Article:

Gimenez, T., Piovesan, C., Braga, M.M. et al. (5 more authors) (2015) Clinical Relevance of Studies on the Accuracy of Visual Inspection for Detecting Caries Lesions: A Systematic Review. Caries Research, 49 (2). pp. 91-98. ISSN 0008-6568

https://doi.org/10.1159/000365948

Visual inspection for caries detection: systematic review and meta-analysis

T. Gimenez[a], C. Piovesan[a,b], M. M. Braga[a], D. P. Raggio[a], C. Deery[c], D. N. Ricketts[d], K. R. Ekstrand[e], F. M. Mendes[a]

[a]Department of Pediatric Dentistry, School of Dentistry, University of São Paulo, São Paulo, Brazil; [b] School of Dentistry, Centro Universitário Franciscano, Santa Maria, Brazil [c] School of Clinical Dentistry, University of Sheffield, Sheffield, UK; [d] Dundee Dental Hospital and School, University of Dundee, Dundee, UK; [e] Department of Cariology and Endodontics and Clinical Genetics, School of Dentistry, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

Running head: Visual inspection for caries detection

Key-words: visual inspection, dental caries, systematic review, performance

Corresponding author

Fausto Medeiros Mendes

Faculdade de Odontologia da Universidade de São Paulo

Av. Lineu Prestes, 2227

São Paulo – SP – Brazil

05508-000

Tel: 55 11 3091-7835 Fax: 55 11 3091-7854

E-mail: fmmendes@usp.br

Abstract word count:

Total word count:

Total number of Tables/figures: 6

Number of references: 38

Abstract

We aimed to perform a comprehensive systematic review with meta-analysis to evaluate the overall accuracy of visual method in detecting caries lesions and identifying possible sources of heterogeneity among the studies. Two independent reviewers searched PubMed, Embase and Scopus through July 2014 to identify articles published in English. Other sources were also evaluated to identify non-published literature. Studies about visual inspection which (1) have assessed accuracy of the method in detecting caries lesions; (2) performed on occlusal, proximal or smooth surfaces in primary or permanent teeth; (3) with a reference standard; and (4) have reported sufficient data about sample size and accuracy of methods were eligible. A diagnostic 2X2 table was extracted from included studies to calculate the pooled sensitivity, specificity and overall accuracy parameters (Diagnostic Odds Ratio and Summary Receiver-Operating Characteristics curve). Methodological quality and heterogeneity of the studies were also assessed. One hundred and two manuscripts met the inclusion criteria from 5,808 articles initially identified. In general, analysis demonstrated that the visual method had good accuracy, but the performance was better for more advanced stages of the carious process. We also observed moderate to high heterogeneity and evidence of publication bias. Through meta-regression analysis, we observed that studies employing widely recognized visual scoring systems presented better performance compared with studies that used their own criteria. In conclusion, visual caries diagnostic method has good overall performance and the use of a visual scoring system improves the accuracy of the method.

Introduction

Caries lesions detection is primarily performed in daily clinical practice by visual inspection (Bader et al., 2002), probably because it is an easy technique with no additional cost to the clinicians. Nevertheless, visual examination has presented some shortcomings, mainly related to its subjective nature (Braga et al., 2010), since different examiners can present inconsistency in the interpretation of the clinical characteristics of the caries lesions (Bader et al., 2002).

Several studies have evaluated the performance of visual inspection in detecting carious lesions and the range of reported results has been extensive and contradictory. This discrepancy may in part be due to the wide assortment of different criteria used for visual inspection and the conditions and methods used to carry out such examinations (Ismail, 2004). In order to overcome this limitation and to reduce the degree of variability in the visual inspection, there has been a move to develop validated meticulous caries diagnostic systems (Ekstrand et al., 1997; Fyffe et al., 2000; Ismail et al., 2007; Nyvad et al., 1999). A further important question is on the actual utility of the visual method. Some authors have suggested that visual inspection performed alone is sufficient for a precise detection of caries lesions (Baelum, 2010; Mendes et al., 2012); hence, studies are necessary to investigate how accurate the visual inspection really is.

With this background in mind, systematic reviews are useful to provide the best evidence on a subject based on the available scientific literature. Furthermore, systematic reviews permit the investigation of possible factors that may influence the performance of the method. To the best of our knowledge, no previous studies have performed a systematic review with meta-analysis to evaluate the overall accuracy of the method and explored possible sources of heterogeneity about visual inspection for detecting caries lesions.

Therefore, the aim of this study was to synthesize the findings regarding the accuracy of visual inspection in detecting caries lesions on occlusal, proximal and smooth surfaces of both permanent and primary unrestored teeth by conducting a comprehensive systematic review including a meta-analysis. Further, we investigated if the utilization of validated visual scoring systems could improve the performance of this method. This is the first study that provided empirical evidence on the importance in using visual scoring systems during the caries lesions detection procedure. Other possible sources of heterogeneity and publication bias were also investigated.

Materials and Methods

To conduct this review, we followed the guideline “Preferred reporting items for systematic reviews and meta-analyses (PRISMA)” (Moher et al., 2009). We registered this systematic review at PROSPERO platform (registration number CRD42013003718). An evaluation related to the clinical relevance of the studies included in this review was published elsewhere (Gimenez et al., 2014).

Information sources

We searched for articles that reported the accuracy of detecting caries lesions by visual inspection published until July 31st, 2014, in MEDLINE (PubMed), Embase and Scopus databases. Unpublished literature was traced through OpenSIGLE and in the Annals of ORCA Congress (European Organisation for Caries Research) from 2003 until 2014. References of the included articles were also checked manually.

Search

The search of electronic databases was designed based on an optimal search strategy for diagnostic studies (Deville et al., 2000), associated with the clinical situation under investigation (caries lesions) and the caries detection method (visual inspection). The syntax was made to search in the MEDLINE database and then, was adjusted for other databases. The entire search strategy is shown in the Appendix 1. Duplicates among databases results were manually eliminated.

Study Selection and Eligibility criteria

All titles and abstracts of studies found were firstly assessed for the inclusion criteria. The inclusion criteria considered were studies which: (1) have some mention on clinical examination or visual inspection for detection of primary coronal caries lesions; (2) have been performed with primary or permanent human teeth, either in vitro or in vivo, and on smooth, proximal or occlusal surfaces; and (3) have been written in English language.

The full papers of included studies were then read to ensure that they presented a clearly defined reference standard (gold standard) and that they reported the absolute numbers of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) or presented sufficient data to derive these figures.

Study selection was performed independently by two reviewers (TG and CP). Doubts or disagreements were resolved by discussion with a third researcher (FMM). Studies that used the same data set for more than 1 publication were included only once in this review. Articles that reported diagnostic performance in artificial or root caries lesions, as well as, caries lesions around restorations, were excluded. Also, studies that made comparisons among methods without a clear definition of a gold standard method were excluded.

Data collection process

One reviewer (TG) collected the data from the selected papers on to structured tables and a second researcher (MMB) independently verified them. Discrepancies were resolved by discussion after rechecking the source.

The following information was extracted from papers: reference standard test used, setting (clinical or laboratorial studies), type of teeth (primary or permanent), surface evaluated (smooth, proximal or occlusal), sample size, examiners´ experience (non-reported, experienced, intermediate with or without training and novices with or without training) and accuracy data (sensitivity and specificity). Undergraduate students were considered as novice examiners and dental practitioners were considered as intermediate level of experience. We also recorded which visual scoring system the authors had used in their research. The following possibilities were considered: no scoring system reported, author’s own criteria, International and Caries Detection Assessment System (ICDAS) (Ismail et al., 2007), World Health Organization criteria (WHO) (WHO, 1997), British Association for the Study of Community Dentistry system (BASCD) (Pitts et al., 1997), Universal Visual Scoring System (Univiss) (Kuhnisch et al., 2009a), ERK (Ekstrand et al., 1997), and those systems described by Nyvad (Nyvad et al., 1999), Nytun (Nytun et al., 1992), Lussi (Lussi, 1993), Downer (Downer, 1975), and Marthaler (Marthaler, 1966).

The values of TP, TN, FP and FN were also recorded, when available. However, if these values were not provided, we derived the numbers from the sample size, caries prevalence of the sample and sensitivity and specificity values reported by the studies. If the study had evaluated the performance of the method with more than one examiner, we considered the values of the first examiner. This strategy was adopted based on a medical systematic reviews aiming to prevent the duplication of sample data (cluster effect), which can lead to bias (Nelemans et al., 2000).

Risk of bias of individual studies

We used the Quality assessment of studies of diagnostic performance included in systematic reviews (QUADAS-2) checklist to assess the risk of bias of the included studies (Whiting et al., 2011). The instrument consists of four key domains: patient selection, index test, reference standard and flow and timing. We used these items to assess possible sources of heterogeneity (Reitsma et al., 2009). Details about the assessment of methodological quality were published elsewhere (Gimenez et al., 2014).

Summary Measures and synthesis of results

The statistical analyses were performed separately at two different thresholds according to reference standard assessment: initial caries lesions (all lesions, independent of the lesion depth or dental surface integrity) and more advanced caries lesions (including only lesions into dentin when the lesion depth was assessed or cavitated lesions when the surface integrity was evaluated). The analyses were performed in subgroups based on the different types of teeth and dental surfaces examined.

First, we perform a qualitative description of included studies and reported their results of sensitivity and specificity in a "Paired Forest Plot" (RevMan Version 5.2, The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen, Denmark) (Macaskill et al., 2010). We also evaluated the individual quality of each study through QUADAS-2 checklist (University of Bristol Resources, Bristol, UK).

Then, statistical pooling of sensitivity, specificity, diagnostic odds ratio (DOR), positive (PLR) and negative likelihood ratios (NLR) were calculated using the DerSimonian Laird method (random effects meta-analysis model). Additionally, we summarized these numbers in receiver operating characteristic (sROC) curves. MetaDisc 1.4 Software (Unidad de Bioestadistica Clínica del Hospital Ramón y Cajal, Madrid, Spain) was used for that.

Publication bias was checked through funnel plots based on the DOR of each study and their respective 95% confidence intervals (95%CI) (Comprehensive Meta-Analysis Software, Statistical Solutions, Boston, USA). The presence of heterogeneity was analyzed via inconsistency (I^2) based on DORs of included studies (MetaDisc 1.4).

The meta-regression analyses intended to compare the effect of methodological differences related to different variables: primary or permanent teeth; clinical or laboratory studies; type of reference standard method used (histological, operative intervention or others, such as tooth separation, radiographic etc.); examiners´ experience (experienced, intermediate, novice and non-reported); and if the authors have used or not visual scoring systems. For this last variable, we considered as reference category studies which the authors did not report any system or when they used their own criteria. Then, we made comparisons with studies which used the ERK criteria, ICDAS, WHO, Nyvad scoring system, Nytun criteria or other visual scoring systems (i.e. Univiss, BASCD, and criteria described by Downer, Marthaler or Lussi). The relative diagnostic odds ratio (RDOR) values and 95%CIs in relation to the reference category were calculated for each condition. The statistical significance was set at $p < 0.05$.

We also performed a sensitivity analysis with the exclusion of each study sequentially. This analysis was performed to determine the robustness of the results.

Results

Study Selection

The study selection flow diagram is shown in Figure 1. Medline (PubMed), Embase and Scopus searches yielded 7,851 studies. After duplication removal, databases identified 5,578 unique studies. Based on the title and abstract, 5,344 articles were excluded. After reading the full text, a further 362 manuscripts were excluded, due to the reasons detailed in Figure 1. Thus, it resulted in 102 papers included in the systematic review. The search of OpenSIGLE and abstracts from Annals of the ORCA Congress yielded 78 investigations (Figure 1), but none were included, mainly due to lack of full data about accuracy.

Study Characteristics

Publication year ranged from 1975 to 2014. The vast majority of studies were conducted in the laboratory using the occlusal surfaces of permanent teeth with a histological reference standard. The characteristics of each included study are provided in Appendix 2.

Risk of bias within studies

The great majority of the studies present a high risk of bias in the selection of patients or teeth (sample). Moreover, nearly 40% of the studies presented high or unclear risk of bias regarding the gold standard method. Contrariwise, most papers revealed a low risk of bias in the domain related to the application of the method and in the timing of exams and flow of patients or teeth.

Regarding applicability considering the sample selection, concerns were raised in relation to most studies. This fact is because the authors did not clearly describe if the spectrum of caries lesions present in the study sample matches the expected prevalence in the target population. Moreover, concerns were raised in relation to almost 40% of studies that the target condition as defined by the reference standard did not match the review question. However, the conduct and interpretation of the index test in all studies did not differ from the review question indicating that there were no concerns regarding the applicability considering this topic. The overview of the QUADAS-2 assessment for all studies is presented in the Figure 2.

Results of individual studies

Paired forest plots summarized the sensitivities and specificities of each study with their 95%CIs. Studies were grouped by permanent or primary teeth, dental surface tested and reference standard method used. We observed a wide range of results across the studies with a tendency to higher specificity than sensitivity values in the detection and diagnosis of both initial and more advanced caries lesions. The paired forest plots of the values of performance are presented for initial caries lesions threshold (Appendix 3) and for more advanced caries lesions threshold (Appendix 4).

Synthesis of results

Pooled sensitivity, specificity, DOR, PLR, NLR, I^2 and sROC curves were calculated separately for combinations of type of each tooth type and dental surface. The results for permanent teeth are presented in Figure 3 and for primary teeth in Figure 4. Within these groups, the area under curves (AUC) of sROC analysis provided more adequate description of the study results.

An overall analysis showed that visual inspection had similar accuracy for all the types of tooth and surfaces. A trend towards better accuracy could be observed at the more advanced caries threshold, except that this was not seen for the occlusal surfaces of permanent teeth. In the same way, a tendency towards higher pooled specificity than the pooled sensitivity could be observed, except for the initial lesions threshold on the occlusal surfaces of primary teeth that showed higher sensitivity value (Figures 3 and 4).

Overall, the studies presented heterogeneity varying from moderate to high. Regarding the occlusal surface of permanent teeth, the values of I^2 were high at initial caries threshold (72.7%), and moderate to high at more advanced lesions threshold (54.5%) (Figure 3). With regard to the occlusal surfaces of primary teeth, I^2 values at the initial caries lesions threshold were 64.2% and at more advanced lesions threshold showed low to moderate heterogeneity (39.9%) (Figure 4). Regarding proximal surfaces of permanent and primary teeth, the method showed high inconsistency in both initial (92.1% and 82.3%, respectively) and more advanced (76.3% and 91.8%, respectively) lesion thresholds (Figures 3 and 4). Heterogeneity analyses were not possible for other situations due to few number of studies.

Evidence of publication bias among the studies

Funnel plots were performed for each type of tooth and tooth surface at each lesion severity threshold (Figure 5). We observed evidence of publication bias considering the following conditions: occlusal surfaces of permanent and primary teeth at more advanced caries lesions threshold (Figure 5; A and B, respectively), and for proximal surfaces of permanent and primary teeth at both caries lesion thresholds (Figure 5; C and D, respectively).

Additional analysis

In the sensitivity analysis, we did not observe any statistically significant difference related to the exclusion of any study.

Meta-regression analyses were performed to compare the effect of methodological differences related to the different situations. Regarding the reference standard, studies about visual examination of the occlusal surfaces of permanent teeth that used operative intervention as reference standard method demonstrated a statistically better performance (RDOR = 4.93; 95%CI = 1.37 to 17.80) than studies using histological examination (reference), but other types of reference standard did not present any significant differences (RDOR = 3.26; 95%CI = 0.46 to 23.11). At initial caries lesions threshold, for occlusal surfaces of permanent teeth, studies using examiners with intermediate experience presented poorer performance (RDOR = 0.21; 95%CI = 0.07 to 0.64) compared to studies using experienced examiners (reference category). Studies that did not report the examiners´ experience also presented lower accuracy (RDOR = 0.28; 95%CI = 0.09 to 0.88), but those using novice trained examiners did not present a significant difference (RDOR = 1.30; 95%CI = 0.05 to 30.98).

When compared to studies that did not report any scoring system or used their own criteria, studies using the ICDAS presented better performance in detecting initial and advanced occlusal caries lesions of primary teeth and advanced caries lesions in occlusal surfaces of permanent teeth (Figure 6). Studies using ERK and Nyvad systems also presented superiority in some conditions (Figure 6). Other meta-regression analyses did not present statistically significant differences and data are not presented.

Discussion

Health professionals are facing a dilemma of being updated through the lack of access and time to keep abreast with the scientific literature (Grimes and Schulz, 2002). In addition, studies can be conducted on the same subject, and they often can present different or even contradictory results. In this way, professionals have difficulty in choosing and implementing the best evidence in clinical practice and due to these difficulties systematic reviews are essential. With regard to the visual inspection for caries detection, a previous systematic review did not perform a meta-analysis (Bader et al., 2002). Since this review, and in part as a consequence of this review, significant advances concerning visual detection and diagnosis of caries have been made; examples of which are the ICDAS initiative (Ismail et al., 2007; Pitts, 2004) and the creation of the Univiss (Kuhnisch et al., 2009b). Therefore, there was a need for an updated systematic review including meta-analysis and meta-regressions for the first time. This is the first study that evaluated different aspects of visual inspection for detecting caries lesions, such as the overall accuracy, source of heterogeneity among the studies, evidence of publication bias, and if differences in the methodology could interfere in the findings.

The most commonly used indicators of diagnostic performance have been sensitivity and specificity. We could see a trend of pooled specificity being greater than the pooled sensitivity in the majority of the analyses considering visual inspection. An ideal diagnostic method should offer high sensitivity and high specificity. However, a high sensitivity is normally obtained at the expense of reduced specificity. This condition would increase the number of false-positive diagnosis which can be dangerous considering dental caries, as it can lead to overtreatment. Thus, it seems to be more appropriate for a method of caries detection to have a high specificity even at the expense of a small reduction in sensitivity (Downer, 1989), as it was observed for the visual inspection in this publication.

The unique exception for this pattern was observed regarding the performance of the method in detecting occlusal initial caries lesions of primary teeth. These findings could be explained by the morphological differences between enamel of primary and permanent teeth, namely the primary enamel being thinner and less mineralized than permanent enamel (Mortimer, 1970;

Shellis, 1984). These characteristics might induce an overestimation in the detection of initial caries lesions in primary teeth, increasing sensitivity with a consequent decrease of the specificity. Since most visual scoring systems are developed and tested firstly in permanent teeth, adaptations for primary teeth should be considered.

Nevertheless, when the results of different studies are pooled, the threshold effect usually occurs since both sensitivity and specificity parameters are not independent (Cota et al., 2012). Thus, the best indicator of accuracy to be used in meta-analysis of diagnostic studies is the DOR, which is a parameter that combines diagnostic values of accuracy in a single parameter. The DOR does not suffer from the influence of the threshold effect among the studies. Considering this parameter, a trend of better performance at more advanced caries lesions could be observed, except for the occlusal surfaces of permanent teeth. This result could be explained by the anatomical characteristics of this surface. As the occlusal surfaces of permanent teeth have a more elaborated morphology than other surfaces, therefore clinicians can experience some difficulty in detecting lesions reaching the dentin. This phenomenon has been referred to as hidden caries (Ricketts et al., 1997). Visual inspection presented an overall accuracy similar to that obtained with radiography (Bader et al., 2002) and fluorescence-based methods (Gimenez et al., 2013). Nevertheless, these adjunct methods tended to present higher sensitivities and lower specificities. The current prevalence of non-evident caries lesions from a visual examination, which could benefit from additional detection methods, is low in most populations. Actually, the prevalence of advanced caries lesions in the sample of studies included in our systematic review that were classified with low risk of selection bias by the QUADAS-2 ranged from 1.32 to 31.25% (mean=12.24%). Due to this reality, the use of visual inspection alone seems to be effective enough for caries detection, confirming previous findings (Baelum et al., 2012; Mendes et al., 2012). However, further systematic reviews should be conducted to compare directly the accuracy of different methods of caries detection using meta-analysis. Moreover, reproducibility of the methods is another important parameter, and it should be consider in additional studies.

Regarding the heterogeneity of the studies, we observed inconsistency ranging from moderate to high in the analyses. The same pattern was observed for fluorescence-based methods (Gimenez et al., 2013). Since systematic reviews analyze together studies that are different in several aspects, heterogeneity is expected. Many possibilities in dealing with this heterogeneity have been described (Dinnes et al., 2005). We chose to consider the heterogeneity using random models and exploring it through subgroup analysis and meta-regressions.

The exclusion of articles published in non-English language can be seen as a possible limitation of our study. However, some studies have shown that the exclusion of articles published in other languages does not seem to bias systematic reviews (Juni et al., 2002; Moher et al., 2000). Furthermore, a systematic review performed by our research group about fluorecence-based methods for caries detection showed that the inclusion of articles in other languages did not influence the results (Gimenez et al., 2013). Another possible limitation was avoided since we have searched for articles in other sources, including grey literature, thus minimizing the occurrence of bias due to missing studies (Whiting et al., 2008). Unfortunately, this search did not add any study in our review, mainly because of the lack of accuracy data. Our suggestion is that authors include a contingency table or the sample size and caries prevalence of their sample in their abstracts.

Concerning the meta-regressions performed in our study, we compared the effect of methodological differences related to the important aspects of the research. Regarding the reference standard methods, at initial caries lesions threshold, we observed a better performance in studies using operative intervention as a reference standard on occlusal surfaces of permanent teeth, probably due to incorporation bias (Lijmer et al., 1999), since teeth thought of as sound in such in vivo studies would not be entered into the study to confirm they were sound by operative intervention due to ethical reasons.

The most important finding of this review concerned the utilization of validated visual scoring systems in detecting and diagnosing carious lesions. The use of detailed visual indices has been proposed as a manner to improve sensitivity and reliability, and to minimize the influence of examiners´ experience (Braga et al., 2010). We found that the studies using some visual scoring systems showed higher accuracy in detecting carious lesions in proximal and occlusal surfaces of both primary and permanent teeth when compared with studies that did not report use of indices or used their own criteria. The utilization of a visual system allows a more accurate diagnosis of caries lesions since it gives practitioners a guideline to characteristics that they have to pay attention to and it provides a rational shortcut associating these characteristics with the type of caries lesions. This is the first robust study that has proved empirically that the utilization of validated and widely used visual scoring systems improves the accuracy in the detection of caries lesions.

In conclusion, visual inspection usually presents higher specificity than sensitivity values and a better performance in detecting advanced caries lesions. The method performed alone shows good accuracy for the detection of occlusal and proximal caries lesions, in both primary and permanent teeth. Furthermore, the utilization of detailed and validated indices improves the performance of the visual inspection for caries lesions detection.

Acknowledgements

The study was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – Process 2012/17888-1) and CAPES. The authors wish also to thank the participants of the Post-Graduation in Pediatric Dentistry Seminar of FOUSP for the critical comments put forth. The authors of this manuscript certify that we have no affiliation with or financial involvement in any organization or entity with a direct financial or personal interest in the subject matter or materials discussed in the manuscript.

References

Bader JD, Shugars DA, Bonito AJ (2002). A systematic review of the performance of methods for identifying carious lesions. *J Public Health Dent* 62(4)::201-213.

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

Baelum V, Hintze H, Wenzel A, Danielsen B, Nyvad B (2012). Implications of caries diagnostic strategies for clinical management decisions. *Community Dent Oral Epidemiol* 40(3):257-266.

Braga MM, Mendes FM, Ekstrand KR (2010). Detection activity assessment and diagnosis of dental caries lesions. *Dent Clin North Am* 54(3):479-493.

Cota GF, de Sousa MR, Demarqui FN, Rabello A (2012). The diagnostic accuracy of serologic and molecular methods for detecting visceral leishmaniasis in HIV infected patients: meta-analysis. *PLoS Negl Trop Dis* 6(5):e1665.

Deville WL, Bezemer PD, Bouter LM (2000). Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. *J Clin Epidemiol* 53(1):65-69.

Dinnes J, Deeks J, Kirby J, Roderick P (2005). A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol Assess* 9(12):1-113, iii.

Downer MC (1975). Concurrent validity of an epidemiological diagnostic system for caries with the histological appearance of extracted teeth as validating criterion. *Caries Res* 9(3):231-246.

Downer MC (1989). Validation of methods used in dental caries diagnosis. *Int Dent J* 39(4):241-246.

Ekstrand KR, Ricketts DN, Kidd EA (1997). Reproducibility and accuracy of three methods for assessment of demineralization depth of the occlusal surface: an in vitro examination. *Caries Res* 31(3):224-231.

Fyffe HE, Deery C, Nugent ZJ, Nuttall NM, Pitts NB (2000). In vitro validity of the Dundee Selectable Threshold Method for caries diagnosis (DSTM). *Community Dent Oral Epidemiol* 28(1):52-58.

Gimenez T, Braga MM, Raggio DP, Deery C, Ricketts DN, Mendes FM (2013). Fluorescence-based methods for detecting caries lesions: systematic review, meta-analysis and sources of heterogeneity. *PLoS One* 8(4):e60421.

Gimenez T, Piovesan C, Braga MM, Raggio DP, Deery C, Ricketts DN *et al.* (2014). Clinical relevance of studies on accuracy of visual inspection for detecting caries lesions: a systematic review. *Caries Res (in press)*

Grimes DA, Schulz KF (2002). An overview of clinical research: the lay of the land. *Lancet* 359(9300):57-61.

Ismail AI (2004). Visual and visuo-tactile detection of dental caries. *J Dent Res* 83 Spec No C(C56-66.

Ismail AI, Sohn W, Tellez M, Amaya A, Sen A, Hasson H *et al.* (2007). The International Caries Detection and Assessment System (ICDAS): an integrated system for measuring dental caries. *Community Dent Oral Epidemiol* 35(3):170-178.

Juni P, Holenstein F, Sterne J, Bartlett C, Egger M (2002). Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 31(1):115-123.

Kuhnisch J, Goddon I, Berger S, Senkel H, Bucher K, Oehme T *et al.* (2009a). Development, methodology and potential of the new Universal Visual Scoring

System (UniViSS) for caries detection and diagnosis. *Int J Environ Res Public Health* 6(9):2500-2509.

Kuhnisch J, Ifland S, Tranaeus S, Heinrich-Weltzien R (2009b). Comparison of visual inspection and different radiographic methods for dentin caries detection on occlusal surfaces. *Dentomaxillofac Radiol* 38(7):452-457.

Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH *et al.* (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 282(11):1061-1066.

Lussi A (1993). Comparison of different methods for the diagnosis of fissure caries without cavitation. *Caries Res* 27(5):409-416.

Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y (2010). Chapter 10: Analysing and Presenting Results. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 10. In: J Deeks, P Bossuyt and C Gatsonis editors: The Cochrane Collaboration.

Marthaler TM (1966). A standardized system of recording dental conditions. *Helv Odontol Acta* 10(1):1-18.

Mendes FM, Novaes TF, Matos R, Bittar DG, Piovesan C, Gimenez T *et al.* (2012). Radiographic and Laser Fluorescence Methods Have No Benefits for Detecting Caries in Primary Teeth. *Caries Res* 46(6):536-543.

Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR *et al.* (2000). What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 53(9):964-972.

Moher D, Liberati A, Tetzlaff J, Altman DG (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 6(7):e1000097.

Mortimer KV (1970). The relationship of deciduous enamel structure to dental disease. *Caries Res* 4(3):206-223.

Nelemans PJ, Leiner T, de Vet HC, van Engelshoven JM (2000). Peripheral arterial disease: meta-analysis of the diagnostic performance of MR angiography. *Radiology* 217(1):105-114.

Nytun RB, Raadal M, Espelid I (1992). Diagnosis of dentin involvement in occlusal caries based on visual and radiographic examination of the teeth. *Scand J Dent Res* 100(3):144-148.

Nyvad B, Machiulskiene V, Baelum V (1999). Reliability of a new caries diagnostic system differentiating between active and inactive caries lesions. *Caries Res* 33(4):252-260.

Pitts N (2004). "ICDAS"--an international system for caries detection and assessment being developed to facilitate caries epidemiology, research and appropriate clinical management. *Community Dent Health* 21(3):193-198.

Pitts NB, Evans DJ, Pine CM (1997). British Association for the Study of Community Dentistry (BASCD) diagnostic criteria for caries prevalence surveys-1996/97. *Community Dent Health* 14 Suppl 1(6-9.

Reitsma J, Rutjes A, Whiting P, Vlassov V, Leeflang M, Deeks J (2009). Chapter 9: Assessing methodological quality. Cochrane Handbook for

Systematic Reviews of Diagnostic Test Accuracy Version 100. In: J Deeks, P Bossuyt and C Gatsonis editors: The Cochrane Collaboration.

Ricketts D, Kidd E, Weerheijm K, de Soet H (1997). Hidden caries: what is it? Does it exist? Does it matter? *Int Dent J* 47(5):259-265.

Shellis RP (1984). Relationship between human enamel structure and the formation of caries-like lesions in vitro. *Arch Oral Biol* 29(12):975-981.

Whiting P, Westwood M, Burke M, Sterne J, Glanville J (2008). Systematic reviews of test accuracy should search a range of databases to identify primary studies. *J Clin Epidemiol* 61(4):357-364.

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB *et al.* (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 155(8):529-536.

WHO (1997). Oral health surveys, basic methods. 4ª ed. Geneva: World Health Organization.

Legends

Figure 1. Flow diagram with the information through the phases of studies selection

Figure 2. Analysis of methodological study quality considering the Quality assessment of studies of diagnostic performance included in systematic reviews (QUADAS-2) checklist

Figure 3. Summary Receiver-operating characteristics (sROC) curves and synthesis of the results obtained with studies of accuracy performed in occlusal and proximal surfaces of permanent teeth

Figure 4. Summary Receiver-operating characteristics (sROC) curves and synthesis of the results obtained with studies of accuracy performed in occlusal and proximal surfaces of primary teeth

Figure 5. Funnel plots to evaluate evidence of publication bias of studies performed on occlusal surfaces of permanent teeth (A), occlusal surfaces of primary teeth (B), proximal surfaces of permanent teeth (C) and proximal surfaces of primary teeth (D)

Figure 6. Graphical summary of meta-regression analysis to compare the effect of differences of the studies regarding the utilization of visual scoring systems. DOR = Diagnostic Odds Ratio. CI = Confidence interval. * difference statistically significant ($p < 0.05$)

Online supplementary material - legends

Appendix 1. Chart containing the search strategy for electronic databases

Appendix 2. Table with the summary of characteristics of included studies

Appendix 3. Paired forest plot summarizing the results of the studies at initial caries lesions threshold

Appendix 4. Paired forest plot summarizing the results of the studies at more advanced caries lesions threshold