J Clin Periodontol 2007; 34: 756–761 doi: 10.1111/j.1600-051X.2007.01118.x

Journal of Clinical Periodontology

Review Article

Reported methodological quality of split-mouth studies

Emmanuel Lesaffre[1], Maria-José Garcia Zattera[1], Carol Redmond[2], Heidi Huber[3] and Ian Needleman[4] on behalf of the ISCB Subcommittee on Dentistry

[1]Biostatistical Centre, Catholic University Leuven, Leuven, Belgium; [2]Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA; [3]School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; [4]International Centre for Evidence-Based Oral Health, UCL Eastman Dental Institute, London, UK

Lesaffre E, Garcia Zattera M-J, Redmond C, Huber H, Needleman I. Reported methodological quality of split-mouth studies. J Clin Periodontol 2007; 34: 756–761. doi: 10.1111/j.1600-051X.2007.01118.x.

Abstract

Background/Aim: Hujoel & Moulton previously questioned the reported quality of split-mouth studies. Since then, there has been little enquiry into the methodology of this study design. The aim was to conduct a systematic review of the reported methodology of clinical studies using a split-mouth design published in dental journals over a 1-year period (2004).

Material and Methods: An extension of the CONSORT guidelines for cluster-randomized designs was used to evaluate quality. We evaluated the methods used and quality of reporting split-mouth studies.

Results: Thirty-four studies were eligible for this review. The results showed that many papers lack essential qualities of good reporting, e.g. five of 34 papers gave the rationale for choosing a split-mouth design, 19 of 34 (56%) used appropriate analytical statistical methods and only one of 34 presented an appropriate sample size calculation. Of the five studies that used survival analysis, none of them used a paired approach.

Conclusions: Despite some progress in statistical analysis, if the reporting of studies represents the actual methodology of the trial, this review has identified important aspects of split-mouth study design and analysis that would benefit from development.

Key words: CONSORT guidelines; quality reporting; split mouth; statistics

Accepted for publication 18 June 2007

In a split-mouth design, divisions of the mouth (dental arches, quadrants, sextants or smaller subdivisions) constitute the experimental units randomly assigned to treatments. Because the patient serves as his/her own control, which can increase statistical efficiency, on average, fewer patients are needed. However, Hujoel and colleagues (Hujoel & Loesche 1990, Hujoel & DeRouen 1992, Hujoel 1998), who evaluated the design, analysis and reporting of periodontal split-mouth studies, concluded that, despite the potential gain in efficiency over a whole-mouth design, one should carefully consider whether the split-mouth design is to be preferred in practice.

In 1988, Hujoel and Moulton reported that only five of 22 periodontal split-mouth studies used an appropriate statistical analysis. Either the statistical method was not reported or the authors used an unpaired instead of a paired test.

Our aim was to evaluate the methodology of split-mouth studies in periodontology after nearly two decades since the last assessment and to extend this across a wide range of clinical research areas in oral health using a preliminary version of the extended CONSORT guidelines for cluster-randomized trials. We further assessed the statistical quality of the papers according to considerations in Hujoel & Loesche (1990), Hujoel & DeRouen (1992) and Hujoel (1998).

Conflict of interest and source of funding statement

The authors declare that they have no conflict of interests.

The second author is supported by the Research Grant OT/00/35 and OT/05/60, Catholic University Leuven. The first two authors also acknowledge the partial support from the Interuniversity Attraction Poles Program P5/24, Belgian State, Federal Office for Scientific, Technical and Cultural Affairs.

Material and Methods

Web-based selection of split-mouth studies

We initiated our review in 2005, using PubMed to identify papers published in 2004 that contained the keyword ''split mouth''. We included studies in all areas of dentistry, but considered only papers written in English, which excluded one non-English-language paper.

Evaluation criteria

Table 1 presents a subset of the preliminary version of the extended CONSORT guidelines for cluster-randomized clinical trials (Campbell et al. 2004) selected for their statistical nature. This version of CONSORT was used due to the similarities in design and analysis between split-mouth and cluster-randomized trials. Utilizing these guidelines, we scored the papers for appropriate statistical methodology and reporting,

© 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

with a slight change made in guideline 6. Each of the initially selected 88 papers was randomly allocated to two among four scorers (first four authors). Each paper was scored on all criteria with comments on whether the paper satisfied the requirement specified in the guidelines. In a second round, the scores and comments of the raters were compared and the papers were re-scored for specific criteria to reconcile discrepancies. The frequency of papers that satisfied each guideline to an acceptable degree was recorded. The guideline was not followed to an acceptable degree when either the procedure followed did not take into account the split-mouth design or insufficient details were given to make a clear judgement. For instance, a paper did not comply with guideline 2 when the rationale for a paired design was not given. Guideline 7 was not followed when there was no sample size calculation or no explicit statement that the sample size calculation took into account the split-mouth design. We also recorded references to the papers of Hujoel, as a surrogate for awareness of possible pitfalls of split-mouth studies. Finally, we reviewed (a) descriptive statistics reported, (b) statistical procedures and (c) how the main statistical results were summarized.

Results

From the 88 papers selected by PubMed, one paper could not be retrieved and 53 were not split-mouth studies. Six of these 53 papers were review papers or an abstract without adequate details for evaluation. In one paper claiming a split-mouth design, data for one patient, who received treatment allocated by non-random methods, could not be isolated in the results. Therefore, we reviewed comprehensively 34 of the 88 papers identified initially. In Table 2, the selected papers are listed.

We classified the dental research domain (number of papers) as follows: periodontology (11), orthodontics (10), cariology (6) and others (7). The leading author was affiliated in Europe in 27 papers; for United States and Canada, there were four papers and three papers originated from Asia. Six papers acknowledged financial support from a scientific grant, 10 papers reported industry support and three papers reported support from both industry and a scientific grant. There was no mention of a funding source in 15 papers.

The split-mouth designs (number of papers) included: full or half contra-lateral (19), diagonal quadrants (8), all quadrants (3), maxillary *versus* mandible (1) and unspecified (3). Thus, the contra-lateral is the most popular split-mouth design. We classified the papers according to primary response variables (with some overlap when the choice of the primary endpoint was unclear) as continuous (23), ordinal (4), binary (2), survival (''time-to-event'') (6) or count (1) outcomes.

When scoring papers according to the extended CONSORT guidelines, some discrepancies occurred between the reviewers. Overall, 28 papers needed a resolution and obtained a consensus score. In 14 papers, there was disagreement on one or more guidelines between the dentist and the statisticians. An equal number of disagreements were seen among the statisticians. In Table 1, the last column reports on the guidelines that were differently scored at the first reading. Overall, there was no clear pattern in the different scoring behaviour, except for guideline 6 (selection of primary and secondary endpoints). For this guideline, the statisticians emphasized whether the choice of the primary endpoint was clear, whereas the oral health researcher checked whether the endpoints were clearly defined from a dental viewpoint. Few papers (Table 1) gave the rationale for choosing a split-mouth design (guideline 2). Most often, the paper stated that this design had been chosen without further clarification. Table 1 also indicates that many papers did not clearly specify the primary endpoint (guideline 6). Four papers contained a sample size calculation, but only one paper took into account the correlation among the responses (guideline 7). Less than half of the papers gave details of randomization (guideline 8), concealment of

Table 1. Selection* of the extended CONSORT guidelines (preliminary version) to clustered data

	Aspect	Guideline	*F*	*D*
2	Introduction and background	Reason; rationale for paired design	5	2
6	Outcomes	Clearly defined primary and secondary outcome measures	10	6
7	Sample size	Sample size determination taking into account the correlation between units (range of possible correlations) and treatment crossover effect (range of possible treatment crossover effects)?	1	4
8	Randomization: sequence generation	Methods used to allocate units within a single individual (e.g. how was the first unit to be randomized decided)	14	4
9	Randomization: allocation concealment	Method used to implement the random allocation sequence (e.g. numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned	9	3
11	Blinding (masking)	Whether or not participants, those administering the interventions and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated	15	7
12	Statistical methods	Use of statistical methods appropriate for a paired design	19	7
13	Participant flow	Number of patients and number of units at each stage	7	3
15	Baseline data	Baseline demographic and clinical characteristics of patients (one column); baseline characteristics of units (two columns)	5	4
16	Numbers analysed	Number of randomized units in each group included in each analysis	15	12
17	Outcomes and estimation	Observed correlation in outcomes between sites within individuals for primary outcome and important secondary outcomes	1	3
18	Ancillary analyses	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory	3	8
20	Interpretation	Potential correlation and treatment crossover effect noted	5	6

The column ''*F*'' reports the frequency of papers (of 34) that satisfied the guideline to a sufficient extent. The column ''*D*'' reports the number of papers that were scored differently by the four scorers.

*The guidelines were selected because of their statistical nature.

© 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

Table 2. Web-selected papers published in 2004 with the key word ''split mouth''

1. Foley, J., Evans, D. & Blackwell, A. (2004) Partial caries removal and cariostatic materials in carious primary molar teeth: a randomised controlled clinical trial. *British Dental Journal* **197**, 697–701; discussion 689.
2. Skold-Larsson, K., Fornell, A. C., Lussi, A. & Twetman S. (2004) Effect of topical applications of a chlorhexidine/thymol-containing varnish on fissure caries assessed by laser fluorescence. *Acta Odontologica Scandinavica* **62**, 339–342.
3. Zerbo, I. R., Zijderveld, S. A., de Boer, A., Bronckers, A. L., de Lange, G., ten Bruggenkate, C. M. & Burger E. H. (2004) Histomorphometry of human sinus floor augmentation using a porous beta-tricalcium phosphate: a prospective study. *Clinical and Oral Implants Research* **15**, 724–732.
4. Graveland, M. P., Rosema, N. A., Timmerman, M. F. & Van der Weijden, G. A. (2004) The plaque-removing efficacy of a finger brush (I-Brush). *Journal of Clinical Periodontology* **31**, 1084–1087.
5. Hanna, R., Trejo, P. M. & Weltman, R. L. (2004) Treatment of intrabony defects with bovine-derived xenograft alone and in combination with platelet-rich plasma: a randomized clinical trial. *Journal of Periodontology* **75**, 1668–1677.
6. Palm, A. M., Kirkegaard, U. & Poulsen S. (2004) The wand versus traditional injection for mandibular nerve block in children and adolescents: perceived pain and time of onset. *Pediatric Dentistry* **26**, 481–484.
7. Elaut, J. & Wehrbein, H. (2004) The effects of argon laser curing of a resin adhesive on bracket retention and enamel decalcification: a prospective clinical trial. *European Journal of Orthodontics* **26**, 553–560.
8. Geitel, B., Kwiatkowski, R., Zimmer, S., Barthel, C. R., Roulet, J. F. & Jahn K. R. (2004) Clinically controlled study on the quality of class III, IV and V composite restorations after two years. *Journal of Adhesive Dentistry* **6**, 247–253.
9. Benson, P. E., Douglas, C. W. & Martin, M. V. (2004) Fluoridated elastomers: effect on the microbiology of plaque. *American Journal of Orthodontics and Dentofacial Orthopedics* **126**, 325–330.
10. Kotsanos, N. & Dionysopoulos, P. (2004) Lack of effect of fluoride releasing resin modified glass ionomer restorations on the contacting surface of adjacent primary molars. A clinical prospective study. *European Journal of Paediatric Dentistry* **5**, 136–142.
11. Summers, A., Kao, E., Gilmore, J., Gunel, E. & Ngan P. (2004) Comparison of bond strength between a conventional resin adhesive and a resin-modified glass ionomer adhesive: an in vitro and in vivo study. *American Journal of Orthodontics and Dentofacial Orthopedics* **126**, 200–206; quiz 254–5.
12. Polimeni, G., Koo, K. T., Qahash, M., Xiropaidis, A. V., Albandar, J. M. & Wikesjo U. M. (2004) Prognostic factors for alveolar regeneration: effect of a space-providing biomaterial on guided tissue regeneration. *Journal of Clinical Periodontology* **31**, 725–729.
13. Cacciafesta, V., Sfondrini, M. F. & Scribante A. (2004) Plasma arc versus halogen light-curing of adhesive-precoated orthodontic brackets: a 12-month clinical study of bond failures. *American Journal of Orthodontics and Dentofacial Orthopedics* **126**, 194–199.
14. Cacciafesta, V., Sfondrini, M. F., Melsen, B. & Scribante, A. (2004) A 12 month clinical study of bond failures of recycled versus new stainless steel orthodontic brackets. *European Journal of Orthodontics* **26**, 449–454.
15. Xu, Y., Hofling, K., Fimmers, R., Frentzen, M. & Jervoe-Storm P. M. (2004) Clinical and microbiological effects of topical subgingival application of hyaluronic acid gel adjunctive to scaling and root planing in the treatment of chronic periodontitis. *Journal of Periodontology* **75**, 1114–1118.
16. Jokstad, A. (2004) A split-mouth randomized clinical trial of single crowns retained with resin-modified glass-ionomer and zinc phosphate luting cements. *International Journal of Prosthodontics* **17**, 411–416.
17. Deinzer, R., Waschul, B. & Herforth, A. (2004) Effects of experimental gingivitis on crevicular PGE2 in a split mouth trial. *Journal of Clinical Periodontology* **31**, 501–505.
18. Sculean, A., Schwarz, F., Berakdar, M., Romanos, G. E., Arweiler, N. B. & Becker, J. (2004) Periodontal treatment with an Er:YAG laser compared to ultrasonic instrumentation: a pilot study. *Journal of Periodontology* **75**, 966–973.
19. Hodge, T. M., Dhopatkar, A. A., Rock, W. P. & Spary, D. J. (2004) A randomized clinical trial comparing the accuracy of direct versus indirect bracket placement. *Journal of Orthodontics* **31**, 132–137.
20. Lampa, E., Brechter, A., van Dijken, J. W. (2004) Effect of a nonrinse conditioner on the durability of a polyacid-modified resin composite fissure sealant. *Journal of Dentistry for Children* (Chicago) **71**, 152–157.
21. Kavvadia, K., Kakaboura, A., Vanderas, A. P. & Papagiannoulis, L. (2004) Clinical evaluation of a compomer and an amalgam primary teeth class II restorations: a 2-year comparative study. *Pediatric Dentistry* **26**, 245–250.
22. Groenendijk, E., Dominicus, J. J., Moorer, W. R., Aartman, I. H. & van Waas, M. A. (2004) Microbiological and clinical effects of chlorhexidine enclosed in fixtures of 3I-Titamed implants. *Clinical and Oral Implants Research* **15**, 174–179.
23. Vandana, K. L., Shah, K. & Prakash, S. (2004) Clinical and radiographic evaluation of Emdogain as a regenerative material in the treatment of interproximal vertical defects in chronic and aggressive periodontitis patients. *International Journal of Periodontics and Restorative Dentistry* **24**, 185–191.
24. Kim, T. S., Klimpel, H., Fiehn, W. & Eickholz P. (2004) Comparison of the pharmacokinetic profiles of two locally administered doxycycline gels in crevicular fluid and saliva. *Journal of Clinical Periodontology* **31**, 286–292. Erratum in: *Journal of Clinical Periodontology* (2004) **31**, 412.
25. Trombelli, L., Tatakis, D. N., Scapoli, C., Bottega, S., Orlandini, E. & Tosi, M. (2004) Modulation of clinical expression of plaque-induced gingivitis. II. Identification of ''high-responder'' and ''low-responder'' subjects. *Journal of Clinical Periodontology* **31**, 239–252.
26. Duran, I. & Sengun, A. (2004) The long-term effectiveness of five current desensitizing products on cervical dentine sensitivity. *Journal of Oral Rehabilitation* **31**, 351–356.
27. Sfondrini, M. F., Cacciafesta, V., Scribante, A. & Klersy, C. (2004) Plasma arc versus halogen light curing of orthodontic brackets: a 12-month clinical study of bond failures. *American Journal of Orthodontics and Dentofacial Orthopedics* **125**, 342–347.
28. Obeid, P. R., D'Hoore, W. & Bercy, P. (2004) Comparative clinical responses related to the use of various periodontal instrumentation. *Journal of Clinical Periodontology* **31**, 193–199.
29. Aimetti, M., Romano, F., Torta, I., Cirillo, D., Caposio, P. & Romagnoli, R. (2004) Debridement and local application of tetracycline-loaded fibres in the management of persistent periodontitis: results after 12 months. *Journal of Clinical Periodontology* **31**, 166–172.
30. Gillam, D. G., Newman, H. N., Davies, E. H., Bulman, J. S., Troullos, E. S. & Curro, F. A. (2004) Clinical evaluation of ferric oxalate in relieving dentine hypersensitivity. *Journal of Oral Rehabilitation* **31**, 245–250. Erratum in: *Journal of Oral Rehabilitation* (2004) **31**, 827.
31. Benson, P. E., Shah, A. A. & Campbell, I. F. (2004) Fluoridated elastomers: effect on disclosed plaque. *Journal of Orthodontics* **31**, 41–46; discussion 16.
32. Kalia, S., Melsen, B. & Verna, C. (2004) Tissue reaction to orthodontic tooth movement in acute and chronic corticosteroid treatment. *Orthodontics and Craniofacial Research* **7**, 26–34.
33. Dalstra, M. & Melsen, B. Does the transition temperature of Cu-NiTi archwires affect the amount of tooth movement during alignment? *Orthodontics and Craniofacial Research* **7**, 21–25.
34. Touyz, L. Z., Lamontagne, P. & Smith, B. E. (2004) Pain and anxiety reduction using a manual stimulation distraction device when administering local analgesia oro-dental injections: a multi-center clinical investigation. *Journal of Clinical Dentistry* **15**, 88–92.

© 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

the randomization procedure (guideline 9) or blinding of the participants of the study (guideline 10). Roughly, one-half of the papers reported a statistical approach appropriate for paired designs (guideline 12). Seven papers provided a flow diagram (guideline 13). Most papers provided baseline information only on age and gender (guideline 15). Less than half of the papers reported the number of patients and units (guideline 16). Only one paper reported on the correlation among the primary response variables within a subject (guideline 17). Five papers indicated the possibility of a carry-across effect (guideline 20). Finally, as in many dental and medical papers, no mention was made of multiple testing (guideline 18) and hence no correction was envisaged, although the appropriateness of a correction mechanism depends on the type of study (confirmatory or exploratory).

In the next stage, we examined the statistical procedures applied in the papers. Of 14 papers that specified the statistical software package, SPSS was used in eight, Statistica in three and SAS in two papers. Because split-mouth studies generate paired responses, both the descriptive statistics as well as the statistical tests should reflect this paired nature. Regarding descriptive statistics, most papers reported the mean and standard deviation irrespective of whether a parametric or non-parametric statistic was used. Only a few reported medians, which may be more appropriate. Discrete responses were generally reported as percentages.

For a continuous response, 17 papers accounted for the paired nature of the data: five papers used a paired *t*-test, five papers a repeated-measures ANOVA approach (including mixed-effects ANOVA), five papers a Wilcoxon's signed-rank test and two papers a Friedman test. However, two papers utilized a one-way ANOVA approach, one paper an ANCOVA approach and two papers used the Mann–Whitney test. One paper did not mention the test used. For binary outcomes, one paper used the McNemar's test, which takes into account the paired data, but one paper used the χ^2 test. For ordinal responses, one paper used an extension of McNemar's test, two used the Wilcoxon's signed-rank test and one the Mann–Whitney test. For survival outcomes, none of the five papers took into account pairing. Instead, they used Kaplan–Meier plots, the log-rank test or the Wilcoxon–Gehan test. Further, three papers reported only descriptive statistics without mentioning *p*-values. Only two papers reported confidence intervals. Thirteen papers interpreted non-significant results as evidence of no treatment effect, a classical misconception in medical papers. In sum, 19 (56%) papers used appropriate statistical methods. Subdivided into the four different clinical areas introduced above, appropriate statistical methodology was used in the following clinical areas: (a) periodontology: seven of 11 papers, (b) orthodontics: five of 10 papers, (c) cariology: three of six papers and (d) other areas: four of seven papers.

Discussion

Our investigation indicated that many split-mouth papers showed deficiencies in reporting and in the application of correct statistical procedures. This problem is not unique to split-mouth designs, but has been noted in many dental and medical papers (e.g. Hujoel & DeRouen 1995).

Since Ramfjord et al. (1968) introduced the split-mouth design, it has become increasingly popular in oral health research. The split-mouth design is a dental version of an agricultural split-plot design where the geographical plots are replaced by regions in the mouth. The split-mouth design is also related to a 2×2 crossover study where patients are, in the case of two treatments, randomly allocated to treatment sequences A–B and B–A. On the other hand, in the case of two sites and two treatments, patients in a split-mouth design are randomly allocated to the treatment combinations A–B and B–A, where in the first (second) combination treatment A (B) is given to the first site. Thus, there are similarities between the split-mouth and the crossover design but they also differ fundamentally, as will be seen below.

A split-plot design evaluates treatments within plots (subjects) controlling for a plot (subject) effect. The larger the plot (subject) effect, the more the responses in a plot (subject) will be correlated and the larger the gain in efficiency that can be expected by using this design in comparison with a between-plot (subject) design. Given the rationale for its use, Hujoel & Moulton's (1988) finding that the correlation of the responses within a mouth was ignored in reports of many split-mouth studies is surprising.

While the gain in efficiency of a split-mouth design is proportional to the within-subject correlation, Hujoel and colleagues warned that heterogeneity in the disease of interest within the mouth lowers this correlation and often implies only a modest to moderate gain in efficiency. Efficiency can only be increased by averaging the treatment effect over many sites per individual.

Identifying sub-plots in a split-mouth design may be more difficult than in an agricultural split-plot design. To apply a split-mouth design, more than one site in the mouth must be affected by the disease. Thus, only patients with multiple affected sites are eligible, which often leads to recruitment problems.

When there is a leakage of the treatment effect from one site to another site, called a carry-across effect, the split-mouth design is seriously handicapped to provide an unbiased estimate of the treatment effect. In an agriculture split-plot design, leakage can usually be overcome by creating physical barriers between the sub-plots. In a crossover design, this leakage is called the carry-over effect and arises because the effect of the treatment administered first has not completely worn out in the second period. Also, here the leakage is unidirectional, i.e. it occurs only from the first to the second period. By increasing the duration between the first and second period in a crossover design, i.e. the washout period, the carry-over effect can be minimized or eliminated. However, eliminating or controlling a carry-across effect in a split-mouth study is more complicated, if not impossible, because physical barriers cannot be implemented in the mouth. Further, the carry-across effect is typically bidirectional, i.e. the treatment administered at site 1 can affect the measurements made at site 2, but equally so the treatment administered at site 2 can affect measurements made at site 1. Finally, while the carry-over effect in a crossover design can be statistically estimated and tested (although with a relatively large uncertainty), it is almost impossible to estimate/test the carry-across effect in a split-mouth study because of the bidirectional effect of the leakage. As a result, in practice, one almost always has to assume away

© 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

the carry-across effects in split-mouth studies.

For these reasons, Hujoel and colleagues were sceptical that the split-mouth design in periodontal diseases will often realize its expectations. Moreover, they warned that the statistical analysis of split-mouth designs is, in general, more complicated than the analysis of a classical whole-mouth study. Hence, an individual with statistical expertise relevant for split-mouth designs is generally needed.

Our review comprised a wide range of split-mouth studies in dentistry. While excluding only one non-English-language publication suggests a comprehensive search, possibly trials reported in other databases, such as Embase, have different methodological characteristics. Furthermore, limiting the search to the keyword might have overlooked some papers not coded as such. However, for the purposes of this review, the search has identified studies that are informative.

The quality of reporting is questionable in many papers. In addition to the classical statistical errors generally made in medical papers, there were specific errors for split-mouth studies. Our examination revealed that, although a paired test was often used for continuous measurements, for discrete measurements the pairing of the data was frequently ignored. For survival outcomes, none of the papers used a valid approach.

Given these specific analytic deficiencies, we reflected on what is an appropriate analysis for a split-mouth study. In the absence of a carry-across effect in a simple design involving two sites and two treatments, the procedures developed for a cross-over analysis are available (see e.g. Jones & Kenward 2003). This involves replacing the ''period'' effect by a ''site'' effect. It is noteworthy that none of the papers in our review tested for a site effect. When there is a balanced design with more than two sites and/or more than two treatments, crossover procedures can again be borrowed.

For ''time-to-event'' or survival outcomes, survival methods, such as Kaplan–Meier life tables, log-rank tests or Cox regression, are not appropriate because these methods do not take into account the fact that the survival times from a same individual, as in a split-mouth study, are correlated. A possible way to deal with this problem is to use survival techniques that treat the patient as a stratum, i.e. compare the survival times within an individual and then summarize the results over the patients, see e.g. Wei (1980). Another possibility is to use frailty models (Therneau & Grambsch 2000). These models are extensions of the survival models for independent survival times by including a subject-specific term, called the frailty, and assuming that this frailty has a distribution over the patients. Software for frailty models is available in the R software system (function *coxph*, R development Core Team 2005) and as SAS-macro's (SAS 2001) written by Klein and Shu downloadable from http://www.biostat.mcw.edu/software/SoftMenu.html. Yet another possibility is to use approaches for multilevel modelling, because the mouth with its teeth and surfaces has a hierarchical structure (see e.g. Goldstein 2003). Analysis of multilevel data can be performed using the MLWIN software; see e.g. Yang & Goldstein (2003).

Unfortunately, when a carry-across effect is present, none of the procedures above is, in principle, valid. Moreover, there is no (easy) or reliable way to detect the presence of a carry-across effect. Thus, in split-mouth studies the situation is more problematic than in a classical crossover study.

The findings in our review were similar to previous investigations of the reported methodological quality of randomized clinical trials (Esposito et al. 2001, Montenegro et al. 2002). Failure to report key aspects of the trial design, such as the allocation procedure, hinders appraisal of the internal validity and therefore the degree of protection from bias of such studies.

We believe that part of the problem is insufficient collaboration between oral health researchers and statisticians. A statistician appeared as a co-author on only three papers and was acknowledged in two additional papers. Further, none of the papers cited the seminal work of Hujoel. Unfortunately, the refereeing process is not correcting these statistical and methodological flaws.

Our review evaluated only the quality of reported methodology, because we did not contact the authors to clarify incomplete or unclear information. Although it is uncertain whether reported quality mirrors actual study conduct, it is noteworthy that studies with unclear methodology have been shown to produce biased estimates of treatment effects (Schulz et al. 1995). Thus, incomplete reporting is an important concern. Adherence to guidelines, such as the CONSORT statement, would help ensure complete reporting. Several dental journals have recently adopted the CONSORT guidelines, but only for reporting parallel-arm trials.

In conclusion, despite some progress in analysing and reporting split-mouth studies, there remains a substantial need for improvement. We believe that the optimal way to improve the statistical quality in oral health research is via collaborative efforts involving epidemiologists, statisticians and researchers from other relevant disciplines.

Acknowledgements

We acknowledge Barbara Hawkins for inviting the ISCB Subcommittee on Statistics in Dentistry to evaluate the extended CONSORT guidelines and for comments on a preliminary version of the paper. The authors also thank Philippe Hujoel, Jorgen Hilden, Dave DeMets and Eric Cobo for insightful remarks. Finally, we acknowledge the useful comments of two anonymous referees, which improved the quality of our review.

References

Campbell, M. K., Elbourne, D. R. & Altman, D. G., for the CONSORT Group (2004) The CONSORT statement: extension to cluster randomized trials. *BMJ* **328**, 702–708.

Esposito, M., Coulthard, P., Worthington, H. & Jokstad, A. (2001) Quality assessment of randomized controlled trials of oral implants. *International Journal of Oral and Maxillofacial Implants* **16**, 783–792.

Goldstein, H. (2003) *Multilevel Statistical Methods*, 3rd edition. London: Kendall's Library of Statistics 3, Arnold Publishing Inc.

Hujoel, P. P. (1998) Design and analysis issues in split mouth clinical trials. *Community Dentistry and Oral Epidemiology* **26**, 85–86.

Hujoel, P. P. & DeRouen, T. A. (1992) Validity issues in split-mouth trials. *Journal of Clinical Periodontology* **19**, 623–627.

Hujoel, P. P. & DeRouen, T. A. (1995) A survey of endpoint characteristics in periodontal clinical trials published 1988–1992, and implications for future studies. *Journal of Clinical Periodontology* **22**, 397–407.

Hujoel, P. P. & Loesche, W. J. (1990) Efficiency of split-mouth designs. *Journal of Clinical Periodontology* **17**, 722–728.

© 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

Hujoel, P. P. & Moulton, L. H. (1988) Evaluation of test statistics in split-mouth clinical trials. *Journal of Periodontal Research* **23**, 378–380.

Jones, B. J. & Kenward, M. G. (2003) *Design and Analysis of Cross-Over Trials*, 2nd edition. London: Chapman & Hall.

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

R Development Core Team (2005) The R Project for Statistical Computing. Available at http://www.r-project.org.

Ramfjord, S. P., Nissle, R. R., Shick, R. A. & Cooper, H. (1968) Subgingival curettage versus surgical elimination of periodontal pockets. *Journal of Periodontology* **39**, 167–175.

SAS Institute (2001) *The SAS System for Windows*. Cary: SAS Institute.

Schulz, K., Chalmers, I., Hayes, R. J. & Altman, D. (1995) Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* **273**, 408–412.

Therneau, T. M. & Grambsch, P. M. (2000) *Modelling Survival Data. Extending the Cox Model*. New York: Springer.

Wei, L. J. (1980) Observations that are subject to arbitrary right censorship. *Journal of the American Statistical Association* **75**, 634–637.

Yang, M. & Goldstein, H. (2003) *Modelling Survival Data in MLwiN 1.20*. London: Centre for Multilevel Modelling, Institute of Education, University of London.

Address:
Emmanuel Lesaffre
Biostatistical Centre
Catholic University Leuven
Kapucijnenvoer 35
B-3000 Leuven
Belgium
E-mail: emmanuel.lesaffre@med.kuleuven.be

Clinical Relevance

Scientific rationale for the study: Split-mouth studies are popular in periodontology and dental research due to the potentially greater efficiency and the need for fewer subjects. Previous reports in the late 1980s suggested problems with study design that could undermine the validity of the findings of such trials. *Principal findings:* Although there have been some improvements in methodology since the earlier assessment, reported methodology often contains serious problems. *Practical implications:* The errors in these studies can undermine the value and validity of these studies. Researchers are encouraged to follow guidance on the design and analysis of split-mouth studies.

© 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.