European Journal of Orthodontics 25 (2003) 159–165

Quantifying enamel demineralization from teeth with orthodontic brackets—a comparison of two methods. Part 2: validity

Philip E. Benson*, Neil Pender** and Susan M. Higham**

*Department of Child Dental Health, University of Sheffield and **Cariology Research Group, Department of Clinical Dental Sciences, University of Liverpool, UK

SUMMARY This is Part 2 of an *in vitro* study investigating two techniques for recording and quantifying demineralization surrounding orthodontic brackets. In Part 1 the repeatability of computerized image analysis from digitally converted photographic slides and quantitative light-induced fluorescence (QLF) was explored. In Part 2 of the investigation the validity of each technique was examined.

Thirty halved human molars, shaped to look like incisors and with an orthodontic bracket bonded to the buccal surface were used. A small area of each of the four edges of the bracket was variously exposed to a demineralizing gel for 0, 3, 7, or 14 days. Images of the teeth were taken and analysed using the two techniques. Repeat images and analysis were carried out blind. The sensitivity, specificity, positive, and negative predictive values were calculated.

A negative result confirmed that there was no demineralization present in the majority of cases. A positive result was less reliable particularly for the occlusal and gingival regions. The 7- and 14-day demineralization patterns were detected every time using the photographic technique. The discrimination of the 0- and 3-day patterns was less reliable. The results of this study were extrapolated to allow for the prevalence of the condition following orthodontic treatment, and the implications of this for a putative clinical trial are discussed, together with the advantages and disadvantages of each method. Either technique would be applicable to use in a clinical trial.

Introduction

In Part 1 of this study the repeatability of two techniques for quantifying the area and severity of mineral loss from white spot lesions surrounding an orthodontic bracket was discussed. A technique for recording and measuring enamel demineralization should demonstrate validity, as well as repeatability. If repeat readings show little variation, but they include recorded information from non-relevant outcomes, the data are worthless.

Part 2 compares the ability of the two techniques to record and measure the presence or absence of white spots that have been created *in vitro* surrounding an orthodontic bracket.

Methods

The preparation of the teeth for this investigation has been fully described in Part 1.

Statistics

The validity of the two techniques was assessed by calculating the sensitivity, specificity, positive predictive value, and negative predictive values (Altman, 1991). For each site it was recorded whether the observer had carried out a reading. This was compared with the actual demineralization pattern for that tooth. It was assumed that if no reading was recorded then no demineralization could be detected. The results were placed in a 2×2 contingency table and the sensitivity and specificity were calculated. Altman (1991) defined these as:

Sensitivity—the proportion of positives that are correctly identified by the test.
Specificity—the proportion of negatives that are correctly identified by the test.

The probability of a correct diagnosis of demineralization from the photograph was calculated using the positive and negative predictive values. Altman (1991) defined these as:

Positive predictive value—the proportion of patients with positive test results who are correctly diagnosed.
Negative predictive value—the proportion of patients with negative test results who are correctly diagnosed.

The positive and negative predictive values were calculated for the four sites around the bracket of each tooth. Thus each tooth was imaged using the two techniques on two occasions (60 images for each technique), and two assessments of each site were

Table 1 Relationship between results of the detection of demineralization from photographs and whether there was demineralization present.

		Site subject to demineralization		
		Yes	No	Total
Site at which demineralization was detected	Yes	169	47	216
	No	15	249	264
	Total	184	296	480

Table 2 Relationship between results of the detection of demineralization from QLF and whether there was demineralization present.

		Site subject to demineralization		
		Yes	No	Total
Site at which demineralization was detected	Yes	158	14	172
	No	26	282	308
	Total	184	296	480

carried out (eight sites per image). Therefore, there were 480 assessments of the photographic images and 480 assessments of the quantitative light-induced fluorescence (QLF) images.

A diagnostic test's predictive values will vary widely depending on the proportion of individuals with the condition to whom the test is applied (Haynes, 1981). If the test is applied to a group of individuals with a high prevalence of the condition, the predictive values are likely to be better than if it is applied to a group with a low prevalence (Altman, 1991). Data from previous studies (Goreleick *et al.*, 1982; Mizrahi, 1982, 1983; Årtun and Brobakken, 1986; Øgaard, 1989) were used to provide an estimated proportion of patients who would have demineralization due to orthodontic treatment. It has been estimated that 56 per cent of subjects who have not had orthodontic treatment or who are about to start orthodontic treatment have white spot lesions. This compares with 70 per cent of subjects following orthodontic treatment. If it were possible to exclude all white spot lesions that are present at the start of treatment then, hypothetically, with a baseline of 56 per cent of subjects with white spots, it can reasonably be estimated that the prevalence of white spots due to orthodontic treatment is 14 per cent.

The association between the mean grey level from the photographic technique and the mean percentage change in fluorescence from the QLF technique was investigated using a scatterplot. The mean of both repeat readings from the two images for each technique was calculated. There were 47 readings when there was at least one reading from both techniques. Pearson's correlation coefficient was calculated.

Finally, a graph of the means and confidence limits of the grey levels and the mean change in fluorescence for the different time periods of demineralization was produced to examine the association between these two variables.

Results

Table 1 shows a contingency table for the results of detecting demineralization compared with the actual demineralization pattern using the photographic technique, and Table 2 shows the same results for the QLF technique. The calculated results for the sensitivity, specificity, positive, and negative predictive values for the two techniques are given in Table 3. This table also shows the hypothetical effect upon these figures of extrapolating the results of the present experiment to the population, taking into account an estimated prevalence of enamel demineralization following orthodontic treatment.

The ability to detect demineralization from a photograph when present was 0.92 or 92 per cent. In other words, 92 times out of 100 readings the presence of demineralization was correctly identified from the photograph. The same result for QLF was 0.86 (Table 3).

The ability to correctly identify the absence of demineralization was 0.84 for the photographic technique and 0.95 for the QLF (Table 3). The figures for the positive and negative predictive values are also shown

Table 3 Values for sensitivity, specificity, positive predictive value, and negative predictive value for the detection of enamel demineralization from photographs and QLF derived from the data in Tables 1 and 2.

	Photographs		QLF	
	Current experiment	Prevalence of 0.14	Current experiment	Prevalence of 0.14
Sensitivity	0.92	0.93	0.86	0.87
Specificity	0.84	0.84	0.95	0.95
Positive predictive value	0.78	0.52	0.92	0.74
Negative predictive value	0.94	0.99	0.92	0.98

in Table 3. According to this investigation, a positive identification of demineralization was correct in 78 per cent of cases from photographs and 92 per cent of QLF images. A negative test result correctly predicted the absence of lesions in 94 per cent of cases when using the photographic technique, and 92 per cent of cases using QLF (Table 3).

Table 3 also shows the results that have been calculated taking into account an estimated prevalence in the population. The sensitivity and specificity are not affected by this hypothetical calculation. The negative predictive values improved for both the photographic and QLF techniques. Therefore, if a negative result is registered there is almost certainly no demineralization present. The certainty of a positive result was adversely affected by this calculation. The positive predictive value for the photographic technique reduced to 0.52 and the QLF technique to 0.74. The results of this study suggest that if the estimated prevalence of the condition were to be correct, in a clinical trial one-half of the areas identified from the photographic technique and one-quarter from the QLF technique would be false positives.

The sensitivity, specificity, positive predictive value, and negative predictive values calculated for the four surfaces of the bracket for the two techniques are shown in Table 4. The readings from the left and right sides of the bracket showed good sensitivity and specificity, as well as positive and negative predictive values for both the photographic and QLF technique. This suggests that demineralization in these areas is accurately recorded. The occlusal and gingival recordings were less reliable, particularly the positive predictive value using the photographic technique. The QLF technique showed good results for the gingival area, but poor results for the occlusal area.

Table 5 shows the number of correct diagnoses made by the two techniques for the different time periods. It can be seen that both the photographic and the QLF techniques accurately diagnosed the 7- and 14-day demineralization. The slightly poorer result for the 14-day demineralization with QLF was due to several occlusal sites being missed. The 0- and 3-day demineralization were less reliably diagnosed by the photographic technique, and although the 'no demineralization' areas were accurately diagnosed by QLF, the 3-day areas were not.

Figure 1 is a scatterplot of the mean grey level from the photographic images and the mean change in fluorescence from the QLF images. No obvious association between the results was apparent. Two data points representing 41 and 61 grey levels were considered to be outliers. They represented points when only one measurement out of the four recordings was carried out from the photographs. They were excluded from the calculation of the Pearson's correlation coefficient, the result of which was –0.176. This was non-significant, confirming that there was no linear correlation between the mean grey level from the photographic technique and the mean change in fluorescence.

The mean and confidence limits of the mean grey level with time of exposure to the demineralizing gel are shown in Figure 2. Figure 3 is a similar graph for the mean and confidence level for the change in

Table 4 Values for sensitivity, specificity, positive predictive value, and negative predictive value for the detection of enamel demineralization for the different edges of the bracket for the photographic and QLF recordings.

	Gingival		Left		Occlusal		Right	
	Photo	QLF	Photo	QLF	Photo	QLF	Photo	QLF
Sensitivity	0.93	0.84	0.98	0.89	0.75	0.54	0.75	1.00
Specificity	0.69	0.94	0.84	1.00	0.93	0.95	0.93	0.93
Positive predictive value	0.68	0.92	0.85	1.00	0.72	0.72	0.88	0.91
Negative predictive value	0.91	0.87	0.92	0.91	0.94	0.89	0.94	1.00

Table 5 Number of correct and incorrect diagnoses for the two techniques for the time periods of 0-, 3-, 7-, and 14-day demineralization periods.

Actual demineralization		Correctly identified		Incorrectly identified		Percentage correct	
Days	*n*	Photo	QLF	Photo	QLF	Photo	QLF
0	296	249	282	47	14	84	95
3	80	65	60	15	20	81	75
7	52	52	51	0	1	100	98
14	52	52	47	0	5	100	90
Total	480	418	440	62	40	87	92

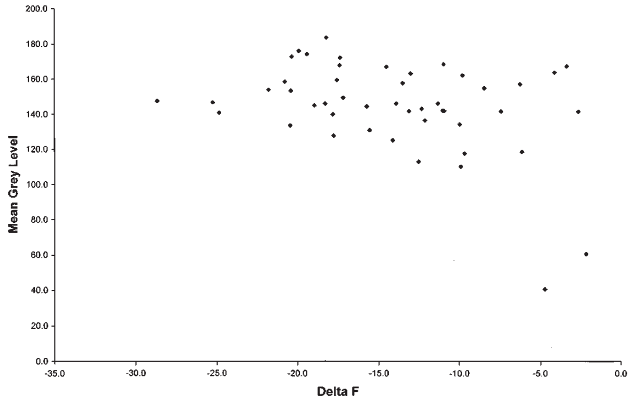

Figure 1 Scatterplot of mean grey level from the photographic technique and mean change in fluorescence (ΔF) from the QLF technique.

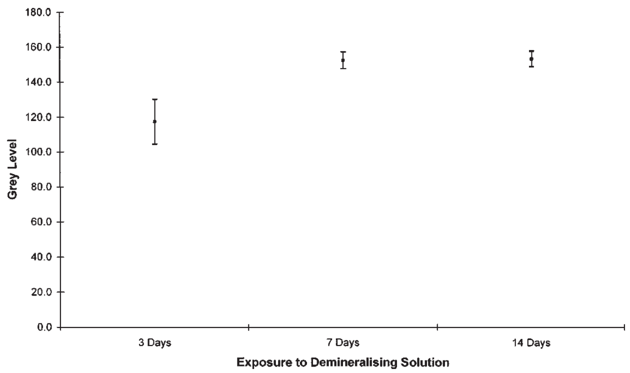

Figure 2 Mean and confidence intervals for the grey levels from the photographic technique (n = 80 for 3 days and n = 52 for 7 and 14 days).

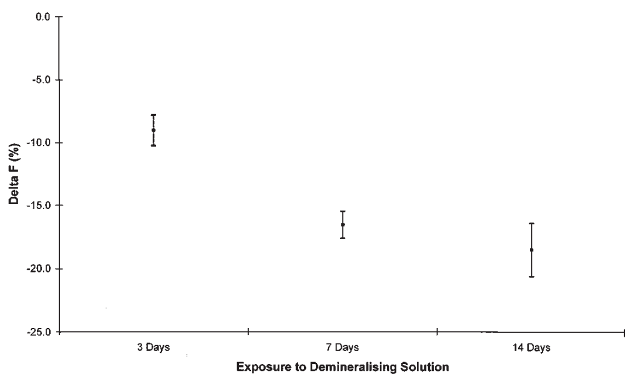

Figure 3 Mean and confidence intervals for mean change in fluorescence from the QLF technique (n = 80 for 3 days and n = 52 for 7 and 14 days).

fluorescence with exposure time. It can be seen that both graphs follow a similar pattern, albeit in the opposite direction because of the way demineralization is expressed by the two techniques. In Figure 2 the mean grey level is significantly larger (whiter) for the 7- and 14-day demineralization patterns compared with the 3-day pattern. However, there was no difference between the 7- and 14-day patterns. There was significantly greater fluorescence loss from the 7- and 14-day patterns compared with the 3-day pattern, but there was no difference between the 7- and 14-day patterns, although the 14-day pattern showed a wider confidence interval (Figure 3).

Discussion

The aim of the investigation described was to examine the repeatability and validity of two techniques for recording and quantifying enamel demineralization surrounding an orthodontic bracket. These techniques could be applied to investigations into the effectiveness of preventive agents. The most clinically definitive method of establishing the effectiveness of agents designed to prevent orthodontic demineralization is through a clinical trial. However such results will be dependent upon the accuracy of the technique or techniques for both recording and measuring the relevant outcomes. The more accurate and reliable the technique, the greater the power of the study and fewer participants need to be recruited. If, however, there is some variability then the power of the study is reduced and increased numbers are required to test the result to a significant level of determination.

In Part 1 it was found that both techniques demonstrated similar and satisfactory repeatability. However, from 92 sites that were exposed to the demineralizing solution, the number of measurements carried out from the photographs was 108 from the first assessment and 108 from the second assessment. The number of measurements carried out using the QLF technique was 87 from the first reading and 85 from the second reading. Therefore, this would suggest that the prevalence of demineralization was slightly overestimated when the photographs were examined. This agrees with Nunn *et al.* (1992), who found that a higher prevalence of opacities is scored from photographs compared with a clinical examination.

Somewhat more surprising was that assessments using QLF slightly underestimated the prevalence of demineralization. One of the advantages outlined for the fluorescent techniques is the greater contrast between sound and demineralized enamel, which makes the detection of demineralization easier (Angmar-Månsson and ten Bosch, 1987). One assumption made during this study was that areas that had been exposed to a demineralizing environment were actually demineralized. It is known that demineralization not only between teeth, but within the surface of one tooth is variable (Schäfer *et al.*, 1992). QLF may be reflecting the true situation, and little if any mineral loss had occurred in some of the 3-day exposure sites. No recordings were taken from two of the 3-day sites with either technique; there were no visible white spots and it is possible that these two sites were not actually demineralized. A previous study has shown a good correlation between the results of QLF and destructive methods of assessing mineral loss (Al-Khateeb *et al*., 1997).

The ability to detect demineralization from a photograph when it is present (Table 3; sensitivity = 0.92) was good. This figure was slightly lower for QLF (0.86), but several areas of obvious demineralization in the occlusal region were missed by QLF and it was noted that the orientation of the camera when capturing the QLF image was important in visualizing different aspects of the tooth surface. It may be necessary to capture several images of one tooth surface, analyse them, and calculate a mean reading.

The ability to correctly identify the absence of demineralization was not quite as accurate for the photographic technique (Table 3; specificity = 0.84), but better for QLF (0.95). The fact that this was lower than the ability to detect the lesions with photographs was probably due to reflections from the camera flash being confused with demineralization. The QLF system is optimized to reduce reflections and this proved a distinct advantage in reducing false positive readings. Methods of reducing the false positives produced by reflections on photographs require further investigation.

The concepts of sensitivity and specificity are not very useful in clinical practice. The positive and negative predictive values give a more practical interpretation of the value of a diagnostic test, because they describe the probability of the test giving the correct diagnosis (Altman, 1991). In this study the negative predictive value for the photographic technique was 0.94 (Table 3). In other words, in 94 per cent of instances when it is determined from a digitally captured photograph of a tooth with an orthodontic bracket that demineralization is not present, it will indeed not be present. In only 6 per cent of instances will it be determined that demineralization is not present, when it is actually present. The same figure for QLF was 0.92. This is a very satisfactory outcome for a clinical test.

The positive predictive value was 0.78. Therefore in 78 per cent of instances when it is determined from a digitally captured photograph of a tooth with an orthodontic bracket that demineralization is present, it will truly be present. In 22 per cent of instances it will be determined that demineralization is present, when it is actually not present. Therefore, nearly one-quarter of readings will provide false positives. Many of these false positives will be due to reflection from the flash and further investigations into methods of reducing reflection should be explored. QLF was much more accurate than the photographic technique at correctly detecting demineralization with a positive predictive value of 0.92, and this indicates that the system is optimized to reduce reflection.

The prevalence of the condition will be a factor in the accuracy of the results. If a condition has a high prevalence and the measuring technique is moderately reliable, a few participants with the condition will be missed, but because more participants suffer from it, sufficient will be recorded to produce a meaningful result. If there is a low prevalence and the method of recording or measuring is unable to detect the condition, then larger numbers of participants will be required to obtain a statistically valid result. For many conditions it is not possible to predict who will suffer the condition and target those for special investigation, so larger numbers need to be studied.

Accounting for the prevalence had no effect on the sensitivity and specificity (Table 3), and the accuracy of a negative result was slightly improved. However, the effect of prevalence had a profound effect on the predicted accuracy of a positive result. Approximately one-half of the positive results from the photographic technique were predicted to be accurate when the prevalence of the condition was taken into account. The result for QLF was more satisfactory with three-quarters predicted to be accurate.

The values for the different aspects of the bracket showed important differences between the two techniques (Table 4). Both techniques showed good negative predictive values for all four corners of the bracket, suggesting that if demineralization was not recorded it was unlikely to be present.

The positive predictive values for the left and right edges of the bracket were also high for both techniques (Table 4). However, the results were poorer for the gingival and occlusal edges of the bracket. The values for the photographic technique suggest that just over two-thirds of positive results were predicted to be accurate for the gingival and occlusal edges. The same results were better for the gingival edge using QLF, but again only two-thirds of positive outcomes for the occlusal edges were predicted to be accurate. The occlusal region is an area that has a low prevalence of demineralization (Mizrahi, 1982, 1983). However, the gingival area is a site with a high prevalence of demineralization during orthodontic treatment. There are two problems with the gingival area. First, it was small in comparison with the other areas, which made it more difficult to measure. Secondly, it was noted that the reflections tended to be concentrated in this region. Ways of improving the readings from this site need to be sought. This may involve taking several images of each tooth surface from different angles and averaging the results.

Areas of obvious (7 and 14 days) demineralization were recorded every time using the visual examination of the photograph (Table 5). Figure 2 demonstrates that there was a significant difference between the grey levels of the 3-day demineralization pattern and the 7- and 14-day patterns, and this was reliably detected during the subjective assessment of the image. QLF recorded the 7-day demineralization on 98 per cent and the 14-day on 90 per cent of occasions. Figure 3 shows similar results to the photographic readings, with a significant difference between the mean change in fluorescence of the 3-day and the 7- and 14-day demineralization patterns; however, this was not detected so successfully during the subjective assessment of the QLF image. As previously noted, the poorer result for the 14-day pattern using QLF was due to a number of 14-day demineralization patterns on the occlusal edge of the bracket being missed during the subjective assessment of the QLF image. To improve the detection of lesions using QLF it may be necessary to average the results from several images taken from different angles. QLF recorded the no demineralization group more successfully than the visual examination of the photographs (84 compared with 95 per cent).

The high potential for producing a false positive due to reflections from the flash is a disadvantage of the photographic technique. Another disadvantage of this technique is the interpretation of the grey level. It was noted that the differences in grey levels between demineralized areas, particularly the 3-day patches and non-demineralized areas were small and certainly too subtle for the computer to distinguish. In addition, grey levels for normal and demineralized enamel varied for different teeth with different lighting levels. Because of this variation, it is not possible to state a threshold grey level above which it can be said that demineralization is present. Indeed, the difference in grey levels between areas of normal and demineralized enamel was smaller than expected. Therefore, a grey level that would represent demineralized enamel on one image could be in the range of normal enamel on another image, because the image itself is generally brighter. It was clear that the visual assessment was looking at the buccal surface as a whole and subjective comparisons were being made about areas that were considered lighter (demineralized) than other areas.

QLF uses a method of calculating the change in fluorescence of the lesion by reconstructing the lesion to the same level as the surrounding sound enamel. It therefore has the advantage of producing a useful measurement, which is comparable not only with other teeth, but also with one tooth longitudinally (Al-Khateeb *et al.*, 1998). QLF is free of the problems of the flash and development of the photographic image, although care needs to be taken to exclude extraneous light when capturing the image. The correlation between the mean decrease in fluorescence with the QLF and the mean grey level measured from the photograph is shown in Figure 1. The correlation coefficient was non-significant showing that there was no linear correlation between the two techniques. The Delta F (DF) measurement of mineral loss of the QLF technique has been found to correlate with mineral loss using destructive techniques (Al-Khateeb *et al.*, 1997). This would also indicate that DF is a more useful measurement of mineral loss than the grey level from a photograph.

QLF has the ability to measure mineral loss longitudinally (Al-Khateeb *et al.*, 1998). The teeth in this study were carefully examined to exclude hypomineralized and developmental white spots before demineralization, which would confuse the results. In a clinical study, it would be necessary to exclude defects that are present at the start. If this were possible, then when the effect of estimating the prevalence of demineralization around a bracket, on the predictive capacity of the photographic technique is taken into account there is the potential for more false positives. It is possible, using digital technology, to create an image showing the differences between the grey levels of two images, by placing one image on top of another and subtracting the grey levels of one from the other. Thus the picture of a tooth before treatment could be subtracted from the picture of the same tooth after treatment and the resulting image would represent the change in the optical properties of the enamel during treatment. This is an area for further investigation.

Although the photographic technique has the advantage of being cheap and portable, QLF would appear to have distinct advantages for the quantification of mineral loss. In addition, images are instantaneous and could be repeated immediately, rather than waiting for a film to be developed, although with the advent of high-resolution digital cameras this is possible with the photographic technique. However, QLF is more expensive, needs to be used in a darkened room, and requires some experience to enable satisfactory images to be taken within the mouth.

Conclusions

This *in vitro* study carried out under ideal conditions has shown that:

1. The measurement of an area of demineralization on a tooth with an orthodontic bracket was found to be comparable in terms of repeatability and validity between the two techniques studied.
2. Quantification of mineral loss was more meaningful using QLF than photographs.
3. A negative result, using either technique, confirms that there is no demineralization present in the majority of cases. A positive result is less reliable, particularly for the occlusal and gingival regions from photographs.
4. Suggestions for further work have been made to improve the repeatability and validity of the recording and quantification of demineralization.

Address for Correspondence

Dr P. E. Benson
Department of Child Dental Health
School of Clinical Dentistry
Claremont Crescent
Sheffield S10 2TA, UK

References

Al-Khateeb S *et al*. 1997 Quantification of formation and remineralization of artificial enamel lesions with a new portable fluorescence device. Advances in Dental Research 11: 502–506

Al-Khateeb S, Forsberg C M, de Josselin de Jong E, Angmar-Månsson B 1998 A longitudinal laser fluorescence study of white spot lesions in orthodontic patients. American Journal of Orthodontics and Dentofacial Orthopedics 113: 595–602

Altman D G 1991 Practical statistics for medical research. Chapman and Hall, London, pp. 409–419

Angmar-Månsson B, ten Bosch J J 1987 Optical methods for the detection and quantification of caries. Advances in Dental Research 1: 14–20

Årtun J, Brobakken B O 1986 Prevalence of carious white spot lesions after orthodontic treatment with multibonded appliances. European Journal of Orthodontics 8: 229–234

Goreleick L, Geiger A M, Gwinnett A J 1982 Incidence of white spot formation after bonding and banding. American Journal of Orthodontics 81: 93–98

Haynes R B 1981 How to read clinical journals: II. To learn about a diagnostic test. Canadian Medical Association Journal 124: 703–710

Mizrahi E 1982 Enamel demineralisation following orthodontic treatment. American Journal of Orthodontics 82: 62–66

Mizrahi E 1983 Surface distribution of enamel opacities following orthodontic treatment. American Journal of Orthodontics 83: 323–331

Nunn J H, Murray J J, Reynolds P, Tabari D, Breckon J 1992 The prevalence of developmental defects of enamel in 15–16-year-old children residing in three districts (natural fluoride, adjusted fluoride, low fluoride) in the north east of England. Community Dental Health 9: 235–247

Øgaard B 1989 Prevalence of white spot lesions in 19 year-olds: a study on untreated and orthodontically treated persons 5 years after treatment. American Journal of Orthodontics and Dentofacial Orthopedics 96: 423–427

Schäfer F, Raven S J, Parr T A 1992 The effect of lesion characteristic on remineralization and model sensitivity. Journal of Dental Research 71: 811–813