Journal of Clinical Periodontology
Official Journal of The European Federation of Periodontology
Founded by the British, Dutch, French, German, Scandinavian, and Swiss Societies of Periodontology

J Clin Periodontol 2008; 35: 775–782 doi: 10.1111/j.1600-051X.2008.01291.x

Journal of Clinical Periodontology

The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study

Julie Fenwick[1], Ian G. Needleman[2] and David R. Moles[2]

[1]Defence Dental Services, London, UK; [2]Unit of Periodontology, International Centre for Evidence-Based Oral Health, Division of Restorative Dental Sciences, UCL Eastman Dental Institute, London, UK

Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. J Clin Periodontol 2008; 35: 775–782. doi: 10.1111/j.1600-051X.2008.01291.x.

Abstract

Aim: To investigate potential effect of bias on magnitude of outcomes.

Material and Methods: Randomized-controlled trials (RCTs) from the Cochrane Database of Systematic Reviews were searched. Methodological quality of RCTs was assessed in terms of allocation concealment and examiner masking. Meta-regression analyses were used to determine associations between the quality assessments and magnitude of treatment outcomes on probing depth and attachment level.

Results: Thirty-five RCTs were identified from five systematic reviews. Adequate allocation concealment and examiner masking were found in 24% and 64% of trials respectively. There were no statistically significant differences in the magnitude of treatment outcomes comparing adequate *versus* inadequate or unclear allocation concealment, nor comparing adequate and inadequately examiner masked trials. However, a retrospective power calculation indicated 265 RCTs would be needed to demonstrate a statistically significant effect for the impact of bias on CAL as an outcome measure for a 0.5 mm exaggeration of mean difference between test and control.

Conclusions: There is insufficient evidence to support or refute the theory that the bias from improper methods of allocation concealment and examiner masking affect the magnitude of clinical outcomes in periodontology trials. The pilot data provide a baseline for sample size calculations in future research.

Key words: allocation concealment; bias; clinical outcome; examiner masking; periodontal disease

Accepted for publication 6 June 2008

The randomized-controlled trial (RCT) is currently the most important tool available to objectively assess the effect of new treatments, particularly where effect sizes are modest and where outcomes are subjective (Schulz 2000). The explanation of this status is that RCTs (if properly conducted) can always provide unpredictable selection of subjects to experimental groups, preventing investigators from recruiting subjects likely to respond better to one group or other and therefore resulting in a biased experiment. Random allocation also distributes known and unknown confounders between groups which is a further strength of this design. Randomization consists of two components; firstly, generation of a true random sequence and secondly, concealment of the random sequence from investigators selecting individuals for the trial (allocation concealment). Measurement bias is a further potential problem when outcomes are subjective, i.e. can be influenced by the examiner. Masking (blinding) of patients and examiners can eliminate such a bias, although this is difficult to achieve unless an identical placebo is used as a comparison to the test intervention. Whereas masking can sometimes be difficult to achieve, allocation concealment is always possible (Higgins & Green 2008).

Adequate methods of allocation concealment include centralized or remote randomization schemes; randomization schemes controlled by a pharmacy; numbered or coded containers in which capsules from identical-looking, numbered bottles are administered sequentially; on-site computer systems, where allocations are in a locked unreadable

Conflicts of interest and source of funding statement

None known.
External sources of funding: No sources of funding supplied.
Internal sources of funding: J. F. is employed by the Defence Dental Services, UK. I. G. N. and D. R. M. are employed by University College London, UK.

file; and sequentially numbered opaque, sealed envelopes (Higgins & Green 2008). However the use of sequential identical envelopes or containers is potentially open to corruption (Hewitt et al. 2005) unless designed to be opaque, tamper proof and used with an appropriate audit trail (Beller et al. 2002).

Investigations of the medical and dental literature have consistently indicated that reported methods and use of allocation concealment and examiner masking are not optimal and are similar comparing these fields (Antczak et al. 1986, Moher et al. 1999, Montenegro et al. 2002, Sjőgren & Halling 2002, Doig & Simpson 2005, Needleman et al. 2008). Typically, adequate allocation concealment is reported in $<25\%$ of RCT publications.

Over the last decade, it has become clear that allocation concealment and investigator masking can affect the size of the treatment effect. Studies have demonstrated that improper methods of allocation concealment are associated with an exaggeration of the effect size by up to 41% compared with studies employing adequate methods (Schulz et al. 1995, Pildal et al. 2005). Examiner masking has been associated with a lesser impact (17% exaggeration of outcome size) although these studies have included both objective and subjective outcomes, which might lessen the overall impact.

It is clear that the reported methods of RCTs in periodontology are not optimal and therefore, the question arises, what is the effect of improper trial methods on outcome size in these studies. There has been no previous evaluation in dentistry and therefore, the aim of this study was to investigate the impact of allocation concealment and examiner masking on the size of clinical outcomes. The null hypothesis was that allocation concealment and examiner blinding had no effect on the magnitude of clinical outcomes.

Material and Methods

Search method

We searched for RCTs from within systematic reviews published in the Cochrane Database of Systematic Reviews. This strategy was chosen because clarification of unclear trial methodology with the original trial authors is often conducted by Cochrane reviewers and therefore we anticipated more complete information than examining trial publications themselves. Systematic review selection criteria were; published up to January 2007, reviews including RCTs and with probing depth (PD) and/or clinical or probing attachment level (CAL or PAL) as outcome measures.

From the systematic reviews identified, the original included articles were retrieved and scrutinized in full. Only articles that were published in English, and used either PD or CAL/PAL as outcome measures were included. Where an author was listed more than once within a systematic review each article listed was scrutinized to determine if this was due to multiple publishing of the same data. If so, only the article with the most inclusive data available was utilized and the other papers were excluded. Where the same articles were utilized in more than one systematic review, the data for both PD and CAL/PAL were selected randomly by toss of a coin, to prevent bias by replication of a particular study's sample data.

Data abstraction and quality assessment

Data abstraction was conducted by one investigator (J. F.) and included the numbers of test and control subjects, the mean results and measures of precision of the estimates (from the included systematic review tables for each included original article). Where data were absent from the systematic review tables they were abstracted directly from the relevant original article. Assessment of allocation concealment for each article was taken directly from the tables of each systematic review. Where data on concealment were not in the Cochrane review, the original papers were scrutinized by two authors (J. F. and D. R. M.) and assessed in accordance with the Cochrane Handbook (2006) into ''adequate'', ''unclear'' or ''inadequate'' categories. Any disagreement between assessors was resolved by discussion.

''Adequate'' allocation concealment included centralized or pharmacy-controlled randomization; coded identical containers administered serially; on-site computer system combined with allocations kept in a locked unreadable computer file; sequentially numbered, sealed, opaque envelopes and similar schemes ensuring that patient and clinician were unaware of the allocation, along with reassurance that the person who generated the allocation scheme did not administer it (Higgins & Green 2008). ''Inadequate'' allocation concealment included alternation of patients; use of patient data to assign patients to a treatment group, such as the use of case record numbers or dates of birth. Similarly any procedure that was entirely transparent before allocation such as using the day of the week or an open list of random numbers to allocate a patient to a treatment group was considered inadequate. ''Unclear'' allocation concealment included studies that did not report any concealment approach.

Assessment of explicit documentation of examiner masking was taken directly from the original article as explicit ''adequate'' examiner masking or ''inadequate'' examiner masking, where it was not stated. Where examiner masking was deemed not to be possible it was also graded as ''inadequate'' (as the examiner was not masked). Assessment of examiner masking was repeated in 10% of studies to determine intra-examiner reproducibility. This was not conducted for allocation concealment as it was intended that this would be taken from the review authors' assessment.

Data analysis

The data were synthesized using random effects meta-regression analyses using the statistical software package Stata version 8.2, 2005 (STATA CORP LP., College Station, TX, USA).

The dependant variables for the statistical models were the magnitudes of treatment effects for PD or CAL/PAL respectively in millimetres. The independent variables (predictors) were dummy variables to indicate the adequacy of allocation concealment or examiner masking.

Using each outcome measure (PD or CAL) the results were analyzed to compare ''adequate'' *versus* ''unclear'', and ''adequate'' *versus* ''inadequate'' to determine if the status of the allocation concealment was associated with the magnitude of the clinical outcome. Also, comparisons of ''adequate'' *versus* ''unclear or inadequate'' allocation concealment were made to determine the potential influence of the unclear allocation concealment grading, where the unclear assessment could also include those studies where allocation concealment may have been undertaken but was not reported (Higgins & Green 2008). For examiner masking ''ade-

quate'' *versus* ''inadequate'' results from the studies were compared by meta-regression analysis to determine if this was associated with the magnitude of the clinical outcome for each outcome measure.

Power calculations were undertaken following the meta-regression analyses using n-Query Advisor, version 4, (Statistical Solutions, MA, Saugus, USA), to determine whether the sample size of this pilot study was appropriate and to give relevance to the results obtained. Calculations were undertaken for 2, 1 and 0.5 mm mean differences to determine the sample sizes required for the various mean treatment differences between adequately and inadequately concealed studies and adequately and inadequately examiner masked studies for each outcome measure (PD and CAL/PAL). The number of studies that would have been required for the meta-regressions to have 80% power were also calculated for each magnitude of mean difference.

Results

Search

The search of the Cochrane Database of Systematic Reviews revealed five systematic reviews fulfilling the inclusion criteria (Supplementary Appendix S1). These reviews included 50 eligible original articles available for data abstraction (Fig. 1). From the 50 original papers available for data abstraction, 35 were available for assessment of allocation concealment bias. Reasons for exclusion of the original papers were principally due to papers not being published in English and/or not using PAL/CAL or PD as a measurement outcome. Publication dates of the original articles ranged from 1973 to 2005.

A discrepancy arose between the total numbers of papers available for allocation concealment and examiner masking assessment. Scrutiny of the articles revealed that this was due to one paper being unobtainable in English for examiner masking assessment although the data for assessment of allocation concealment assessment was available via the Cochrane systematic review (Tang et al. 2002). This resulted in only 34 original papers being available for assessment of examiner masking (Table 1).

Further scrutiny of the articles revealed that three papers (Pontoriero et al. 1999, Silvestri et al. 2000, Zucchelli et al. 2002) were duplicated in two systematic reviews (Esposito et al. 2005, Needleman et al. 2006), even though the actual data presented in the systematic reviews varied (due to the data being handled differently according to the aim of the systematic review in question). To prevent bias due to replication of specific population data, one set of each duplicated data was randomly chosen (toss of a coin) to be deleted. This resulted in 34 articles available for assessment of allocation concealment bias, with 29 articles utilizing PD measurements and 31 articles utilizing CAL/PAL measurements. For examiner masking, 28 articles utilized PD and 31 CAL/PAL outcome measurements. The results were recalculated and meta-analyses were repeated without

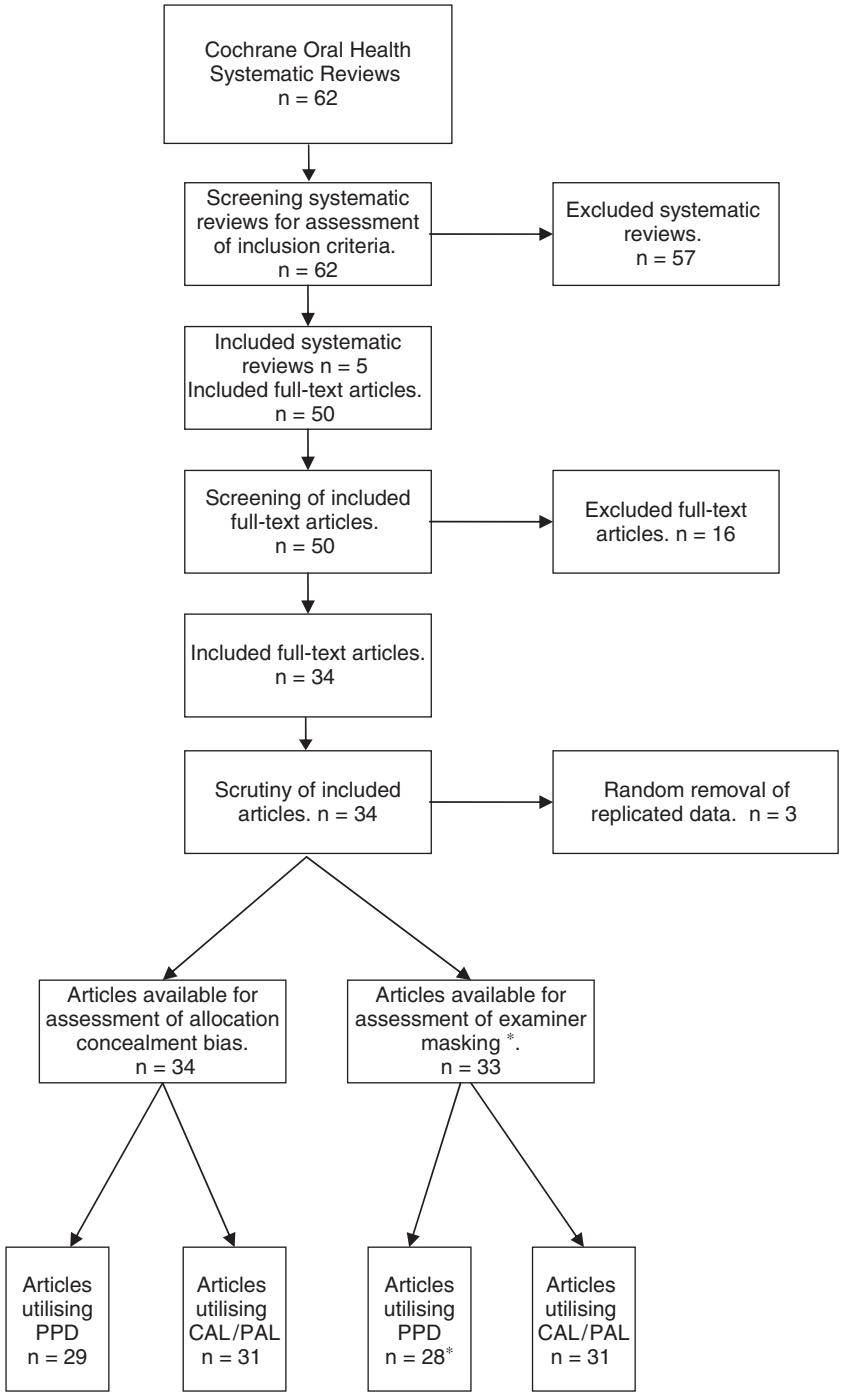

Fig. 1. Flow of articles (number) through study. *One article was not available for examiner masking assessment in English (Tang et al. 2002).

Table 1. Summary of types of interventions in included studies

Study*	Interventions
Blumenthal & Steinberg (1990)	T1: GTR T2: GTR+biomaterial C: Surgery
Büchter et al. (2004)	T: Scaling C: Scaling+antibiotic
Chung et al. (1990)	T: GTR C: Surgery
Ciancio et al. (1995)	T: Mouthwash C: Placebo mouthwash
Cortellini et al. (1995)	T1: GTR (titanium reinforced) T2: GTR (non-resorbable) C: Surgery
Cortellini et al. (1996)	T1: GTR (resorbable) T2: GTR (non-resorbable) C: Surgery
Cortellini et al. (1998)	T: GTR C: Surgery
Cortellini et al. (2001)	T: GTR C: Surgery
Francetti et al. (2004)	T: EMD C: Surgery
Glavind (1977)	T: Monthly professional plaque removal C: Routine care
Heijl et al. (1997)	T: EMD C: Surgery+placebo
Listgarten et al. (1986)	T: Variable frequency maintenance programme C: Three monthly maintenance programme
Loos et al. (2002)	T1: GTR T2: GTR+antibiotics C1: Surgery C2: Surgery+antibiotics
Mayfield et al. (1998)	T: GTR C: Surgery
Mora et al. (1996)	T: GTR C: Surgery
Nyman et al. (1975)	T: Periodontal surgery then maintenance every 2/52 C: Periodontal surgery then maintenance every 6/12
Okuda et al. (2000)	T: EMD C: Surgery+placebo
Pontoriero et al. (1999)	T1: GTR (resorbable) T2: GTR (resorbable) T3: GTR (non-absorbable) T4: EMD C: Surgery
Pritlove-Carson et al. (1995)	T: GTR C: Surgery
Ratka-Kruger et al. (2000)	T: GTR C: Surgery
Rösling et al. (2005)	T: EMD C: Surgery+placebo
Rosling et al. (1976)	T: Periodontal surgery then maintenance every 2/52 C: Periodontal surgery then maintenance once per year
Sanz et al. (2004)	T: EMD C: GTR
Schwarz et al. (2005)	T: Laser C: Scaling
Sculean et al. (2001)	T1: GTR T2: EMD T3: GTR+EMD C: Surgery
Silvestri et al. (2000)	T1: GTR T2: EMD C: Surgery
Silvestri et al. (2003)	T: EMD C: GTR
Strooker et al. (1998)	T: Gel C: Scaling

inclusion of the replicated data and only the de-duplicated data are presented can be found at Tables 2 and 3.

Allocation concealment and examiner masking characteristics

For allocation concealment, the grading could not be derived from the relevant systematic review in 23 articles therefore the original article was assessed in accordance with the Cochrane Handbook guidelines (2006). Of the 34 articles included, eight (23%) were classified as having adequate allocation concealment; 24 (71%) classified as unclear and two (6%) as having inadequate allocation concealment.

Of the 33 articles assessed for examiner masking, 21 (64%) were found to have adequate descriptions of examiner masking while 12 (36%) were assessed as inadequate. The inadequate category included two trials where examiner blinding was assessed as not being possible due to the methodology of the RCT. Intra-examiner reproducibility was perfect ($\kappa = 1.00$) for a 10% sample of replicated assessments of examiner masking.

Frequency of outcome measures (PD and CAL/PAL)

The frequency of the outcome measures (PD and CAL/PAL) in the articles available for data abstraction was very similar, with 48% for PD and 52% for CAL/PAL.

Effect of allocation concealment assessment on PD (Table 2)

The meta-analysis results showed no difference in mean difference between test and control groups in terms of PD (treatment effect size) between studies that were classified as having adequate allocation concealment and those classified as having unclear allocation concealment (mean difference: 0.22 mm, 95% CI: −0.58 mm, 1.03 mm, $p = 0.59$) or for the comparison of adequate *versus* inadequate allocation concealment (mean difference: 0.6 mm, 95% CI: −1.70 mm, 1.89 mm, $p = 0.37$) on the magnitude of clinical outcomes. Similarly, when adequate allocation concealment was compared with either inadequate or unclear allocation concealment there was no difference in magnitude of clinical outcomes was not statistically significant (mean difference:

Table 1. (Contd.)

Study*	Interventions
Suomi et al. (1973)	T1: Maintenance every 4/12 T2: Maintenance every 6/12 C: Maintenance every 12 months
Tang et al. (2002)	T: Antibiotic C: Scaling
Tonetti et al. (1998)	T: GTR C: Surgery
Tonetti et al. (2002)	T: EMD C: Surgery
Wolff et al. (1998)	T: Scaling C: Toothbrushing
Zucchelli et al. (2002)	T1: GTR T2: EMD C: Surgery

*References available in Supplementary Appendix S1.
EMD, surgery with placement of enamel matrix derivative; GTR, guided tissue regeneration surgery; Surgery, non-regenerative surgery.

Table 2. Results of allocation concealment meta-analysis

Effect of allocation concealment on PD	Effect size difference (mm)*	95% CI	SE	*p* value
Adequate *versus* unclear	0.22	− 0.58, 1.03	0.41	0.59
Adequate *versus* inadequate	0.60	− 1.70, 1.89	0.66	0.37
Adequate *versus* (either inadequate or unclear)	0.25	− 0.53, 1.03	0.40	0.53
Effect of allocation concealment on CAL/PAL				
Adequate *versus* unclear	0.05	− 0.95, 1.06	0.51	0.92
Adequate *versus* inadequate	− 0.09	− 2.0, 1.82	0.98	0.93
Adequate *versus* (either inadequate or unclear)	0.10	− 1.01, 1.04	0.53	0.98

*Mean difference in treatment effect size according to study quality classification. Positive mean difference in treatment effect size indicates a tendency for poor quality studies to obtain greater treatment effect sizes. Conversely, a negative mean difference indicates good quality studies obtain greater treatment effect sizes.

Table 3. Results of examiner masking meta-analysis

Effect of examiner masking on PD	Effect size difference (mm)*	95% CI	SE	*p* value
Adequate *versus* inadequate	− 0.20	− 0.76, 0.36	0.29	0.49
Effect of examiner masking on CAL/PAL				
Adequate *versus* inadequate	− 0.19	− 1.05, 0.68	0.44	0.67

*Mean difference in treatment effect size according to study quality classification.

0.25 mm, 95% CI: − 0.53 mm, 1.03 mm, $p = 0.53$).

Effect of allocation concealment assessments on CAL/PAL (Table 2)

Similar results were found when using CAL/PAL as the outcome measure. Again, comparisons of adequate *versus* unclear (mean difference: 0.05 mm, 95% CI: − 0.95, 1.06, $p = 0.92$); adequate *versus* inadequate assessments of reported allocation concealment (mean difference: − 0.09 mm, 95%CI: − 2.0, 1.82, $p = 0.93$) and adequate *versus* either unclear or inadequate allocation concealment (mean difference: 0.1 mm, 95% CI: − 1.01, 1.04, $p = 0.98$) showed no difference in mean difference between test and control groups in terms of CAL/PAL (treatment effect size).

Results of examiner masking assessments meta-analyses (Table 3)

No difference in mean difference between test and control groups in terms of effect size between studies that were classified are having adequate examiner masking and those classified as having inadequate examiner masking were shown for PD as an outcome measure (mean difference: − 0.20 mm, 95% CI − 0.76, 0.36, $p = 0.49$) or for CAL/PAL as an outcome measure (mean difference: 0.44 mm, 95% CI: − 1.05, 0.68, $p = 0.67$).

Power calculations (Table 4)

The retrospective power calculations for a mean difference of 2 mm in favour of either test or control for PPD or CAL/PAL between adequately and inadequately concealed studies showed that the current synthesis would have had 95–99% power to detect such a difference if it existed for allocation concealment and 99% power for examiner masking. However for a mean difference of 1 mm the power dropped to between 45% and 92%, according to the outcome measure used and quality indicator investigated. Similarly, to detect a 0.5 mm mean difference in the effectiveness of the interventions, in terms of PPD or CAL, between adequately and inadequately concealed studies the power of this pilot study dropped to 28–31%. Calculations to determine the number of trials required for the meta-regression to have 80% power show that at least 265 RCTs would be needed for this synthesis when assessing allocation concealment using CAL as an outcome measure for a 0.5 mm mean difference between test and control between adequately and inadequately concealed trials. See Table 4 for illustration.

Discussion

The effect of allocation concealment and examiner masking on the magnitude of treatment effects has not previously been investigated in periodontology, despite previous descriptive studies finding problems with the reporting of these aspects (Antczak et al. 1986, Montenegro et al. 2002, Sjögren & Halling 2002).

Principal findings

The results of this pilot study have shown that in the sample of trials investigated, there is insufficient evidence to support or refute an effect of adequacy of allocation concealment methods or

Table 4. Effect of outcome measure mean difference on study power and size

Treatment effect size	2 mm mean difference between test and control				1 mm mean difference between test and control				0.5 mm mean difference between test and control			
Assessment criterion	allocation concealment		examiner masking		allocation concealment		examiner masking		allocation concealment		examiner masking	
Outcome measure	PD	CAL	PD	CAL	PD	CAL	PD	CAL	PD	CAL	PD	CAL
Effect size	0.93	0.68	1.32	0.82	0.46	0.34	0.66	0.41	0.23	0.17	0.33	0.20
% Observed power	99	95	99	99	67	45	92	59	22	15	39	19
No. of studies included in meta-regression	29	31	28	31	29	31	28	31	29	31	28	31
No. of studies required for 80% power of meta-regression	N/A	N/A	N/A	N/A	39	70	N/A	50	148	265	74	191

N/A, not applicable as the current meta-regression already had over 80% with the existing number of studies for this effect size.

examiner on the magnitude of the clinical outcome, when using PD or CAL. Within this sample, only 24% were assessed as adequate for allocation concealment. Similarly, only 64% of trials reviewed had adequate examiner masking assessments.

It is possible that allocation concealment and examiner masking do not affect clinical outcomes in periodontology because this is not a consistent finding throughout all fields of medicine (Balk et al. 2002, Kunz et al. 2007). However there may be other explanations for our results.

There are many factors apart from adequate allocation concealment and examiner masking may explain variability of treatment effect, such as heterogeneity of study populations, treatment types and study design. Other forms of bias and random error may also affect the magnitude of treatment results. For these reasons, a particular quality measure such as allocation concealment or examiner masking, may not necessarily explain the magnitude of treatment result alone and this might be the sum of many variables within a trial (Balk et al. 2002). A recent meta-epidemiological study which investigated study quality and outcome inflation concluded that type of outcome was an important factor (Wood et al. 2008); exaggeration of outcome size was associated with subjective outcomes, but not objective measures such as mortality. Probing measures, especially when conducted with a manual and therefore non-pressure sensitive probe must be considered subjective as their output can be affected by examination characteristics such as probing pressure, probe angulation and rounding to nearest millimetre calibration mark (Listgarten 1980, Caton 1989, Zappa et al. 1995).

Limitations and strengths of the study

Power of the study

The results of our study may be due to the limited number of studies reviewed which may have prevented the detection of significant differences between the magnitude of study outcomes and quality of the papers i.e. a type II error. To calculate the sample size before the study we required numerical information to provide an estimate of the variation in a numerical variable from other published studies with similar outcomes (Petrie & Sabin 2005). As this was a pilot study in periodontology and because medical studies had used odds ratios rather than continuous data we were not able to calculate a sample size of sufficient power before the study.

Distribution of data

In addition to the limited numbers of studies reviewed, the spread of the data was skewed, with only 24% of studies having adequate allocation concealment. This may have had an impact on the results due to the excessive distribution of potentially biased trials overwhelming the results of the trials of higher quality. Skewing of the data was less apparent for the assessment of bias in examiner masking, where 64% were assessed as adequately masked.

Methodology

For this study we chose to examine ''included'' papers of systematic reviews that were part of the Cochrane Database of Systematic Reviews. The reason for this was that the original papers had already been scrutinized for quality and relevant outcome measures. In addition, data from the original papers had been analyzed to detect treatment effects and were presented in tabular form which we utilized for our study analyses. It is possible that RCTs included in Cochrane reviews were of higher bias protection due to their rigorous methodology and critical appraisal (Glenny et al. 2003). Therefore differences in quality might be less marked. Our study excluded unpublished and non-English language trials due to difficulty in evaluating the methodological quality. This may have exacerbated any publication bias already present from using the systematic reviews as a source of trials.

Analysis of data

Random effects meta-regression analysis models were used in this study rather than a logistic regression approach. This avoided the necessity of converting treatment effects into odds ratios, which is not always possible with published data (Sterne et al. 2002) and has been shown to increase bias (Deeks et al. 2003). In addition, using the logistic regression approach assumes the effects of bias are constant across meta-analyses (Sterne et al. 2002). The lack of an effect of bias on outcome magnitude for the studies we used may have been due to the fact that aspects of study quality are not associated with treatment outcomes in a predictable way (Kunz et al. 2007) or that the study method may have failed to characterise methodological quality adequately.

Clinical heterogeneity

Previous investigations have included a wide variety of interventions and outcomes (Chalmers et al. 1983, Schulz et al. 1995), increasing the risk of

confounding in the analysis. The current study was limited to five systematic reviews in periodontology and focussed on two related subjective probing outcomes, although investigating both periodontitis and peri-implantitis. As a result, confounding is less likely to have affected our data than the studies in the medical literature with a much greater range of outcomes and health conditions.

As discussed, examiner masking is not always possible in many surgical procedures (Deyo et al. 1990) and by assessing those studies where examiner masking was not possible as ''inadequate'' may have skewed our results. However as only two studies in this pilot study were assessed as being not possible to mask, it was judged inappropriate to undertake a meta-analysis to compare the results due to the small sample size.

Current study and totality of the evidence

No previous studies in oral health research have provided data for this topic. Indirect evidence for an effect of bias in periodontology on outcome size was reported in a systematic review (Needleman et al. 2006) which found that when studies without both operator and examiner blinding were excluded from a meta-analyses, the difference between test and control became smaller and non-statistically significant. The small number of included trials suggests caution in drawing conclusions. However, when comparing the authors' meta-analysis of GTR *versus* open flap debridement with previous meta-analyses which included studies at greater risk of bias, the greater the potential for bias in the previous meta-analyses, the greater the apparent benefit of the test treatment (difference in CAL gain was twice that comparing meta-analyses with most *versus* least risk of bias studies included).

Suggestions for future research

We recommend further studies to investigate the potential impact of bias on effect estimates in trials of oral health interventions. The power calculations undertaken in this study indicate a much larger sample of studies will need to be included. This could be achieved from searching standard databases such as MEDLINE, EMBASE, LILACS and Cochrane CENTRAL. However, such a study would need to include a variety of interventions to achieve adequate power. Whether outcomes could be limited to PD and CAL and still achieve sufficient power is not clear.

One barrier to evaluating oral health research remains incomplete trial reporting. The CONSORT statement is an international guideline to facilitate adequate reporting and is available for a range of RCT designs. We strongly urge all trials authors to employ such a standard whether it is a requirement for publication or not (Needleman et al. 2008).

Conclusions

In summary, this pilot study has found insufficient evidence to support or refute an effect of allocation concealment or examiner masking on the magnitude of the treatment effect. Retrospective power calculations suggest that at least 265 trials would be needed to demonstrate an effect if the magnitude of the overestimation of the magnitude of clinical effect was 0.5 mm. Future definitive research on this topic will therefore need to examine a much larger sample of trials.

References

Antczak, A. A., Tang, J. & Chalmers, T. C. (1986) Quality assessment of randomized controlled trials in dental research. II. Results: periodontal research. *Journal of Periodontal Research* **21**, 315–321.

Balk, E. M., Bonis, P. A. L., Moskowitz, H., Schmid, C. H., Ionnidis, J. P. A., Wang, C. & Lau, J. (2002) Correlation of quality measures with estimates of treatment effect in meta-analyses of random controlled trials. *Journal of American Medical Association* **287**, 2973–2982.

Beller, E. M., Gebski, V. & Keech, A. C. (2002) Randomisation in clinical trials. *Medical Journal of Australia* **177**, 565–567.

Caton, J. (1989) Periodontal diagnosis and diagnostic aids. American association of periodontology. *Proceedings of the World Workshop in Clinical Periodontology* 1–22.

Chalmers, T. C., Celano, P., Sacks, H. S. & Smith, H. (1983) Bias in treatment assignment in controlled clinical trials. *New England Journal of Medicine* **309**, 1358–1361.

Deeks, J. J., Dinnes, J., D'Amico, R., Sowden, A. J., Sakarovitch, C., Song, F., Petticrew, M. & Altman, D. G. (2003) International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technology Assessments* **7**, iii–x, 1–173.

Deyo, R. A., Walsh, N. E., Schoenfeld, L. S. & Ramamurthy, S. (1990) Can trials of physical treatments be blinded? The example of transcutaneous electrical nerve stimulation for chronic pain. *American Journal Physical Medical Rehabilitation* **69**, 6–10.

Doig, G. S. & Simpson, F. (2005) Randomisation and allocation concealment: a practical guide for researchers. *Journal of Critical Care* **20**, 187–191.

Esposito, M., Grusovin, M. G., Coultard, P. & Worthington, H. V. (2005) Enamel matrix derivative (Emdogain) for periodontal tissue regeneration in intrabony defects. *Cochrane Database of Systematic Reviews*, Issue 4.

Glenny, A. M., Eposito, M., Coultard, P. & Worthington, H. V. (2003) The assessment of systematic reviews in dentistry. *European Journal of Oral Sciences* **111**, 85–92.

Hewitt, C., Hahn, S., Torgerson, D., Watson, J. & Bland, J. M. (2005) Adequacy and reporting of allocation concealment: review of recent trials published in four general medical journals. *British Medical Journal* **330**, 1057–1058.

Higgins, J. P. T. & Green, S., editors. Cochrane Handbook for Systematic Reviews of Interventions 5.0.0 [updated February 2008]. The Cochrane Collaboration, 2008. www.cochrane-handbook.org [Accessed 24 April 2008].

Kunz, R., Vist, G. & Oxman, A. D. (2007) Randomisation to protect against selection bias in healthcare trials. *Cochrane Database of Systematic Reviews*, Issue 2.

Listgarten, M. A. (1980) Periodontal probing: what does it mean? *Journal of Clinical Periodontology* **7**, 165–176.

Moher, D., Cook, D. J., Jadad, A. R., Tugwell, P., Moher, M. & Jones, A. (1999) Assessing the quality of reports of randomised trials: implications for the conduct of meta-analysis. *Health Technology Assessments* **3**, 1–98.

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

Needleman, I., Worthington, H., Moher, D., Altman, D. & Schulz, K. (2008) Improving the completeness and transparency of reports of randomised trials in oral health: the CONSORT Statement. *American Journal of Dentistry* **21**, 7–12.

Needleman, I. G., Worthington, H. V., Giedrys-Leeper, E. & Tucker, R. J. (2006) Guided tissue regeneration for periodontal infra-bony defects. *Cochrane Database of Systematic Reviews*, Issue 2.

Petrie, A. & Sabin, C. (2005) Sample size calculations. *Medical Statistics at a Glance*, 2nd edition, pp. 96–97. Oxford: Blackwell Publishing Ltd.

Pildal, J., Chan, A. W., Hrobjartsson, A., Forfang, E., Altman, D. G. & Gotzsche, P. C. (2005) Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *British Medical Journal* **330**, 1049 epub.

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

Schulz, K. F. (2000) Assessing allocation concealment and blinding in randomised controlled trials: why bother? *Evidence-based medicine* **5**, 36–38.

Schulz, K. F., Chalmers, I., Hayes, R. J. & Altman, D. G. (1995) Empirical evidence of bias. Dimensions of methodology quality associated with estimates of treatment effects in controlled trials. *Journal of American Medical Association* **273**, 408–412.

Silvestri, M., Ricci, G., Rasperini, G., Sartori, S. & Cattaneo, V. (2000) Comparison of treatment of intrabony defects with enamel matrix derivative, guided tissue regeneration with a nonresorbable membrane and modified Widman flap. A pilot study. *Journal of Clinical Periodontology* **27**, 603–610.

Sjőgren, P. & Halling, A. (2002) Medline search validity for randomised controlled trials in different areas of dental research. *British Dental Journal* **192**, 97–99.

Sterne, J., Juni, P., Schulz, K., Altman, D., Bartlett, C. & Egger, M. (2002) Statistical methods for assessing the influence of study characteristics on treatment effects in ''meta-epidemiological'' research. *Statistics in Medicine* **21**, 1513–1524.

Tang, Z., Cao, C., Sha, Y., Lin, Y. & Wang, X. (2002) Effects of non-surgical treatment modalities on peri-implantitis. *Chinese Journal of Stomatology* **37**, 173–175.

Wood, L., Egger, M., Gluud, C., Schulz, K., Juni, P., Altman, D. G., Gluud, C., Martin, R. M., Wood, A. J. G. & Sterne, J. A. C. (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* **336**, 601–605.

Zappa, U., Simona, C., Graf, H., Case, D. & Thomas, J. (1995) Reliability of single and double probing attachment level measurements. *Journal of Clinical Periodontology* **22**, 764–771.

Zucchelli, G., Bernardi, F., Montebugnoli, L. & De, 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:
Ian Needleman
Unit of Periodontology
International Centre for Evidence-Based Oral Health
Division of Restorative Dental Sciences
UCL Eastman Dental Institute
256 Gray's Inn Road
London WC1X 8LD
UK
E-mail: i.needleman@eastman.ucl.ac.uk

Clinical Relevance

Scientific rationale for the study: Inadequate protection from bias in trials in medicine has been shown to lead to inflated estimates of treatment effect of the magnitude of up to 41%. This pilot study attempted to determine if RCTs in periodontology are similarly affected.

Principal findings: Insufficient evidence was found to support or refute an *effect of inadequate protection from bias* on the magnitude of the treatment effect in periodontology RCTs within the Cochrane Database of Systematic Reviews.

Practical implications: A definitive assessment of RCT quality and magnitude of treatment outcome is required with from a study with greater power.

Supplementary Material

The following material is available for this article online:

Appendix S1. List of included systematic reviews from Cochrane Database of Systematic Reviews. Included randomized controlled trials from included systematic reviews.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1600-051X.2008.01291.x (This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.