A Systematic Review of the Performance of Methods for Identifying Carious Lesions

James D. Bader, DDS, MPH; Daniel A. Shugars, DDS, PhD; Arthur J. Bonito, PhD

Abstract

This systematic review evaluates evidence describing histologically validated performance of methods for identifying carious lesions. A search identified 1,407 articles, of which 39 were included that described 126 assessments of visual, visual/tactile, radiographic (film and digital), fiber optic transillumination, electrical conductance, and laser fluorescence methods. A subsequent update added four studies contributing 10 assessments. The strength of the evidence was judged to be poor for all applications, signifying that the available information is insufficient to support generalizable estimates of the sensitivity and specificity of any given application of a diagnostic method. The literature is problematic with respect to complete reporting of methods, variations in histological validation methods, the small number of in vivo studies, selection of teeth, small numbers of examiners, and other factors threatening both internal and external validity. Future research must address these problems as well as expand the range of assessments to include primary teeth and root surfaces. [J Public Health Dent 2002;62(4):201-13]

Key Words: dental caries, diagnosis, systematic review.

The principal methods dentists currently use to diagnose carious lesions—visual and visual/tactile examinations and radiographic assessments—have been employed for decades with only incremental changes in equipment and techniques. Refinement, rather than development of new technology, characterized these methods over the years. Illumination has improved and magnification is more easily employed for visual examinations, while radiation doses have decreased for radiographic assessments as both equipment and film have been improved. Recently, however, a wider variety of new methods and refinements in existing methods for the detection of carious lesions has become available, including fiber optic transillumination, direct digital imaging, electrical conductance, and most recently laser fluorescence.

The diagnosis of dental caries is, in reality, an exhaustive search for evidence of demineralization on individual tooth surfaces. Using the traditional methods, it has been primarily a visual process, based principally on clinical inspection and review of radiographs, although some tactile information may be considered as well. Chiefly because these methods depend on subjective interpretation of subtle visual and tactile cues, variation among dentists' diagnoses had tended to be extensive (1). Some of the alternative diagnostic methods, such as fiber optic transillumination (FOTI), and direct digital imaging continue to rely on dentists' interpretation of visual cues, while other emerging methods, such as electrical conductance (EC) and computer analysis of digitized radiographic images, offer the first "objective" assessments, where visual and tactile cues are either supplemented or supplanted by quantitative measurements.

The various methods have been the subject of numerous investigations examining the effects of refinements in techniques on performance in identifying carious lesions. Additionally, a relatively robust literatures comparing the performance of two or more of the methods on the same samples of teeth. However, the number of reviews, meta-analyses, and other syntheses addressing comparative performance among methods is much smaller, and generally restricted to the diagnosis of one type of carious lesion, or lesions on one type of tooth surface (2-5).

This paper reports the findings of a rigorous systematic review of the relevant scientific literature on the performance of methods for the identification of carious lesions. The review was conducted under the auspices of the Research Triangle Institute–University of North Carolina Evidence-based Practice Center, sponsored by the Agency for Healthcare Research and Quality (AHRQ), supported by the National Institute for Dental and Craniofacial Research (NIDCR), and intended to anchor portions of the March 2001 Consensus Development Conference on Diagnosis and Management of Dental Caries Through Life. The specification of the review question was guided by the Steering Committee for the Consensus Development Conference.

Methods

We developed our review methods with the help of a Technical Expert Advisory Group (see *Acknowledgments*). The question we addressed

Send correspondence to Dr. Bader, Sheps Center, CB#7590, University of North Carolina, Chapel Hill, NC 27599. E-mail: jim_bader@unc.edu. Web site: http://www.unc.edu. Dr. Shugars is affiliated with the School of Dentistry and the Sheps Center for Health Services Research, University of North Carolina at Chapel Hill. Dr. Bonito is affiliated with the Research Triangle Institute. Reprints will not be available. This study was developed by the RTI/UNC Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (contract #290-97-0011), Rockville, MD. This paper was presented at the NIH Consensus Develop Conference on Diagnosis and Management of Dental Caries Throughout Life, March 26, 2001, Bethesda, MD. The authors of this article are responsible for its contents, including any clinical or treatment recommendations. No statement in this article should be construed as an official position of the Agency for Healthcare Research and Quality or the US Department of Health and Human Services. Manuscript received: 8/7/01; returned to authors for revision: 10/2/01; final version accepted for publication: 11/13/01.

was "What are the validities of the available diagnostic methods for detecting carious lesions in primary and permanent teeth?" We conducted a detailed search of the relevant English language literature published between 1966 and October 1999 using MEDLINE and EMBASE. We did not pursue reports in the gray literature, defined as theses, dissertations, product reports, and unpublished studies. We did hand search the most fruitful journals from 1998 to the end of 1999 to accommodate for the lag in MEDLINE postings. The search used exploded key words for lesions, diagnostic methods, and study designs, and limited the focus to human studies. Our initial search of MEDLINE plus hand searching identified 1,328 citations, with 79 additional citations identified through EMBASE (Figure 1). Subsequent to presentation of the results of the systematic review at the Consensus Development Conference, we searched MEDLINE in May 2001 for the year 2000, and added all studies meeting the original criteria.

We applied several inclusion and exclusion criteria to the reports identified in our literature search. Most importantly, we limited studies to those in which the performance of the diagnostic method was compared to a reference or gold standard determined histologically—i.e., by visual or microradiologic inspection of tooth sections for all evaluated surfaces. We excluded all studies in which one diagnostic method was compared to another with one of the two methods declared as the reference standard. We were unwilling to include studies using such a "silver" standard; no currently available clinical diagnostic method is perfectly valid, i.e., has 100 percent sensitivity and specificity compared to histological evaluation. Although a mathematical "correction" for radiograph performance has been proposed, it has not been validated (6). We also excluded studies in which exploratory surgical intervention was employed to confirm positive diagnoses, as this validation method cannot detect false negatives (7). Similarly, studies in which only suspicious surfaces were evaluated histologically were excluded. One exception to the histological validation criterion was made for studies of methods for the identification of cavitated lesions: here direct visual or visual-tactile inspection was accepted as a validation method.

FIGURE 1
Strategy and Results of MEDLINE and EMBASE Searches

MEDLINE	
Wide search of early caries literature	
1 exp dental caries/pa,di,ra	2,846
2 limit to human, English, 1966–75	219
Defining studies of caries	
3 exp tooth demineralization/pa,di,ra	2,928
4 exp dental caries	21,830
5 3 or 4	21,904
Limiting 5 to diagnostic methods	
6 exp diagnosis/oral diagnosis	2,420
7 exp radiography/dental radiography/digital dental radiology	816
8 exp pathology/ oral pathology	4
9 1 or 6 or 7 or 8	2,539
10 limit to human, English	1,776
Limiting 10 to various study types	
11 controlled clinical trial	21
12 meta-analysis	4
13 randomized controlled trial	50
14 epidemiologic study characteristics	244
15 epidemiologic research design	333
16 comparative study	457
17 2 or 11 or 12 or 13 or 14 or 15 or 16	1,266
Adding all root caries	
18 exp root caries/pa,di,ra	62
Total	1,328
EMBASE	1,554
1 dental adj caries	
2 diagnosis	248,652
3 dental radiology	121
4 2 and 3	248,677
Total (1 and 4)	87
Unduplicated MEDLINE and EMBASE	1,407

We included both in vivo and in vitro studies, and we included studies of all primary diagnostic methods that were commercially available at the time of the review. The methods included visual, visual/tactile, radiographic (using D or E speed film, and phosphor storage plates or charge-coupled devices, with or without computer-based image analysis), fiber optic transillumination, electrical conductance (EC), and laser fluorescence (LF).

We included studies in which the outcomes were expressed in terms of sensitivity and specificity, or in which these statistics could be calculated from the information presented. This criterion resulted in the exclusion of several studies in which results were expressed only as receiver-operating characteristic (ROC) curves. Such outcomes typically are obtained when observers indicate their level of certainty about a diagnosis on a five-point scale. The argument for using such an analytic approach is that asking an observer to state a level of certainty helps disassociate the observer's degree of leniency from the implications of any given decision criterion, thus permitting diagnostic performance to be reflected independent of an observer's perceived "cost" of an incorrect diagnosis (8). However, the ROC statistic, as well as the D_z statistic (3), obviate separate consideration of two functions of a diagnostic procedure: to identify lesions when they are present (sensitivity), and to rule out the presence of lesions when they are not present (specificity). Many studies that re-

ported ROC analyses also reported sensitivity and specificity outcomes for the combined levels of "reasonably certain" and "certain" that a lesion is present. If these outcomes were reported, we included the study in the evidence table. However, studies were excluded in which it was necessary to estimate values for sensitivity and specificity outcomes directly from ROC curves because no data were reported in text or table.

We applied the inclusion and exclusion criteria by examining titles, abstracts, and, where necessary, full papers using dual independent reviews for the 1,407 papers identified in the search (Figure 2). The two reviewers agreed on inclusion status for 97 percent of the papers, with discussion leading to consensus where disagreement occurred. A total of 39 studies were included. Four were added in the subsequent update.

We abstracted data (single abstraction, subsequent independent review) from the studies, achieving a 100 percent agreement rate on results, and 88 percent for other study descriptors. From most studies, we were able to abstract information describing more than one evaluation of a diagnostic method. We defined each evaluation as an assessment. Typically, a study would yield assessments involving multiple methods (listed above) and/or multiple types of lesions (cavitated lesion, lesion penetrating into dentin, lesion confined to enamel, any lesion). In addition, we classified each assessment according to the teeth assessed (permanent posterior only, permanent anterior and posterior, primary), setting (in vivo, in vitro), and the tooth surfaces assessed (occlusal, proximal, root). Thus, we organized each individual assessment in the evidence table by its specific application, which we defined as a unique combination of diagnostic method, tooth type and surface, and type of lesion involved. For example, a specific application might be visual inspection of posterior occlusal surfaces for the detection of cavitated lesions.

We also computed a quality score for each included study using a quality rating form covering 11 elements of internal validity, including issues involving sample size, selection of teeth and surfaces, setting, validation method, validation criteria, validation reliability, lesion prevalence, number of examiners, examiner reliability, and lesion criteria. The criteria and weighting scheme for the quality score are shown in Table 1. As originally calculated, quality scores could range from 0–20. All scores were rescaled to a 0 to 100 scale. We did not exclude any study based on its quality score; rather, we judged the overall strength of the evidence in terms of the extent to which it offered a clear, unambiguous assessment of the validity for a specific application. The three possible ratings for an application were:

- *Good:* The number of studies is larger than three, the quality of the studies is generally high, and the results of the studies represent narrow ranges of observed sensitivity and specificity.
- *Fair:* There are at least three studies, the quality of the studies is at least average, and the results represent moderate ranges of observed sensitivity and specificity.
- *Poor:* There are fewer than three studies, or the quality of the available studies is generally lower than average, and/or the results represent wide ranges of observed sensitivities and/or specificities.

For purposes of these strength ratings, a narrow range was defined as no more than 0.15 on a scale of 0.0 to 1.00, a moderate range was no more than 0.35, and a wide range was more than 0.35. High quality was defined as most study scores at or above 60, average quality was defined as most study scores at or above 45.

FIGURE 2
Flow Diagram of Identification and Inclusion Steps

Step	Number of Articles
Initial MEDLINE search	1,328
⇩	
Initial EMBASE search (nonduplicates)	79
⇩	
Total articles for review	1,407
⇩	
Surviving title review	285
⇩	
Surviving abstract/paper review	50
⇩	
Included in final review	39

Results

From the 43 studies, we identified 136 separate assessments addressing 36 specific applications. Table 2 summarizes the distribution of the assessments across all possible applications. Three assessments that evaluated combined methods are not represented in the table. The coverage of the assessments was extremely uneven. Only five assessments evaluated diagnostic performance in primary teeth, and seven in both anterior and posterior teeth. No studies of diagnosis of carious lesions on root surfaces were included in the review. Thirty-seven assessments reported applications for proximal surfaces, while 99 focused on occlusal surfaces. While more than half (69) of the assessments evaluated radiographic methods, there were only nine assessments of the visual/tactile method, four for FOTI, and two for laser fluorescence.

Table 3 summarizes the information describing the 25 assessments of the visual method for the detection of carious lesions. These results are grouped into five specific applications with different combinations of type of lesion and surface. All but one study examined occlusal surfaces. For the applications in which there were three or more assessments, the variation in sensitivity scores was extensive, while specificity scores tended to be somewhat more stable across assessments. For example, for the detection of dentinal lesions on occlusal surfaces, sensitivity ranged from 0.03 to 0.95, with six scores falling in a lower range up to 0.40 and six in an upper range from

0.45. Specificity scores ranged from 0.41 to 1.0, with all but two of the 12 scores being above 0.80. Almost two-thirds of the visual assessments had sample sizes larger than 100 sites, but fewer than one-half featured four or more examiners. In most studies the prevalence of lesions was higher than what is encountered in clinical practice; for almost half of the assessments, more than 50 percent of the sites examined had a lesion. The sites that were evaluated varied among studies from a specific individual site on an occlusal surface indicated by special marking, through all pit and fissure sites on an occlusal surface, to an entire occlusal or proximal surface. The quality scores for this group of assessments were average, ranging from 30 to 75, with a mean of 51. We rated the strength of the evidence for estimating the validity of the visual method for detecting carious lesions as poor for all five applications. Where there were sufficient numbers of assessments, the variation in the assessments was wide, rendering any attempt to establish a stable estimate of performance problematic.

Table 4 summarizes the nine assessments of the visual/tactile method in five different applications. The variation in sensitivity and specificity scores was similar to that seen for the visual method, with a range from 0.17 to 0.93 for sensitivity, and a more restricted range of from 0.71 to 1.00 for specificity. Again, one-third of the studies employed four or more examiners. However, lesion prevalence was lower, above 50 percent in only one assessment. Quality scores were average, with a mean of 50 and a range of 35–70. We rated the strength of the evidence as poor primarily due to the limited number of studies available for any application, but also for cavitated lesions on proximal surfaces, due to the variation in the sensitivity scores.

Table 5 summarizes the 69 assessments of radiographic methods for detection of carious lesions. The pattern of performance was similar for each of the applications for which there were three or more assessments. A moderate or narrow range of specificity scores was accompanied by a wider range of sensitivity scores. Of necessity, these assessments all included the entire tooth surface. Lesion prevalence was generally high, with prevalences being greater than 50 percent for 27 of the 54 assessments where prevalence was reported. More than one-half of the assessments involved three or fewer examiners. The pattern of results for digital methods was similar to that for film-based methods, with consistently high specificities accompanied by a wide range of sensitivity values for any specific surface/lesion combination. We rated the strength of the evidence for stable evidence of performance as poor for all applications due to the extensive variation in sensitivity values among studies.

Table 6 presents the results of 24

TABLE 1
Quality Score Criteria*

Points	Criterion
Tooth selection	
3	Both posterior and anterior teeth
2	Only anterior or only posterior teeth
1	Selected posterior or anterior teeth
0	Single tooth type (e.g., max. or mand. 3rd molar)
# of sites assessed	
3	150 or more
2	75–149
1	40–74
0	Fewer than 40
Caries prevalence	
2	Less than 20%
1	20%–49%
0	50% or greater
Study setting	
2	In vivo
0	In vitro
Number of evaluators/observers	
2	4 or more
1	2–3
0	1
Test reliability reported	
2	Interevaluator and intraevaluator reliability reported
1	Inter- or intraevaluator reliability reported
0	No evaluator reliability reported
Validation method	
2	Light microscopy (stereo/mono) with/without dye of sectioned tooth
1	Other visual or radiographic assessment of sectioned tooth
0	Assessment of unsectioned tooth
Validation criteria	
1	Criteria explicitly stated
0	Criteria not explicitly stated
Validation reliability reported	
1	Interevaluator or intraevaluator reported
0	No validation reliability reported
Area assessed for any site	
1	Entire surface
0	Specific site on surface
Criteria for caries	
1	Specified prior to evaluation
0	Developed post hoc

*20 total points available: scores subsequently rescaled to a 0–100 scale by multiplying assigned points by 5.

TABLE 2
Number of Diagnostic Assessments by Tooth Surface, Tooth Type, Lesion Type, Diagnostic Method, and Study Setting

Tooth Type	Lesion Type	Radio		Visual/Tact.		Visual		FOTI*		ECM†		LS‡	
		Vivo	Vitro	Vivo	Vitro	Vivo	Vitro	Vivo	Vitro	Vivo	Vitro	Vivo	Vitro
Proximal surfaces													
Posterior	Cavitation	5	1	2	1		1	1					
	Dentin		7		1								
	Any lesion		9										
	Enamel only		1										
	Root surface												
Anterior	Cavitation												
and	Dentin		2										
posterior	Any lesion		3										
	Enamel only		2										
	Root surface												
Primary	Cavitation	1											
	Dentin												
	Any lesion												
	Enamel only												
Occlusal surfaces													
Posterior	Cavitation				1	1	2						
	Dentin		26		2	1	10		2	2	12		2
	Any lesion		7		2	1	4			1	7		
	Enamel only		4				3		1		1		
	Root surface												
Anterior	Cavitation												
and	Dentin												
posterior	Any lesion												
	Enamel only												
	Root surface												
Primary	Cavitation						1						
	Dentin		1				1				1		
	Any lesion												
	Enamel only												

*Fiberoptic transillumination.
†Electrical conductance.
‡Laser fluorescence.

assessments of three applications of the electrical conductance method for detection of carious lesions. The bulk of the assessments dealt with the detection of dentinal lesions and any lesions on occlusal surfaces. Here the pattern of results was somewhat different from that seen for visual, visual/tactile, and radiographic methods. The variation in sensitivity values among the studies was reduced, with a range from 0.58 to 0.97 for dentinal lesions, and from 0.61 to 0.92 for any lesions, while the variation in specificity values for dentinal lesions is more pronounced, with a range from 0.56 to 1.00. Variation among specificity values for detection of any lesion was less evident, remaining in the narrower range of high values typical for other detection methods (0.74 to 1.00). If only studies of the most recently marketed device were considered, specificity for dental lesions was less variable, with all studies around .80. The studies of EC methods generally used fewer examiners, only one assessment reported was based on four or more examiners, and 19 of the 24 assessments used only one examiner. In 13 instances, this single examiner was the same person. Caries prevalence for these assessments was high for the assessments of the detection of any lesion on occlusal surfaces, with all eight assessments involving sites with more than 50 percent lesion prevalence. The mean quality score for the dentinal lesion assessments was 43, while the mean for the any lesion assessments was 29. We rated the strength of the evidence as poor for all applications for reasons of variation of both sensitivity and specificity values for assessments of dentinal lesions, as well as the generally low quality of scores for the assessments of detection of any lesions.

Table 7 summarizes the assessments for FOTI, laser fluorescence, and combined visual and radio-

TABLE 3
Studies of Visual Methods for Detection of Carious Lesions

Source (Ref.)	Setting	Sites (*N*)	Teeth/Site*	Specific Method	Examiner (*N*)	Lesion Prevalence	Quality Score†	Sensitivity (SD)	Specificity (SD)
Cavitated lesions on occlusal surfaces									
Downer, 1975 (9)	in vivo	230	all post/p&f	direct visual	1	74%	60	.84‡	.78‡
Downer, O'Mullane, 1975 (10)	in vitro	230	all post/p&f	direct visual	1	¶	50	.91‡	.81‡
Kelly, Holt, 1993 (11)	in vitro	100	1st molar/ss	direct visual	1	51%	35	.31‡	.98‡
Kelly, Holt, 1993 (11)	in vitro	100	pri 2nd mol/ss	direct visual	1	44%	35	.45‡	1.00‡
Dentinal lesions on occlusal surfaces									
Ricketts et al., 1995 (12)	in vitro	100	3rd molar/ss	direct visual	¶	30%	45	.63¶	.97¶
Nytun et al. 1992 (13)	in vitro	30	all molar/surf	direct visual	10	77%	40	.72¶	.41¶
Wenzel, Fejerskov, 1992 (14)	in vitro	78	3rd molar/surf	direct visual	1	67%	45	.54‡	.81‡
Lussi, 1993 (15)	in vitro	63	all post/ss	direct visual	26	44%	45	.12¶	.93¶
Lussi, 1993 (15)	in vitro	63	all post/ss	magnification	26	44%	45	.20¶	.89¶
Verdonschot et al., 1993 (16)	in vitro	81	3rd molar /p&f		4	67%	30	.48¶	.89¶
Deery et al., 1995 (17)	in vitro	112	all post/surf	direct visual	7	54%	65	.12¶	.97¶
Ekstrand et al. 1997 (18)	in vitro	100	all post/p&f	direct visual	3	39%	60	.95¶	.90¶
Ashley et al., 1998 (19)	in vitro	103	all post/surf	direct visual	1	36%	60	.24‡	.97‡
Huysmans et al., 1998 (20)	in vitro	107	all post/p&f	direct visual	2	41%	60	.27¶	1.00¶
Cortes et al., 2000 (21)	in vitro	59	molars/ss	direct visual	4	32%	65	.95¶	.53¶
Ashley, 2000 (22)§	in vitro	58	pri molars/surf	direct visual	1	64%	50	.78‡	.95‡
Any lesions on occlusal sufaces									
Ricketts et al., 1995 (12)	in vivo	100	3rd molar/ss	direct visual	¶	64%	40	.27¶	.89¶
Wenzel et al., 1990 (23)	in vitro	45	all post/surf	x10 photos	4	89%	40	.89¶	.65¶
Lussi, 1991 (24)	in vitro	61	all post/p&f	direct visual	26	61%	45	.63¶	.83¶
Deery et al., 1995 (17)	in vitro	112	all molar/surf	direct visual	7	97%	65	.60¶	.50¶
Fyffe et al., 2000 (25)	in vitro	421	all post/surf∞	direct visual	20	50%	70	.48¶	.76¶
Enamel lesions on occlusal surfaces									
Wenzel et al., 1990 (23)	in vitro	45	all post/surf	x10 photos	4	17%	40	.72¶	.66¶
Ashley et al., 1998 (19)	in vitro	103	all post/surf	direct visual	1	24%	55	.60‡	.73‡
Fyffe et al., 2000 (25)	in vitro	421	all post/surf∞	direct visual	20	35%	75	.10¶	.98¶
Cavitated lesions on proximal surfaces									
Downer, O'Mullane, 1975 (10)	in vitro	185	all post/surf	direct visual	1	¶	50	.94‡	.92‡

*Teeth/site abbreviations: post=posterior teeth; surf=entire surface; p&f=pit and fissure sites; ss=specific sites; pri=primary tooth.
†Quality score is based on a scale of 0 to 100.
‡Not applicable.
¶Not reported.
§From among 4 possible criteria sets.
∞Includes buccal and lingual surfaces.

graphic methods. For FOTI and LF there were a total of six assessments reported in four studies. Three independent reports of combined visual and radiographic methods again presented a wide range of sensitivity scores, from 0.49 to 0.86. We rated the strength of the evidence as poor for each application for reasons of insufficient numbers and/or extent of variation in sensitivity scores.

Discussion

This review evaluated the evidence describing the performance of diagnostic methods for carious lesions in permanent and primary teeth. The results of the evaluation were uniform across all applications for which assessments were reported: the strength of the evidence was rated as poor with respect to providing information with which to establish expected performance levels. It must be stressed that these ratings reflect the subjective judgments of the review team concerning the strength of the evidence describing the performance of the various diagnostic methods, rather than ratings of the diagnostic methods themselves. The purpose of the review was to identify and evaluate all available evidence describing the histologically validated performance of the diagnostic methods, thereby letting the evidence speak for itself in terms of describing performance. It is our judgment that the evidence does not speak clearly in describing expected ranges of performance, or in comparing the performance of different methods.

The vast majority of assessments of

TABLE 4
Studies of Visual/Tactile Methods for Detection of Carious Lesions

Source (Ref.)	Setting	Sites (*N*)	Teeth/Site*	Specific Method	Examiner (*N*)	Lesion Prevalence	Quality Score†	Sensitivity (SD)	Specificity (SD)
Cavitated lesions on occlusal surfaces									
Downer, O'Mullane, 1975 (10)	in vitro	230	all post/p&f	visual/tactile	1	‡	50	.92¶	.85¶
Dentinal lesions on occlusal surfaces									
Penning et al., 1992 (26)§	in vitro	1,140	3 teeth/surf	automatic probing	1	13%	45	.24¶	1.00¶
Lussi, 1993 (15)	in vitro	63	all post/ss	visual/tactile	23	44%	45	.14‡	.93‡
Any lesions on occlusal surfaces									
Penning et al., 1992§	in vitro	1,140	3 teeth/surf	automatic probing	1	19%	45	.17¶	1.00¶
Lussi, 1993 (15)	in vitro	63	all post/ss	visual/tactile	8	61%	45	.61‡	.87‡
Cavitated lesions on proximal surfaces									
Mejare et al., 1985 (27)	in vivo	598	all post/surf	visual/tactile	3	5%	65	.29‡	.89‡
Hintze et al., 1998 (28)	in vivo	338	all post/surf	visual/tactile	4	6%	70	.34‡	.99‡
Downer, O'Mullane, 1975 (10)	in vitro	185	all post/surf	visual/tactile	1	‡	50	.93¶	.97¶
Dentinal lesions on proximal surfaces									
Verdonschot et al., 1991(40)	in vitro	21	mand pre/surf	visual/tactile	3	‡	35	.50‡	.71‡

*Teeth/site abbreviations: post=posterior teeth; surf=entire surface; p&f=pit and fissure sites; ss=specific sites; mand=mandibular..
†Quality score is based on a scale of 0 to 100.
‡Not reported.
¶Not applicable.
§Study involved standardized mechanical probing of complete occlusal surfaces of three teeth.

diagnostic methods involved occlusal and proximal surfaces of posterior permanent teeth. Too few assessments addressed diagnosis of carious lesions on primary teeth and anterior teeth to permit any definitive conclusions about performance of any of the diagnostic methods to be drawn. Further, no assessments addressed diagnosis on root surfaces. For posterior permanent teeth the evidence was unevenly distributed among methods, with assessments of radiographic methods accounting for over half of all reports. The assessments were also unevenly distributed with respect to the type of lesion to be identified, with just over half diagnosing lesions penetrating into dentin, and only 10 examining diagnosis of enamel lesions. The small number of assessments of a specific application of a diagnostic method effectively precluded any definitive conclusions about sensitivity and specificity levels for most applications. When a larger number of assessments was available, the extent of the variation among assessments and, occasionally, the collective quality of the evidence available led to the same result.

Two potential weaknesses of our review should be noted. First, data from the included studies were abstracted for inclusion in the evidence table by a single abstractor, whose efforts were reviewed but not duplicated by a second abstractor. Possibly the absence of an independent second abstraction resulted in errors being introduced into the evidence table. Second, we used a composite scale to assess study quality, and the scores were assigned by a single reviewer. Use of composite scores is controversial in that currently there is no consensus on scale items or their weights. Thus, different scales may well produce different quality ratings. In addition, without independent assignment of scores by a second reviewer, it is possible that reviewer bias was introduced. We note that the scores were not used in any synthesis of the results, and figured only in the overall assessment of quality for one diagnostic method. Thus, any overall effect of the quality scores on the results of the review must be viewed as limited.

Limitations in the Evidence Base. The literature describing the histologically determined validity of methods for diagnosing carious lesions has a variety of limitations, many of which represent potentially serious threats to internal validity, and most of which represent barriers to generalization of the reported results to dental practice. The most obvious limitation has already been noted, the virtual absence of any assessments of diagnostic methods applied to primary teeth, and to root surfaces of permanent teeth. The breadth of reported studies also seriously restricts any conclusions about differences in the validities of visual and visual/tactile examinations, and possible advantages of combining examination methods. A minority of method/surface/lesion type combinations are represented by more than three studies, and for most of these combinations the variation among reported performances is extensive for either sensitivity or specificity.

Although we considered meta-analyzing the results for several applications, two characteristics of the assessments discourage the use of meta-analysis. First, not all of the assessments within an application category are independent evaluations. Many reflect the use of common examiners and sample teeth, but slightly different techniques. Second, the studies do not all assess the same "outcomes," since criteria for diagnosis are

TABLE 5
Studies of Radiographic Methods for Detection of Carious Lesions *[cont. on page 209]*

Source (Ref.)	Setting	Sites (*N*)	Teeth/ Site*	Specific Method†	Examiner (*N*)	Lesion Prevalence	Quality Score‡	Sensitivity (SD)	Specificity (SD)
Dentinal lesions on occlusal surfaces									
Wenzel et al., 1990 (23)	in vitro	46	all post/surf	D	4	72%	40	.64¶	.94¶
Wenzel et al., 1990 (23)	in vitro	46	all post/surf	di(D)	2	72%	40	.68¶	.98¶
Wenzel et al., 1991 (29)	in vitro	81	3rd molar/surf	E	4	67%	45	.63¶	.85¶
Wenzel et al., 1991 (29)	in vitro	81	3rd molar/surf	di(E)	4	67%	45	.72¶	.83¶
Wenzel et al., 1991 (29)	in vitro	81	3rd molar/surf	di(E)/edge	4	67%	45	.62¶	.83¶
Wenzel et al., 1991 (29)	in vitro	81	3rd molar/surf	RVG	4	67%	45	.69¶	.84¶
Wenzel et al., 1991 (29)	in vitro	81	3rd molar/surf	RVG/den	4	67%	45	.64¶	.82¶
Wenzel, Fejerskov, 1992 (14)	in vitro	78	3rd molar/surf	E	1	67%	45	.48§	.81§
Wenzel, Fejerskov, 1992 (14)	in vitro	78	3rd molar/surf	di(E)	1	67%	45	.71§	.85§
Wenzel, Fejerskov, 1992 (14)	in vitro	78	3rd molar/surf	di(E)edge	1	67%	45	.55§	.77§
Nytun et al, 1992 (13)	in vitro	30	all post/surf	NR	10	77%	40	.66¶	.50¶
Kelly, Holt, 1993 (11)	in vitro	100	1st molar/surf	D	1	51%	35	.67§	.92§
Kelly, Holt, 1993 (11)	in vitro	100	2nd molar/surf	D	1	44%	35	.93§	.89§
Russell, Pitts, 1993 (30)	in vitro	120	all post/surf	RVG	3	28%	50	.21¶	.97¶
Russell, Pitts, 1993 (30)	in vitro	120	all post/surf	D	3	28%	50	.18¶	.98¶
Russell, Pitts, 1993 (30)	in vitro	120	all post/surf	E	3	28%	50	.21¶	.99¶
Lussi, 1993 (15)	in vitro	63	all post/surf	D	24	44%	45	.45¶	.83¶
Verdonshot et al. 1993 (16)	in vitro	81	3rd molar/surf	E	4	67%	30	.61¶	.79¶
Lussi et al., 1995 (31)	in vitro	26	3rd molar/surf	D	6	42%	50	.62¶	.77¶
Ricketts et al., 1994 (32)	in vitro	48	all molar/surf	D	12	67%	40	.62¶	.76¶
Ekstrand et al., 1997 (18)	in vitro	100	all post/surf	E	3	39%	60	.54¶	100¶
Huysmans et al., 1997 (33)	in vitro	198	all post/surf	Digora	3	55%	65	.60¶	.94¶
Ricketts et al., 1997 (34)¶	in vitro	96	all post/surf	D	5	39%	60	.14¶	.95¶
Ashley et al., 1998 (19)	in vitro	103	all post/surf	E	1	36%	60	.24§	.89§
Ashley et al., 1998 (19)	in vitro	103	all post/surf	Digora	1	36%	60	.19§	.89§
Huysmans et al., 1998 (20)	in vitro	96	all post/surf	E	2	41%	60	.58¶	.87¶
Cortes et al., 2000 (21)	in vitro	59	molars/ss	E	1	32%	55	.84¶	.83¶
Enamel lesions on occlusal sufaces									
Wenzel et al., 1990 (23)	in vitro	46	all post/surf	D	4	17%	40	.44¶	.70¶
Wenzel et al., 1990 (23)	in vitro	46	all post/surf	di(D)	2	17%	40	.31¶	.72¶
Ashley et al., 1998 (19)	in vitro	103	all post/surf	Digora	1	25%	55	.24§	.80§
Ashley et al., 1998 (19)	in vitro	103	all post/surf	E	1	25%	55	.19§	.80§
Any lesions on occlusal surfaces									
Wenzel et al., 1990 (23)	in vitro	46	all post/surf	D	4	89%	40	.79¶	.90¶
Wenzel et al., 1990 (23)	in vitro	46	all post/surf	di(D)	2	89%	40	.73¶	.80¶
Russell, Pitts, 1993 (30)	in vitro	120	all post/surf	RVG	3	¶	50	.12¶	.96¶
Russell, Pitts, 1993 (30)	in vitro	120	all post/surf	D	3	¶	50	.12¶	.95¶
Russell, Pitts, 1993 (30)	in vitro	120	all post/surf	E	3	¶	50	.15¶	.97¶
Lazarchik et al., 1995 (35)	in vitro	100	all post/surf	D	15	79%	50	.58 (.12)	.70 (.16)
Ricketts et al., 1997 (34)¶	in vitro	96	all post/surf	D	5	70%	60	.27¶	.77¶
Cavitated lesions on proximal surfaces									
Rugg-Gunn, 1972 (36)	in vivo	370	all post/surf°°	D	¶	9%	60	.35¶	.70¶
Downer, 1975 (9)	in vivo	185	all post/surf	D	¶	36%	70	.73¶	.97¶
Mejare et al, 1985 (27)	in vivo	598	all post/surf	D	3**	5%	65	.36¶	.98¶
Pitts, Rimmer, 1992 (37)	in vivo	1,468	all post/surf	D	1	1%	60	.87‡	.99‡
Pitts, Rimmer, 1992 (37)	in vivo	775	pri post/surf	D	1	¶	60	.99‡	.87‡
Hintze et al., 1998 (28)	in vivo	338	all post/surf	E	4	6%	70	.63¶	.93¶
Espelid, Tveit, 1986 (38)	in vitro	151	all post/surf	D	7	19%	55	.69¶	.98¶
Enamel lesions on proximal surfaces									
White, Yoon, 1997 (41)	in vitro	320	ant&post/surf	Shick	10	50%	60	.35¶	.80¶
White, Yoon, 1997 (41)	in vitro	320	ant&post/surf	E	10	50%	60	.46¶	.76¶
White, Yoon, 2000 (42)‡‡	in vitro	¶	all post/surf	E	12	50%‡‡	55	.42¶	.78¶

TABLE 5 *[cont. from page 208]*
Studies of Radiographic Methods for Detection of Carious Lesions

Source (Ref.)	Setting	Sites (*N*)	Teeth/Site*	Specific Method†	Examiner (*N*)	Lesion Prevalence	Quality Score‡	Sensitivity (SD)	Specificity (SD)
Dentinal lesions on proximal surfaces									
Mileman, van de Weele, 1990 (39)	in vitro	105	all post/surf	D	276	43%	60	.54 (.14)	.97 (.05)
Verdonschot et al., 1991(40)	in vitro	21	mand post/surf	D	3	¶	35	.50¶	.94¶
Russell, Pitts, 1993 (30)	in vitro	240	all post/surf	RVG	3	¶	50	.16¶	.96¶
Russell, Pitts, 1993 (30)	in vitro	240	all post/surf	D	3	¶	50	.29¶	.92¶
Russell, Pitts, 1993 (30)	in vitro	240	all post/surf	E	3	¶	50	.30¶	.96¶
Ricketts et al., 1997 (34)¶	in vitro	¶#	all post/surf	D	5	15%	60	.16¶	.99¶
White, Yoon, 1997 (41)	in vitro	320	ant&post/surf	Shick	10	25%	60	.52¶	.95¶
White, Yoon, 1997 (41)	in vitro	320	ant&post/surf	E	10	25%	60	.58¶	.94¶
White, Yoon, 2000 (42)**	in vitro	¶**	all post/surf	E	12	25%**	55	.63¶	.92¶
Any lesions on proximal sufaces									
Heaven et al., 1992 (43)	in vitro	16	pre/surf	di(D) auto read††	1	75%	20	100§	100§
Russell, Pitts, 1993 (30)	in vitro	240	all post/surf	E	3	¶	50	.25¶	.90¶
Russell, Pitts, 1993 (30)	in vitro	240	all post/surf	D	3	¶	50	.26¶	.90¶
Russell, Pitts, 1993 (30)	in vitro	240	all post/surf	RVG	3	¶	50	.15¶	.92¶
Ricketts et al., 1997 (34)¶	in vitro	¶#	all post/surf	D	5	37%	60	.27¶	.97¶
Firestone et al., 1998 (44)	in vitro	102	pre/surf	Caries Finder	1	66%	40	.78§	.74§
Firestone et al., 1998 (44)	in vitro	102	pre/surf	D	16	66%	50	.61 (.18)	.86 (.20)
Firestone et al., 1998 (44)	in vitro	102	pre/surf	di(D)	16	66%	50	.73 (.11)	.82 (.21)
White, Yoon, 1997 (41)	in vitro	320	ant&post/surf	Shick	10	75%	55	.49¶	.80¶
White, Yoon, 1997 (41)	in vitro	320	ant&post/surf	E	10	75%	55	.58¶	.76¶
Huysmans et al., 1997 (33)	in vitro	410	can&post/surf	Digora	3	36%	70	.33¶	.76¶
White, Yoon, 2000 (42)**	in vitro	¶**	all post/surf	E	12	75%**	55	.54¶	.78¶

*Teeth/site abbreviations: post=posterior teeth; ant=anterior teeth; can=canine teeth; surf=entire surface; p&f=pit and fissure sites; ss=specific sites.
†Film speed or image type: assume visual reading unless indicated: di=digital imaging; edge=edge enhancement; den=density manipulation.
‡Quality score is based on a scale of 0 to 100.
¶Not reported.
§Not applicable.
°°All posterior proximal surfaces out of contact for less than one year.
• Consensus of three examiners.
96 teeth, surfaces not reported, teeth sectioned and reassembled before assessment.
**80 teeth, surfaces not reported.
††Automatic image analysis of digitized image of D-speed film.

different in different assessments of the same application. These differences in criteria help explain some of the heterogeneity in the assessments; nevertheless, the usefulness of any synthesized estimate of performance will be limited by the technique/criteria dependency of the diagnostic method.

The quality scores for these studies tended to cluster in the lower middle to middle of the 0–100 scale of possible scores. The mean for in vivo studies was 61; for in vitro studies, 46. The range of scores was 5–75. Lower scoring studies typically had several features that represented threats to either internal or external validity. Most studies had sample sizes of 75 or more sites or surfaces, but the choice of sites rather than surfaces may pose a threat to external validity because most occlusal surfaces will present multiple sites for assessment. The results of site assessment do not summarize the status of the entire surface, as is routinely done in clinical practice. As noted, most of the studies were performed in vitro, a practical necessity if histological validation is to be accomplished easily. However, systematic differences between results obtained in vivo and in vitro have been reported (4). Many in vivo studies were limited to premolars and/or third molars, where extraction can be scheduled, and where the teeth are in good clinical condition prior to extraction. In vitro studies also often relied on these teeth for the same reason, that they are more frequently available with unrestored, noncarious crowns. The problem is that the teeth most frequently experiencing occlusal and proximal surface decay—the first and second molars—differ from the premolars and third molars in ways that may affect the performance of diagnostic methods. For example, occlusal surfaces of third molars tend to have more fissures (53), which are often less well coalesced. In addition, proximal enamel thicknesses, both bucco-lingually and mesio-distally are usually less in both premolars and third molars.

A variety of histological validation

TABLE 6
Studies of Electrical Conductance Methods for Detection of Carious Lesions

Source (Ref.)	Setting	Sites (*N*)	Teeth/ Site*	Specific Method	Examiner (*N*)	Lesion Prevalence	Quality Score‡	Sensitivity (SD)	Specificity (SD)
Dentinal lesions on occlusal surfaces									
Lussi et al., 1995 (31)	in vivo	41	3rd molar/ss	ECM	1	37%	40	.93‡	.77‡
Ricketts et al., 1995 (12)	in vivo	100	3rd molar/ss	Vanguard	1	30%	45	.97‡	.56‡
Verdonschot et al., 1993 (16)	in vitro	81	3rd molar/p&f	Vanguard	4	67%	30	.67¶	.82¶
Ricketts et al., 1995 (16)	in vitro	100	3rd molar/ss	Vanguard	1	30%	45	.93‡	.63‡
Ricketts et al., 1995 (16)	in vitro	100	3rd molar/ss	Caries L	1	30%	45	.96‡	.62‡
Ricketts et al., 1997 (45)*	in vitro	76	3rd molar/ss	ECMII-10 l/min§	1	32%	20	.92‡	.89‡
Ricketts et al., 1997 (46)†	in vitro	76	3rd molar/ss	ECMII-stable∞	1	32%	25	.92‡	.89‡
Ricketts et al., 1997 (47)‡	in vitro	96	all post/surf	ECMII•	1	39%	40	.76‡	.76‡
Ekstrand et al., 1997 (18)	in vitro	100	all post/p&f	ECM	3	39%	60	.90¶	.85¶
Huysmans et al., 1998 (20)	in vitro	107	all post/p&f	ECM	2	41%	50	.58¶	.94¶
Huysmans et al., 1998 (20)	in vitro	107	all post/p&f	ECM w/gel	2	41%	55	.76¶	.90¶
Huysmans et al., 1998 (20)	in vitro	107	all post/p&f	Caries L	2	41%	60	.78¶	.79¶
Ashley et al., 1998 (19)	in vitro	103	all post/surf	ECMIIb	1	36%	55	.78‡	.80‡
Lussi et al., 1999 (48)	in vitro	105	all post/p&f	ECM	1	36%	35	.92‡	.78‡
Ashley, 2000 (22)0)	in vitro	58	pri molar/surf	ECM	1	64%	45	.90‡	.81‡
Any lesions on occlusal surfaces									
Ricketts et al., 1995 (12)	in vivo	100	3rd molar/ss	Vanguard	1	64%	40	.81‡	.78‡
Rock, Kidd, 1988 (49)	in vitro	50	all pos//ss	Vanguard	1	74%	35	.70‡	.85‡
Ricketts et al., 1995 (12)	in vitro	100	3rd molar/ss	Vanguard	1	64%	40	.70‡	.83‡
Ricketts et al., 1995 (12)	in vitro	100	3rd molar/ss	Caries L	1	64%	40	.74‡	.74‡
Ricketts et al., 1996 (50)	in vitro	30	3rd molar/ss	ECM	1	80%	50	.92‡	100‡
Ricketts et al., 1997 (45)*	in vitro	76	3rd molar/ss	ECMII-10 l/min§	1	64%	15	.71‡	.78‡
Ricketts et al., 1997 (46)†	in vitro	76	3rd molar/ss	ECMII-stable∞	1	64%	20	.61‡	.96‡
Ricketts et al., 1997 (47)‡	in vitro	96	all post/surf	ECMII•	1	78%	40	.61‡	.86‡
Enamel lesions on occlusal sufaces									
Ashley et al., 1998 (19)	in vitro	103	all post/surf	ECMII†	1	24%	50	.65‡	.73‡

*Teeth/site abbreviations: post=posterior teeth; surf=entire surface; p&f=pit and fissure sites; ss=specific sites; pri=primary tooth.
†Quality score is based on a scale of 0 to 100.
‡Not applicable.
¶Not reported.
§Airflow 1 liter/minute.
∞Stable reading.
•With gel coating occlusal surface.

methods are represented in the included studies, with little assurance that different methods are equivalent (54-56). A little less than one-half of the studies relied on light microscopy, with an identical number using other methods for evaluating the extent of caries on sectioned teeth. The remainder used visual criteria to confirm cavitated lesions. Further, a majority of studies supplied no explicit criteria for the validation, and a large majority did not report reliability information for the validation despite known variability in this procedure (55). The extent to which differences in validation methods contribute to differences in reported diagnostic performances is unknown.

The percent of surfaces evaluated that actually included a carious lesion was less than 20 percent in only 5 percent of in vitro assessments compared to 53 percent of in vivo assessments. Further, most of the in vivo sites with lesion prevalences above 20 percent were selected sites on third molars. It is rare to encounter a patient with detectable lesions on more than 20 percent of all surfaces (a DS score of 40 in a fully dentate individual) in US clinical practice, let alone a population of such patients. Elevated frequencies of occurrence in assessment samples raise issues about examiner bias, as unusual presentations may alter examiner alertness and behavior, albeit in an unknown manner. The criteria used for selecting teeth for the samples also raise issues about both the comparability of studies and the generalization of results to clinical practice. The criteria described for selection, which were intended to minimize possible confounding factors such as noncarious enamel defects and restorations, were dissimilar across studies, and certainly are not representative of the range of presentations encountered in clinical practice. When this limitation is coupled with the previously noted limitations due to restricted tooth types and the use of sites rather than surfaces, generalization to clinical practice is again problematic. It should be noted that, ideally, the sample for any given assessment should be a spe-

TABLE 7
Studies of Other Methods for Detection of Carious Lesions

Source (Ref.)	Setting	Sites (*N*)	Teeth/Site*	Specific Method	Examiner (*N*)	Lesion Prevalence	Quality Score†	Sensitivity (SD)	Specificity (SD)
FOTI method, cavitated lesions on proximal surfaces									
Hintze et al., 199 (38)	in vivo	338	all post/prox	FOTI	4	6%	70	.04‡	100‡
FOTI method, dentinal lesions on occlusal surfaces									
Ashley et al., 1998 (19)	in vitro	103	all post/surf	FOTI	1	36%	60	.14¶	.95¶
Cortes et al., 2000 (21)	in vitro	59	molars/ss	FOTI	4	32%	65	.74‡	.85‡
FOTI method, enamel lesions on occlusal surfaces									
Ashley et al., 1998 (19)	in vitro	103	all post/surf	FOTI	1	24%	55	.21¶	.95¶
Laser fluorescence method, dentinal lesions on occlusal surfaces									
Lussi et al., 1999 (48)§	in vitro	105	all post/p&f	LF-moist	1	36%	30	.76¶	.87¶
Lussi et al., 1999 (48)§	in vitro	105	all post/p&f	LF-dry	1	36%	30	.84¶	.79¶
Shi et al., 2000 (51)	in vitro	76	all post/p&f	LF/wet	1	39%	40	.80¶	1.00¶
Shi et al., 2000 (51)	in vitro	76	all post/p&f	LF/dry	1	39%	40	.82¶	1.00¶
Laser fluorescence method, enamel lesions on occlusal surfaces									
Shi et al., 2000 (51)	in vitro	76	all post/p&f	LF/wet	1	34%	40	.42¶	.95¶
Shi et al., 2000 (51)	in vitro	76	all post/p&f	LF/dry	1	34%	40	.46¶	.95¶
Combination method, dentinal lesions on occlusal surfaces									
Nytun et al, 1992 (12)	in vitro	30	all post/surf	visual/radio	10	77%	40	.86‡	.64‡
Lussi, 1993	in vitro	63	all post/surf	visual/radio	10	44%	45	.49‡	.87‡
Cayley, Holt, 1997 (52)	in vitro	60	all post/surf	visual/radio	11	‡	55	.65 (.11)	.74 (.09)

*Teeth/site abbreviations: post=posterior teeth; surf=entire surface; p&f=pit and fissure sites; ss=specific sites; mand=mandibular; pre=premolar.
†Quality score is based on a scale of 0 to 100.
‡Not reported.
¶Not applicable.
§Studies are identical with exception of moisture status of teeth.

cific tooth type—e.g., lower first molars—and that multiple assessments for each tooth type would be available from which to determine performance. While such an approach would eliminate the inevitable variation among studies introduced by simple differences in proportional representation of tooth types within samples, it would increase the complexity of reporting and require larger samples.

Only one-half of the studies reported the combined performance of four or more evaluators. The studies relying on a smaller number of evaluators may present difficulties in generalization due to positive influence of particularly skilled investigators/examiners. Single-examiner studies, of which there were 17 (46 percent of studies), are especially vulnerable to this phenomenon. Further, 17 studies reported no reliability information for the examiner(s). When interexaminer reliability was reported, the values often were low enough to underscore the threat to external validity represented by the use of single examiners.

Finally, 14 of the assessments described post hoc determination of the optimal criteria for lesion designation. While the development of new diagnostic techniques requires such analyses, it is usually expected that the criteria will then be tested in a second "validation" sample. Such a procedure was not reported in any of the 14 assessments.

Recommendations for Future Research. Our review revealed two principal shortcomings of the existing literature describing histologically validated assessments of diagnostic methods. First, the coverage is spotty in terms of combinations of lesion types, tooth surfaces, tooth types, and diagnostic methods for which assessments are available. Second, the literature is characterized by designs that are open to threats to internal validity and are problematic in terms of external validity. Efforts must be made to increase the number of assessments for all specific applications except visual-tactile methods, where concern over harms militate against use of the method (57-59), and at the same time address the design characteristics that limit the applicability of existing studies. Merely acquiring additional studies similar to those currently available is at best an inefficient approach to advance our understanding of the performance of methods for dental caries diagnosis.

Perhaps one of most limiting aspects of the literature is that a majority of the studies have been performed in vitro. This characteristic of current research is eminently understandable for practical reasons; however, in vitro studies have serious limitations. They are more difficult to generalize to the environment of dental practice for several reasons, not the least of which is that they permit careful selection of individual teeth or surfaces for assessment, rather than forcing the inclusion of a more representative set of teeth or surfaces. In vitro studies also minimize many limitations imposed by working within the oral cavity, arguably leading to improved perform-

ance, and they tend to emphasize certain tooth types. Because most in vivo studies also have had this latter limitation, rethinking the source of research material for in vivo studies of dental caries diagnosis may be necessary. One possibility might be to conduct such studies post-mortem, although this approach is not strictly speaking in vivo and would tend to overrepresent elderly subjects. Another possibility might be to expend more effort recruiting subjects from among patients in dental care systems where extractions are part of planned treatment. The former approach would facilitate the use of multiple examiners, thereby addressing another substantial limitation of the current literature.

Attention to the method used for histological validation of the sample is also necessary. Work is needed to determine an acceptable standard technique for determination of the presence of a carious lesion from among techniques based on microscopy, stereomicroscopy, and microradiography. Standards for sectioning method and thickness, number of sections surveyed, magnification, dyes, and criteria for identification of a lesion all must be identified. Minimum expectations for number of examiners, reliability, and reporting methods should also be specified as a part of the standard. A conference or workshop of invited experts would represent a possible mechanism for setting standards.

We found the majority of studies deficient in terms of complete descriptions of important study characteristics, including the criteria for positive diagnoses of carious lesions, the criteria for selection of the sample of teeth or surfaces to be diagnosed, the background and training of the examiners, and examiner reliability. All reports should include a minimum set of descriptions in a standard format to facilitate comparisons among studies. The development of such a standard could be undertaken by one or more dental organizations sponsoring journals in which carious lesion diagnostic studies appear. The standards might be developed along the lines of the CONSORT statement for reports of clinical trials (60).

The issues of outcome measures and disease prevalence in diagnostic studies should be addressed in the standards document. This review included only studies reporting outcomes in terms of sensitivity and specificity. This inclusion criterion was limiting in that some studies reporting results in terms of areas under receiver-operating characteristic (ROC) curves were excluded. ROC results permit comparison of studies on the basis of a single number reflecting the trade-off between sensitivity and specificity across a range of examiner confidence levels, but require collection of additional information from examiners regarding their certainty for each diagnostic decision. Although the utility of providing this type of outcome data for dental caries diagnosis studies has yet to be demonstrated, the standards should address the circumstances in which one or both outcomes might be reported. The prevalence of carious lesions in the samples of studies included in the review represents a barrier to generalization of the results of these studies, if not a threat to internal validity. Prevalence of lesions in any in vivo or in vitro sample should be reasonably representative of the population prevalence for the same type of lesion on the same surfaces.

Once a set of standards is in place to guide investigators in designing studies and preparing reports that will facilitate the assessment of the validities of diagnostic methods for carious lesions, some attention to the coverage of those assessments will be beneficial. Clearly, more assessments of newer methods are necessary. FOTI and digital radiographic methods are two obvious candidates. EC methods also would benefit from stronger assessments, and laser fluorescence methods also will require more assessment in the near future. Equally important, studies must include assessments on primary teeth and on root surfaces of permanent teeth.

Because identification of a lesion at one examination may not furnish sufficient information to provide an accurate assessment of the lesion's activity status and prognosis, consideration must be given to how various diagnostic methods facilitate longitudinal assessment of changes in lesion volume. Finally, diagnostic studies must begin to evaluate more than just the immediate outcomes of the use of the particular method or methods being assessed. While the validity of diagnosis must be the principal concern in such assessments, some attention should be paid to the outcome in terms of the appropriateness of the treatment provided in response to the diagnosis. These longer-term considerations represent the ultimate outcomes of diagnostic procedures. To this end the procedures must be evaluated in terms of their benefit to patients, an outcome mediated by dentists' application of the information provided by the diagnostic method.

Acknowledgments

We thank Jacqueline Besteman and Tina Murray, AHRQ, and Isabel Garcia, NIDCR, for their administrative expertise. We thank the Technical Expert Advisory Committee members, Craig Amundson, Ken Anusavice, Brian Burt, John Featherstone, David Pendrys, Nigel Pitts, and Jane Weintraub, and our consultants, Jan Clarkson, Amid Ismail, Gary Rozier, and Alex White, for their helpful comments throughout the review process. Finally, we thank Kathy Lohr and Jessica Nelson at RTI and Anne Jackman, Lynn Whitner, Donna Curasi, Sally Mauriello, Teg Hughes, and Jessica Lee at UNC for their invaluable assistance with the project.

References

1. Bader J, Shugars D. Variation in dentists' clinical decisions. J Public Health Dent 1995;55:181-8.
2. Kidd E. Occlusal caries diagnosis: a changing challenge for clinicians and epidemiologists. J Dent 1993;21:323-31.
3. Ie Y, Verdonschot E. Performance of diagnostic systems in occlusal caries detection compared. Community Dent Oral Epidemiol 1994;22:187-91.
4. van Rijkom H, Verdonschot E. Factors involved in validity measurements of diagnostic tests for approximal caries—a meta-anlysis. Caries Res 1995;29:364-70.
5. Pine C, ten Bosch J. Dynamics and diagnostic methods for detecting small carious lesions Caries Res 1996;30:381-8.
6. Verdonschot E, van de Rijke J, Brouwer W, ten Bosch J, Truin G. Optical quantitation and radiographic diagnosis of incipient approximal caries lesions. Caries Res 1991;25:359-64.
7. Woodward G, Leake J. The use of dental radiographs to estimate the probability of cavitation of various interproximal lesions. Part I. Evidence from the literature. J Can Dent Assoc 1996;62:731-6.
8. Swets J. ROC analysis applied to the evaluation of medical imaging techniques. Invest Radiol 1979;14:109-21.
9. Downer M. Concurrent validity of an epidemiological diagnostic system for caries with the histological appearance of extracted teeth as validating criterion. Caries Res 1975;9:231-46.
10. Downer M, O'Mullane D. A comparison of the concurrent validity of two epidemiologic diagnostic systems for caries evaluation. Community Dent Oral Epidemiol 1975;3:20-4.
11. Ketley C, Holt R. Visual and radio-

graphic diagnosis of occlusal caries in first permanent molars and in second primary molars. Br Dent J 1993;174:364-70.
12. Ricketts D, Kidd E, Wilson R. A reevaluation of electrical resistance measurements for the diagnosis of occlusal caries. Br Dent J 1995;178:11-17.
13. Nytun R, Raadal M, Espelid I. Diagnosis of dentin involvement in occlusal caries based on visual and radiographic examination of the teeth. Scand J Dent Res 1992;100:144-8.
14. Wenzel A, Fejerskov O. Validity of diagnosis of questionable caries lesions in occlusal surfaces of extracted third molars. Caries Res 1992;26:188-94.
15. Lussi A. Comparison of different methods for the diagnosis of fissure caries without cavitation. Caries Res 1993;27: 409-16.
16. Verdonschot E, Wenzel A, Truin G, Konig K. Performance of electrical resistance measurements adjunct to visual inspection in the early diagnosis of occlusal caries. J Dent 1993;21:332-7.
17. Deery C, Fyffe H, Nugent Z, Nuttall N, Pitts N. The effect of placing a clear pit and fissure sealant on the validity and reproducibility of occlusal caries diagnosis. Car Res 1995;29:377-81.
19. Ashley P, Blinkhorn A, Davies R. Occlusal caries diagnosis: an in vitro histological validation of the Electronic Caries Monitor (ECM) and other methods. J Dent 1998;26:83-8.
20. Huysmans M, Longbottom C, Pitts N. Electrical methods in occlusal caries diagnosis: an in vitro comparison with visual inspection and bite-wing radiography. Caries Res 1998;32:324-9.
21. Cortes D, Ekstrand K, Elias-Boneta A, Ellwood R. An in vitro comparison of the ability of fibre-optic transillumination and radiographs to detect occlusal caries and evaluate lesion depth. Caries Res 2000;34:443-7.
22. Ashley P. Diagnosis of occlusal caries in primary teeth. Int J Paediatr Dent 2000;10:166-71.
23. Wenzel A, Fejerskov O, Kidd E, Joyston-Bechal S, Groeneveld A. Depth of occlusal caries assessed clinically, by conventional film radiographs, and by digitized, processed radiographs. Caries Res 1990;24:327-33.
24. Lussi A. Validity of diagnostic and treatment decisions of fissure caries. Cares Res 1991;25:296-303.
25. Fyffe H Deery C, Nugent Z, Nuttall N, Pitts N. In vitro validity of the Dundee selectable threshold method for caries diagnosis (DSTM). Community Dent Oral Epidemiol 2000;28:52-8.
26. Penning C, van Amerongen J, Seef R, ten Cate J. Validity of probing for fissure caries diagnosis. Caries Res 1992;26:445-9.
27. Mejare I, Grondahl H, Carlstedt K, Grever A, Ottosson E. Accuracy at radiography and probing for the diagnosis of proximal caries. Scand J Dent Res 1985; 93:178-84.
28. Hintze H, Wenzel A, Danielsen B, Nyvad B. Reliability of visual examination, fibre-optic transillumination, and bite-wing radiography, and reproducibility of direct visual examination following tooth separation for the identification of cavitated carious lesions in contacting approximal surfaces. Caries Res 1998;32: 204-9.
29. Wenzel A, Hintze H, Mikkelsen L, Mouyen F. Radiographic detection of occlusal caries in noncavitated teeth. A comparison of conventional film radiographs, digitized film radiographs, and RadioVisioGraphy. Oral Surg Oral Med Oral Pathol 1991;72:621-6.
30. Russell M, Pitts N. Radiovisiographic diagnosis of dental caries: initial comparison of basic mode videoprints with bitewing radiography. Caries Res 1993; 27:65-70.
31. Lussi A, Firestone A, Schoenberg V, Hotz P, Stich H. In vivo diagnosis of fissure caries using a new electrical resistance monitor. Caries Res 1995;29:81-7.
32. Ricketts D, Kidd E, Smith B, Wilson R. Radiographic detection of occlusal caries: effect of X-ray beam factors on diagnosis. Eur J Prosthodont Restor Dent 1994;2:149-54.
33. Huysmans M, Hintze H, Wenzel A. Effect of exposure time on in vitro caries diagnosis using the Digora system. Eur J Oral Sci 1997;105:15-20.
34. Ricketts D, Whaites E, Kidd E, Brown J, Wilson R. An evaluation of the diagnostic yield from bitewing radiographs of small approximal and occlusal carious lesions in a low prevalence sample in vitro using different film types and speeds. Br Dent J 1997;182:51-8.
35. Lazarchik D, Firestone A, Heaven T, Filler S, Lussi A. Radiographic evaluation of occlusal caries: effect of training and experience. Caries Res 1995;29:355-8.
36. Rugg-Gunn A. Approximal carious lesions. A comparison of the radiological and clinical appearances. Br Dent J 1972; 133:481-4.
37. Pitts N, Rimmer P. An in vivo comparison of radiographic and directly assessed clinical caries status of posterior approximal surfaces in primary and permanent teeth. Caries Res 1992;26:146-52.
38. Espelid I, Tveit A. Clinical and radiographic assessment of approximal carious lesions. Acta Odontol Scand 1986;44: 31-7.
39. Mileman P, van der Weele L. Accuracy in radiographic diagnosis: Dutch practitioners and dental caries. J Dent 1990;18: 130-6.
40. Verdonschot E, Bronkhorst E, Wenzel A. Approximal caries diagnosis using fiberoptic transillumination: a mathematical adjustment to improve validity. Community Dent Oral Epidemiol 1991;19: 329-32.
41. White S, Yoon D. Comparative performance of digital and conventional images for detecting proximal surface caries. Dent Maxillofac Radiol 1997;26:32-8.
42. White S, Yoon D. Comparison of sensitometric and diagnostic performance of two films. Compend Contin Educ Dent 21;]2000:530-2, 534 536.
43. Heaven T, Firestone A, Feagin F. Computer-based image analysis of natural approximal caries on radiographic films. J Dent Res 1992;71:846-9.
44. Firestone A, Sema D, Heaven T, Weems R. The effect of a knowledge-based, image analysis and clinical decision support system on observer performance in the diagnosis of approximal caries from radiographic images. Caries Res 1998;32: 127-34.
45. Ricketts D, Kidd E, Wilson R. The effect of airflow on site-specific electrical conductance measurements used in the diagnosis of pit and fissure caries in vitro. Caries Res 1997;31:111-18.
46. Ricketts D, Kidd E, Wilson R. The electronic diagnosis of caries in pits and fissures: site-specific stable conductance readings or cumulative resistance readings? Caries Res 1997;31:119-24.
47. Ricketts D, Kidd E, Wilson R. Electronic diagnosis of occlusal caries in vitro: adaptation of the technique for epidemiological purposes. Community Dent Oral Epidemiol 1997;25:238-41.
48. Lussi A, Imwinkelried S, Pitts N, Longbotton C, Reich E. Performance and reproducibility of a laser fluorescence system for detection of occlusal caries in vitro. Caries Res 1999;33:261-6.
49. Rock W, Kidd E. The electronic detection of demineralisation in occlusal fissures. Br Dent J 1988;164:243-7.
50. Ricketts D, Kidd E, Liepins P, Wilson R. Histological validation of electrical resistance measurements in the diagnosis of occlusal caries. Caries Res 1996;30:148-55.
51. Shi X, Welander W, Angmar-Mansson B. Occlusal caries detection with KaVo DIAGNOdent and radiography: an in vitro comparison. Caries Res 2000;34:151-8.
52. Cayley A, Holt R. The influence of audit on the diagnosis of occlusal caries. Caries Res 1997;31:97-102.
53. Brand R, Isselhard D. Anatomy of orofacial structures. 6th ed. St. Louis: Mosby, 1988.
54. Wenzel A, Verdonschot EH, Truin GJ, Konig KG. Impact of the validator and the validation method on the outcome of occlusal caries diagnosis. Caries Res 1994;28:373-7.
55. Hintze H, Wenzel A, Larsen M. Stereomicroscopy, film radiography, microradiography and naked-eye inspection of tooth sections as validation for occlusal caries diagnosis. Caries Res 1995;29:359-63.
56. Wenzel A, Hintze H. The choice of gold standard for evaluating tests for caries diagnosis. Dentomaxillofac Radiol 1999; 28:132-6.
57. Ekstrand K, Qvist V, Thylstrup A. Light microscope study of the effect of probing in occlusal surfaces. Caries Res 1987;21: 368-74.
58. van Dorp C, Exterkate R, ten Cate J. The effect of dental probing on subsequent enamel demineralization. ASDC J Dent Child 1988;55:343-7.
59. Yassin O. In vitro studies of the effect of a dental explorer on the formation of an artificial carious lesion. ASDC J Dent Child 1995;62:111-17.
60. Moher D, Schultz K, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials JAMA 2001;285:1987-91.