

SCIENTIFIC REPORTS



OPEN

Reliability of colour and hardness clinical examinations in detecting dentine caries severity: a systematic review and meta-analysis

Larry Hon¹, Ahmed Mohamed² & Edward Lynch³

Dental caries is the most common human infectious disease and is caused by microorganisms producing acids, resulting in changes in dental tissue hardness and colour. However, the accuracy and reliability of dentine colour and hardness as indicators for carious lesion severity has never been assessed in a systematic review. By applying strict criteria, only seven papers (five randomized control trials and two diagnostic studies) were considered for full text qualitative and quantitative assessment. Only three studies produced high quality evidence and only four articles were considered for meta-analysis, as these provided \log_{10} colony forming units (CFU) data from caries biopsies following colour and hardness clinical examinations. When comparing the amount of CFU isolated from carious biopsies from different colour and hardness categories, hardness clinical examination was found to be a statistically more discriminate test than colour clinical examination. Therefore, hardness clinical examination is more specific and reliable than colour to detect dentine carious lesion severity. Further large carefully designed clinical studies are needed to consolidate the findings of this systematic review.

Dental caries is the most common chronic disease amongst all oral conditions¹. Caries experience is found in 21% of children between the ages of 6–11 years, and in 91% of adults older than 20 years of age in the United States². While odds for untreated decay increases with decreasing income, dental caries prevalence in the primary dentition is 10.8% in the Philippines, 64.2% in India, and 30% in Spain, signifying the varying effects of geographic location and socioeconomic status on onset and spread of dental caries^{3–6}.

Demineralized tooth structure and the formation of dental caries is associated with acid production by microbial metabolism of sugar⁷. Studies have shown that the initial colonizing microorganisms are mainly *Streptococcus sanguinis*, *Streptococcus oralis* and *Streptococcus mitis*⁸. Eventually, *Streptococcus mutans* proportionally increases when compared to other decay bacterial species, while surface enamel changes to a frosty white colour caused by the acid damage⁸. This domination of *S. mutans* will subsequently reduce at the advancing stages of dental caries due to the increasing numbers of other microorganisms, such as *Lactobacilli*, *Prevotella*, and *Bifidobacterium*⁹. Therefore, the detection of microorganisms and the number of microbial colonies formed inside carious lesions can be considered as definitive and sensitive indicators for carious lesion severity. Carious lesion severity could be essentially defined as the level of clinical treatment needs (e.g. intrusive or minimal) of a carious lesion based on the amount of microbial colonies isolated from the carious lesion as dentinal lesions that contain large amounts of isolated microbial colony forming units (CFUs) often require a more intrusive clinical treatment (e.g. caries debridement and restoration) if compared to those with less isolated CFUs which may either require less intrusive treatments such as chemotherapeutic treatments (e.g. topical fluoride treatment) or no treatment need at all^{10–14}. Furthermore, by identifying the number and types of microbial species isolated from carious lesions, the stages of caries progression might be distinguished with great specificity, sensitivity, and reliability.

¹Pediatric Dental Resident, Pediatric Dentistry Department, University of Nevada, Las Vegas, UNLV, 89106, USA.

²Visiting Faculty, Biomedical and Clinical Research, University of Nevada, Las Vegas, UNLV, 89106, USA. ³Professor and Principal Director of Biomedical and Clinical Research, University of Nevada, Las Vegas, UNLV, 89106, USA. Correspondence and requests for materials should be addressed to E.L. (email: edward.lynch@unlv.edu)

The most common methods used for dental caries detection in clinical practice are visual and tactile examinations¹⁵. Under normal daylight, the colour of a lesion is categorized by visual comparison with a standard guide of four shades (yellow, light brown, dark brown, and black) which is prepared from photographs of primary dentinal carious lesions. On the other hand, the texture or hardness of lesions are classified into three grades (hard, medium or leathery, soft) as described by Hellyer *et al.*¹⁶. Briefly, under standard dental lighting, hard lesions are as hard as the surrounding tooth tissue, leathery lesions are penetrated by a new Ash No. 6 probe under modest pressure but displayed resistance to its withdrawal, while soft lesions are easily penetrated by a new Ash No.6 probe under modest pressure and displayed no resistance to withdrawal of the probe^{10,17}. However, visual and tactile examinations result in low reproducibility and low sensitivity due to their subjective nature, but produce highly specific outcomes¹⁸. Studies indicated that cariogenic microorganisms produce acids that destroy tooth structure, resulting in changes of colour, consistency, and moisture content of dental tissue. For instance, darker and softer carious lesions contained larger numbers of microorganisms^{10,14,19,20}. Thus, to improve the sensitivity, specificity, and reliability of clinical dentine caries examination detection methods, it is suggested that visual and tactile criteria for caries detection should be assessed in relation to the carious microbial activity as this would provide an accurate consistent method to assess caries quantitatively and qualitatively.

Therefore, the aim of this study was to systematically review the literature into colour and hardness of dentine caries and their association with microbial activity, particularly the amount of microbial colony forming units (CFU) that are isolated clinically from biopsies of these lesions following colour and hardness clinical examinations. This potentially proves which of the categories of colour or hardness is more specific and reliable to reflect the severity of carious lesions.

Methods

The Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) guidelines were adopted for the current study²¹.

Search strategy. Studies that assessed the association between colour and hardness with the quantification of CFUs from biopsies of these carious lesions as an indicator of dentinal carious lesion severity were accessed using a defined search strategy in the following electronic databases from 1950 to April 1st, 2018: PubMed, Medline via Ovid, and Web of Science. Handsearching was also performed by accessing the following journals: Community Dentistry and Oral Epidemiology, Caries Research, Journal of Dental Research, Journal of Pediatric Dentistry, Journal of Dentistry, Journal of Oral Health and Preventive Dentistry, and Journal of Dentistry for Children to April 1st, 2018. The following search terms were used: “color” or “colour” or “hardness” or “texture” or “consistency” and “dentin caries” or “dentine caries” and “microbiology” or “microflora” or “bacteria”.

Eligibility criteria. Search strategy and literature search findings were reviewed by two authors (Hon and Mohamed) to determine whether the identified studies met the inclusion criteria. The inclusion criteria were as follows: (1) randomized controlled trial (RCT) or diagnostic tests studies; (2) colour and hardness scored clinically; (3) CFU reported; (4) dentine caries; (5) open primary dentine caries in deciduous or permanent teeth; (6) articles published in English; and (7) human studies *in vivo*. Animal studies, *in vitro* studies, reviews, comments, editorials, and non-English studies were excluded. Additionally, research that contained irreversible pulpitis, secondary caries, hidden caries or caries that could only be seen radiographically were excluded. Disagreements between the two authors were resolved by an independent reviewer (Lynch). For detailed information, please refer to Fig. 1.

Data extraction and quality assessment. Two authors (Hon and Mohamed) extracted data using a pre-set data extraction sheet. The data extraction sheet included the following information: author, title, population, intervention, comparison, sample size, confidence level, results and outcomes. For the data extraction details, please refer to Table 1. The quality of evidence in the included studies was assessed by the Critical Appraisal Skills Programme (CASP) tools^{22,23}. The included studies were randomized controlled trials and diagnostic tests studies. Thus, the respective CASP tools were used to assess each study type. Each article was independently assessed by two authors (Hon and Mohamed), while disagreements were resolved by an independent reviewer (Lynch). Inter-examiner variability among the 2 examiners was measured by calculating the percentage of agreement (%) among the 2 examiners in each CASP checklist criteria. This is due to the low number of included studies (diagnostic studies $n = 2$ and RCTs $n = 5$)^{24–27}. For detailed information, please refer to Fig. 2.

Statistical analysis. Clinical recordings using colour and hardness of carious dentine have been proposed as indicators of lesion severity. Microbiological biopsies of carious samples have been used to enumerate the numbers of microorganisms expressed as CFUs. The primary outcome measures analyzed were: total microbial load in CFU in each category of colour (yellow, light brown, dark brown, black) and hardness classification (soft, medium hard or leathery, hard). Differences (d_{colour}) in means of total microbial load between categories of the colour scale were assumed as a measure of “discriminant power” using this test. Three differences between adjacent categories were computed as: yellow vs. rest, yellow/light brown vs. dark brown/black, black vs. rest. Differences (d_{hardness}) in means of total microbial load between categories of the hardness scale were also assumed as a measure of “discriminant power” using this test: hard vs. rest, soft vs. rest. If the difference obtained from the colour test is significantly higher than those from the hardness test, it means that the colour test is more discriminating than the hardness test. Inversely, if the difference obtained from the colour test is significantly lower than those from the hardness test, it means that the colour test is less discriminating than the hardness test. Therefore, mean differences between all colour and hardness category differences were computed and the weighted mean difference (WMD) was the global effect measure in a random-effects model.

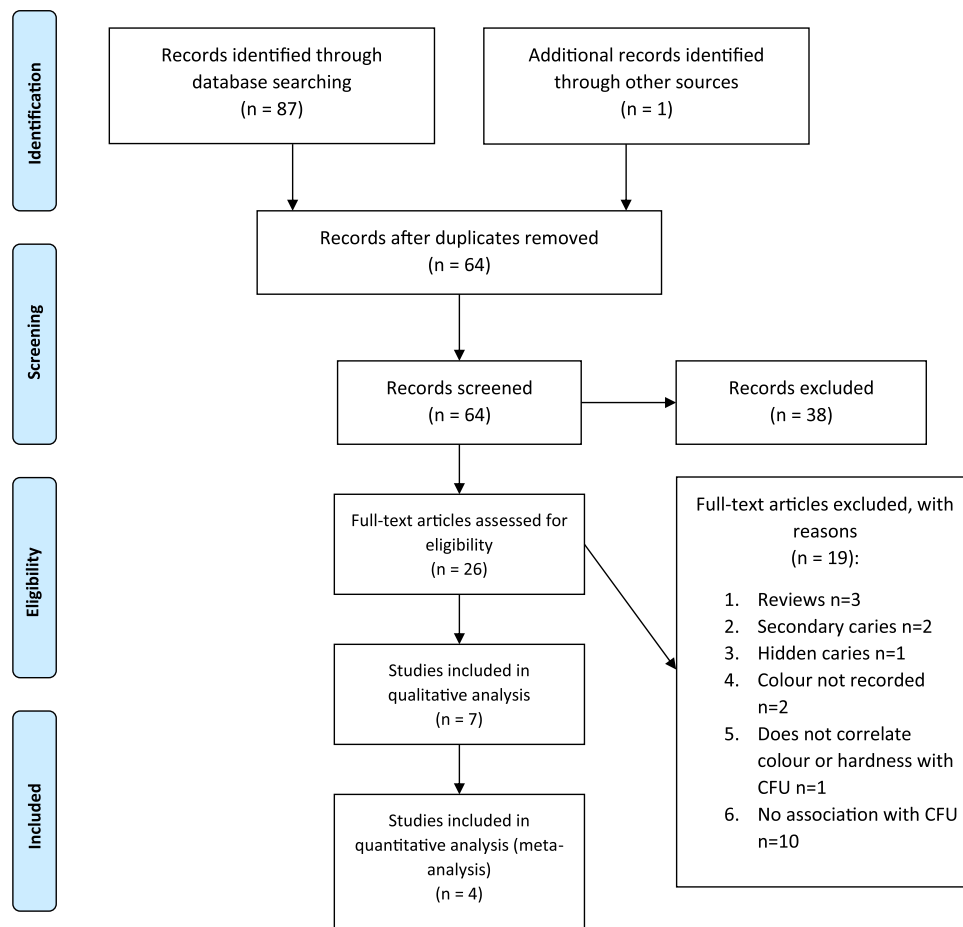


Figure 1. Flow diagram of study selection process.

In addition, the effect size of the differences between colour (or hardness) categories regarding mean CFU was also calculated. The effect size index for a conventional one-way ANOVA is... $f = \sqrt{\frac{\sigma^2}{1 - \sigma^2}}$, whereby σ^2 is the ratio of the between-groups variance to the total variance. The larger the effect size “ f ” is, the larger the difference between mean total microbial load of categories is, meaning there is a higher power to discriminate. As a general convention, small $f = 0.10$, medium $f = 0.25$, and large $f = 0.40$. An overall effect size was estimated weighting the individual numbers by the sample size of each study. The level of significance used in the analysis was 5% ($\alpha = 0.05$). The software used to perform this meta-analysis was R 3.0.2 and its ‘metafor’ package. The software Gpower 3.1.3 was used to estimate effect sizes.

Results

Study characteristics and quality assessment. Fig. 1 shows the detailed steps used for the literature search. Of the 64 potentially relevant articles, 26 articles were eligible for full text screening. However, 19 articles were further excluded because they did not meet the inclusion criteria. Please refer to Table 2 for detailed information on the excluded papers and the rationale for their exclusion. Therefore, only seven articles^{10,14,19,28–31} were included for full text quantitative and qualitative assessment in this systematic review. These seven articles were published between 1950 and 2018. Among these articles five were randomized controlled trials^{19,28–31} and two were diagnostic test studies^{10,14}.

A-Randomized controlled trials. Five randomized controlled trial articles were included^{19,28–31}. One article disclosed the gender of the participants: 51 females and 60 males²⁸. Three articles detailed the age of participants age: 4–15 years old²⁸, 12–23 years old²⁹, and 5–8 years old³¹; therefore, the range of the age groups of the participants within the latter three studies was 4–23 years old. Three articles specified the type of teeth that received intervention and comparator treatments: 94 mandibular second primary molars and 60 mandibular first permanent molars²⁸ and primary molars^{30,31}. Only one article provided the number of subjects recruited²⁸, whilst the other studies only reported the number of teeth investigated^{29–31}. Sample size calculations were only carried out by Lula *et al.*, who conducted sample size calculations based on a pilot study where 16 teeth obtained an 80% power at a 5% statistical significance. It was noticeable that the article by Bjorndal *et al.* did not give details as to the gender, age, or type of teeth used (incisor, canine, premolars, molars, primary dentition, and permanent dentition).

Study	Population	Intervention	Comparison	Outcome of Interest	Sample Size	Confidence Level	Results and Findings
Orhan <i>et al.</i>	154 teeth	One-visit Indirect pulp treatment (IPT).	Two-visit IPT, and direct complete excavation (DCE)	CFU (CFU/ml)	No sample size calculations.	Confidence level was reported.	Lesions classified by colour did not show difference in CFU of bacteria ($p > 0.05$), but harder lesions showed less CFU of bacteria than softer lesions ($p < 0.05$).
Bjørndal <i>et al.</i>	31 teeth	Step-wise excavation	Paired study design	CFU count	No sample size calculations.	Confidence level was not reported.	No statistical data was provided.
Maltz <i>et al.</i>	32 teeth	Step-wise excavation	Paired study design	CFU ($\log_{10}(\text{cfu} + 1)$)	No sample size calculations.	Confidence level was reported.	Authors did not statistically analyze and compare CFU of hardness and colour categories.
Bönecker <i>et al.</i>	40 teeth	ART removal of carious dentin and immediately restored with Glass Ionomer.	Paired study design	CFU (CFU/ml $\times 10^3$)	No sample size calculations.	Confidence level was reported.	No statistical significant difference results reported.
Lula <i>et al.</i>	16 teeth	Removal of carious dentin; only a superficial layer of carious dentin was removed from the pulpal wall.	Paired study design	CFU (\log_{10} (CFU/mg))	Sample size was calculated based on pilot study with 6 teeth (80% power test at 5% level of significance)	Confidence level was reported.	No difference in bacteria counts between colour categories. Also, no difference in bacteria counts between hardness categories. Authors did not compare bacteria count of colour to hardness.
Author	Population	Exposure	Comparison	Outcome of Interest	Sample Size	Confidence Level	Results and Findings
Lynch <i>et al.</i>	117 patients	Caries examination by colour	Caries examination by texture	CFU (\log_{10} (CFU + 1))	No sample size calculations.	Confidence level was reported.	Black soft and black leathery lesions had higher CFU counts. Soft and leathery lesions generally had a higher count in microorganisms regardless of color. There was no indication that lesion color was significantly associated with total bacteria or proportions of species.
Beighton <i>et al.</i>	59 patients	Caries examination by colour	Caries examination by texture	CFU (\log_{10} (CFU + 1))	No sample size calculations.	Confidence level was reported.	Soft lesions contained significantly ($p < 0.001$) greater CFU than leathery lesions, and samples from leathery lesions contain greater CFU ($p < 0.001$) than hard lesions.

Table 1. Studies used for qualitative analysis.

As we investigated the intervention and comparison groups, colour and hardness of lesions were related to the total CFU each contained. Two articles scored colour and hardness of lesions whilst evaluating different levels of carious dentine prior to final restoration in a stepwise approach^{19,28}, which was using incomplete removal of dentine caries to try to prevent pulpal exposure. Three articles related total CFU in carious dentine with colour and hardness at different time points, at intervals of 6–7 months²⁹, 4–6 months³⁰, and 3–6 months³¹. Refer to Table 1 for further details on study characteristics. The CASP quality assessment for these included randomized controlled trials is presented in Fig. 2. According to the CASP analysis, the study of Lula *et al.* produced the highest quality of evidence and therefore its findings carried more weight. The inter-examiner variability for quality assessment was 72.7% for the five RCT's, signifying high reproducibility between the two examiners.

B-Diagnostic test studies. Two articles were diagnostic test studies^{10,14}. Both articles examined the colour and hardness of carious lesions and compared these to their respective total CFU counts^{10,14}. Both these articles reported the number of participants and gender: 45 females and 72 males¹⁴, 25 females and 34 males¹⁰. Therefore, the total number of participants in each gender was 70 females and 106 males. Both articles examined the same age range: 29–80 years old^{10,14}. Both articles^{10,14} reported the number of teeth included in each study but neither article detailed the type of teeth that had been examined (incisor, canine, premolars, molars, primary dentition, permanent dentition). Neither article reported any sample size calculation^{10,14}. For further details on study characteristics, please refer to Table 1. According to the CASP quality assessment, both articles produced equally high-quality evidence. For further detail CASP analysis of the aforementioned articles, please refer to Fig. 2. The inter-examiner variability for quality assessment was 83.3% for the diagnostic test studies, signifying high reproducibility between the two examiners.

Comparison of CFU's Between Colour and Hardness Examination Methods

A-Randomized controlled trials. One article compared CFUs in one-visit indirect pulp treatment (IPT) with two-visit IPT and direct complete excavation (DCE)²⁸. Four articles were paired study designs that compared samples from different time points^{19,29–31}. These four articles measured the CFUs taken for colour or hardness categories. The total CFU counts were only measured and compared within the categories of caries colour or categories of caries hardness independently^{19,28–30}. Only one article related both colour and hardness to total CFU. This study also compared the CFU counts within each caries colour category (i.e. yellow, light brown, dark brown) and within caries hardness categories (i.e. soft, medium hard, hard)³¹.

(a)

	KEY													
	Yes	Unclear	No	Did the trial address a clearly focused issue?	Was the assignment of patients to treatments randomized?	Were all of the patients who entered the trial properly counted for at its conclusion?	Were patients, health workers and study personnel blind to treatment?	Were the groups similar at the start of the trial?	Aside from the experimental intervention, were the groups treated equally?	How large was the treatment effect?	How precise was the estimate of the treatment effect?	Can the results be applied to the local population, or in your context?	Were all clinically important outcomes considered?	Are the benefits worth the harms and costs?
Bönecker et al., 2003	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Bjorndal et al., 1997	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Lula et al., 2011	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Maltz et al., 2002	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Orhan et al., 2008	●	●	●	●	●	●	●	●	●	●	●	●	●	●

(b)

	KEY														
	Yes	Unclear	No	Was there a clear question for the study to address?	Was there a comparison with an appropriate reference standard?	Did all patients get the diagnostic test and reference standard?	Could the results of the test have been influenced by the results of the reference standard?	Is the disease status of the tested population clearly described?	Were the methods for performing the test described in sufficient detail?	What are the results?	How sure are we about the results? Consequences and cost of alternatives performed?	Can the results be applied to your patients/the population of interest?	Can the test be applied to your patient or population of interest?	Were all outcomes important to the individual or population considered?	What would be the impact of using this test on your patients/population?
Beighton et al., 1993	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Lynch et al., 1994	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●

Figure 2. Qualitative analysis with CASP tools for randomized controlled trials. Summary review of the qualitative assessment of the included studies by using CASP tools for randomized controlled trials (RCTs) consisting of 11 quality criteria (a) and for diagnostic studies consisting of 12 quality criteria (b). Green-coded circle indicates that the study satisfactorily met the respective quality criterion, yellow-coded circle indicates that the study partially met the respective quality criterion, and the red-coded circle indicates that the study did not meet the respective quality criterion.

Study	Reason for exclusion
Kidd et al. ³⁴	Colour not recorded
Weerheijm et al. ³⁵	Hidden caries, not gross occlusal caries
Kidd et al. ³⁶	Includes secondary caries
Ayna et al. ²⁰	No results correlating hardness and CFU
Loesche et al. ³⁷	Incipient caries
Manji et al. ³⁸	No CFU
Bönecker et al. ³⁹	No CFU
Iwami et al. ⁴⁰	No CFU
Fusayama et al. ⁴¹	No CFU
Iwami et al. ⁴²	No CFU
Torii et al. ⁴³	No CFU
Iwami et al. ⁴⁴	No CFU
Nyvad et al. ⁴⁵	No CFU
Milnes et al. ⁴⁶	No CFU
Nyvad et al. ⁴⁷	No CFU
Fejerskov et al. ⁴⁸	Review Article
Kidd et al. ⁴⁹	Review Article
Takashashi et al. ⁵⁰	Review Article
Kidd et al. ⁵¹	Secondary Caries

Table 2. Excluded studies and rationale for exclusion.

Author		Yellow	Light brown	Dark brown	Black	Soft	Medium-hard (leathery)	Hard
Orhan 2008	CFU/ml							
Bjorndal 1997	log ₁₀ (CFU)			2.70 (1.24)	2.97 (0.93)		2.59 (1.07)	3.23 (1.28)
Maltz 2002	log ₁₀ (CFU + 1)	4.50 (0)	4.07 (0.93)	4.25 (0.71)		4.17 (0.84)		
Bonecker	CFU/ml × 10 ³							
Lula 2011	log ₁₀ (CFU/mg)	4.77 (0.32)	2.29 (2.97)	2.66 (3.08)		4.49 (3.00)	2.50 (2.65)	0.00 (0)
Lynch 1994	Percentages							
Beighton 1993	log ₁₀ (CFU)	4.60 (2.22)	3.50 (1.81)	3.40 (1.89)	6.00 (0.99)	6.80 (0.37)	4.30 (0.56)	1.80 (0.82)

Table 3. Shows the initial summary data in the included studies: means (sd).

(a)

AUTHOR	COLOUR (B - DB/LB/Y)			HARDNESS (S - M/H)		
	nTX	mTX	sTX	nCT	mCT	sCT
ORHAN, 2008	135			135		
BJORNDAL, 1997	13	0.27	1.11	13		
MALTZ, 2002	30			30		
BÖNECKER, 2003	40			40		
LULA, 2011	16			16	2.41	2.78
LYNCH, 1994	395			395		
BEIGHTON, 1993	301	2.32	1.66	301	3.72	0.61

(b)

AUTHOR	COLOUR (B - DB/LB/Y)			HARDNESS (S/M - H)		
	nTX	mTX	sTX	nCT	mCT	sCT
ORHAN, 2008	135			135		
BJORNDAL, 1997	13	0.27	1.11	13	-0.64	1.19
MALTZ, 2002	30			30		
BÖNECKER, 2003	40			40		
LULA, 2011	16			16	3.07	2.74
LYNCH, 1994	395			395		
BEIGHTON, 1993	301	2.32	1.66	301	3.26	0.72

(c)

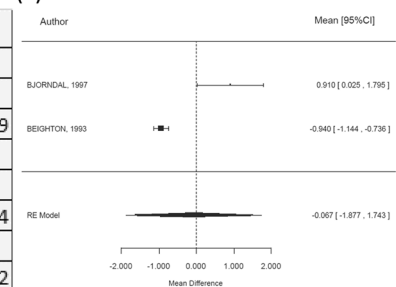


Figure 3. Comparison B-DB/LB/Y versus S-M/H and S/M-H. Comparing mean differences for log(CFU) between categories B-DB/LB/Y (colour test) and categories S-M/H (hardness test) (a) and comparison between black against other colour categories with hard and other hardness categories (b,c). (b) Table shows mean differences for log(CFU) between categories B-DB/LB/Y (colour test) and categories S/M-H (hardness test). (c) Forest plot showing WMD value for the difference is -0.067 . The effect size on population is estimated between -1.88 and 1.74 with a CI of 95%. It should be noted that the interval includes zero, so no significant difference was reached ($p = 0.942$). TX = treatment/exposure group (colour test); CT = control group (hardness test); n = number of samples; m = mean; s = standard deviation.

Findings of statistical significance are summarized in the study characteristic Table 1. Two articles^{28,31} found no statistical significance in total CFU, *S. mutans* or for *Lactobacillus* spp., recovered from different colour categories but harder lesions contained less total CFU than softer lesions^{28,31}. One article did not carry out any statistical analysis between the microflora associations with colour and/or hardness categories because the study did not include sufficient number of samples³⁰.

B-Diagnostic test studies. Both articles assessed the relationships between total CFU with colour and hardness^{10,14}. Total CFUs was highest in all soft colour categories compared to all colour categories of leathery lesions which in turn contained more total CFU than all hard colour categories^{10,14}, and black soft lesions contained more lactobacilli than black leathery lesions¹⁰. For further details, please refer to Table 1.

Statistical analysis for pooled CFU's outcomes. Some of the included papers had to be excluded from the meta-analysis because they did not have a mean load for each category²⁸, where microbial levels were only measured by turbidity methods^{28,30}, or where data were only shown as percentages of specific species instead of the total CFU¹⁴.

The microbial load in each colour and hardness category (Black [B], Dark Brown [DB], Light Brown [LB], Yellow [Y], Soft [S], Medium [M], Hard [H]) was calculated by measuring the log₁₀CFU in the studies by Bjorndal

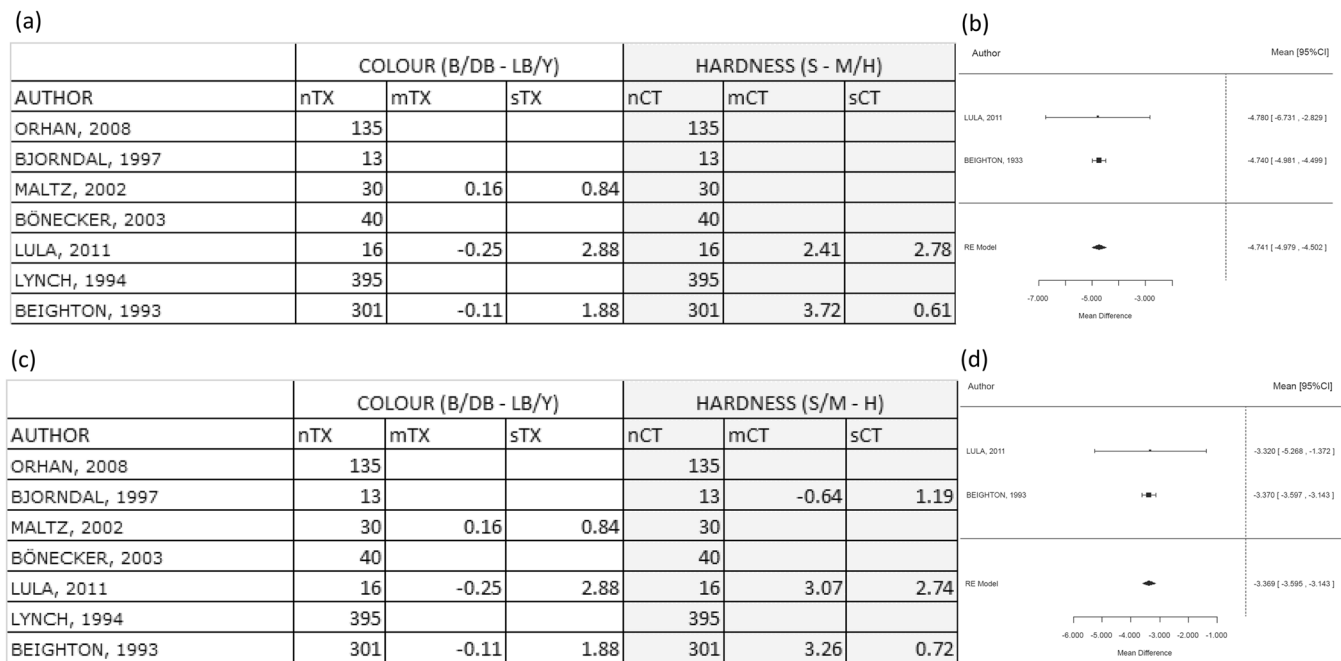


Figure 4. Comparison B/DB - LB/Y versus S - M/H and comparison B/DB - LB/Y versus S/M - H. Comparison between dark brown/black against other yellow/light brown with soft and other hardness categories (a,b) and comparison between dark brown/black against other yellow/light brown with hard and other hardness categories (c,d). (a) Table shows mean differences for log(CFU) between categories B/DB - LB/Y (colour test) and categories S-M/H (hardness test). (b) WMD value for the difference is -3.67 ($p < 0.001$), favoring the hypothesis that the hardness test is more discriminant. (c) Table shows mean differences for log(CFU) between categories B/DB - LB/Y (colour test) and categories S/M - H (hardness test). (d) WMD value for the difference is -3.37 ($p < 0.001$), favoring the hypothesis that the hardness test is more discriminant. TX = treatment/exposure group (colour test); CT = control group (hardness test); n = number of samples; m = mean; s = standard deviation.

et al., Maltz *et al.*, Lula *et al.*, and Beighton *et al.* as shown in Table 3. When lesions were described as leathery these were classified as medium for this analysis.

When comparing B-DB/LB/Y vs. S-M/H, the only possible comparison was by using the study by Beighton *et al.* It was found that there were more CFU (2.32 units of \log_{10} CFU in average) in black samples than any of the other colour groups. In addition, it was found that there were on average 3.72 units more of \log_{10} CFU in soft samples than in leathery samples and in turn these had significantly more CFU than in hard lesions. The differences in CFU between the hardness categories were considerably more than any differences between the CFU contained in the different colour categories. Therefore, the hardness test was more discriminant, but no meta-analysis was conducted because only 1 study was involved in this comparison as shown in Fig. 3.

When comparing B-DB/LB/Y vs. S/M - H, Beighton *et al.* and Bjorndal *et al.* provided data in relation to CFU differences between these categories, so a meta-analysis was performed, and a forest plot is shown in Fig. 3. The WMD value for the difference was -0.067 . The effect size on the population was estimated to be between -1.88 and 1.74 with a confidence of 95%. It should be noted that the interval includes zero, so no significant difference was reached ($p = 0.942$). Therefore, none of these tests were more discriminant than the other.

Because of the large sample size of Beighton *et al.*, ($n = 301$) in the comparisons between B/DB - LB/Y vs. S - M/H, its weight was very dominant compared to Lula *et al.* ($n = 16$). The overall effect measure of the meta-analysis was similar to the Beighton *et al.* individual results. The WMD value for the difference was -3.67 ($p < 0.001$), indicating that the hardness test was more discriminant than colour. The analysis for this forest plot is shown in Fig. 4.

When comparing B/DB - LB/Y vs. S/M - H, the results are similar to the previous comparison. The WMD value for the difference was -3.37 ($p < 0.001$), suggesting that the hardness test was again more discriminant than colour and the forest plot of this analysis is presented in Fig. 4.

When comparing B/DB/LB - Y vs. S - M/H, the WMD value for the difference was -4.74 ($p < 0.001$), supporting that the hardness test was again more discriminate than colour. The analysis for this forest plot is shown in Fig. 5.

When comparing B/DB/LB - Y vs. S/M - H, the WMD value for the difference was -4.45 ($p < 0.001$), also indicating that the hardness test was more discriminant than colour. The forest plot of this analysis is shown in Fig. 5.

For Bjorndal *et al.*, Lula *et al.* and Beighton *et al.* the effect size for the hardness category is larger than the colour categories; for Bönecker *et al.* the results again favored hardness over colour. Consequently, the weighted total effect size showed a large advantage for hardness as a discriminant rather than colour. However, this difference is

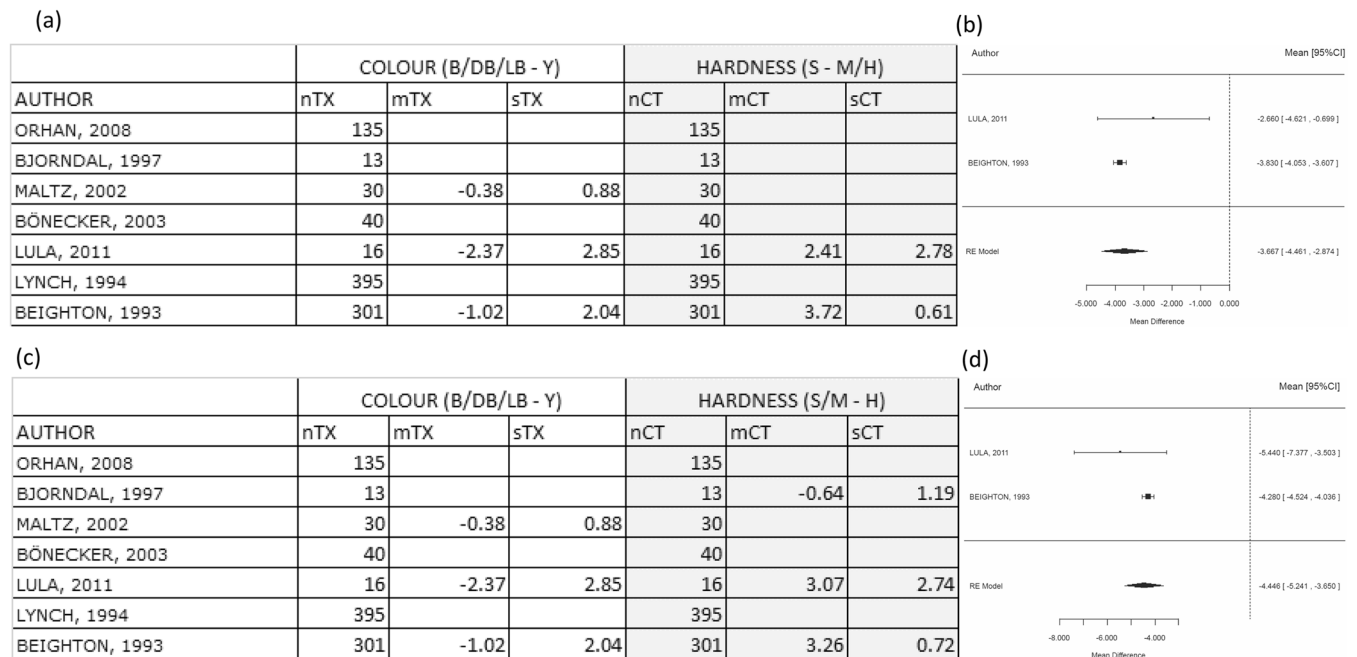


Figure 5. Comparison B/DB/LB - Y versus S - M/H and comparison B/DB/LB - Y versus S/M - H. Comparison between yellow against other colours with soft and other hardness categories. (a) Table shows mean differences for log(CFU) between categories B/DB/LB - Y (colour test) and categories S - M/H (hardness test). (b) WMD value for the difference is -4.74 ($p < 0.001$), favoring the hypothesis that the hardness test is more discriminant. Comparison between yellow against other colour with hard and other hardness categories. (c) Table shows mean differences for log(CFU) between categories B/DB/LB - Y (colour test) and categories S/M - H (hardness test). (d) WMD value for the difference is -4.45 ($p < 0.001$), favoring the hypothesis that the hardness test is more discriminant. TX = treatment group (colour test); CT = control group (hardness test); n = number of samples; m = mean; s = standard deviation.

caused, especially, by the huge dominance of the Beighton *et al.* data in the overall numbers. The effect size table is shown in Table 4.

Discussion

Dental caries involve microorganisms excreting acid resulting in changes in dental tissue hardness and colour. However, the accuracy and reliability of dentine colour and hardness as indicators for carious lesion severity has never been assessed in a systematic review. Thus, the aim of this study was to systematically analyze published research investigating whether colour or hardness of dentine caries was a more accurate, reliable, and valid method in detecting carious lesion severity when related to the amount of detectable CFU in biopsies of these lesion categories.

Seven papers met the inclusion criteria for this systematic review from searching multiple electronic databases and hand searching multiple journals: five articles were RCTs and two were diagnostic test studies. The five included RCTs were not primarily conducted to investigate the relationship between colour and hardness with the numbers of microorganisms. They were rather trial studies to investigate certain clinical interventions (e.g. one-visit IPT, two-visit IPT, DCE, step-wise excavation, and atraumatic restorative treatment). For the purpose of this systematic review, only the RCT data related to the CFU counts obtained from carious biopsies following the colour and hardness measurements were reported.

The quality of each study was critically appraised using CASP tools. The RCT and diagnostic test studies have separate CASP tools to systematically examine and appraise the evidences of each article. Using CASP tools, it was noticeable that many articles lacked details of patient's age, gender, type of teeth, socioeconomic status, ethnicity, diet, or sample size calculations^{19,30,31}. This would question the quality of evidence in the included papers. In fact, the unreported randomization protocols^{10,14,19,29-31}, sample size calculations^{10,14,19,28-30}, and confidence limits^{10,14,19,28-31} are additional evidence of the low quality of evidence in most of the included studies. This study shows that only the studies by Lula *et al.*, Lynch *et al.* and Beighton *et al.* reported acceptable levels of evidence.

From the meta-analysis, only 4 articles were considered for meta-analysis, as these provided CFU data for different categories of colour and hardness. Differences in log₁₀CFU between categories of hardness were significantly greater than between categories of colour. However, CFU results were also not consistently reported. For instance, all 30 samples for Maltz *et al.* were categorized as 'soft', thus no comparison was possible. Also, Bjorndal *et al.* provided CFU for samples ($n = 31$) but they all were classified exclusively in only 2 colour categories (dark brown and black) and 2 hardness categories (medium hard and hard). Therefore, comparisons had to be focused to results from Lula *et al.* and Beighton *et al.* The difference between their sample sizes was extreme, so overall

	Colour test	Hardness test
ORHAN, 2008		
BJORN DAL, 1997	0.10	0.28
MALTZ, 2002	0.05	—
BÖNECKER, 2003 ^a	0.26	0.23
LULA, 2011	0.38	0.69
LYNCH, 1994		
BEIGHTON, 1993	0.55	3.14
Total effect size	0.46	2.62

Table 4. Summary of the effect size (f) of the multiple comparisons of bacterial load means in colour and hardness examinations. Effect size (f) for a one-way analysis of variance to estimate differences in total bacteria load between categories of the colour and hardness test. Total effect size is the weighted (by sample size) mean from the authors. It is possible to estimate the effect size for Bönecker because it does not depend on the original units. The effect size is a standardized value.

results from the meta-analysis inherited mostly properties and results from Beighton *et al.* As with weighted effect size calculations, results data also showed a large advantage for hardness examination compared to colour examination. However, unlike pooled meta-analysis calculations, the study by Bönecker *et al.* was added to the weighted total effect size calculations. This was possible because effect size can be calculated regardless of the differences in the reported units of CFUs as effect size calculations measure ratio of variability between groups to within-groups³². Nevertheless, the meta-analyses and effect size calculations data in this study indicate that the texture (or hardness) tactile clinical examination has more discriminatory power in comparison to the colour visual clinical examination. Therefore, hardness or texture categories are more reliable in reflecting the amount of isolated cariogenic microorganisms and thereby more reliable in detecting the severity of carious lesions. Thus, based on the findings of this study, the hardness tactile clinical examination is a more specific and sensitive clinical examination method as hardness categories would be more reliable indicators for treatment planning in which soft carious lesions would harbor more cariogenic microorganisms, if compared to hard carious lesions, and therefore may require a more intrusive dental clinical intervention. These findings are consistent with previously reported data in which hard carious lesions were found to harbour no cariogenic bacteria (e.g. neither streptococci nor lactobacilli) and therefore required no dental clinical interventions. On the other hand, a large proportion of soft carious lesions contained the latter species (63.6% and 48.4% respectively) and required both caries debridement and restoration¹⁰.

The findings of this systematic review can be viewed as additional evidence in support of the International and Caries Detection Assessment System (ICDAS), which is currently considered as the recommended dental caries examination scoring index in dental practice³³. In this study, hardness clinical examination was found to be more specific and reliable than colour examination. Thus, changes in dental tissue hardness should warrant more scores in dental caries classification indices, which match the current setup of the ICDAS that emphasizes on the importance of hardness clinical examination by allocating more scores to changes in hardness³³.

The findings of this systematic review must be interpreted with caution as there are a number of limitations. Firstly, it was not possible to group study data and findings in all of the included 7 studies due to the inconsistency and heterogeneous approaches in which the authors recorded teeth colour, hardness, and CFU. For instance, the articles reported various methodologies for detecting and analyzing microbial CFU. Also, the authors used dissimilar approaches to report CFU data, such as: some studies presented the results with “total CFU counts”^{19,28}, while other studies used “turbidity tests”^{28,30}. Other articles normalized the CFU data by using $\log_{10}(\text{CFU}/\text{mg})$ ³¹ or $\log_{10}(\text{CFU} + 1)$ ^{10,14,29}. Secondly, all data tables presented by the diagnostic test studies^{10,14} showed CFU percentages of different bacterial species instead of stating the exact numerical actual counts, which further made it difficult to group and compare data and findings across the studies and therefore made it difficult for conducting a meta-analysis of all included studies which also hampered our ability to have an overall systematic understanding of the data. Finally, most of the included studies were RCTs, a design that is more vulnerable to sampling bias without accurate sample size calculations. It was noticeable in this systematic review that sample size calculation was only reported in one study³¹.

In summary, this study presents systematically reviewed evidence in support of hardness clinical examination being a more reliable and specific dental caries detection method. Therefore, it is recommended to consider the use of tactile examination over visual inspection for caries detection during routine and treatment dental visits. However, due to the limitations in this study, further research is needed to consolidate the findings in this systematic review.

In conclusion, the microbial differences between hardness types showed a weighted total effect size with a large advantage for hardness compared to colour and colour alone also yielded inconsistent microbial results. Hardness is more reliable than colour to detect dentine caries severity.

References

1. Marcenes, W. *et al.* Global burden of oral conditions in 1990–2010: A systematic analysis. *J. Dent. Res.* **92**, 592–597 (2013).
2. Dye, B. A., Thornton-Evans, G., Li, X. & Jafolla, T. J. Dental Caries and Sealant Prevalence in Children and Adolescents in the United States, 2011 – 2012. *NCHS. Data Brief* **191**, 1–8 (2015).
3. Kassebaum, N. J. *et al.* Global burden of untreated caries: A systematic review and metaregression. *J. Dent. Res.* **94**, 650–658 (2015).

4. Capurro, D. A., Iafolla, T., Kingman, A., Chattopadhyay, A. & Garcia, I. Trends in income-related inequality in untreated caries among children in the United States: Findings from NHANES I, NHANES III, and NHANES 1999–2004. *Community Dent. Oral Epidemiol.* **43**, 500–510 (2015).
5. Reddy, K. *et al.* Prevalence of dental caries among 6–12 years school children of Mahbubnagar District, Telangana State, India: A cross-sectional study. *Indian J. Dent. Sci.* **9**, 1–7 (2017).
6. Almerich-Silla, J., Boronat-Ferrer, T., Montiel-Company, J. & Iranzo-Cortes, J. Caries prevalence in children from Valencia (Spain) using ICDAS II criteria, 2010. *Med. Oral Patol. Oral Cir. Bucal* **19**, e574–e580 (2014).
7. Usha, C. & Sathyanarayanan, R. Dental caries - A complete changeover (Part I). *J. Conserv. Dent.* **12**, 46–54 (2009).
8. Nyvad, B. & Kilian, M. Microbiology of the early colonization of human enamel and root surfaces *in vivo*. *Eur. J. Oral Sci.* **95**, 369–380 (1987).
9. Aas, J. A. *et al.* Bacteria of dental caries in primary and permanent teeth in children and young adults. *J. Clin. Microbiol.* **46**, 1407–1417 (2008).
10. Beighton, D., Lynch, E. & Heath, M. R. A Microbiological Study of Primary Root-caries Lesions with Different Treatment Needs. *J. Dent. Res.* **72**, 623–629 (1993).
11. Beighton, D. & Lynch, E. Comparison of Selected Microflora of Plaque and Underlying Carious Dentine Associated with Primary Root Caries Lesions. *Caries Res.* **29**, 154–158 (1995).
12. Beighton, D. & Lynch, E. Relationships between yeasts and primary root-caries lesions. *Gerodontology* **10**, 105–108 (1993).
13. Lynch, E. & Beighton, D. Relationships between mutans streptococci and perceived treatment need of primary root-caries lesions. *Gerodontology* **10**, 98–104 (1993).
14. Lynch, E. & Beighton, D. A comparison of primary root caries lesions classified according to colour. *Caries Res.* **28**, 233–239 (1994).
15. Pitts, N. B. Current methods and criteria for caries diagnosis in Europe. *J. Dent. Educ.* **57**, 409–414 (1993).
16. Hellyer, P. H., Beighton, D., Heath, M. R. & Lynch, E. J. Root caries in older people attending a general dental practice in East Sussex. *Br. Dent. J.* **169**, 201–6 (1990).
17. Chong, M. J., Seow, W. K., Purdie, D. M., Cheng, E. & Wan, V. Visual-tactile examination compared with conventional radiography, digital radiography, and Diagnodent in the diagnosis of occlusal occult caries in extracted premolars. *Pediatr. Dent.* **25**, 341–349 (2003).
18. Bader, J. D., Shugars, D. A. & Bonito, A. J. A systematic review of the performance of methods for identifying carious lesions. *J. Public Health Dent.* **62**, 201–213 (2002).
19. Bjorndal, L., Larsen, T. & Thylstrup, A. A clinical and microbiological study of deep carious lesions during stepwise excavation using long treatment intervals. *Caries Res.* **31**, 411–417 (1997).
20. Ayna, B., Celenk, S., Atakul, F., Sezgin, B. & Ozekinci, T. Evaluation of clinical and microbiological features of deep carious lesions in primary molars. *J Dent Child* **70**, 15–18 (2003).
21. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J. Clin. Epidemiol.* **62**, 1006–1012 (2009).
22. CASP. CASP Tools & Checklists: CASP Randomised Controlled Trial Checklist. CASP Checklists 1–5 Available at: <http://www.casp-uk.net/#!casp-tools-checklists/c18f8>. (Accessed: 6th January 2017) (2013).
23. Critical Appraisal Skills Programme. CASP Diagnostic Checklist. Available at: <http://www.casp-uk.net/checklists>. (Accessed: 6th January 2017) (2017).
24. Kim, H.-Y. Statistical notes for clinical researchers: Evaluation of measurement error 2: Dahlberg's error, Bland-Altman method, and Kappa coefficient. *Restor. Dent. Endod.* **38**, 182–185 (2013).
25. Fleiss, J. L. & Cohen, J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ. Psychol. Meas.* **33**, 613–619 (1973).
26. Banerjee, M., Capozzoli, M., McSweeney, L. & Sinha, D. Beyond kappa: A review of interrater agreement measures. *Can. J. Stat.* **27**, 3–23 (1999).
27. Wiwanitkit, V. Authorship for leading journals: Issue on ethical problems. *Am. J. Roentgenol.* **206**, W61–W61 (2016).
28. Orhan, A. I., Oz, F. T., Ozcelik, B. & Orhan, K. A clinical and microbiological comparative study of deep carious lesion treatment in deciduous and young permanent molars. *Clin. Oral Investig.* **12**, 369–378 (2008).
29. Maltz, M., de Oliveira, E. F., Fontanella, V. & Bianchi, R. A clinical, microbiologic, and radiographic study of deep caries lesions after incomplete caries removal. *Quintessence Int.* **33**, 151–159 (2002).
30. Bönecker, M., Toi, C. & Cleaton-Jones, P. Mutans streptococci and lactobacilli in carious dentine before and after Atraumatic Restorative Treatment. *J. Dent.* **31**, 423–428 (2003).
31. Lula, E. C. O., Almeida, L. J. S., Alves, C. M. C., Monteiro-Neto, V. & Ribeiro, C. C. C. Partial caries removal in primary teeth: Association of clinical parameters with microbiological status. *Caries Res.* **45**, 275–280 (2011).
32. Grissom, R. J. & Kim, J. J. *Effect Sizes for Research: A Broad Practical Approach*. Mahwah, NJ: Lawrence Erlbaum. (2005).
33. Braga, M. M., Mendes, F. M. & Ekstrand, K. R. Detection Activity Assessment and Diagnosis of Dental Caries Lesions. *Dent. Clin. North Am.* **54**, 479–493 (2010).
34. Kidd, E. A. M., Ricketts, D. N. J. & Beighton, D. Criteria for caries removal at the enamel-dentine junction: A clinical and microbiological study. *Br. Dent. J.* **180**, 287–291 (1996).
35. Weerheijm, K. L., de Soet, J. J., de Graaff, J. & van Amerongen, W. E. Occlusal hidden caries: a bacteriological profile. *ASDC J. Dent. Child.* **57**, 428–32 (1990).
36. Kidd, E. A. M., Joyston-Bechal, S. & Beighton, D. Microbiological validation of assessments of caries activity during cavity preparation. *Caries Res.* **27**, 402–408 (1993).
37. Loesche, W. J., Eklund, S., Earnest, R. & Burt, B. Longitudinal investigation of bacteriology of human fissure decay: Epidemiological studies in molars shortly after eruption. *Infect. Immun.* **46**, 765–772 (1984).
38. Manji, F., Fejerskov, O., Nagelkerke, N. J. & Baelum, V. A random effects model for some epidemiological features of dental caries. *Community Dent Oral Epidemiol* **19**, 324–328 (1991).
39. Bonecker, M., Grossman, E., Cleaton-Jones, P. E. & Parak, R. Clinical, histological and microbiological study of hand-excavated carious dentine in extracted permanent teeth. *Sadj* **58**, 273–278 (2003).
40. Iwami, Y. *et al.* Evaluating the objectivity of caries removal with a caries detector dye using color evaluation and PCR. *J. Dent.* **35**, 749–754 (2007).
41. Fusayama, T., Okuse, K. & Hosoda, H. Relationship between Hardness, Discoloration, and Microbial Invasion in Carious Dentin. *J. Dent. Res.* **45**, 1033–1046 (1966).
42. Iwami, Y., Hayashi, N., Takeshige, F. & Ebisu, S. Relationship between the color of carious dentin with varying lesion activity, and bacterial detection. *J. Dent.* **36**, 143–151 (2008).
43. Torii, Y., Shimizu, A. & Tsuchitani, Y. The relationship between hardness and infection rate of carious dentin. *JOsaka Univ Dent Sch* **26**, 139–143 (1986).
44. Iwami, Y., Shimizu, A., Narimatsu, M., Kinomoto, Y. & Ebisu, S. The relationship between the color of carious dentin stained with a caries detector dye and bacterial infection. *Oper. Dent.* **30**, 83–89 (2005).
45. Nyvad, B., Machiulskiene, V. & Baelum, V. Construct and predictive validity of clinical caries diagnostic criteria assessing lesion activity. *J. Dent. Res.* **82**, 117–122 (2003).
46. Milnes, A. R. & Bowden, G. H. W. The microflora associated with developing lesions of nursing caries. *Caries Res.* **19**, 289–297 (1985).

47. Nyvad, B., Machiulskiene, V. & Baelum, V. Reliability of a New Caries Diagnostic System Differentiating between Active and Inactive Caries Lesions. *Caries Res.* **33**, 252–260 (1999).
48. Fejerskov, O. Concepts of dental caries and their consequences for understanding the disease. *Community Dent. Oral Epidemiol.* **25**, 5–12 (1997).
49. Kidd, E. A. M. & Fejerskov, O. What constitutes dental caries? Histopathology of carious enamel and dentin related to the action of cariogenic biofilms. *J. Dent. Res.* **83**, (2004).
50. Takahashi, N. & Nyvad, B. Caries ecology revisited: Microbial dynamics and the caries process. *Caries Res.* **42**, 409–418 (2008).
51. Kidd, E. A. M., Joyston-Bechal, S. & Beighton, D. Marginal Ditching and Staining as a Predictor of Secondary Caries Around Amalgam Restorations: A Clinical and Microbiological Study. *J. Dent. Res.* **74**, 1206–1211 (1995).

Acknowledgements

The authors would like to thank Juan Luis Gomez Martinez for his assistance with statistical analysis. This study was funded by the University of Nevada, Las Vegas (UNLV) School of Dental Medicine.

Author Contributions

L.H. and A.M. participated in study design, literature search, data collection, analysis, quality assessment, statistical analysis, interpretation, and in drafting and revising the manuscript. E.L. participated in study designing, analysis, interpretation, discussion, critical revisions, and manuscript revision.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019