy and scores in enamel and dentin. Scores were dichotomized 1 2 3 4 and 5.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Therapy</td>
<td>Scores in enamel</td>
<td>0.708</td>
<td>0.164</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td>Scores in dentin</td>
<td>0.750</td>
<td>0.144</td>
<td>0.178</td>
</tr>
</tbody>
</table>

Fig. 1: Distribution of radiographic caries diagnoses and therapy suggested for lesions with small cavities (\( < 1 \) mm).

Fig. 2: Distribution of radiographic caries diagnoses and therapy suggested for lesions with big cavities (\( > 1 \) mm).
predictions to treat surfaces which were sound. This "overtreating" group would also restore more of the surfaces with true lesions ($P < 0.001$). In addition, they judged the "true" lesions to penetrate dentin more often than the other dentists. This tendency was statistically very highly significant ($P < 0.001$). In enamel, however, the difference between the number of lesions detected by the two groups of dentists was not statistically significant ($P = 0.1$).

The variation in treatment proposals for sound surfaces was in part explained by variation in treatment proposals for surfaces having cavities and the frequency of true positive (TP) scores in enamel and dentin (cumulated scores 4 and 5; Table 4). The number of false positive (FP) diagnoses in dentin gave a greater contribution to the total variance of restorative decisions for sound surfaces than FP diagnoses in enamel (Table 5).

Figs. 6 and 7 show ROC curves representing diagnoses made in outer enamel, inner enamel, and dentin. The graphs in

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Percent of explained variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP treatment</td>
<td>59.2%</td>
</tr>
<tr>
<td>TP dentin</td>
<td>30.1%</td>
</tr>
<tr>
<td>TP enamel</td>
<td>10.7%</td>
</tr>
<tr>
<td></td>
<td>$R^2 = 23.1%$, adjusted for D.F.</td>
</tr>
</tbody>
</table>

Table 4. Multiple regression analysis of prediction of TP treatment proposals (dependent variable). Independent variables were frequency of TP treatment proposals, and frequencies of TP diagnoses in dentin and in TP diagnoses in enamel. True state of the surfaces was verified during direct inspection and probing on surface. The 5-graded ratings were dichotomized.

![ROC curves](image-url)
Table 1: Multiple regression analysis for prediction of I P treatment proposals (dependant variables) and I P diagnoses in dentin and enamel (independent variables). True state of the surfaces was verified during direct inspection and probing on surface. The 5 graded ratings were dichotomized.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Percent of explained variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I P dentin</td>
<td>99.7%</td>
</tr>
<tr>
<td>I P enamel</td>
<td>81%</td>
</tr>
<tr>
<td>R = 70.0%</td>
<td>Adjusted for 111.</td>
</tr>
</tbody>
</table>

Indicate what consequences a shift along each ROC might have on the true positive and false positive frequencies. Cavitatior (Fig. 6) and cavitatior > 1 mm (Fig. 7) are used as positive findings, while the other lesions and sound surfaces are denoted as negative.

Discussion

In general, the participating dentists agreed that lesions, which on radiographs extend into dentin, should be restored.

Fillings were proposed for more than 90% of the lesions judged to penetrate dentin. On the other hand, the dentists wanted to restore only 22% of lesions which they believed were confined to the outer half of the enamel, and more than 75% of the deeper enamel lesions. These findings indicate that halfway through enamel may constitute a borderline for restorative treatment among the participating dentists.

About 65% of the dentists who participated in a previous questionnaire study responded that they would prefer to restore lesions before they extended into dentin on radiographs, while about 30% would postpone restorative therapy until the radiographic outline of the lesion had reached outer third of dentin. In this study the participants were not asked about their criteria for restorative therapy since this might have influenced their level of diagnosis and treatment decision. Although a direct comparison cannot be made between these investigations, the present results support the results obtained in the questionnaire study.

The fact that very few dentists consistently wanted to fill every dentin-lesion they diagnosed may in part be explained by the total extent of the radiolucency due to caries, not only the depth. The number of proposals for restorative therapy increased with the severity of the lesion.

It is important to know the frequency of over-treatment when evaluating the usefulness of restorative therapy. It is reasonable to avoid restoration of sound surfaces and incipient lesions (without cavitation). The present results indicate that when the dentists are more confident about caries being present in dentin compared to enamel, they should consider whether this might be an I P diagnosis and try to verify the diagnosis by other means. Statistically, confidence that caries is present seems to be more or less correlated with the severity of lesion (12).

Brouwer et al. (13) found that a major part of the error variance in radiographic caries diagnoses when several observers diagnose the same surfaces, is random error rather than systematic error. Variability concerning filling proposals is also reported not to be systematic (14).

Diagnostic exercises based on radiographs alone might be objected to, because several clinical variables are excluded. It is not possible to predict the observer performance in the clinic from such a study, but in this type of simulated diagnostic situation it is possible to isolate the variables which are linked to the radiographic interpretation. The radiographic diagnosis of approximal caries does play a crucial role in the treatment decision (15), and inclusion of clinical variables seems to increase the variation between observers in dental treatment decisions compared to in vitro settings (16, 17).

Appropriate strategy for avoiding restoration of incipient lesions and sound surfaces, might be outlined from Figs. 6 and 7 depending on what type of lesion it is important to restore. In dentistry, as well as in medical care in general, it is useful to have utility and preference scales for different outcomes associated with diagnosis and treatment (18). The "cost" of a false negative versus a false positive diagnosis has to be considered when the treatment philosophy is outlined. Figs. 6 and 7 could be used to find appropriate strategy to minimize the problem of over-treatment. In both fig-
therapy have changed during the last decade. Great differences calculated for diagnostic quality give poor reliability. It has and treatment decisions in addition to exeat interobserver variation in diagnosis of or collection disturbed. Weak basis for epidemiological surveys implies that prevalence and treatment serve as a stated that clinicians' reports about caries is present in dentin will serve as an influence that public's dental health. They conclude that the more frequent the patient is examined, the more likely it will be that teeth are unnecessarily filled.

The prevalence of caries in the present material was extremely high (0.71) and this might force the observer performance toward a restorative line. The interpretation of the results should therefore be done with some care when clinical implications are discussed. It can be stated that clinicians' reports about caries prevalence and treatment serve as a weak basis for epidemiological surveys or collection of national caries data. The great interobserver variation in diagnosis and treatment decisions in addition to great differences calculated for diagnostic quality give poor reliability. It has been reported that criteria for restorative therapy have changed during the last decade and this gives support to the hypothesis that different DMF data cannot be compared directly. Criteria for diagnosis and treatment, caries prevalence and reliability of the diagnostic method should be taken into consideration when different DMF values are compared.

In conclusion, it can be stated that diagnostic quality differed widely between dentists and great interobserver variation existed with respect to caries diagnosis and restorative proposals based on radiographic interpretation; very few dentists applied a restorative philosophy which was based only on the depth of penetration of caries; and most of the proposals to fill sound surfaces were suggested by relatively few dentists.

References

PAPER VII
Variation in radiographic interpretation and restorative treatment decisions on approximal caries among dentists in Norway


Abstract Of a random sample of dentists in Norway (n = 741), 83% responded to a questionnaire about their use of radiographs in and their opinions and knowledge about the diagnosis of approximal caries. There was great disparity in criteria for initiation of restorative treatment of approximal caries based on radiographic appearance. Two thirds of the dentists would do restorative treatment of lesions confined to enamel, the others would wait until lesions had reached the dentin before treatment was commenced. The criteria for restoration based on radiographic appearance was best correlated with the dentists' opinions about cavity formation. About 20% of the variation in criteria among dentists was explained by the radiographic appearance dentists associated with the presence of a cavity. The majority of the dentists believed that approximal caries progressed faster than is indicated by progression studies. The interval between recall examinations using radiographs was not consistently shorter for dentists who decided to carry out operative caries treatment at a relatively advanced stage.

A wide disparity exists between treatment plans made by different dentists (14). In one study the number of tooth surfaces in 18 patients judged by 15 dentists to need operative treatment ranged from 28 to 153 (1). In another study, 843 dentists examined a single bitewing radiograph: the number of surfaces recorded as requiring restoration varied from 0 to 13 (4).

This great interobserver disagreement indicates a high proportion of errors in planning dental treatment. The errors in diagnosis may lead to either overregistration or underregistration of pathology. There is an inverse relationship between the two types of error and it is not possible to eliminate them both simultaneously (5). For this reason the dentist has to weigh the risk and cost of restoring a sound surface against the consequences of leaving a carious lesion untreated.

Some variation in radiographic diagnosis of caries will always exist; this and a lack of standardized treatment criteria may explain a large part of the variation in dentists' treatment decisions. Every decision about approximal caries is a decision made under conditions of uncertainty, and the observer's level of confidence that a lesion is present may be important. Weighting of the various elements in the logic analysis, and the valued outcomes of a decision may differ from dentist to dentist. To influence decisions about treatment of caries or, possibly, to minimize variations in treatment planning, knowledge about the decision-analytic process of dentists is necessary.

Therefore, the aim of this study was to describe criteria for restorative treatment based on the radiographic appearance of approximal caries used by dentists in Norway; to examine the relationship between these criteria and opinions about the caries process; and to determine whether these factors could explain variations in radiographic routines among dentists.

Material and Methods

A pre-coded questionnaire was sent to a random sample of 741 dentists during March 1983. They were drawn from the Norwegian Dental Association's register of authorized dentists. Those 69 yr of age and older, specialists, full-time teachers and administrators were excluded.

Seven questions asked about opinions, experience and routines in connection with radiographic examination of approximal carious lesions. Question 1: The figure shows different radiologic appearances of approximal carious lesions (Fig. 1). Which lesion or lesions should he restored immediately? We assume that the patient's caries activity is low and his oral hygiene adequate. Question 2: Based on your clinical experience, what or which of these radiologic appearances (Fig. 1) do you associate with cavitation (loss of continuity in the enamel surface)? Question 3: Do you think that the radiographic appearance of approximal caries as compared to clinical findings in the prepared cavity usually: a: shows overestimation of depth? b: depicts the true depth? c: shows overestimation of depth?

Accepted for publication 3 March 1984
Radiographic examinations for approximal caries

Fig. 1. Stages of approximal caries (denoted A to F). Used to determine respondents' criteria for initiation of restorative treatment and opinion on-stage at which a cavity is present. A not more than half enamel depth, B between outer half and outer two thirds of enamel, C to dentin-enamel junction, D in outer third of dentin, E between outer third and outer half of dentin, F not more than two thirds of dentin depth.

The age rate of progress of an approximal carious lesion developing from outer enamel to dentin (permanent dentition).

Question 3: How frequently do you take radiographs as part of the examination of patients aged 3-5 yr, 6-12 yr, 13-18 yr and those who are older than 18 yr?

A total of 616 dentists (83%*) responded to the questionnaire. Their mean age was 42 yr (SE 10.8). Their distribution by age and type of practice was in close agreement with that of professionally active dentists in Norway (6, 7). The dentists were asked to write their names on the reply envelope, and those who did not reply within 2 wk were sent a reminder. Returned questionnaires were treated anonymously. Data were analyzed employing standard programs available in the statistical packages HM13P (8) and MINITAB (9). For the multiple regression analysis the following dichotomized independent variables were used: Criterion for restorative treatment: 1 - enamel, 2 - dentin (original response alternatives are shown in Fig. 1). Type of practice: 1 - private practice, 2 - public services. Presence of a cavity: 1 - two thirds of enamel, 2 - deeper than two thirds of enamel (original response alternatives are shown in Fig. 1). Clinical vs radiographic lesion depth: 1 = radiographic underestimation, 2 = other. Caries progression: 1 = through enamel ≤ 11 months, 2 = 12 months or more. Age was dichotomized: 1 = born 1914-35, 2 = born 1936-47, 3 = born 1948 or later. The dependent variable in the regression analysis was the treatment criterion.

Table 1: Respondents' estimates of time required for a caries lesion to progress from outer enamel to dentin

<table>
<thead>
<tr>
<th>Rate in months</th>
<th>n (%) (n = 555)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.3</td>
</tr>
<tr>
<td>1-11</td>
<td>41.6</td>
</tr>
<tr>
<td>12-23</td>
<td>48.5</td>
</tr>
<tr>
<td>≥ 24</td>
<td>7.4</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 2: Distribution of dentists according to time interval they used between radiographic examinations for caries (percent)

<table>
<thead>
<tr>
<th>Patient's age</th>
<th>Months between examinations</th>
<th>Occasional</th>
<th>Never</th>
<th>Total</th>
<th>n*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>6</td>
<td>12</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>0.4</td>
<td>11.7</td>
<td>45.5</td>
<td>8.9</td>
<td>17.3</td>
</tr>
<tr>
<td>6-12</td>
<td>0.2</td>
<td>14.4</td>
<td>71.2</td>
<td>6.2</td>
<td>6.5</td>
</tr>
<tr>
<td>13-18</td>
<td>0.2</td>
<td>17.2</td>
<td>78.1</td>
<td>3.7</td>
<td>0.5</td>
</tr>
<tr>
<td>≥ 18</td>
<td>0.2</td>
<td>9.7</td>
<td>75.6</td>
<td>13.6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

* Total numbers of replies within each patient category.

Results

Fig. 2 presents the distribution of dentists according to treatment criteria, based on the radiographic image of approximal caries. Two thirds of the dentists would have performed restorative treatment while the lesions were confined to the enamel only and the majority of the dentists would have carried out operative treatment on observation of any radiolucency in dentin. More than 80% of the dentists believed that cavitation occurred before the lesions were apparent radiographically in dentin (Fig. 2).

That radiographs usually underestimate the carious lesion compared to clinical findings was the opinion of 69% of the dentists, while 25% responded that lesion depth registered radiographically was closely correlated with its clinical extent. Only 3% felt that radiographs usually overestimated true lesion depth.

Table 1 presents the responses on the rate of progress of caries. Nearly half of the dentists (49%) believed that the mean time for a carious lesion to progress from outer enamel to dentin was between 12 and 23 months.

According to the majority of dentists, radiographs should be taken annually as part of the dental examination of persons aged 6 yr or older, but there was wide disagreement about patients aged 3-5 yr. Table 2 shows how frequently dentists used radiographs for caries examinations according to patient's age. Sixty-nine percent of dentists would take radiographs more frequently when caries was under observation in a patient; 78% chose half-year intervals between examinations of these patients.

To test for correlation between the criteria for restorative therapy for approximal caries and other variables, a stepwise multiple regression analysis was performed using the original response categories of each variable. The best predictor was the dentist's opinion about the
who would initiate retro- morphic therapy if they had found a lesion in an earlier stage. Thus, the reexamination routines were not based on the level at which dentists preferred to restore a lesion. In addition, public dental awareness may have increased (12). Whether these factors have encouraged a change of attitude among dentists from a restorative to a preventive approach in treatment of approximal caries is uncertain. To find out it would be necessary to outline the different philosophies used in treatment planning. Since 50-60% of treatment decisions of approximal carious lesions are based on radiographic examination alone (13), the dentists in the present study were asked to state the criteria they used when deciding to perform restorative treatment on the basis of radiographs.

Differences in treatment criteria were registered between dentists in the public dental services and in private practice. The reason why private practitioners decide to restore lesions at an earlier stage is not known. One suggestion has been that dentists restore teeth unnecessarily, to extract a fee (14). However, economic considerations do not seem to influence the number of fillings which are produced (14), based on the small difference observed between public dental officers (salaried) and private practitioners (fee for service) in England. The differences in treatment criteria between dentists in public service and in private practice could, therefore, be explained partly by the fact that public dental officers tend to be younger than private practitioners rather than speculation about overtreatment for economic reasons. If the new concepts about caries treatment are accepted, namely, not to treat every radiographically detectable lesion, the results indicate that some overtreatment of caries occurs, particularly by older dentists.

### Table 4

<table>
<thead>
<tr>
<th>Variable</th>
<th>Age group</th>
<th>Type of practice</th>
<th>Treatment criteria</th>
<th>Cavity criteria</th>
<th>Relationship clin. radiogr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of a cavity</td>
<td>0.106**</td>
<td>0.057</td>
<td>0.408***</td>
<td>0.039</td>
<td>0.030</td>
</tr>
<tr>
<td>Relationship clin. radiogr.</td>
<td>0.017</td>
<td>0.107*</td>
<td>0.176***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progress rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** P < 0.05

### Table 4

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>Beta</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity criteria</td>
<td>0.379</td>
<td>&lt;0.0000</td>
</tr>
<tr>
<td>Relationship between clinical and radiographic findings</td>
<td>0.135</td>
<td>&lt;0.0011</td>
</tr>
<tr>
<td>Progress rate</td>
<td>0.094</td>
<td>&lt;0.0109</td>
</tr>
</tbody>
</table>

R^2 = 0.19, P = 0.0000.

Correlation coefficients for some of the dichotomized variables are given in Table 3. Most were low, but some were highly significant. In addition to the correlation between criteria for treatment and presence of a cavity, strong correlations were found between presence of a cavity and year of birth and type of practice. In Table 4 the best predictors of treatment criteria are listed with their beta coefficients. Results suggest that opinions about cavitation and about relationship between radiographic and clinical findings made the greatest contribution when all other variables were held constant.

### Discussion

The incidence of caries has decreased markedly during the last decade (10, 11), and the rate of progress of lesions through enamel and dentin may have slowed down. In addition, public dental awareness may have increased (12). Whether these factors have encouraged a change of attitude among dentists from a restorative to a preventive approach in treatment of approximal caries is uncertain. To find out it would be necessary to outline the different philosophies used in treatment planning. Since 50-60% of treatment decisions of approximal carious lesions are based on radiographic examination alone (13), the dentists in the present study were asked to state the criteria they used when deciding to perform restorative treatment on the basis of radiographs.

Differences in treatment criteria were registered between dentists in the public dental services and in private practice. The reason why private practitioners decide to restore lesions at an earlier stage is not known. One suggestion has been that dentists restore teeth unnecessarily, to extract a fee (14). However, economic considerations do not seem to influence the number of fillings which are produced (14), based on the small difference observed between public dental officers (salaried) and private practitioners (fee for service) in England. The differences in treatment criteria between dentists in public service and in private practice could, therefore, be explained partly by the fact that public dental officers tend to be younger than private practitioners rather than speculation about overtreatment for economic reasons. If the new concepts about caries treatment are accepted, namely, not to treat every radiographically detectable lesion, the results indicate that some overtreatment of caries occurs, particularly by older dentists.
Nowadays it is likely that dentists whose philosophy is "when in doubt fill," will do some unnecessary fillings.

The correlation between radiographic criteria for restorative therapy of approximal caries and for cavitation in the enamel surface (Table 3) indicates that dentists find it important to restore approximal carious lesions with cavitation. Although radiographic criteria for cavitation could explain about 20% of the variation in radiographic treatment criteria, the other replies only explained a minor part of this variation. This may give the impression that the choice of treatment criteria is only partly based on current scientific knowledge or on clinical experience. It is likely that a tendency exists to use generally accepted criteria for caries treatment, rather than to develop a personal philosophy. The first visible breakdown of the enamel surface is commonly accepted as a proper criterion for restorative therapy (15). According to comparisons made on radiographic and clinical appearances of approximal caries, there is about a 50% probability that a clinical cavity is present when the lesion has reached the dento-enamel border radiographically, while all lesions diagnosed radiographically into dentin are likely to have a clinical cavity (16). This finding is contrary to the majority opinion among dentists. Most of them believed that cavitation occurred while lesions were radiographically confined to the enamel.

The variation in recall routines for dental examinations showed no correlation with replies concerning rate of caries progression or treatment criterion. Thus a dentist who resorts to restorative treatment at a relatively advanced stage of disease progression is likely among dentists in Norway. It is likely that outdated restorative treatment criteria and unnecessarily frequent radiographic caries examinations are used by some. Practitioners should adopt a critical approach to caries diagnosis using radiographs, and should adjust their recall intervals and their radiographic treatment criteria to the prevailing caries activity.

References