J Clin Periodontol 2006; 33: 151–156 doi: 10.1111/j.1600-051X.2005.00885.x

Journal of Clinical Periodontology

Evaluating the quality of active-control trials in periodontal research

Yu-KangTu[1,2], Ian Maddick[3], Margaret Kellett[1], Valerie Clerehugh[1] and Mark S. Gilthorpe[2]

[1]Department of Periodontology, Leeds Dental Institute, University of Leeds, Clarendon Way, Leeds, UK; [2]Biostatistics Unit, Centre for Epidemiology & Biostatistics, University of Leeds, Leeds, UK; [3]Department of Paediatric Dentistry, Eastman Dental Institute, University College London, London, UK

Tu Y-K, Maddick I, Kellett M, Clerehugh V, Gilthorpe MS. Evaluating the quality of active-control trials in periodontal research. J Clin Periodontol 2006; 33: 151–156. doi: 10.1111/j.1600-051X.2005.00885.x. © Blackwell Munksgaard, 2006.

Abstract

Aim: The increasing popularity of randomized-controlled trials (RCTs) has raiseed the issue of their quality. Frequently overlooked are the differences between superiority and equivalence trials. The purpose of this study was to apply specific methodological criteria to evaluate the quality of active-control trials using studies that compared guided tissue regeneration (GTR) with enamel matrix derivatives (EMD).

Materials and Methods: Seven RCTs were identified in the literature. Standard methodological criteria and seven additional criteria for trials using active-control groups were used to evaluate the quality of the seven RCTs.

Results: Two trials were considered as superiority trials. The remaining five provided no clear statement of their research aim. However, two claimed that EMD and GTR were equally effective, because their results failed to show a significant difference between EMD and GTR. Most trials did not meet the majority of the design criteria.

Conclusions: The general lack of compliance with quality criteria might place doubt on the value of these trials and may render any conclusions questionable. It is therefore important to distinguish clearly between superiority trials and equivalence trials, and to incorporate appropriate additional criteria in the design of future RCTs with active-control groups.

Key words: equivalence trials; randomized-controlled trials; superiority trials

Accepted for publication 8 November 2005

A recent systematic review examined the ''quality'' of randomized-controlled trials (RCTs) in dental research (Montenegro et al. 2002). One hundred and seventy-seven trials reported in the literature were assessed using the following criteria: (1) the randomization process; (2) the concealment of treatment allocation; (3) the blinding of examiners in the assessment of treatment outcomes; and (4) the follow-up of patients. Results indicated that the quality of these RCTs frequently failed to reach recommended standards.

An important issue in evaluating the quality of RCTs is the distinction between superiority and equivalence. This distinction has been discussed previously, but frequently overlooked in dental research (Fleiss 1992, Koch & Paquette 1997, Gunsolley et al. 1998, Burns & Elswick 2001). A clinical trial to test whether a new treatment modality has a genuine effect, or can give rise to better outcomes than the placebo or conventional treatment, is known as a superiority trial. A clinical trial to test whether the performance of a new treatment, which might be cheaper or easier to use, is comparable with that of the established treatment in current practice, is known as an equivalence trial.

Equivalence trials have become more and more prevalent in medical research as the use of placebo controls within trials has been claimed to be unethical when an established treatment is available (Rothman & Michels 1994). Although superiority and equivalence trials share common characteristics in study design – for example random allocation, concealment of randomization, blindness, intention to treat, etc – there are additional considerations for study designs of equivalence trials when an active-control is used (Greene et al. 2000, McAlister & Sackett 2001). This is important, because the design and interpretation of equivalence trials require a different methodology if the results and conclusions are to be valid. Failure to show the superiority of one treatment over the other does not prove that these two treatments are equivalent.

An RCT using an established treatment as an active-control group can be a superiority trial, if the aim of the trial is to show that the new treatment is better than the established treatment. In contrast, an RCT using an established treatment as an active-control group can be an equivalence trial (or a non-inferiority trial), if the aim of the trial is to show

that the new treatment is as good as (or at least as good as) the established treatment. However, the research hypotheses and study designs are entirely different for these two dissimilar research strategies. For instance, the required sample size for an equivalence trial might be substantially greater than that for a superiority trial (Tu et al. 2005). Underpowered active-control trials might give rise to a false impression that the new treatment is as good as the established one, if there is no distinction between superiority and equivalence in their research hypotheses.

Consider a hypothetical study, in which a new model of a powered toothbrush is tested against an existing model of a powered toothbrush and a manual toothbrush. The study might show no significant difference in the performance between the two powered toothbrushes, yet find a significant difference between the powered toothbrushes and the manual toothbrush. Therefore, the study demonstrates the superiority of the new powered toothbrush over the manual one. However, no definite conclusion can be arrived at regarding the comparison between the two powered toothbrushes; the failure to show that the powered toothbrushes are different does not necessarily imply they are equally effective, because by increasing the sample size, even a small difference in the performance between the two powered toothbrushes can be shown to be statistically significant.

The aim of this study is to demonstrate how to apply additional methodological considerations anticipated for equivalence trials (Greene et al. 2000, McAlister & Sackett 2001) to evaluate the quality of clinical trials when an established treatment is used as an active-control. An example in periodontology, chosen for illustration, is the comparison between guided tissue regeneration (GTR) and enamel matrix protein derivatives (EMD) as used in regenerative surgery. According to a recent systematic review (Esposito et al. 2003), no clinically important difference in the treatment outcomes was found between the two techniques. Nevertheless, no additional methodological criteria, such as distinction between superiority/equivalence trials, power calculation, and correct null hypothesis testing, for evaluation of active-control trials were taken into consideration within this systematic review.

Material and Methods

An electronic search of the database MEDLINE from the year 1997 (when the first case report on using EMD was published) up to March 2004 was undertaken to identify studies that compared the treatment effects of GTR and EMD. The reference lists cited by the two recent reviews on EMD (Kalpitis & Ruben 2002, Esposito et al. 2003) were crosschecked to ensure that the search strategy was comprehensive.

Standard methodological criteria for evaluation of the quality of RCTs have been discussed in detail by a recent systematic review (Montenegro et al. 2002). We describe seven additional criteria specifically designed for evaluation of quality of RCTs in which an active-control group is used, and apply them to clinical trials that compare treatment effects between GTR and EMD. The first six of these additional criteria have been used already in medical research (Greene et al. 2000, McAlister & Sackett 2001); they are modified here slightly to accommodate their use in this particular study. Criterion 7, which we introduce here to evaluate the general quality of clinical trials, is overlooked by most systematic reviews, despite being critical in assessing the conduct of RCTs.

Additional Methodological Criteria for Active-Control Trials

Criterion 1: Did the Research Aim Describe Test Equivalence or Superiority?

A clearly stated research aim is necessary for researchers to choose the pertinent variables for their study, i.e. boundaries for the magnitude of equivalence or difference between treatments, and is essential to make appropriate statistical and clinical inferences from the results. If the research question has been described as testing whether using EMD (or GTR) can achieve better treatment outcomes than GTR (or EMD), the study will be considered a superiority trial. If the research question has been described as testing whether using EMD (or GTR) can achieve comparable treatment outcomes with GTR (or EMD), the study will be considered an equivalence trial. When there is no clear statement of the research aims, such as ''the aim of this study is to compare EMD with GTR'', and if a conclusion is made in the study that the treatment effects of EMD and GTR are equally effective, the study will be assumed to test the therapeutic equivalence between the two regenerative treatments. Otherwise, the study test remains indeterminate.

Criterion 2: Was the Superiority/ Equivalence Margin Specified Quantitatively before the Study Commenced?

In a superiority trial, a pre-specified margin of difference between the two treatments that is deemed large enough to be clinically significant has to be specified *before* the trial commences. In an equivalence trial, the equivalence margin that is deemed sufficiently small by clinicians not to differentiate the treatment effects as clinically significant must also be specified before the trial commences. For the latter, when the confidence interval of the difference in the efficacy between the treatments falls within the pre-specified margin, these two treatments are considered equivalent.

Criterion 3: Was the Appropriate Null Hypothesis Tested?

For a superiority trial, the null hypothesis is that there is no difference between treatments. The trial is designed to reject the null hypothesis, thereby showing that the test group achieves a statistically better outcome. In contrast, equivalence trials test the null hypothesis that any difference that occurs between the treatments is greater than the pre-specified equivalence margin. If the results reject the null hypothesis, where the difference in treatment efficacy falls within the pre-specified margin, the new treatment (such as EMD) is shown not to be inferior to or is ''as good as'' the active-control (such as GTR). It is important to note that the failure to reject the null hypothesis in a superiority trial cannot be used directly as evidence that the two treatments are equivalent (Duke & Garrett 1998). A common error in superiority trials is to assume that the two treatments are equivalent if the results fail to reject the null hypothesis.

Criterion 4: Was the Required Sample Size Calculated?

The required sample size should be calculated according to the appropriate null hypothesis and the pre-specified superiority or equivalence margin. This

Table 1. Summary of the clinical findings from the seven studies comparing enamel matrix derivatives (EMD) and guided tissue regeneration (GTR)

Study	Sample size in EMD group	Sample size in GTR group	Mean PPD (mm) reduction in EMD	Mean PPD (mm) reduction in GTR	Mean CAL (mm) gain in EMD	Mean CAL (mm) gain in GTR
I* (Pontoriero et al. 1999)	10	10	4.4	4.7	2.9	2.9
II (Silvestri et al. 2000)	10	10	4.9	5.7	4.5	4.8
III (Sculean et al. 2001b)	14	14	4.1	4.2	3.4	3.1
IV (Sculean et al. 2001a)	12	12	3.4	3.4	3.0	2.9
V (Zucchelli et al. 2002)	30	30	5.1	6.5	4.2	4.9
VI† (Minabe et al. 2002)	22	23	5.4	4.6	3.0	3.0
VII (Silvestri et al. 2003)	48	48	5.3	5.6	4.1	4.3

*The sample size in the GTR group is the group using non-resorbable membrane as the other two resorbable membranes are no longer available.
†The unit of analysis for sample size was site, even though some patients contributed more than one site. The sample size in the GTR group is the group using membrane only, and there is a third group treated with membrane and EMD simultaneously.

Table 2. Quality assessment of the studies using the standard randomized-controlled trials methodological criteria (Montenegro et al. 2002)

Study	Described as randomized	Randomization methods	Allocation concealment method	Patient blinding	Caregiver blinding	Examiner blinding	All patients accounted for at the end of the study	Analysis accounts for patient loss
I (Pontoriero et al. 1999)	Yes	Unclear	Unclear	Unclear	Not applicable†	Unclear	Yes	Not applicable
II (Silvestri et al. 2000)	Inadequate*	Unclear	Unclear	Unclear	Not applicable	Unclear	Yes	Not applicable
III (Sculean et al. 2001b)	Yes	Unclear	Unclear	Unclear	Not applicable	Yes	Yes	Not applicable
IV (Sculean et al. 2001a)	Yes	Yes	Unclear	Unclear	Not applicable	Yes	No	No
V (Zucchelli et al. 2002)	Yes	Yes	Unclear	Unclear	Not applicable	Yes	Yes	Not applicable
VI (Minabe et al. 2002)	Yes	Unclear	Unclear	Unclear	Not applicable	Unclear	Yes	Not applicable
VII (Silvestri et al. 2003)	Yes	Unclear	Unclear	Unclear	Not applicable	Unclear	No	No

*This study was described as a clinical trial, although random allocation was not explicitly described.
†As applications of guided tissue regeneration and EMD need different procedures, caregiver blinding was considered not applicable.

ensures that the trial will have sufficient power to detect a difference in the treatment efficacy that exceeds the specified margin.

Criterion 5: Was the Active Control Used Shown to be Effective?

This is important because positive results of equivalence trials are usually interpreted as the new treatment being as effective as the standard one. However, if the active control has not been shown to be effective, then positive results of equivalence trials would show nothing more than both treatments being equally ineffective.

Criterion 6: Were Both Treatment Regimens Applied in an Optimal Fashion?

For a treatment to be effective, it needs to be applied in an optimal way that has been established previously, otherwise a false-negative result may occur. For instance, close follow-up and a strict oral hygiene regime have been shown to be key factors for the success of GTR (Cortellini & Tonetti 2000). If inadequate details of the regimen are provided, it is not clear whether the treatments were applied in an optimal fashion.

Criterion 7: Was the Appropriate Statistical Analysis Chosen, and was its Interpretation Correct?

The choice of statistical test is an important part of the research design of the study. Inappropriate analyses and incorrect interpretation of their results may lead to misleading or incorrect conclusions be drawn from the research. The statistical analyses of the studies identified in this article are verified to observe whether the statistical tests adopted are appropriate, and that the statistical test results are interpreted correctly.

Results

Nine studies were identified from the literature review. One did not state whether the study was designed as a clinical trial (Pietruska 2001) and another one was not a randomized-controlled trial (Parashis et al. 2004). Hence, seven studies were included for assessment of their methodological quality. The main outcome variables were probing pocket depth (PPD) and clinical attachment level (CAL). Findings from the seven trials are summarized in Table 1. Studies were listed (and hence labelled) according to their publication date. The quality assessment of the seven clinical trials using the standard criteria is summarized in Table 2 (Montenegro et al. 2002).

Quality Assessment of Additional Methodological Criteria for Active-Control Trials Additional Criteria (See Table 3)

Criterion 1: Was the Research Aim Described as Testing Equivalence or Superiority?

Study IV claimed that it was designed to test superiority, and Study III to test ''difference''. There was no clear state-

Table 3. Quality assessment of the studies using the additional methodological criteria for active-control trials

Study	Described as testing equivalence or superiority	Superiority/ equivalence margin specified quantitatively before the study	Appropriate null hypothesis tested	Required sample size calculated	Active control previously shown to be effective	Both regimens applied in an optimal fashion	Statistical analysis or its interpretation appropriate
I (Pontoriero et al. 1999)	Equivalence*	No	Unclear	No	Yes	Yes	No
II (Silvestri et al. 2000)	Unclear	No	Unclear	No	Yes	Yes	No
III (Sculean et al. 2001b)	Superiority†	Yes	Unclear	Unclear‡	Yes	Yes	Yes
IV (Sculean et al. 2001a)	Superiority	Yes	Unclear	Unclear‡	Yes	Yes	Yes
V (Zucchelli et al. 2002)	Equivalence*	No	Unclear	No	Yes	Yes	No
VI (Minabe et al. 2002)	Unclear	No	Unclear	No	Yes	Yes	No
VII (Silvestri et al. 2003)	Unclear	No	Unclear	No	Yes	Yes	No

*These two studies had no clear statement of research aim, but they claimed in their conclusion that the two treatments were equally effective or satisfactory.

†This study claimed that it was designed to test the difference between the treatments.

‡These two studies stated that their study power was 72% and 70%, respectively, which were lower than the generally accepted 80%. It is unclear whether or not their power was calculated prospectively or retrospectively.

ment in the remaining publications whether the research tested superiority or equivalence. However, two studies claimed in their conclusions that EMD and GTR were ''equally effective'' (Study I) or ''equally satisfactory'' (Study V). These two studies were therefore assumed to test the therapeutic equivalence between the two regenerative treatments. The research aim in the remaining three studies was unclear.

Criterion 2: Was the Superiority/ Equivalence Margin Specified Quantitatively Before the Study?

Studies III and IV pre-specified a margin of 1 mm difference, although no explanation was given as to why. The other five trials did not pre-specify either the equivalence or superiority margin.

Criterion 3: Was the Appropriate Null Hypothesis Tested?

No study gave a clear description of their research hypothesis to be tested. None of the two equivalence trials stated their null hypothesis (or their alternative hypothesis) clearly.

Criterion 4: Was the Required Sample Size Calculated?

None of the trials clearly indicated how the sample size was determined, although trials III and IV stated that the power of their studies was 72% and 70%, respectively. The general accepted criterion for minimum study power is 80% (Dawson & Trapp 2001). It is unclear whether trials III and IV were designed to have a lower power than is standard practice, or whether these were calculated retrospectively.

Criterion 5: Was the Active Control Previously Shown to be Effective?

All seven trials tested the treatment efficacy of EMD and GTR in infrabony lesions, since GTR has previously been shown to be effective in the treatment of infrabony lesions (Needleman et al. 2002). All seven trials satisfied this criterion.

Criterion 6: Were Both Regimens Applied in an Optimal Fashion?

Surgical techniques and the post-treatment regimen for infection control and maintenance care varied from trial to trial. In all seven studies, these were considered to have been applied in a satisfactory manner. Re-examination was performed in most trials at least one year post-operatively; in Study VI re-examination was at six months and one year post-operatively. All seven trials satisfy this criterion.

Criterion 7: Was the Appropriate Statistical Analysis Chosen and was its Interpretation Correct?

In terms of choosing an appropriate statistical method, all seven studies used statistical analyses intended for testing superiority. Negative results (i.e. failure to show the difference) were interpreted by the two equivalence trials as evidence of comparable treatment effects between EMD and GTR. In Study V, statistically significant better outcomes were achieved by GTR compared with EMD, although the authors concluded that the two treatments were almost equally satisfactory. Notwithstanding any confusion over the study design being superiority or equivalence, all the analyses of covariance and multiple regression used by four trials (II, V, VI, VII) suffered statistical methodological problems, such as mathematical coupling and/or collinearity, which render some of their conclusions questionable (Tu et al. 2002, 2004a, b, c).

For instance, although no statistical difference in treatment effects was found, Study II suggested that, based on the regression analyses, GTR seemed to work better than EMD in patients with $\geqslant 9$ mm probing attachment level and vice versa in patients with <9 mm probing attachment level in terms of percentage probing attachment level gain. However, this conclusion is questionable because of mathematical coupling between baseline probing attachment level and percentage change of probing attachment level. Nevertheless, Study V recognized that the lack of significance of the depth of baseline infrabony component in the explanation of the attachment gain is due to baseline pocket depth and the depth of infrabony component being highly correlated. A detailed examination of the problem of mathematical coupling in Study II can be found in our previous study (Tu et al. 2004b).

Discussion

The aim of this study is to show how to apply additional methodological criteria to assess the quality of active-control

clinical trials, and improve the awareness of the distinction between equivalence and superiority trials among periodontal researchers. It should be emphasized that it is not our intention to criticize specific studies or researchers. Our results show that the quality of clinical trials that have compared EMD with GTR should be improved, and it will help clinicians in the selection of treatment modalities when using such information as the source for ''evidence based practice''. Our study also indicates that the quality of RCTs would be potentially over estimated if only standard methodological criteria were used.

The dental literature is not short of discussions around the distinction between superiority and equivalence trials (Duke & Garrett 1998, Gunsolley et al. 1998, Burns & Elswick 2001). Nevertheless, dental research is still plagued by the misconception that if no difference in the treatment effect is found, then treatments are shown to be equivalent. This study clearly shows that some researchers remain unaware that testing therapeutic equivalence needs a completely different approach from testing superiority, with regard to both null hypothesis and sample size calculation (Tu et al. 2005). Failing to reject the null hypothesis in a superiority trial does not mean that the null hypothesis is true; it is plausible that a larger study would always reach a smaller *p*-value, thereby rejecting the null hypothesis given that the difference in treatment effect remains the same (Altman & Bland 1995). The confusion around the differences in objectives between superiority and equivalence trials might be responsible for the design of these trials and their inappropriate interpretation.

It has been suggested that although a trial was designed to test superiority or compare a new treatment with an active-control and a placebo group, the equivalence between two active treatments can be tested post hoc by examining whether the confidence intervals of the differences in the treatments include zero and are less than the equivalence margin accepted generally (Duke & Garrett 1998). For instance, although the two equivalence trials (Studies I and V) and the three trials (Studies II, VI and VII) with unclear objectives adopted statistical methods for testing superiority, testing the differences in pocket reduction and attachment gain between EMD and GTR, equivalence could still be claimed if the confidence intervals of the differences included zero and these differences were also less than the equivalence margin generally accepted by most clinicians. Therefore, when a fixed equivalence margin of $\delta = 0.5$ mm is used, if confidence intervals of the differences in the treatment effects between EMD and GTR include zero, and their upper limit is less than 0.5 mm in favour of GTR, then EMD and GTR could be considered equivalent. A recent meta-analysis on EMD (Esposito et al. 2003) provided confidence intervals of differences in pocket reduction and attachment level for six of the trials in this study (except VI). If $\delta = 0.5$ mm was accepted as an equivalence margin, none of the six trials showed equivalence between EMD and GTR. If $\delta = 1$ mm, the treatment effects of EMD and GTR in Study VII would be considered comparable. As noticed in Table 1, Study VII is the largest trial among the seven, and therefore its confidence intervals are the narrowest. This indicates that, notwithstanding whether the original intention of these clinical trials is to test superiority or equivalence, the available evidence has not yet demonstrated that the two treatments are comparable. To show that the treatment effect of EMD is comparable with GTR, we need a large (sufficiently powered), well-designed equivalence trial with a pre-defined, widely accepted equivalence margin, and analysed using appropriate statistical analyses.

We strongly advocate that the evaluation of statistical analyses, such as criteria 4 and 7, be adopted in the assessment of the quality of all RCTs. The standard methodological criteria mainly evaluate the quality of RCTs regarding their conduct of randomization, blinding and intention to treat. We totally agree that these are very important criteria in the evaluation of the validity of results from RCTs, although we also observe that these criteria alone are insufficient to certify the quality of RCTs. For instance, although Study V seemed to be of better quality than the other six studies using standard methodological criteria (Table 2), it did not report any power calculation and its interpretation of the statistical analyses was questionable, notwithstanding the uncertainty over its design being a superiority or equivalence trial. Another example is that Study VII claimed that GTR yielded better results than EMD in deeper defects. However, Study VII did not report explicitly its results of a multivariable regression analysis, which could provide partial regression coefficients and associated *p*-values; this therefore casts doubt as to the reliability of the conclusions reached by the authors.

Use of set of methodological criteria to assess critically the quality of published trials has been made in both medical (Concato et al. 1993, Moss et al. 2003, Schumm et al. 1999) and dental research (Montenegro et al. 2002). Some focused upon general design and the reporting of clinical trials (Schumm et al. 1999, Montenegro et al. 2002), while others focused upon the adequacy of statistical analyses (Concato et al. 1993, Moss et al. 2003). One recent survey (Schumm et al. 1999) suggested an improvement in reporting clinical trials in general surgery, compared with a previous survey in the 1980s. It is our sincere expectation that our study will also bring about a similar improvement in periodontal research.

This study demonstrates that the differences in design and statistical testing between superiority trials and equivalence trials are still not fully appreciated by dental researchers. Therefore, the additional criteria proposed should be used to design future clinical trials with active-control groups, and to assess the quality of evidence from clinical trials when active controls are used.

Acknowledgements

We would like to thank Dr. Ian Needleman at the Eastman Dental Institute, University College London, London, for his suggestion on how to apply Criterion 1 to studies without clear research aims. All but the second author were funded by the United Kingdom government's Higher Education Funding Council for England (HEFCE).

References

Altman, D. G. & Bland, J. M. (1995) Statistics notes: absence of evidence is not evidence of absence. *British Medical Journal* **311**, 485.

Burns, D. R. & Elswick, R. K. (2001) Equivalence testing with dental clinical trials. *Journal of Dental Research* **80**, 1513–1517.

Concato, J., Feinstein, A. R. & Holford, T. R. (1993) The risk of determining risk with multivariable models. *Annals of Internal Medicine* **118**, 201–210.

Cortellini, P. & Tonetti, M. S. (2000) Focus on intrabony defects: guided tissue regeneration. *Periodontology 2000* **22**, 104–132.

Dawson, B. & Trapp, R. G. (2001) *Basic and Clinical Biostatistics*, 3rd Edition. New York: McGraw-Hill.

Duke, S. P. & Garrett, S. (1998) Equivalence in periodontal trials: a description for the clinician. *Journal of Periodontology* **69**, 650–654.

Esposito, M., Coulthard, P. & Worthington, H. V. (2003) Enamel matrix derivative (Emdogain®) for periodontal tissue regeneration in intrabony defects. *The Cochrane Database of Systematic Review* - 2003, Issue 2.

Fleiss, J. L. (1992) General design issues in efficiency, equivalency and superiority trials. *Journal of Periodontal Research* **27**, 306–313.

Greene, W. L., Concato, J. & Feinstein, A. R. (2000) Claims of equivalence in medical research: are they supported by the evidence? *Annals of Internal Medicine* **132**, 715–722.

Gunsolley, J. C., Elswick, R. K. & Davenport, J. M. (1998) Equivalence and superiority testing in regeneration clinical trials. *Journal of Periodontology* **69**, 521–527.

Kalpitis, C. D. R. & Ruben, M. P. (2002) Treatment of intrabony periodontal defects with enamel matrix derivative: a literature review. *Journal of Periodontology* **73**, 1360–1376.

Koch, G. G. & Paquette, D. W. (1997) Design principles and statistical considerations in periodontal clinical trials. *Annals of Periodontology* **2**, 42–63.

McAlister, F. A. & Sackett, D. L. (2001) Active-control equivalence trials and antihypertensive agents. *American Journal of Medicine* **111**, 553–558.

Minabe, M., Kodama, T., Kogou, T., Takeuchi, K., Fushimi, H., Sugiyama, T. & Mitarai, E. (2002) A comparative study of combined treatment with a collagen membrane and enamel matrix proteins for the regeneration of intraosseous defects. *International Journal of Periodontics and Restorative Dentistry* **22**, 595–605.

Moher, D., Schulz, K. F. & Altman, D. G. (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Medical Research Methodology* **1**, 2.

Montenegro, R., Needleman, I. G., Moles, D. R. & Tonetti, M. S. (2002) Quality of RCTs in periodontology-a systematic review. *Journal of Dental Research* **81**, 866–870.

Moss, M., Wellman, D. A. & Cotsonis, G. A. (2003) An appraisal of multivariable logistic models in the pulmonary and critical care literature. *Chest* **123**, 923–928.

Needleman, I. G., Giedrys-Leepper, E., Tucker, R. J. & Worthington, H. V. (2002) Guided tissue regeneration for periodontal infra-bony defects (Cochrane Review). Cochrane Data Base Systematic Review. *Journal Periodontal Research* **37**, 380–388.

Parashis, A., Andronikaki-Faldami, A. & Tsiklakis, K. (2004) Clinical and radiographic comparison of three regenerative procedures in the treatment of intrabony defects. *International Journal of Periodontics Restorative Dentistry* **24**, 81–90.

Pietruska, M. D. (2001) A comparative study on the use of Bio-Oss® and enamel matrix derivative (Emdogain®) in the treatment of periodontal bone defects. *European Journal of Oral Sciences* **109**, 178–181.

Pontoriero, R., Wennstrom, J. & Lindhe, J. (1999) The use of barrier membranes and enamel matrix proteins in the treatment of angular bone defects. A prospective controlled clinical study. *Journal of Clinical Periodontology* **26**, 833–840.

Rothman, K. J. & Michels, K. B. (1994) The continuing unethical use of placebo controls. *New England Journal Medicine* **331**, 394–398.

Schumm, L. P., Fisher, J. S., Thisted, R. A. & Olak, J. (1999) Clinical trials in general surgical journals: are methods better reported? *Surgery* **125**, 41–45.

Sculean, A., Donos, N., Miliauskaite, A., Arweiler, N. & Brecx, M. (2001a) Treatment of intrabony defects with enamel matrix proteins or bioresorbable membranes. A 4-year follow-up split-mouth study. *Journal of Periodontology* **72**, 1695–1701.

Sculean, A., Windisch, P., Chiantella, G. C., Donos, N., Brecx, M. & Reich, E. (2001b) Treatment of intrabony defects with enamel matrix proteins orguided tissue regeneration. A prospective controlled clinical study. *Journal of Clinical Periodontology* **28**, 397–403.

Silvestri, M., Ricci, G., Rasperini, G., Sartori, S. & Catteneo, V. (2000) Comparison of treatments of infrabony defects with enamel matrix derivative, guided tissue regeneration with a nonresorbable membrane and Widman modified flap. *Journal of Clinical Periodontology* **27**, 603–10.

Silvestri, M., Sartori, S., Rasperini, G., Ricci, G., Rota, C. & Catteneo, V. (2003) Comparison of infrabony defects treated with enamel matrix derivative *versus* guided tissue regeneration with a nonresorbable membrane. *Journal of Clinical Periodontology* **30**, 386–393.

Tu, Y.-K., Gilthorpe, M. S. & Griffiths, G. S. (2002) Is reduction of pocket probing depth correlated with the baseline value or is it ''mathematical coupling''? *Journal of Dental Research* **81**, 722–726.

Tu, Y.-K., Maddick, I. H., Griffiths, G. S. & Gilthorpe, M. S. (2004a) Mathematical coupling still undermines the statistical assessment of clinical research: illustration from the treatment of guided tissue regeneration. *Journal of Dentistry* **32**, 133–142.

Tu, Y.-K., Clerehugh, V. & Gilthorpe, M. S. (2004b) Ratio variables in regression analysis can give rise to spurious results: a lesson from guided tissue regeneration. *Journal of Dentistry* **32**, 143–151.

Tu, Y.-K., Clerehugh, V. & Gilthorpe, M. S. (2004c) Collinearity in linear regression is a serious problem in oral health research. *European Journal of Oral Sciences* **112**, 389–397.

Tu, Y.-K., Blance, A., Clerehugh, V. & Gilthorpe, MS. (2005) Statistical power for analyses of change in randomized controlled trials. *Journal of Dental Research* **84**, 283–287.

Zucchelli, G., Bernardi, F., Montebugnoli, L. & De Sanctis, M. (2002) Enamel matrix proteins and guided tissue regeneration with Titanium-reinforced expanded polytetrafluoroethylene membranes in the treatment of infrabony defects: a comparative controlled clinical trial. *Journal of Periodontology* **73**, 3–12.

Address:
Yu-Kang Tu
Biostatistics Unit
Centre for Epidemiology & Biostatistics
University of Leeds
30/32 Hyde Terrace
Leeds LS2 9LN
UK
E-mail: y.k.tu@leeds.ac.uk

Clinical relevance

Scientific rationale for study: Clinical trials using active-control groups may be a superiority trial or an equivalence trial. The failure to distinguish these two designs might cause confusions in the interpretation of results, and may yield questionable conclusions.

Principal findings: On evaluating seven clinical trials on the comparison between GTR and EMD using additional methodological criteria it was found that most trials did not clearly specify their study design and failed to meet the majority of the criteria.

Practical implication: The additional criteria proposed in this study will improve the quality of clinical trials if periodontal researchers use them to design their studies.