Quality Assessment of Randomized Controlled Trials of Oral Implants

Marco Esposito, DDS, PhD\textsuperscript{1}/Paul Coulthard, BDS, MFGDP, MDS, FDSRCS, PhD\textsuperscript{2}/
Helen V. Worthington, BSc, MSc, PhD, FIS\textsuperscript{3}/Asbjørn Jokstad, DDS, PhD\textsuperscript{4}

The aim of this study was to assess the quality of randomized controlled trials (RCTs) concerned with the effectiveness of oral implants and to create a trial register. A multilayered search strategy was used to identify all RCTs published by the end of 1999 in any language. The Cochrane Oral Health Group specialist register, PubMed, and personal libraries were searched. Seventy-four RCTs were identified. Forty-three articles, not presenting the same patient material, were independently assessed by 3 researchers using a specially designed form. A statistician assessed all trials for the appropriateness of statistics. The quality of each study was assessed on 7 items, including 3 key domains. Randomization and concealment allocation procedures were not described in 30 articles (70%). Reasons for withdrawals were not given in 10 reports (23%). No attempt at blinding was reported in 31 studies (72%). The quality of RCTs of oral implants is generally poor and needs to be improved. (INT J ORAL MAXILLOFAC IMPLANTS 2001;16:783–792)

Key words: dental implants, randomized controlled trial, registries, research design, review literature

The rehabilitation of patients with missing teeth is one of the most important tasks in dentistry. It would be of great benefit to know whether current therapeutic interventions are effective and, among alternative treatments, which is the best option. Such knowledge should be derived from clinical research of the highest quality. Clinical trials are designed to assess the effectiveness of an intervention in comparison with alternative interventions or no treatments. Different study designs are used to evaluate the magnitude of gains attributed to therapeutic interventions. However, well-designed, large, randomized clinical trials (RCTs) are considered the most scientifically sound method to minimize bias (systematic error).\textsuperscript{1} Proper randomization and allocation concealment minimize bias in treatment allocation, and a large sample size ensures improved precision of estimated treatment effects.\textsuperscript{2} Other important factors that should be taken into consideration to limit bias are well-defined inclusion and exclusion criteria and proper recording of the reason(s) for withdrawals of study subjects (attrition bias); also, whenever possible, all measures should be taken to blind the study subjects and the researchers to the treatment allocation (performance bias).

Identification of published RCTs is difficult and time-consuming,\textsuperscript{3,4} and their methodologic quality shows considerable variation.\textsuperscript{5,6} Thus, it is important to assess their quality before basing any changes in clinical practice on their findings. Therefore, the creation of a register of RCTs on oral rehabilitation procedures that also includes an objective quality assessment would be of value for improving patient care and for planning relevant research. There are already several RCT registers, among which the most complete is the Oral Health controlled clinical

\textsuperscript{1}Researcher, Institute of Anatomy and Cell Biology, Göteborg University, Göteborg, Sweden; and Researcher, NIOM (Scandinavian Institute of Dental Materials), Haslum, Norway.

\textsuperscript{2}Head, Oral and Maxillofacial Surgery, University Dental Hospital of Manchester, Manchester, United Kingdom.

\textsuperscript{3}Reader, Cochrane Oral Health Group, University of Manchester, Manchester, United Kingdom.

\textsuperscript{4}Associate Professor, Institute of Clinical Dentistry, University of Oslo, Oslo, Norway.

Reprint requests: Dr Marco Esposito, Institute of Anatomy and Cell Biology, Göteborg University, P.O. Box 420, SE 405 30 Göteborg, Sweden. Fax: +46-0-7733308. E-mail: marco.esposito@odontologi.gu.se
trials register of the Cochrane Collaboration (http://www.cochrane-oral.man.ac.uk/). A register of RCTs published in United States prosthodontic journals was recently initiated. However, the quality of RCTs included in such registers has not yet been evaluated.

The general aim of the authors was to create and maintain a register of published and unpublished RCTs involving rehabilitation of edentulism, to assess the methodologic quality of the studies, and to conduct a Cochrane systematic review of oral implants. The aim of the present investigation was to assess the methodologic quality of published RCTs of oral implants in an objective and reproducible way.

MATERIALS AND METHODS

Literature Search

A literature search strategy appropriate for a Cochrane systematic review was undertaken. The Cochrane Oral Health Group (OHG) specialist register was searched using the key word (“implant”). In September 2000, this database contained more than 8,900 RCTs, controlled clinical trials, and related material published on oral health. Trials included in this register are identified either by hand-searching or from various databases, including MEDLINE and EMBASE. Thirty-five journals were and are being hand-searched by the OHG. PubMed was independently searched for RCTs using the “related articles” feature. Two personal indexed databases containing over 3,000 (ME) and 1,500 (AJ) references on topics related to oral implants and prosthetics were also searched. Bibliographies of RCTs and relevant review articles were also searched. Bibliographies of RCTs and relevant review articles were checked for studies outside the hand-searched journals. Randomized controlled trials were also identified through correspondence and personal contacts with experts in the field. The present search was limited to RCTs published through the end of 1999 and was not restricted to the English language.

Quality Assessment

An evaluation form was designed to assess the quality of the study design and statistical analysis using 7 items (A to G in Fig 1). Also recorded were the country of origin, the funding source, the setting of the study, and the study design. This form was adapted from a validated source. Methodologic issues such as the relevance of the hypothesis tested, the choice of outcome measures, and the interpretation of results were not evaluated, since these are difficult to quantify objectively.

Articles were evaluated only for the information that they included, and no additional reference or information was sought. Since there were several follow-up RCTs presenting the same patient material, the last published of the series was analyzed under the rationale that it would contain the most complete information. Trials were not appraised for quality if they included fewer than 10 patients for a parallel study design or fewer than 5 for split-mouth or crossover designs.

Four nonblinded assessors (3 clinical researchers and 1 statistician) independently evaluated the quality of selected RCTs. Each article was assessed by 2 clinical researchers. The statistician evaluated all articles for the quality of statistical analysis (question G in Fig 1 and Table 1) and recorded any reason that statistical analyses were performed incorrectly. The final quality score of each article was determined in a consensus meeting by the 3 clinical researchers. In cases of inability to reach consensus, the dental statistician was consulted to make the final judgment.

RESULTS

Literature Search

Seventy-four RCTs investigating oral implant treatment were identified. All identified RCTs were published in English. After RCTs presenting the same patient population were excluded, 43 articles remained and were assessed in the present investigation (Table 1).

Interrater Agreement

For funding, setting, design, and items A to F in Fig 1, the percentage agreement was generally high, ranging from 87% to 100% between raters 1 and 2, from 69% to 100% between raters 1 and 3, and from 53% to 100% for raters 2 and 3 (Table 2). Kappa values were also generally high, with the comparison between raters 1 and 2 ranging from 0.72 to 1.00, with a median value of 1.00 and perfect agreement on 6 of the 9 criteria (Table 2). The kappa values between raters 1 and 3 ranged from 0.28 to 1.00, with a median value of 0.83 and perfect agreement on 3 criteria. The kappa values between raters 2 and 3 were low (~0) for 2 criteria; however, there was perfect agreement for 2 other criteria, and the median kappa value was 0.68. Nearly all disagreement could be attributed to reading errors or to differences in interpretation of the published material. All but 1 disagreement among clinicians were solved during a consensus meeting.
<table>
<thead>
<tr>
<th>Completion date:</th>
<th>Reviewer:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>Year of publication</td>
</tr>
<tr>
<td>Journal</td>
<td>Country</td>
</tr>
<tr>
<td>Funding source</td>
<td>Commercial</td>
</tr>
<tr>
<td>Setting</td>
<td>University</td>
</tr>
<tr>
<td>Study design</td>
<td>Parallel</td>
</tr>
</tbody>
</table>

Is the sample size ≥ 10 (≥ 5 for split-mouth and crossover studies)?

- No STOP HERE
- Yes Continue to complete form

A. Was a sample size calculation undertaken?
- No/not mentioned
- Yes, but not confirmed by calculation
- Yes, confirmed

B. Randomization and allocation concealment method
- Not described
- Clearly inadequate: Transparent before assignment (tossing coin, quasi-randomization such as sequential randomization)
- Possibly adequate: Sealed envelopes
- Clearly adequate: Centralized randomization and third party contact for group code

C. Were inclusion/exclusion criteria clearly defined?
- No
- Yes

D. Was reason for withdrawal specified by study group?
- No/not mentioned
- Yes, or not applicable as no withdrawals

E. Were the control and treatment groups comparable at entry for important prognostic factors?
- No
- Unclear
- Yes

F. Was there any attempt at blinding (for example, independent assessor)?
- No
- Yes

G. Was the statistical analysis appropriate?
- No
- Unclear
- Yes

Fig 1 Data collection form.
Table 1 Methodologic Scoring of RCTs

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Funding</th>
<th>Setting</th>
<th>Design</th>
<th>A (0–2)</th>
<th>B (0–3)</th>
<th>C (0–1)</th>
<th>D (0–1)</th>
<th>E (0–2)</th>
<th>F (0–1)</th>
<th>G (0–2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>El Charkawi et al</td>
<td>Egypt</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dahlin et al</td>
<td>Sweden</td>
<td>Independent</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Friberg et al</td>
<td>Sweden</td>
<td>Commercial</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Lundqvist et al</td>
<td>Sweden</td>
<td>Independent</td>
<td>Non-university</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Feine et al</td>
<td>Canada</td>
<td>Commercial</td>
<td>University</td>
<td>Crossover</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Gher et al</td>
<td>USA</td>
<td>Commercial</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Lavigne et al</td>
<td>Canada</td>
<td>Independent</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Palmer et al</td>
<td>USA</td>
<td>Independent</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Boerrigter et al</td>
<td>The Netherlands</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Burns et al</td>
<td>USA</td>
<td>Commercial</td>
<td>University</td>
<td>Crossover</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ciancio et al</td>
<td>USA</td>
<td>Commercial</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Jeffcoat et al</td>
<td>USA</td>
<td>Commercial</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Barber et al</td>
<td>USA</td>
<td>Commercial</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Bollen et al</td>
<td>Belgium</td>
<td>Commercial</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Hämmerle et al</td>
<td>Switzerland</td>
<td>Independent</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hunt et al</td>
<td>USA</td>
<td>Independent</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Quirynen et al</td>
<td>Belgium</td>
<td>Commercial</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Simion et al</td>
<td>Italy</td>
<td>Independent</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Bouma et al</td>
<td>The Netherlands</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Felo et al</td>
<td>USA</td>
<td>Commercial</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Kemppainen et al</td>
<td>Finland</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Trulhar et al</td>
<td>USA</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Zitzmann et al</td>
<td>Switzerland</td>
<td>Independent</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Batenburg et al</td>
<td>The Netherlands</td>
<td>Commercial</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Batenburg et al</td>
<td>The Netherlands</td>
<td>Commercial</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Bergendal and Engquist</td>
<td>Sweden</td>
<td>Independent</td>
<td>Non-university</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jemt et al</td>
<td>Sweden</td>
<td>Unclear</td>
<td>Non-university</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Khamis et al</td>
<td>USA</td>
<td>Commercial</td>
<td>University</td>
<td>Crossover</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Kwakman et al</td>
<td>The Netherlands</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Strooker et al</td>
<td>The Netherlands</td>
<td>Independent</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Wolff et al</td>
<td>USA</td>
<td>Commercial</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Andersson et al</td>
<td>Sweden</td>
<td>Unclear</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Åstrand et al</td>
<td>Sweden</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Davis and Packer</td>
<td>UK</td>
<td>Commercial</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Geertman et al</td>
<td>The Netherlands</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Gunne et al</td>
<td>Sweden</td>
<td>Independent</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jones et al</td>
<td>USA</td>
<td>Commercial</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Kapur et al</td>
<td>USA</td>
<td>Independent</td>
<td>Non-university</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Majzoub et al</td>
<td>Italy</td>
<td>Commercial</td>
<td>University</td>
<td>Split-mouth</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Meijer et al</td>
<td>The Netherlands</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Naert et al</td>
<td>Belgium</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Tang et al</td>
<td>Canada</td>
<td>Independent</td>
<td>University</td>
<td>Crossover</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Wissemijer et al</td>
<td>The Netherlands</td>
<td>Independent</td>
<td>University</td>
<td>Parallel</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

No. of studies with maximum score

| (of 43 reports assessed) | 0% | 2% | 79% | 77% | 49% | 28% | 65% |

A = sample size calculation; B = method of randomization and allocation concealment; C = inclusion/exclusion criteria; D = reason for withdrawals; E = comparability of control and treatment groups; F = attempt at blinding; G = appropriateness of statistical analysis.
Setting, Funding, and Study Design
Twenty-seven RCTs (63%) were conducted in Europe, 15 (35%) were conducted in North America, and 1 (2%) took place in Egypt. Most of the European studies were set in The Netherlands (9) and Sweden (8). Twenty RCTs (44%) were determined to have been commercially supported according to the information presented in the article. Twenty-one (49%) were determined to be independently funded, and for 2 studies (5%), the source of support was unclear. Thirty-eight studies (88%) were undertaken in universities or shared between university departments and other publicly funded institutions (eg, government health services). Twenty-five trials (58%) were designed as parallel, 14 (33%) as split-mouth, and 4 (9%) as crossover.

Quality Assessment
There were no RCTs with a sample size of fewer than 5 patients for split-mouth and crossover designs and 10 for parallel design. Results of the methodologic assessment of RCTs are summarized in Table 1. Only 1 study (2%) indicated that a sample size calculation (question A) was undertaken, although no figures were given. A clearly adequate randomization and allocation concealment (question B) was described in only 1 paper (2%). Seven articles (16%) scored 2 on question B, indicating a possibly adequate randomization. Six papers (14%) scored 1, indicating a clearly inadequate procedure, and in 30 investigations (70%), no information was provided. Inclusion/exclusion criteria (question C) were clearly defined in 34 studies (79%). The reason for withdrawals was specified by study group (question D) in 33 reports (77%). Control and treatment groups were comparable at entry for important prognostic factors (question E) in 20 articles (46%). Seventeen papers (40%) were unclear, and 6 studies (14%) were judged to have baseline groups that were not comparable. No attempts at blinding (question F) were described in 31 papers (72%). Twelve investigations (28%) described some sort of blinding procedure.

The appropriateness of the statistical analysis (question G) was assessed by a single rater. Of 43 reports, 2 (5%) included no statistical analysis, 9 (21%) were considered to have an inappropriate statistical analysis applied, in 4 (9%) articles it was unclear whether the analysis was appropriate, and for the 28 remaining reports (65%) the statistical methods were considered to be adequate. The statistical analysis was considered inappropriate in 1 paper, as the calculations were incorrect and it was stated that “there was no significant difference between groups with \( P < .01. \)” In 5 papers, the clustering of implants within patients was ignored in the analysis. In several studies in which the statistical methods were considered appropriate, the actual \( P \) values were not given, only indications of ranges were provided, including \( P \) values, eg, “.01 < \( P < .05. \)” Currently, with the use of computers it is advisable to quote actual \( P \) values. Thirteen of the 28 reports (46%) in which the statistical analysis was considered to be appropriate included a statistician as an author or acknowledged the help of a statistician, compared with only 2 (15%) of the reports considered inappropriate or unclear. Although not significant, this trend suggests that it is helpful to involve a statistician in the design and

### Table 2 Assessment of Interexaminer Agreement

<table>
<thead>
<tr>
<th>Factor</th>
<th>Rater 1 versus rater 2 (n = 15)</th>
<th>Rater 1 versus rater 3 (n = 13)</th>
<th>Rater 2 versus rater 3 (n = 15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Kappa SE 95% CI</td>
<td>% Kappa SE 95% CI</td>
<td>% Kappa SE 95% CI</td>
</tr>
<tr>
<td>Funding</td>
<td>87 0.72 0.14 0.49 to 1.00</td>
<td>92 0.85 0.13 0.59 to 1.00</td>
<td>80 0.66 0.16 0.36 to 0.96</td>
</tr>
<tr>
<td>Setting</td>
<td>93 0.85 0.14 0.58 to 1.00</td>
<td>92 0.76 0.23 0.31 to 1.00</td>
<td>87 N/A* N/A N/A</td>
</tr>
<tr>
<td>Design</td>
<td>100 1.00 N/A N/A</td>
<td>92 0.83 0.16 0.52 to 1.00</td>
<td>100 1.00 N/A N/A</td>
</tr>
<tr>
<td>A</td>
<td>100 1.00 N/A N/A</td>
<td>100 1.00 N/A N/A</td>
<td>100 1.00 N/A N/A</td>
</tr>
<tr>
<td>B</td>
<td>100 1.00 N/A N/A</td>
<td>69 0.55 0.17 0.22 to 0.87</td>
<td>87 0.71 0.17 0.38 to 1.00</td>
</tr>
<tr>
<td>C</td>
<td>100 1.00 N/A N/A</td>
<td>69 0.28 0.28 –0.28 to 0.83</td>
<td>60 –0.25 0.10 –0.41 to –0.06</td>
</tr>
<tr>
<td>D</td>
<td>93 0.87 0.13 0.61 to 1.00</td>
<td>100 1.00 N/A N/A</td>
<td>86 –0.07 0.05 –0.17 to 0.03</td>
</tr>
<tr>
<td>E</td>
<td>100 1.00 N/A N/A</td>
<td>69 0.55 0.17 0.21 to 0.88</td>
<td>93 0.14 0.20 –0.25 to 0.53</td>
</tr>
<tr>
<td>F</td>
<td>100 1.00 N/A N/A</td>
<td>100 1.00 N/A N/A</td>
<td>87 0.71 0.19 0.34 to 1.00</td>
</tr>
</tbody>
</table>

*Cannot calculate, as one rater gave same category for all papers.

% = percent agreement; SE = standard error; CI = confidence interval; A–F = questions A through F on data collection form (Fig 1).
analysis of RCTs in this area (chi-square = 3.7, 1 degree of freedom, \( P = .055 \)).

**DISCUSSION**

As in other kinds of empirical research, searching and assessment of the literature is susceptible to bias. The major limitation of a register of published RCTs is that it could be biased toward positive and “encouraging” results (publication bias). This is because of the fact that “uninteresting” information is less likely to reach the publication stage, which may lead to erroneous conclusions of therapeutic effectiveness.84 Therefore, it would be of great benefit if unpublished trials could also be identified.85

The fact that no RCTs in languages other than English have been identified may reflect either an inability to access such publications or the preference of researchers in this discipline to use the English language to disseminate their “best” clinical research. It is recommended that systematic literature searches also include articles written in languages other than English.86,87

The aim of the present paper was to attempt an objective and reproducible quality assessment of RCTs published in implant dentistry. To undertake this, a specifically designed checklist was developed from one previously published.9 While there are some differences in the items included, all checklists basically focus on the same sources of potential bias. A summary score was not calculated, as these have been shown to be problematic in identifying trials of high quality.88

The quality of the study design was assessed indirectly by evaluating the quality of reporting. It is important to note that there is a difference between the quality of the presentation and the manner in which the study was actually conducted. However, it has been suggested that failure to report important items is usually the result of these procedures not having been carried out, rather than underreporting.89 Additional information may have been presented in previously published reports that were not assessed in the present article. However, it has been recommended that all information should be presented clearly, allowing the reader to make an informed judgment regarding internal and external validity of the trials.90 In many instances, important information had not been provided by the authors of the papers. The best solution to this problem is to write to the authors asking for the missing information. This task is currently being undertaken for a Cochrane systematic review (http://www.cochrane-oral.man.ac.uk/).

Randomized controlled trials were assessed in a nonblinded fashion, and this may lead to potential bias.91 However, the findings are presented in a reproducible way (Table 1). Thus the critical reader is able to check the scores. In addition, it is very time-consuming and difficult to blind experienced literature assessors.

Independent scoring by 3 reviewers resulted in a relatively high agreement and a subsequent consensus meeting solved all but one of the disagreements in individual interpretation. Two of the reviewers had previously undertaken a similar evaluation together, which may explain the closer agreement of these 2 raters.

The authors were aware from personal contacts that several RTCs were conducted with manufacturers’ financial support, but this was not always disclosed in the published articles. Therefore, the number of RCTs recorded as sponsored by industry in the present article is likely to be underestimated. Only a few dental journals require that authors disclose any conflict of interest; it would be preferable if such a policy were adopted universally.

The majority of trials (58%) were of parallel design; however, a significant number of investigations used the split-mouth (33%) or crossover (9%) design. In many medical disciplines, it is not possible to undertake RCTs using a split-mouth or crossover design. Such designs offer the advantages of limiting the number of variables, thus reducing the number of needed study subjects.92

For evaluation of the methodologic quality, the authors focused on those items that have been shown to be particularly relevant for decreasing bias.85,90 An attempt was made to formulate questions in a way to minimize subjective interpretation. However, questions C, E, and G (Fig 1) were still prone to subjective preferences.

An arbitrary cutoff value of 10 study participants for parallel studies and 5 subjects for split-mouth and crossover design was chosen for inclusion in the present assessment, in accordance with published literature.1,7 Despite no RCT being excluded on this basis, a sample size calculation was mentioned, though not confirmed by calculation, in only 1 trial.73 Sample size calculation estimates the minimal number of patients needed to detect a significant difference among groups to be compared. If the number of subjects included in a study is too small, clinically important effects related to different interventions may not be detected.5,91,94 It should be recognized that such studies may be scientifically useless and thus unethical in their use of patients and other resources.

An RCT is a study in which participants are allocated at random to receive different interventions.
Random allocation means that all participants have the same likelihood to be assigned to each of the study groups. If properly accomplished, randomization minimizes bias in allocating participants to the study groups. To be effective, the randomly generated sequence should be strictly implemented, and maximal attention should be given to avoid any possible source of subversion. This process is called allocation concealment and is meant to prevent foreknowledge of the treatment assignment. The use of central telephone randomization or sequentially numbered sealed opaque envelopes has been recommended as the minimum measure for allocation concealment. Studies that present inadequate or unclear randomization and allocation concealment have been shown to yield larger estimates of treatment effects. Without proper allocation concealment, randomization is lost and bias is likely to distort results. The majority of RCTs published in implant dentistry (70%) did not describe how randomization and allocation concealment were performed. Only 1 paper reported a clearly adequate method of randomization and allocation concealment.

External validity or generalizability denotes the precision and extent to which it is possible to generalize the results of a study to other settings. External validity is relevant to making treatment decisions. Clearly defined inclusion and exclusion criteria will help the reader to decide whether the results of a trial are applicable to his or her own population of patients. The majority of papers (79%) clearly defined the inclusion and exclusion criteria.

It is important to know whether withdrawals or exclusions of study participants occurred and from which group (attrition bias), since this may result in a systematic error that would lead to an incorrect estimate of the treatment effectiveness. For instance, patients may drop out because of intervention side effects or may be deliberately excluded by an investigator because of alleged protocol deviation. The majority of trials (77%) either described the reason for withdrawals or experienced no withdrawals.

If an RCT is truly randomized, systematic bias is (in theory) avoided by selecting participants from a particular population and by allocating them randomly to different groups. The groups should be identical apart from the treatment so that any difference in outcome is attributable to the intervention. Therefore, it was interesting to find that 6 studies (14%) were judged to have baseline groups that were not comparable and that for 40% of the studies, comparability was judged to be unclear.

When clinical judgment is needed, personal preferences of the investigators may intrude. This problem can be prevented if those assessing treatment outcomes are unaware of the treatment that each patient received. Blinding is not always possible for surgical interventions such as oral implant treatment. However, some precautions should be taken to minimize bias, such as the use of independent assessors for measuring outcomes. Only 28% of the assessed RCTs described some sort of blinding procedure.

The statistical methods were considered to be inappropriate or unclear in a third of the RCTs. The 2 main reasons that they were considered inappropriate were: (1) implants within the same patient were assumed to be independent in the analysis; and (2) in split-mouth studies, the analysis ignored the “pairing” because of the study design. Implants are sometimes clustered within patients, and this must be taken into account in the analysis. For RCTs in implant dentistry, this is frequently achieved by using average patient scores, with the patient being the unit of analysis. Other methods, such as generalized estimating equations or multilevel modeling, may be appropriate but were not used in any of these studies. Ignoring the split-mouth design by analyzing the data as though they were from different patients in a parallel group study leads to incorrect estimates of the standard errors of the treatment effects.

Several investigators who explored different dental and medical disciplines concluded that study methodology was generally poor. Therefore, it was not surprising to find that the methodologic quality of RCTs in implant dentistry was poor. Since much effort in terms of resources and time is invested in research, it would be valuable if the methodologic quality of research was of a sufficient level to produce more meaningful results.

To improve the quality of reporting RCTs, a unified statement of a panel of experts (the CONSORT statement) was published. In principle, the requirements are that authors provide enough information for the readers to know how a trial was performed so they can judge whether the findings are likely to be reliable. Several eminent medical and dental journals, including The Lancet, Journal of the American Medical Association, British Medical Journal, and British Dental Journal, have adopted these recommendations for publishing RCTs.

In conclusion, there seems to be considerable potential for improving the design, conduct, statistical analysis, and reporting of RCTs in implant dentistry.
ACKNOWLEDGMENTS

This study was supported in part by the Swedish Medical Research Council (9495), the Swedish National Board for Technical Development (NUTEK), the Hjalmar Svensson Foundation (Sweden), and the PPP Healthcare Trust (United Kingdom). No financial support has been received from any oral implant manufacturer. We wish to thank Sylvia Bickley (Cochrane Collaboration) for her assistance with literature searching. It is possible that the authors failed to mention the work of some investigators. Such omissions do not reflect the importance of the studies but an inability to gain access to this material. The authors wish to thank in advance all those who will provide information regarding any published or unpublished RCT of oral implants not included in the present article.

REFERENCES


