Assessment of the validity and consequences of different methods of expressing the severity of dental fluorosis in a subject

Charles Mugisha Rwenyonyi, Jan Magne Birkeland and Ola Haugejorden
Faculty of Dentistry, University of Bergen, Bergen, Norway

Rwenyonyi CM, Birkeland JM, Haugejorden O. Assessment of the validity and consequences of different methods of expressing the severity of dental fluorosis in a subject. Acta Odontol Scand 2000;58:148–154. Oslo. ISSN 0001-6357.

The aim was to assess the validity and consequences of different methods of expressing severity of dental fluorosis in a subject. The analyses were based on Ugandan children ($n = 481$), aged 10–14 years, with life-long consumption of drinking water with either 0.5 or 2.5 mg fluoride per liter. Fluorosis was assessed using the Thylstrup and Fejerskov (TF) index. All children ($n = 219$) with 28 teeth and fluorosis on at least 1 tooth pair were selected to test methods of expressing fluorosis, e.g. the TF score on the most severely affected tooth, a maxillary central incisor, the median and the mean scores. A *test group* ($n = 40$), the 10 most and the 10 least severely affected children in each district, was used to evaluate the methods and a *reference group* ($n = 179$) to confirm or refute the findings in the test group. To evaluate consequences of the different methods of expressing severity of fluorosis in a subject, children from the low ($n = 130$) and the high fluoride ($n = 132$) districts not included in the test or reference group formed the community comparison groups. Comparison between the median (*gold standard*) and mean scores showed a significant deviation in the reference group only. Most of the partial recording methods, such as the score for the most severely affected tooth, were significantly and systematically higher than the median. While for all recording methods the median score was zero in both communities, the distribution of the subjects according to severity differed significantly between the communities. □ *Dental fluorosis; expressing severity; mean; median; partial recording*

Charles M. Rwenyonyi, Faculty of Dentistry, University of Bergen, Årstadveien 17, NO-5009 Bergen, Norway. Tel. +47 55 586539, fax. +4755586487, e-mail. Mugisha.Rwenyonyi@odont.uib.no

Both partial and full-mouth recording methods, as well as different indices, have been employed or recommended to reflect the severity of dental fluorosis in a subject (1–21). The partial recording methods are based on the severity of either a single tooth or the 2 most affected teeth (Table 1). Full-mouth recordings have been reported in terms of either a measure of location (mean or median) or the frequency of scores at a certain level of severity (Table 1).

Partial recording presupposes a high degree of agreement between full-mouth severity scores and the selected tooth or subsets of teeth. The severity of fluorosis is known to vary within the dentition (3, 22–27). The late mineralizing and erupting teeth are usually the most severely affected. The partial recordings are based on either the early (6–10) or the late mineralizing and erupting teeth (11). Hence, partial recording may under- or over-report dental fluorosis compared to full-mouth recording.

To our knowledge, only two studies (15, 28) have compared different methods of expressing the severity of fluorosis in a subject (Table 2). The comparisons were based on bivariate correlation analyses; a questionable approach as high correlation does not necessarily mean agreement between two methods (29). Moreover, correlation analyses do not reveal systematic under- or over-recording by one method as compared with another. Besides comparing different indices, no attempt seems to have been made to study the validity and consequences of different methods of reporting the severity of fluorosis. The aim of this paper was to assess the validity of different methods of expressing the severity of fluorosis in subjects and to determine any effect on comparisons between communities.

Material

The material used in these analytical assessments has been described previously (20). It comprised children ($n = 481$) aged 10–14 years who were continuous residents of rural small-scale subsistence farming communities in Kasese

Table 1. List of methods used or recommended for expressing the severity of fluorosis in a subject. The severity was based on Dean's index (1, 2), Thylstrup and Fejerskov's index (3, 4), and/or the Tooth Surface Index of Fluorosis (5)

Methods	Source/reference
Partial recording	
Central incisor	(6–9)
Central incisors or 1st molars	(10)
1st premolar	(11)
Two most affected teeth	(1, 12–14)
Full-mouth recording	
Mean	(15–18)
Median	(3, 11, 15)
Frequency of scores at certain level of severity	(19–21)

Table 2. Previous comparisons of methods of expressing the severity of fluorosis in a subject, using bivariate correlation analyses (15, 28). The severity was based on Dean's index (1, 2), Thylstrup and Fejerskov's index (3, 4) and/or Tooth Surface Index of Fluorosis (5)

	Source/reference	
Method	Kingman (15)	Katz and Brizendine (28)
Partial recording		
Highest score	+	–
Two highest scores	+	–
4th highest score	+	+
Mode score	–	+
Full-mouth recording		
Mean score	+	+
Median score	–	+
Percentage of scores for whole dentition	+	+
Percentage of scores for anterior or posterior teeth	–	+

+ Method included, – Method not included.

Table 3. Distribution of children according to group and district

Group	Kasese (0.5 mg F/l) (*n*)	Kisoro (2.5 mg F/l) (*n*)	Total (*n*)
Test – validation of methods	20	20	40
Reference – verify the test	13	166	179
Community comparison	130	132	262
Total	163	318	481

($n = 163$) and Kisoro ($n = 318$) districts in western Uganda. The fluoride concentration of the drinking water was 0.5 and 2.5 mg per liter in the respective districts.

Under field conditions, the children were clinically examined by one dentist (C.M.R.) while seated on a chair outside the school building in the shade. Natural light was the source of illumination. Cotton rolls were used to wipe the teeth clean and to control saliva. Severity of fluorosis was assessed on the vestibular surface of each permanent tooth (excluding 3rd molars) using the modified Thylstrup and Fejerskov (TF) index (4). The prevalence of dental fluorosis (TF score ≥ 1) was 35% in the low and 77% in the high fluoride district. Fluorosis at TF score ≥ 5 was recorded in 6% and 32% of the children, respectively.

In order to assess the validity of different recording methods, each subject had to fulfill the following criteria: all 28 teeth present and at least 1 tooth pair with fluorosis. This left 219 children: 33 and 186 from low and high fluoride districts, respectively. Among the 219 children, a *test group* of 40 children, the 10 most and the 10 least severely affected by fluorosis, was selected (Table 3). The most severely affected children were chosen on the basis of the highest score for any tooth pair, whereas the least affected had at least 1 tooth pair with a TF score ≥ 1. In these children ($n = 40$), severity scores from 0 to 8 were represented. The remaining 179 children: 13 and 166 from the low and high fluoride communities, respectively, were used as the *reference group* (Table 3) to confirm or refute the findings in the test group.

† FDI tooth nomenclature.

Children not included in the test and reference groups: 130 and 132 from the low and high fluoride districts, respectively, were employed to ascertain whether or not the different methods of expressing the severity of dental fluorosis in a subject affected inter-community comparisons. These children constituted the *community comparison groups* in Table 3 and do not represent their original districts. They had a varying number of permanent teeth and their TF scores ranged from 0 to 9.

Methods

Reporting dental fluorosis in a subject

To reflect the severity of dental fluorosis within a subject, either *partial* or *full-mouth recording* methods were used. The partial mouth recording employed the score for the maxillary right central incisor (*tooth 11*†), the score for the maxillary right first premolar (*tooth 14*†), the score for the most affected tooth (MAT) and the mean score for the 2 most severely affected teeth in a subject represented by *tooth 15*† and *24*† (mean 15 and 24). Full-mouth recording methods included the median and the mean TF scores for a subject, as well as the tooth prevalence of fluorosis (TPF %) at scores ≥ 1.

Statistical analyses

Data were analyzed using the Statistical Package for Social Sciences (SPSS). The frequency distribution of teeth

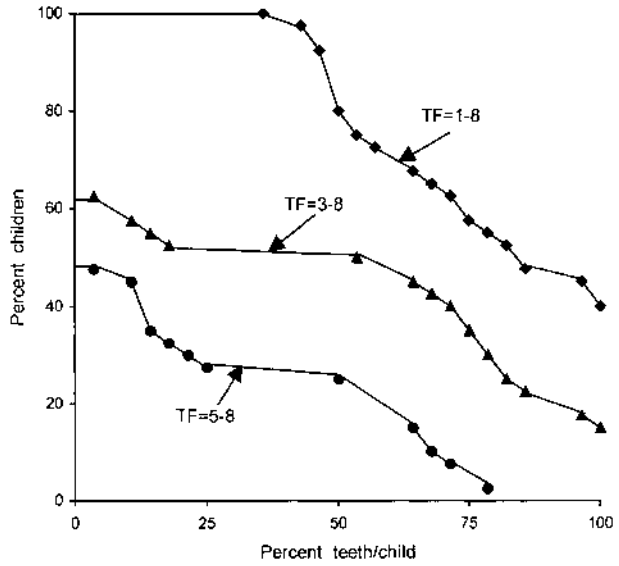

Fig. 1. Distribution of children according to the proportion of teeth per child exhibiting fluorosis of TF scores ≥1, ≥3, and ≥5 ($n = 40$).

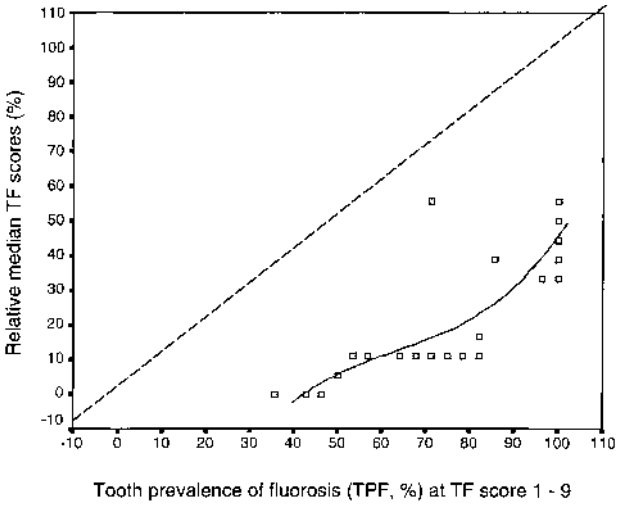

Fig. 3. Scatter plots and cubic regression between the relative median TF score of a subject (%) and the tooth prevalence of fluorosis (TPF, %) at scores ≥1 ($n = 40$). The broken line indicates perfect agreement between the methods.

per child with fluorosis at different levels of severity was employed to describe the material (Fig. 1). Dental fluorosis indices such as the TF index (3, 4) provide graduated scores of severity on an ordinal scale of measurement. This dictates use of non-parametric statistical analyses (30). However, for samples of at least 20 (31, 32) or 30 observations (33), parametric procedures can justifiably be employed for skewed data typical of fluorosis scores.

To compare different methods, a useful first step is to plot a scatter diagram of one method against a true status or *gold standard* (29). In order to meet the assumptions of the ordinal scale of measurement, and to provide for complete inventory of the dentition, the *median* was chosen as a gold standard. The scatter diagram and the cubic regression line were used to show the relationship between the median score and any other method of expressing severity of fluorosis in a subject (Figs 2–5). The number of plots may be less than the number of paired observations because of ties (identical scores). The cubic regression line (continuous line) described the observations better than the linear regression line, and was therefore chosen. The line of equality (broken line in Figs 2–5) was drawn to indicate the location of perfect agreement, and hence any

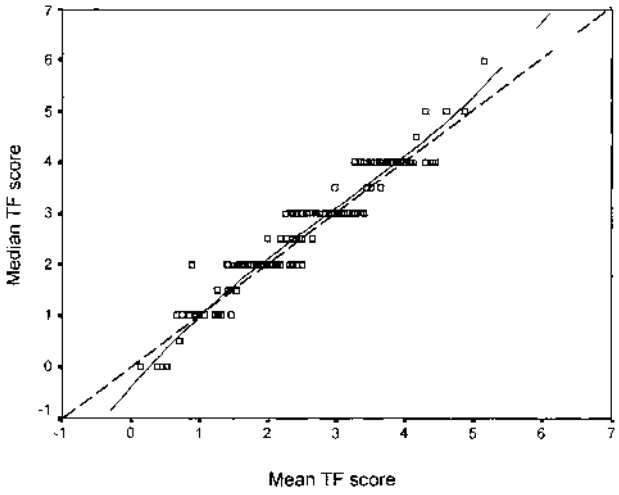

Fig. 2. Scatter plots and cubic regression between the median and the mean TF score of a subject ($n = 179$). The broken line indicates perfect agreement between the methods.

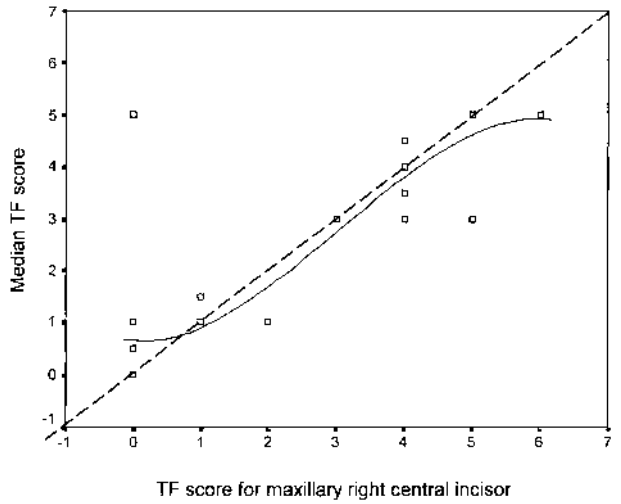

Fig. 4. Scatter plots and cubic regression between the median TF score of a subject and the TF score for the maxillary right central incisor ($n = 40$). The broken line indicates perfect agreement between the methods.

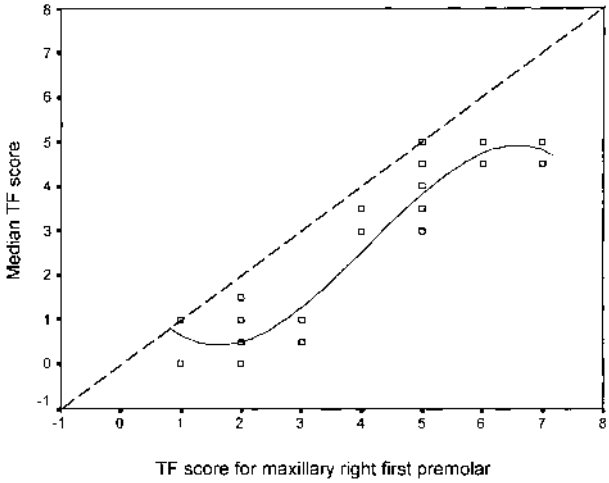

Fig. 5. Scatter plots and cubic regression between the median TF score of a subject and the TF score for the maxillary right first premolar ($n = 40$). The broken line indicates perfect agreement between the methods.

Table 4. Comparison of the severity of dental fluorosis based on various recording methods of expressing the severity in a subject according to group of children

Methods	Test ($n = 40$)	Reference ($n = 179$)
Median vs mean	ns	*
Median¶ vs TPF at score ≥ 1	*	*
Median vs tooth 11	ns	ns
Median vs MAT, tooth 14, mean 15 and 24	*	*

*$P < 0.01$; ns = not significant; ¶ relative median TF scores (%); TPF = tooth prevalence of fluorosis (%) at score ≥1; Mean 15 and 24 = mean score for tooth 15 and 24; MAT = most affected tooth.

systematic deviation between the methods. Wilcoxon signed ranks test for paired observations was used to check for significant intra-subject differences between the median score and other methods (Table 4).

In order to compare the TPF and the median score, a relative median value was calculated: (median score *x* 100)/9 where 9 is the maximum score of the TF index. To compare the severity of fluorosis between communities, the median and mean values were estimated based on different methods of reporting the severity in a subject (Table 5). However, the median at a community level could not be justified when the median score had been used to reflect the severity in a subject (34). The Mann-Whitney U test was employed to assess the significance of differences between communities when different methods were used to express the severity of fluorosis at a subject level (Table 5).

In the comparison between methods at a subject level, the significance was set at 1% to offset the effect of intra-subject correlation. In the inter-community comparison, 5% was used.

Results

Comparison of severity within a subject

The tooth prevalence of fluorosis (TF scores ≥1) was not significantly different in the test and reference groups ($P > 0.05$, Mann-Whitney U test). In the test group, all the children ($n = 40$) had fluorosis on 36% of their teeth (Fig. 1). The corresponding value in the reference group ($n = 179$) was 32%. The severity at TF scores ≥5, however, was significantly different in the two groups. Twenty-five percent of the children in the test group and only 3% in the reference group had TF scores ≥5 on at least 50% of their teeth.

Comparing methods of expressing severity in the test and the reference groups generally gave similar results (Table 4).

Full-mouth recording: In the intra-subject comparison, there was no significant deviation between the median (gold standard) and the mean TF scores in the test group ($P > 0.01$, Wilcoxon test). In the reference group, however, the median scores generally exceeded the mean scores (Fig. 2). This deviation was significant ($P < 0.01$, Wilcoxon test). The TPF at scores ≥1 overrated the severity compared to the relative median scores in both groups. This deviation

Table 5. The median, mean, and the range for different methods used to express the severity of dental fluorosis in a subject according to fluoride level in the drinking water in 2 communities. Estimates of dental fluorosis based on the modified Thylstrup and Fejerskov index (18)

	0.5 mg F/1; $n = 130$			2.5 mg F/1; $n = 132$		
Method	Median	Mean	Range	Median	Mean	Range
Median	–	0.4	0–3.0	–	1.2	0–7.0
Mean	0	0.3	0–3.1	0	1.1	0–6.4
MAT	0	0.6	0–5.0	0	1.9	0–9.0
Mean 15 and 24	0	0.4	0–4.0	0	1.8	0–8.5
Tooth 11	0	0.4	0–4.0	0	1.2	0–8.0
Tooth 14	0	0.6	0–5.0	0	1.8	0–8.0
TPF at scores ≥1	0	15.2	0–100	0	40.2	0–100

– Two or more median scores cannot be expressed as a median (34). MAT = most affected tooth; mean 15 and 24 = mean score for tooth 15 and 24; TPF = tooth prevalence of fluorosis (%) at score ≥1.

was systematic and significant ($P < 0.01$, Wilcoxon test; Fig. 3).

Partial versus full-mouth recording: Paired comparisons between the median scores, i.e. the gold standard, on the one hand, and the various partial mouth recordings on the other, gave different results (Table 4). There was a non-significant deviation between the median scores and the scores on the maxillary right central incisor (tooth 11) (Table 4, Fig. 4). The scores on the maxillary right first premolar (tooth 14; Fig. 5), the single most affected tooth (MAT), and the scores for the two most affected teeth (mean 15 and 24) were systematically and significantly higher than the median scores ($P < 0.01$, Wilcoxon test).

Comparing communities according to method of expressing the severity

The prevalence of fluorosis (TF scores ≥ 1) in the low ($n = 130$) and high fluoride ($n = 132$) communities was 18% and 45%, respectively. The corresponding tooth prevalence of fluorosis (TPF) was 15% and 40%. In both communities, all methods of expressing the severity of fluorosis in a subject gave a *median* of zero, while the *mean* indicated differences in the severity of fluorosis between the communities. Thus, the mean TF score of the low-fluoride community constituted 22–33% of the mean severity for the high-fluoride community (Table 5). The mean score for the 2 most affected (mean 15 and 24) teeth was lowest at 22%. However, independent of the method used to express the severity of fluorosis in a subject, there were significant differences in the distribution of children according to scores between the communities ($P < 0.05$, Mann-Whitney U test).

Discussion

Comments on methodology

To avoid complications due to missing values and to reduce the effect of ties for TF score 0, the evaluation of methods of expressing the severity of fluorosis in a subject was restricted to children having all 28 teeth present, and fluorosis on at least 1 tooth pair. The *test group* was based on the extreme cases of fluorosis to include a wide range of severity scores: here 0 to 8. Most of the children in the *reference group* were from the high fluoride district (93%) and comprised intermediate severity scores.

Bivariate correlation analyses have frequently been used to compare methods of measurement. These analyses can give misleading results (29) since they do not show whether or not there are systematic deviations. Scatter plots were used to show the relationship between methods. The cubic regression line was employed to summarize and to show the direction of any relationship. It also indicated whether or not there was systematic deviation between two methods in relation to the line of equality. The cubic regression line was chosen because it fitted the observations more closely than the straight line. For validation purposes, the median score was used as the gold standard. Since the TPF values are proportions, the median scores were transformed into relative values when comparing these methods (Fig. 3).

When a partial mouth recording method was compared with the median score for a subject, the score for that particular partial recording was not included when determining the median value. These exclusions, however, did not influence the results. Similarly, including or excluding the score for tooth 14 from the most severely affected tooth (MAT) had no effect on the MAT score. Thus, this study suggests that the score for tooth 14 may substitute for the most severely affected tooth in this material.

When determining the median scores for the communities, the estimates were based on the different recording methods used for a subject. This approach excluded use of the median score for a subject because when groups of observations are pooled the median of the combined group cannot be expressed in terms of the medians of the component groups (34). Due to the selection of the test and reference groups, the community comparison groups may not represent the two districts. However, the groups are composed of children exposed to different levels of fluoride in the drinking water and had varying degrees of severity of fluorosis.

Comments on findings

The analyses based on the test and reference groups showed that the degree of severity of fluorosis depended on the method used to express the severity, even though all recordings were based on the TF index. The findings in the reference group generally corroborated those in the test group, although the level of severity was different in the two groups. In the validation approach, the partial recording methods significantly overrated the severity compared to the median score, except for the score on tooth 11. Among the partial recording methods, the score for tooth 11 provided the closest approximation to the gold standard. Using full-mouth recording to express severity of fluorosis in a subject, the mean scores deviated significantly from the median in the reference group only (Fig. 2).

The median values were zero in both the low and high fluoride communities (Table 5), thus implying no difference in the severity, albeit 15% and 40% of the teeth and 18% and 45% of the children were affected in the respective groups. By definition, the median will be zero unless more than 50% of the observed units are affected. Less than 50% of the teeth or the subjects may be affected by high severity of fluorosis, but the median will not show fluorosis at the subject or at the community level. Although the community median values were zero, there were significant differences between the communities in the distribution of children according to scores. This inter-

community difference, however, was independent of methods used to express the severity in a subject (Table 5).

Provided there are at least 20 (31, 32) or 30 observations (33), the *mean* may be used to summarize ordinal data, such as index scores for the severity of fluorosis. Moreover, the mean score for a subject or a community takes into account all the observations. For public health purposes, the mean has the advantage that it reflects dental fluorosis in a subject or community provided it has been observed. In this material, the differences in the mean values of the recording methods between the two communities varied from 0.8 to 1.4 (Table 5). The 2 most affected teeth (mean 15 and 24) gave the largest mean difference at 1.4. No such differences could be revealed by the median, however. Owing to the ordinal level of the severity scores and regardless of the method used to express the severity in a subject, statistical comparisons between groups should rely on non-parametric tests.

The proportion of teeth with fluorosis at certain levels of severity (TPF) has been recommended for use to summarize such materials (4, 11, 19). The TPF, based on all teeth observed in a subject, is the basic unit of this system and has been used in analytical epidemiology (20, 21). The drawback of the TPF at a chosen level is that the variation in severity of fluorosis is not measured. It may also be influenced by a floor or ceiling effect depending on the cut-off point and the prevalence or the severity of fluorosis in the population. It may also be difficult for public health workers to interpret. From a statistical point of view, when the severity of fluorosis has been expressed as TPF at certain level of severity, parametric as well as non-parametric tests may be applied.

Conclusions

Various methods of expressing the severity of dental fluorosis in a subject might give different results at a subject level. Compared to the full-mouth median score, the partial recording methods generally resulted in a systematic and significant over-recording. By definition, the median does not indicate dental fluorosis unless more than 50% of the teeth or subjects are affected. Irrespective of the method of recording the severity of fluorosis in a subject, the median was zero, but the distributions of subjects according to severity were significantly different in the two communities. Thus the median may give misleading estimates. Since the mean utilizes all the observations, it will be a more informative method of summarizing and expressing the severity of dental fluorosis in a subject, or community, provided there is adequate sample size.

Acknowledgements.—We thank the children for participating in the study. The Norwegian Universities' Committee for Development Research and Education (NUFU, Pro. 61/96) and the Committee for Research and Postgraduate Training, University of Bergen supported the study financially.

References

1. Dean HT. Classification of mottled enamel diagnosis. J Am Dent Assoc 1934;21:1421–6.
2. Dean HT. The investigation of physiological effects by the epidemiological method. In: Moulton FR, editor. Fluoride and dental health. Washington DC: American Association for the Advancement of Science; 1942. p. 23–31.
3. Thylstrup A, Fejerskov O. Clinical appearance of dental fluorosis in permanent teeth in relation to histologic changes. Community Dent Oral Epidemiol 1978;6:315–28.
4. Fejerskov O, Manji F, Bælum V, Møller IJ. Dental fluorosis. A handbook for health workers. Copenhagen: Munksgaard; 1988.
5. Horowitz HS, Driscoll WS, Meyers RJ, Heifetz SB, Kingman A. A new method for assessing the prevalence of dental fluorosis—the Tooth Surface Index of Fluorosis. J Am Dent Assoc 1984; 109:37–41.
6. Riordan PJ, Banks JA. Dental fluorosis and fluoride exposure in Western Australia. J Dent Res 1991;70:1022–8.
7. Riordan PJ. Dental fluorosis, dental caries and fluoride exposure among 7-year-olds. Caries Res 1993;27:71–7.
8. Ibrahim YE, Affan AA, Bjorvatn K. Prevalence of dental fluorosis in Sudanese children from two villages with 0.25 and 2.56 ppm fluoride in the drinking water. Int J Paed Dent 1995;5:223–9.
9. Wang NJ, Gropen A-M, Øgaard B. Risk factors associated with fluorosis in a non-fluoridated population in Norway. Community Dent Oral Epidemiol 1997;25:396–401.
10. Sampaio FC, Fehr FR vd, Arneberg P, Gigante DP, Hatløy A. Dental fluorosis and nutritional status of 6- to 11-year-old children living in rural areas of Paraiba, Brazil. Caries Res 1999;33:66–73.
11. Wenzel A, Thylstrup A. Dental fluorosis and localized enamel opacities in fluoride and non-fluoride Danish communities. Caries Res 1982;16:340–8.
12. Aasenden R, Peebles TC. Effects of fluoride supplementation from birth on human deciduous and permanent teeth. Arch Oral Biol 1974;19:321–6.
13. Aasenden R, Peebles TC. Effects of fluoride supplementation from birth on dental caries and fluorosis in teenaged children. Arch Oral Biol 1978;23:111–5.
14. Angelillo IF, Torre I, Nobile CGA, Villari P. Caries and fluorosis prevalence in communities with different concentrations of fluoride in the water. Caries Res 1999;33:114–22.
15. Kingman A. Current techniques for measuring dental fluorosis: issues in data analysis. Adv Dent Res 1994;8:56–65.
16. Møller IJ, Pindborg JJ, Gedalia I, Roed-Petersen B. The prevalence of dental fluorosis in the people of Uganda. Arch Oral Biol 1970;15:213–25.
17. van Palenstein Helderman HW, Mabelya L, van't Hof MA, König KG. Two types of intraoral distribution of fluorotic enamel. Community Dent Oral Epidemiol 1997;25:251–5.
18. Yoder KM, Mabelya L, Robison VA, Dunipace AJ, Brizendine EJ, Stookey GK. Severe dental fluorosis in a Tanzanian population consuming water with negligible fluoride concentration. Community Dent Oral Epidemiol 1998;26:382–93.
19. Fejerskov O, Ekstrand J, Burt B. Fluoride in dentistry. Copenhagen: Munksgaard; 1996. p. 136–7.
20. Rwenyonyi CM, Bjorvatn K, Birkeland JM, Haugejorden O. Altitude as a risk indicator of dental fluorosis in children residing in areas with 0.5 and 2.5 mg fluoride per litre in drinking water. Caries Res 1999;33:267–74.
21. Awadia AK, Haugejorden O, Bjorvatn K, Birkeland JM. Vegetarianism and dental fluorosis among children in a high fluoride area of Northern Tanzania. Int J Paediat Dent 1999; 9:3–11.
22. Ainsworth JN. Mottled teeth. Br Dent J 1933;55:233–50.
23. Forrest JR. Caries incidence and enamel defects in areas with different levels of fluoride in the drinking water. Br Dent J 1956;100:195–200.

24. Møller IJ. Dental fluorose og caries. En oversigt over litteraturen samt nogle epidemiologiske undersøgelser af danske børn. Thesis. Rhodos, International Science Publishers, Copenhagen; 1965.
25. Larsen MJ, Richards A, Fejerskov O. Development of dental fluorosis according to age at start of fluoride administration. Caries Res 1985;19:519–27.
26. Manji F, Bælum V, Fejerskov O, Gemert W. Enamel changes in two low-fluoride areas of Kenya. Caries Res 1986;20:371–80.
27. Rwenyonyi CM, Birkeland JM, Haugejorden O, Bjorvatn K. Dental variables associated with differences in severity of fluorosis within the permanent dentition. Clin Oral Invest 2000;4:57–63.
28. Katz BP, Brizendine EJ. New forms for measuring dental fluorosis using the TSIF. J Dent Res 1996;75(IADR Abst):142.
29. Bland JM, Altman DG. Comparing two methods of clinical measurement: a personal history. Int J Epidemiol 1995;24:7–14.
30. Siegel S, Castellan NJ Jr. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill Book Company Inc; 1988.
31. Heeren T, D'Agostino RB. Robustness of the two independent samples t-test when applied to ordinal scaled data. Stat Med 1987;44:23–7.
32. Kingman A. Specific statistical considerations relevant to the design and analysis of gingivitis trials demonstrating product superiority or equivalence. J Periodont Res 1992;27:378–89.
33. Bradley JV. Distribution-free statistical tests. Englewood Cliffs: Prentice–Hall; 1968. p. 22.
34. Armitage P, Berry G. Statistical methods in medical research, 3rd ed. Oxford: Blackwell Science; 1996. p. 29.

Received for publication 1 October 1999
Accepted 18 April 2000