J Clin Periodontol 2006; 33: 172–176 doi: 10.1111/j.1600-051X.2006.00877.x

Journal of Clinical Periodontology

Measuring plaque in clinical trials: index or weight?

G. I. McCracken[1], P. M. Preshaw[1], I. N. Steen[2], M. Swan[1], M. deJager[3] and P. A. Heasman[1]

[1]School of Dental Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne, UK; [2]Centre for Health Services Research, University of Newcastle upon Tyne; [3]Philips OralHealthcare, Snoqualmie, WA, USA

McCracken GI, Preshaw PM, Steen IN, Swan M, deJager M, Heasman PA., Measuring plaque in clinical trials: index or weight? J Clin Periodontol 2006; 33: 172–176. doi: 10.1111/j.1600-051X.2006.00877.x. © Blackwell Munksgaard, 2006.

Abstract

Aim: To explore the possibility of using plaque weight rather than plaque index as a more objective, clinical outcome measure in periodontal clinical trials.

Materials and methods: The study initially recruited 12 healthy volunteers who abstained from tooth cleaning for 24 h on each of the three occasions and then for 48 h on each of a further three occasions to accumulate plaque. On a further three visits, the subjects abstained from tooth cleaning for 24 h and then brushed with a powered toothbrush for 2 min. A split-mouth design with target teeth was adopted and plaque was first scored separately on each side of the mouth using the Turesky modification of the Quigley and Hein plaque index. Post-brushing residual plaque was also scored after tooth cleaning. Inter-proximal plaque was dried, removed, pooled and weighed: on one side of the mouth from the entire inter-proximal surfaces; and from beneath the contact points on the contra-lateral side.

Results: Discriminant validity showed the index to have an advantage over weight in discerning between 24- and 48-h plaque deposits, and between 24-h plaque and post-brushing plaque. Test-retest validity confirmed that for repeated plaque growth, variability within subjects was greater than the variability between subjects. There was an association between plaque weight and plaque index although the regression lines were non-linear.

Conclusion: There appears to be no significant advantage in using plaque weight in periodontal clinical trials.

Key words: clinical trials; dental plaque; plaque indices

Accepted for publication 25 October 2005

Clinical trials investigating the efficacy of toothbrushes have conventionally used indices to evaluate the amount of plaque on tooth surfaces, for example, before and after tooth brushing (Heasman & McCracken 1999). The indices most often used to quantify plaque are those of Silness & Löe (1964), O'Leary (1967) and Quigley & Hein (1962) which was later modified by Turesky et al. (1970). Indeed, the index of Quigley & Hein (1962) was devised specifically to score plaque before and after tooth brushing in a comparative powered versus manual toothbrush trial.

There is no doubt that plaque indices have considerable advantages, most notably in that they are quick to use and, therefore, efficient in the clinical trial situation. There are, however,disadvantages also. Firstly, plaque indices are used to score plaque deposits on the clinical crown and do not assess subgingival deposits. This flaw was commented upon in a review by Ash (1964), who noted that while cleaning of the subgingival crevice should be considered a criterion of effectiveness, very few studies have evaluated the ability of toothbrushes to remove subgingival deposits.

Secondly, the indices are entirely subjective, particularly at the hard-to-visualize inter-proximal sites, and their use in clinical trials necessitates intra- and (when there is more than one examiner) inter-examiner calibrations that usually only achieve κ statistics in the range [0.60–0.75] (McCracken et al. 2003). Expected and standardized differences, and power and sample size calculations must allow for this degree of potential subjectivity and variability, thus necessitating a need to recruit relatively large subject cohorts leading to time-consuming and expensive clinical trials.

Thirdly, plaque indices are categorical in that they are based on discrete point scales where a score 2 (for example) does not necessarily indicate twice as much plaque as a score 1 (Quigley & Hein 1962, Silness & Löe 1964, Turesky et al. 1970). The use of means and parametric analyses may be inappropriate and the clinical relevance of reductions in plaque as a consequence of an intervention may be difficult, if not impossible to determine.

The primary aim of this study was:

- To explore the possibility of using plaque weight as the objective,

principal clinical measurement in clinical trials of toothbrushes.

Secondary aims were:

- To compare plaque weight as an outcome measure against the more conventional Turesksy modification of the Quigley and Hein Index (Turesky et al. 1970);
- To evaluate the ability of a cohort of healthy subjects to re-grow plaque in a repeatable and consistent manner over 24 and 48 h; plaque weight as an outcome measure can only be assessed either before or after an intervention such as tooth brushing. Post-brushing data (for example) can only be compared with pre-brushing data from a previous plaque growth episode and intra-subject consistency between episodes would need to be proven.

Material and Methods

The study was given a favourable ethical opinion by the Local Research Ethics Committee of Newcastle and North Tyneside, UK.

The clinical model recruited 12 healthy volunteers who satisfied the following inclusion criteria:

- Excellent general and periodontal health;
- 18–30 years of age;
- had a minimum of 18 natural teeth with at least one premolar and the first molar in each quadrant;
- were willing to participate and available at all times required for participation;
- were willing to abstain from brushing and using oral hygiene aids for 24 or 48 h before each clinical appointment.

This investigation was regarded as a pilot study principally because of the absence of published or in-house plaque weight data that would have provided an estimate of standard difference for this outcome. The sample size of 12 was, therefore, based simply on sample sizes that we have calculated and used in previous clinical trials adopting either a cross-over or parallel group design and with plaque index as the only outcome measure.

Screening visit (visit 1)

Before the screening visit, subjects were asked to refrain from brushing and using floss, mouth rinses or other oral hygiene aids for 24 h.

The screening examination included an intra-oral, soft-tissue examination, periodontal screening (BPE) and recording of pocket depth measurements, and plaque scores using disclosing solution and the Turesky modification of the Quigley–Hein plaque index (Turesky et al. 1970). A full mouth plaque score was calculated to ensure a minimum of 2.2 (as a further inclusion criteria) which confirmed each subject's ability to form dental plaque over the 24 h period. A professional prophylaxis was given and subjects then asked to abstain from tooth cleaning for a further 24 h before visit 2.

Study design

Subjects were requested to abstain from tooth cleaning for 24 h on three occasions (visits 2–4), and for 48 h on a further three occasions (visits 5–7) to allow dental plaque to accumulate. Visits 2–4 were on successive days and visits 5–7 on alternate days within a week.

Target (test) sites were selected as the inter-proximal sites (six tooth surfaces) between the distal surface of the first premolar and the mesial surface of the second permanent molar in all four quadrants. A split-mouth design allowed plaque to be collected from inter-proximal sites with boundaries delineated by the mesiobuccal, and mesiopalatal, distobuccal and distopalatal line angles of the tooth surfaces on one side of the mouth (line angle plaque) and from only the area directly beneath the contact point on the contra-lateral side (contact point plaque). At all visits, inter-proximal dental plaque at the test sites was also scored conventionally before plaque removal using the modified Quigley and Hein index (Turesky et al. 1970). A separate score was calculated for each side of the split-mouth (right and left dentition).

The plaque was then dried thoroughly using an air-syringe for 30 s, removed from the inter-proximal test sites using a dental explorer and then transferred to a pre-weighed square of aluminium foil using the tip of a number 11 scalpel blade to aid removal.

Contact point plaque was removed from inter-proximal target sites on one side of the mouth by inserting the tip of the explorer from the buccal direction only, contacting the target tooth immediately below the contact point at the most apical part of the gingival crevice (i.e., subgingival placement). A single sweep in a coronal direction was used, keeping the explorer on the tooth surface to remove the plaque.

Line angle plaque was removed from the contra-lateral side of the mouth. The first sweep was from the buccal direction and executed in an identical manner as for the removal of contact point plaque. Further sweeps of the explorer were then undertaken from both buccal and lingual directions until the inter-proximal surface between the line angles was visibly plaque free.

At three additional visits (8–10), 24 h plaque growth in 10 of the original subjects was again scored at each visit using the split-mouth design and the modified Quigley and Hein index. The subjects then removed plaque using a powered toothbrush (Sonicare Elite, Philips OralHealth Care, Snoqualmie, WA, USA) for 2 min. and the plaque index was repeated. Plaque was removed for weighing from line angle and contact point sites on the same target teeth using exactly the same protocol as described for visits 2–7.

This part of the study design allowed pre- and post-brushing plaque to be assessed conventionally using a plaque index while enabling post-brushing plaque to be evaluated by weighing.

Plaque weights

The pooled plaque samples were left at room temperature for 1 h before weighing, thus resulting in steady-state plaque weights. Preliminary studies confirmed that 1 h was adequate to allow complete evaporation of water from the deposits before weighing. The plaque from upper and lower target teeth on each side of the mouth was pooled and weighed using an M3 Microbalance with a weighing range of 150 mg, readability of 1 μg. Thus, pooled weights of the plaque from line angles and contact points were determined.

Examiner calibration

Before the screening visit, the single clinical examiner was calibrated for accuracy and repeatability using the Turesky Modification of the Quigley and Hein plaque index on subjects similar to those selected for the study. The teeth were disclosed for plaque before scoring. The calibration exercise involved two subjects with replicate

examinations of 210 sites within each subject. The unit of statistical analysis in the calibration was the site.

Results

There were complete data sets for 11 subjects over visits 2–7 and for 10 subjects over visits 8–10. The subjects who withdrew did so for non-study-related reasons.

The non-weighted κ value for the clinical examiner using the plaque index in the calibration exercise was 0.62 with perfect agreement scores for 73% of sites and 93% of scores being ± 1.

The mean (SD) summary data for line angle and contact point plaque weights (mg) after 24-h plaque growth (visits 2–4), 48-h plaque growth (visits 5–7) and post-brushing after 24 h plaque growth (visits 8–10) are given in Table 1. Table 1 also shows the conventional mean (SD) Quigley and Hein plaque indices for 24- and 48-h plaque growth together with both pre- and post-brushing scores at visits 8–10.

An analysis of variance was undertaken to determine whether there was an overall difference for plaque that was weighed from contact points compared with that weighed from line angles (Table 1). There was significantly more mean plaque weighed from line angle sites than from contact point sites at 24-and 48-h visits ($p = 0.008$). The magnitude of the difference was of the order 0.1 mg (24 h) and 0.3 mg (48 h) (data not shown). This significant difference was not, however, observed at the post-brushing visits (8–10).

Table 1. Mean (SD) plaque weights (mg), and corresponding mean Quigley and Hein plaque indices for subjects at visits 2–10

Visit	24 or 48 h	CPw Mean (SD)	CP PI Mean	LAw Mean (SD)	LA PI Mean
2	24	0.26(0.15)	4.2(0.3)	0.42(0.18)	4.2(0.3)
3	24	0.26(0.13)	4.0(0.4)	0.33(0.17)	4.1(0.4)
4	24	0.16(0.10)	3.9(0.5)	0.28(0.13)	4.0(0.6)
5	48	0.37(0.17)	4.6(0.2)	0.64(0.35)	4.6(0.3)
6	48	0.28(0.14)	4.3(0.2)	0.39(0.18)	4.3(0.4)
7	48	0.36(0.18)	4.4(0.2)	0.61(0.36)	4.4(0.3)
8	24	0.10(0.05)	2.8(0.6)	0.21(0.13)	2.9(0.6)
9	24	0.10(0.04)	3.0(0.5)	0.10(0.05)	3.0(0.5)
10	24	0.09(0.05)	2.8(0.6)	0.09(0.08)	2.7(0.7)

SD, standard deviation; CPw, weight of contact point plaque (mg); LAw, weight of line angle plaque (mg); PI, plaque indices for quadrants corresponding to CPw, and LAw.

Table 2. Discriminant validity. (a) The ability of plaque weight and plaque index to discriminate between 24-hour and 48-hour plaque, and (b) between 24-hour plaque and post-brushing deposits

a	Duration of growth		Paired differences		
Index	24 h Mean (SD)	48 h Mean (SD)	Mean (SD)	95% CI	ES
CPw	0.23 (0.13)	0.34 (0.16)	0.11 (0.19)	0.04, 0.17	0.57
LAw	0.34 (0.17)	0.55 (0.32)	0.21 (0.31)	0.10, 0.32	0.66
PI	4.08 (0.43)	4.43 (0.28)	0.35 (0.33)	0.27, 0.43	1.06
b	Occasion		Paired differences		
Index	Prebrushing Mean (SD)	Postbrushing Mean (SD)	Mean (SD)	95% CI	ES
CPw	0.23 (0.13)	0.14 (0.10)	− 0.10 (0.14)	− 0.15, − 0.04	0.68
LAw	0.34 (0.17)	0.13 (0.10)	− 0.22 (0.19)	− 0.28, 0.15	1.14
PI	4.08 (0.43)	2.84 (0.56)	− 1.27 (0.54)	0.27, 0.43	2.37

SD, standard deviation; CPw, weight of contact point plaque; LAw, weight of line angle plaque; CI, confidence interval; ES, effect size.

Discriminant validity

The discriminant validity (measured by effect size) is the ability of the outcome measures (index and weight) to distinguish between 24-h (visits 2–4) and 48-h (visits 5–7) plaque growth, as well as between 24-h plaque growth (visits 2–4) and 24-h post-brushing plaque (visits 8–10). The means (SDs) for contact point plaque growth, line angle plaque growth and the plaque index data at 24 (pre- and post-brushing) and 48 h are presented in Table 2.

Strictly speaking, the data are not paired but were treated as such so that data from visits 2 to 4 were paired firstly with those from visits 5, to 7, and then with the data from visits 8 to 10 (Table 2). The effect size, defined as the mean difference between plaque weights (mg) divided by the standard deviations of the paired differences was, for each comparison, greater for plaque index than for either of the plaque weight determinants indicating that the plaque index was a better discriminator between 24 and 48 h plaque.

Test–retest reliability

The measure of test–retest reliability is the intra-class correlation coefficient (ICC). The ICC used in this study assumes that there is random variation within subjects (for an individual subject 24 h plaque scores will vary randomly from visit-to-visit about a subject specific mean) and variation between subjects (the subject specific mean 24 h plaque score will vary randomly about some population mean value). These components of variance are denoted as σ_e^2 and σ_u^2, respectively.

The ICC is the proportion of the variation explained by differences between subjects $\frac{\sigma_u^2}{\sigma_u^2+\sigma_e^2}$. Thus, if all the plaque scores were the same on each visit, σ_e^2 would be zero; all the variation would be because of differences between subjects and the ICC would be 1.

The ICC and 95% confidence interval data for 24-h, 48-h and post-brushing plaque weights and indices are shown in Table 3. For 24 and 48 h plaque growth, the ICCs are relatively consistent for both plaque weight and index (0.4–0.6) which, taken together with the wide confidence intervals, suggests that there is little to choose between the methods. For post-brushing, plaque index (ICC, 0.52) appears to be more reliable for test–retest reliability than either line angle (ICC, 0) or contact point plaque weight (ICC, 0.139).

Relationship between plaque index and weight

In Fig. 1, plaque weight is plotted against plaque index for 24 h (pre- and

Table 3. Test–retest validity.

Plaque growth (h)	Index	σ_u^2	σ_e^2	Test–retest reliability		
				ICC	95% confidence interval	
24	CPw	0.095	0.092	0.518	0.213	0.812
	LAw	0.121	0.114	0.531	0.226	0.818
	Plaque index	0.312	0.284	0.548	0.298	0.780
48	CPw	0.130	0.096	0.646	0.355	0.868
	Law	0.232	0.213	0.543	0.238	0.823
	Plaque index	0.187	0.208	0.445	0.204	0.711
Post brushing	CPw	0.001	0.005	0.139	0.002	0.756
	LAw	0	0.005	0	–	–
	Plaque index	0.160	0.147	0.520	0.260	0.772

CPw, weight of contact point plaque; LAw, weight of line angle plaque; ICC, intraclass correlation coefficient; σ_e^2 within-subject variation; σ_u^2 between-subject variation; the proportion of variation in plaque growth and measurement explained by differences between subjects.

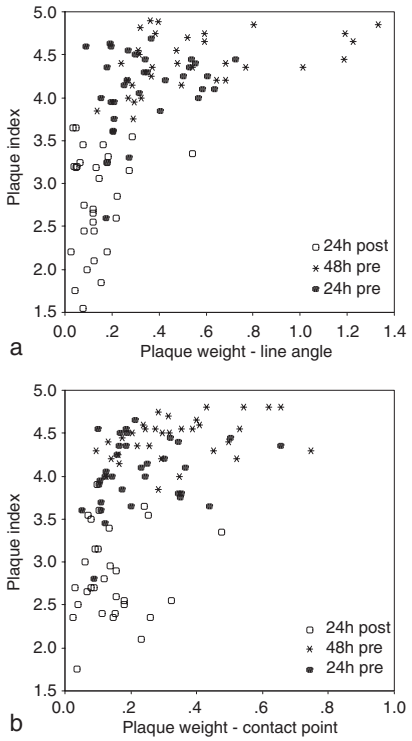

Fig. 1. Correlation plots between plaque weight and plaque index for plaque at (a) line angles and (b) contact points. Data points are shown for 24 h, 48 h and post-brushing plaque. There is a positive correlation between the outcome measures although the regression lines in each case deviate significantly from zero.

post-brushing) and 48 h plaque growth at target line angles (Fig. 1a) and contact points (Fig. 1b). In both cases, there was a strong positive correlation between plaque index and plaque weight. Regression analysis with plaque index as the dependent variable indicated that the relationship was not linear; a model that included a quadratic plaque weight term fitted the data significantly better than when the model contained only a linear term ($p<0.01$). One obvious explanation for the non-linear relationship is that the plaque index is capped (at a maximum value of 5) whereas plaque weight is not. In addition, the plaque index is an ordinal rather than an interval scale.

Discussion

Plaque indices remain the principal assessment of clinical outcome in clinical trials of toothbrushes and other methods of plaque removal. Despite well-recognized flaws with their use, researchers continue to use plaque indices mainly perhaps because of the overwhelming advantage of ease-of-use and, therefore, efficiency. Indeed, Goodson (1986) inferred that their adoption in clinical trials is undoubtedly perpetuated because they are ''tried and tested'' and acceptable to the FDA and other regulatory authorities. The overall objective of this pilot study was to determine whether plaque weight might be a more valuable outcome measure when compared to the Turesky modification of the Quigley and Hein plaque index in clinical periodontal trials.

Weight has been recently been assessed as an objective measure of plaque formation alongside plaque index in a study evaluating the inhibitory properties of three toothpastes. The observations showed that differences in plaque growth (in favour of a positive control) were more apparent with the objective measure; an expected observation with the potentially greater discriminatory power of a quantitative continuous scale (Claydon et al. 2005). The authors did conclude, however, that such potential benefits need to be evaluated against practical aspects of use in the clinical setting such as the difficulty of use, cost and time of using the objective measure.

In our study, the data in Table 1 reflect expected observations. The conventional, mean plaque index scores for the right and left sides of the split-mouth (corresponding to the line angle and contact point plaque weights) are virtually identical at all visits, both pre- and post-brushing. The mean, pooled weight of plaque removed from line angle sites (with greater surface area) was significantly greater than that removed only from beneath the contact points at visits 2–7. At visits 9 and 10 (after brushing 24 h plaque), the plaque weights at contact point and line angle sites were virtually identical; an unexpected observation given the potential variables of first removing, pooling and then weighing the deposits. This finding suggests that post-brushing, the majority of the remaining plaque is situated subgingivally and beneath the contact point. This finding was not observed post-brushing at visit 8, however, when the weight of line angle plaque was approximately twice that of contact point plaque. Further, a consistent finding from the study was that subgingival plaque was often not stained with disclosing solution (Fig. 2), thus adding credibility to the earlier comments of Ash (1964) and suggesting that the true efficacy of inter-proximal cleaning is not assessed with conventional plaque indices.

Neither discriminant validity nor test–retest reliability appeared to favour the introduction of using plaque weight rather than the index as an outcome measure. Indeed, in relative terms, the effect size for the plaque index suggests that the index is better able to distinguish between 24and 48 h as well as between pre-and post-brushing deposits (Table 2).

Test–retest reliability was assessed to determine the ability of subjects to consistently re-grow plaque and to evaluate between- and within-subject variability. One disadvantage of using plaque weight rather than an index as the outcome measure, is that weight may only be determined either before, or after plaque removal by tooth cleaning. Post- and pre-brushing data can therefore only be paired using measurements from plaque growth on different occasions. In terms of test–retest reliability

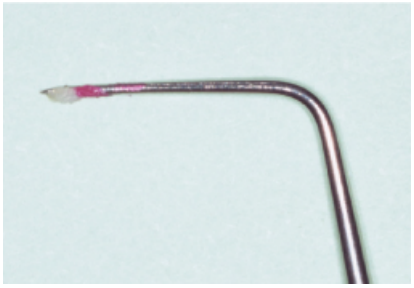

Fig. 2. An example of dental plaque removed from a contact point site. The unstained deposit indicates the deposit that would be included as a weight measurement but not included in the index score.

there was little to choose between the indices (Table 3). The degree of overlap between the confidence intervals for the ICC was large suggesting that it is not possible to distinguish between the methods. The confidence intervals were also wide suggesting that there is limited power to compare the methods.

Considering the post-brushing scores, plaque index had higher test–retest reliability than plaque weight. In conjunction with the plots in Fig. 1, this suggests that the index is much better at discriminating between individuals when the level of plaque is relatively low.

Further, the plaque index is capped at 5. After a certain threshold, visual inspection is not likely to discriminate between higher levels of plaque. Above this threshold, plaque weight is likely to discriminate better between different individuals. Test–retest reliability suggested that the plaque index was as good as plaque weight at discriminating between individuals after 24 h of plaque growth. Although not statistically significant, there may be a suggestion that 48 h plaque growth, plaque index may be slightly less effective at discriminating between individuals than is plaque weight. This trend, in conjunction with the plots in Fig. 1 would suggest that for longer plaque growth periods, plaque weight is likely to be much better at discriminating between individuals than is plaque index.

It may also be seen from Fig. 1a (for example) that the mean plaque indices for line angle plaque range from 1.5 to 4.6 for mean plaque weights less than 0.2 mg. Conversely, mean plaque indices over 4.5 are observed for a full range of mean plaque weights from approximately 0.1 to 1.3 mg. Similar observations are seen in Fig. 1b for contact point plaque weight and plaque index. Such findings may be expected as the Quigley and Hein index is dependent upon the site of the plaque as much as the quantity. A thin layer of plaque extending coronally beyond 2/3 of the clinical crown would score 5 whereas isolated, thick deposits of plaque just above the gingival margin would perhaps score 2. The weights of both deposits may, of course, be similar.

The time taken to carry out the plaque measurement will always be an overriding consideration in deciding which method to use in clinical trials (Claydon et al. 2005). In this study, without exception, the plaque indices for the target sites (six tooth surfaces) of one quadrant took less than 60 s to record. The clinical removal of plaque and the subsequent drying and weighing process took between 80 and 90 min., which, excluding the 60 min. for drying, necessitates up to 30 min. of clinical and laboratory time. It has to be questioned as to whether weighing plaque offers a more objective method and an overall significant advantage over the more efficient and conventional outcome measure.

Conclusion

The threshold at which plaque weight performs better than plaque index is likely to be greater at values that are currently considered to be of clinical importance within the context of a clinical trial. Typically, such a trial would involve measuring baseline plaque and re-growth plaque after 24 or 48 h. Over most of this range, the plaque index performs either better than, or at least as well as plaque weight.

Acknowledgement

The study was supported by a grant from Philips OralHealthcare, Snoqualmie, USA.

References

Ash, M. M. (1964) A review of the problems and results of studies on manual and power toothbrushes. *Journal of Periodontology* **35**, 202–213.

Claydon, N., Yates, R., Labello, R., Hall, C., Armstrong, R. & Addy, M. (2005) A methodology using subjective and objective measures to compare plaque inhibition by toothpastes. *Journal of Clinical Periodontology* **31**, 1106–1109.

Heasman, P. A. & McCracken, G. I. (1999) Powered toothbrushes – a review of clinical trials. *Journal of Clinical Periodontology* **26**, 407–420.

Goodson, J. M. (1986) Conference on clinical trials in periodontal diseases. *Journal of Clinical Periodontology* **13**, 384.

McCracken, G. I., Janssen, J., Swan, M., Steen, N., de Jager, M. & Heasman, P. A. (2003) The effect of brushing force and time on plaque removal using a powered toothbrush. *Journal of Clinical Periodontology* **30**, 409–413.

O'Leary, T. J. (1967) Periodontal diagnosis. *Journal of Periodontology* **38**, 617–624.

Quigley, G. A. & Hein, J. W. (1962) Comparative cleansing efficiency of manual and powered brushing. *Journal of the American Dental Association* **65**, 26–29.

Silness, J. & Löe, H. (1964) Periodontal disease in pregnancy. II Correlation between oral hygiene and periodontal condition. *Acta Odontologica Scandinavica* **22**, 121–135.

Turesky, S., Gilmore, N. D. & Glickman, I. (1970) Reduced plaque formation by chlormethyl analogue of vitamin C. *Journal of Periodontology* **41**, 41–43.

Address:
Professor P.A. Heasman
School of Dental Sciences
University of Newcastle upon Tyne
Framlington Place
Newcastle upon Tyne, NE2 4BW,UK
E-mail: p.a.heasman@newcastle.ac.uk

Clinical relevance

Scientific rational for the study: The clinical measurement of dental plaque is a fundamental necessity in clinical trials of toothbrushes and other plaque removal interventions. The use of conventional, yet subjective plaque indices is certainly efficient but the clinical relevance of statistically significant differences is often extremely difficult to determine.

Principal findings: The results of this pilot study show that plaque weight does not appear to be advantageous over a plaque index in terms of discriminatory ability or test–retest ability. The inefficiency of weighing plaque is also a significant disadvantage.

Practical implication: Researchers should continue to strive to identify other methods of objectively measuring dental plaque for use in clinical trials.