How Well Are Randomized Controlled Trials Reported in the Dermatology Literature?

Kayode Adetugbo, PhD; Hywel Williams, PhD

Objective: To assess the methodological quality of the design and reporting of randomized controlled trials published in one major dermatology specialty journal.

Design and Data Sources: In a survey of all published parallel group randomized controlled trials, we found 73 reports with allocation described as randomized from all issues of *Clinical and Experimental Dermatology* from its inception in 1976 through 1997.

Main Outcome Measures: Direct and indirect measures of the adequacy of randomization, trial sample size, baseline comparisons, and intention-to-treat analysis.

Results: Hand searching identified 73 randomized controlled trials, but only 31 of these were found by searching MEDLINE for the publication type *clinical trials*. Of the 73 randomized controlled trials, 68 contained sufficient information to include in the analysis. Only 1 study (1%) reported the method of random sequence generation, and only 5 studies (7%) reported adequate concealment of allocation. Among 38 trials that used simple randomization, the sample sizes in the comparison groups were identical in 22 occasions, raising the possibility that simple randomization might not have been adequately generated or concealed. Most trials (88%) excluded some randomized participants from their analysis. The median sample size was 23 per trial. Only 1 trial reported sample size and statistical power considerations and had an a priori main hypothesis.

Conclusions: Hand searching is important for locating all relevant trials. There is the need for higher methodological quality in clinical trial reporting in dermatology journals. The adoption of the CONSORT (Consolidated Standards of Reporting Trials) statement and checklist for the reporting of trials should enhance the validity of and strengthen the evidence from clinical trials reports.

*Arch Dermatol.* 2000;136:381-385

WELL-DESIGNED clinical trials yield strong evidence for the effects of health care interventions. The randomized controlled trial (RCT) is the criterion standard for clinical trials and offers the most robust and rigorous way of determining the causal relationship between treatment and effect and for assessing the cost-effectiveness of a treatment. Random assignment greatly reduces the potential for bias in the allocation of interventions. It ensures that comparison groups are as similar as possible to each other in terms of both known and unknown possible predictors of treatment response, and RCTs have been consistently shown to give more reliable evidence of therapeutic effectiveness than non-randomized studies. Even so, there are some limitations and skepticism regarding the use and value of randomized trials. Essential features of the RCT include (1) allocation of participants to intervention groups in a truly unpredictable, randomized sequence (including the concealment of the allocation schedule until the assignment is made); (2) blinding or masking of participants and/or observers as to what treatment an individual is receiving;

From the Cochrane Skin Group, Dermato-Epidemiology Unit, Queen’s Medical Centre, University Hospital, Nottingham, England.

**Section Editors**

Damiano Abeni, MD, MPH, Istituto Dermopatico dell’Immacolata, Rome, Italy  
Michael Bigby, MD, Beth Israel Deaconess Medical Center, Boston, Mass  
Paola Pasquini, MD, MPhil, Istituto Dermopatico dell’Immacolata, Rome, Italy  
Moyses Szklo, MD, MPH, DrPH, Johns Hopkins University, Baltimore, MD  
Hywel Williams, MD, Queens Medical Center, Nottingham, England
METHODS

HAND SEARCHING

We hand searched all issues of Clinical and Experimental Dermatology from its inception in 1976 through 1997 (volumes 1-22) as part of the systematic effort of the Cochrane Collaboration to identify all published trials. Each journal article, review, letter, and meeting abstract was examined to determine if it was a randomized controlled trial or controlled clinical trial (CCT) publication type as defined by the National Library of Medicine. We then downloaded the MEDLINE record for each trial (RCT or CCT) identified by hand searching into a bibliographic database (ProCite, version 3.1; Research Information Systems, Carlsbad, Calif), and we recorded its publication type according to MEDLINE.

In this study, a clinical trial involved at least 1 test treatment and 1 control treatment and concurrent enrollment and follow-up of the treatment groups. A trial was classified as an RCT if there was an explicit statement that the comparison groups were established by random allocation, such as by a table of random numbers. Trials that allocated treatment using coin flips; odd, even, patient Social Security, or medical record numbers; days of the week; or other pseudorandom or quasi-random processes and other trials for which the authors did not explicitly state that random allocation was made were classified as CCTs.

ASSESSMENT OF QUALITY

We assessed each of the RCT reports for the following: the reporting of the methods of random sequence generation, concealment of allocation schedule, adequacy of sample size, use of masking, and intention-to-treat analysis (or inclusion of all randomized participants in the trial analysis); comparability of adequate baseline measures for comparison groups; (it is desirable that the comparison groups in a randomized trial be as similar as possible in factors or characteristics that might influence the participants’ response to the intervention. Stratified randomization is used to ensure that equal numbers of participants with a characteristic thought to affect prognosis or response to the intervention are allocated to each comparison group. For example, in a trial of women with psoriasis, it may be important to have similar numbers of smokers and nonsmokers in each comparison group, and stratified randomization could be used to achieve this balance. Stratified randomization is carried out either by performing separate randomization [often using random permuted blocks] for each stratum or by using minimization. Random permuted blocks [or block randomization] ensures that at any point in a trial roughly equal numbers of participants are allocated to the comparison groups. Permuted blocks are often used in combination with stratified randomization. With small trials, simple randomization often produces unbalanced groups. Minimization is a method of allocation used, particularly in small trials, to ensure that comparison groups are closely similar for several variables [that may be important prognostic factors]. It can be done with or without a component of randomization, and it is best performed centrally with a computer program to ensure allocation concealment.)

Acceptable methods of random sequence generation were the following: computerized random number generator, table of random numbers, and minimization. It is important for the reduction of bias that the allocation schedule is concealed up to the point of the allocation of treatment and we accepted the following approaches to allocation concealment as adequate: central randomization; central dispensation of intervention at a pharmacy; numbered or coded containers; and sequentially numbered, opaque, and sealed envelopes. We recorded the trial size and the sample size of each comparison group. We looked for a prior definition of an acceptable difference in outcome measures or a hypothesis to be tested and the calculation of sample sizes around this definition. Also, we recorded whether all randomized participants were included in the analysis of the trial outcomes (intention-to-treat analysis). We looked for an account of the mechanism to ensure that neither trialist nor trial participant was able to identify the intervention being provided or that comparison treatments could not be distinguished by taste or appearance. Moreover, we also looked in the reports for baseline comparisons of appropriate variables for the intervention groups using measures of SD, range, raw data, or χ² tests for nominally scaled variables.

The data were extracted onto forms, and thereafter, the assessments were entered into a worksheet (Excel; Microsoft Corp, Redmond, Wash) and analyzed using simple descriptive statistics.

(3) use of an intention-to-treat analysis, the principle of which is to include in the analysis all randomized participants in the group into which they were allocated, whether or not they received or completed the intervention, so as to preserve random allocation and thus the comparability of the study groups; and (4) use of a prospectively stated and explicitly defined trial hypothesis. All these features minimize bias in the selection of participants, allocation of intervention, and analysis of trial outcomes.

However, to maximize usefulness, the RCT should be accessible and assessable. Accessibility allows RCTs to be combined with similar trials in a systematic review or in a meta-analysis, often regarded as the highest level of evidence for the effects of health care. For such pooling, the quality of the design and execution of the RCT needs to be evaluated, especially as the quality of the research evidence from meta-analyses depends on the quality of the trials that form them. Complete reporting aids this evaluation. It has been shown that the low quality of randomization in RCT’s is associated with bias and that poor evidence of randomization (such as an inadequate or unclear method of treatment allocation) significantly exaggerates estimates of treatment effects by 30% to 41%. Therefore, the evidence from such studies is that much less accurate. Consequently, the patients and the scientific community can derive maximum benefit from the various contributions to trials only if they are rigorously designed and executed and completely and unambiguously reported. This consideration has led to the formulation of a standard for the reporting of RCTs (Table 1), embodied in the CONSORT (Consolidated Standards of Reporting Trials) statement of 1996.

The Cochrane Skin Group was recently formed as part of the evidence-based health care movement. It pre-
pares, maintains, and disseminates systematic reviews of the effects of skin treatments and maintains a specialized register of skin trials. The aim of its work is to facilitate the development and practice of evidence-based dermatology. Important for this development are high-quality skin trials and the clear and complete reporting of their information, as all skin trials (published or not) may be potentially included in the systematic reviews and meta-analyses. As part of the group’s work of assembling and assessing all the evidence for the effectiveness of skin treatments, we identified, by hand searching, all the clinical trials potentially included in the systematic reviews and meta-analyses. As part of the group’s work of assembling and assessing all the evidence for the effectiveness of skin treatments, we identified, by hand searching, 96 reports of clinical trials. All the clinical trial reports came from 88 full-length journal articles, 2 letters, and 6 meeting abstracts.

Records of 88 of these 96 studies were recoverable from MEDLINE: there were no records for 4 RCTs and 4 CCTs (1 letter and 3 abstracts in each category). Conversely, only 59 of the 96 reports identified by hand searching as CCTs or RCTs were recorded as clinical trials according to MEDLINE (31 RCTs, 13 CCTs, and 15 clinical trials) (Table 2).

**RESULTS**

We found, by hand searching, 96 reports of clinical trials. There were 73 RCTs with 5629 participants and 22 CCTs with 1412 participants. One report was an overview that combined several separate clinical trials. All the trial reports came from 88 full-length journal articles, 2 letters, and 6 meeting abstracts.

**ASSESSMENT OF QUALITY**

Of the 73 RCTs, 68 were reviewed; the 3 abstracts and 1 letter did not contain enough information for assessment, and 1 other publication also was eliminated from this assessment because of inadequate information. The

---

*Adapted from the CONSORT statement.*

---

**Table 1. Modified CONSORT (Consolidation of Standards of Reporting Trials) Checklist**

<table>
<thead>
<tr>
<th>Heading</th>
<th>Subheading</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td></td>
<td>Identify the study as a randomized trial.</td>
</tr>
<tr>
<td>Abstract</td>
<td></td>
<td>Use a structured format.</td>
</tr>
<tr>
<td>Introduction</td>
<td>Protocol</td>
<td>State prospectively defined hypothesis, clinical objectives, and planned subgroup or covariate analyses.</td>
</tr>
<tr>
<td>Methods</td>
<td>Analysis</td>
<td>Planned study population, together with inclusion and exclusion criteria.</td>
</tr>
<tr>
<td></td>
<td>Result</td>
<td>Planned interventions and their timing.</td>
</tr>
<tr>
<td></td>
<td>Analysis</td>
<td>Primary and secondary outcome measure(s) and the minimum important difference(s) and how the target sample size was projected.</td>
</tr>
<tr>
<td></td>
<td>Comment</td>
<td>Rationale and methods for statistical analyses, detailing main comparative analyses and whether they were completed on an intention-to-treat basis.</td>
</tr>
<tr>
<td></td>
<td>Comment</td>
<td>Prospectively defined stopping rules (if warranted).</td>
</tr>
<tr>
<td>Assignment</td>
<td>Description</td>
<td>Describe</td>
</tr>
<tr>
<td>Masking</td>
<td>(blinding)</td>
<td>Unit of randomization (eg, individual, cluster, or geographic).</td>
</tr>
<tr>
<td>Results</td>
<td>Participant flow and follow-up</td>
<td>Method used to generate the allocation schedule.</td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
<td>Method of allocation concealment and timing of assignment.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Method to separate the generator from the executor of assignment.</td>
</tr>
<tr>
<td>Comment</td>
<td></td>
<td>Describe mechanism (eg, capsules or tablets); similarity of treatment characteristics (eg, appearance or taste); allocation schedule control (location of code during trial and when broken); and evidence for successful masking (blinding) among participants, person doing intervention, outcome assessors, and data analysts.</td>
</tr>
<tr>
<td>Comment</td>
<td></td>
<td>State specific interpretation of study findings, including sources of bias and imprecision (internal validity) and discussion of external validity (generalizability of findings to other populations), including appropriate quantitative measures when possible.</td>
</tr>
<tr>
<td>Comment</td>
<td></td>
<td>State general interpretation of the data in light of the totality of the available evidence.</td>
</tr>
</tbody>
</table>

---

**Table 2. Yield of Clinical Trials**

<table>
<thead>
<tr>
<th>Publication Type</th>
<th>Hand Searching</th>
<th>Electronic Searching</th>
</tr>
</thead>
<tbody>
<tr>
<td>All clinical trials</td>
<td>96 (100)</td>
<td>59 (100)</td>
</tr>
<tr>
<td>Randomized controlled trial</td>
<td>73 (78)</td>
<td>31 (53)</td>
</tr>
<tr>
<td>Clinical controlled trial</td>
<td>22 (23)</td>
<td>27 (46)</td>
</tr>
<tr>
<td>Overview</td>
<td>1 (1)</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

*Percentages are rounded.*
results of the quality assessment are based on 68 RCT reports and are summarized in Table 3.

All the studies randomly allocated trial participants and units to comparison groups, but only 1 of the studies reported a method for random sequence generation (use of a table of random numbers). In addition, 5 studies reported using stratified randomization or blocking. Of the 68 studies, 5 (7%) reported an adequate method of concealment of allocation.

Of 38 RCT reports of unblocked parallel studies in which the unit of randomization was the individual, 22 (58%) had identical numbers of participants in each comparison group (Figure). These 22 included reports of 6 trials with 3 or 4 comparison groups. In another 5 studies, the sizes of comparison groups were not different from those achieved through simple alternation. Thus, only 11 (29%) of 38 presumably simply RCTs reported a difference of greater than one individual in the sizes of the comparison groups.

Sample sizes were generally small, with a median of 23 randomized units per intervention group (interquartile range, 15-40 units [containing 50% of the sample size values]) (Figure). Only 1 report included sample size and power considerations. Thirty-two studies (47%) were reported as double masked while 26 (38%) did not report any masking (and were presumably not masked).

Only 4 (6%) of the 68 reports specifically mentioned that they did not exclude randomized participants from the analysis of the study result (intention to treat). In another 4, there were apparently no dropouts, as all were accounted for in the analysis of trial results. In 60 reports, the number of participants analyzed was less than those randomized. Thus, 60 (88%) of the 68 reports excluded (or appeared to have excluded) some randomized participants from the analysis of trial results.

**Table 3. Measures of Methodological Quality of 68 Randomized Controlled Trials**

<table>
<thead>
<tr>
<th>Property or Method</th>
<th>No. (%) of Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adequate method for</td>
<td></td>
</tr>
<tr>
<td>Generating a random sequence</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Concealment of allocation sequence</td>
<td>5 (7)</td>
</tr>
<tr>
<td>Sample size and power calculations</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Masking (blinding)*</td>
<td></td>
</tr>
<tr>
<td>Double masking</td>
<td>32 (47)</td>
</tr>
<tr>
<td>Single masking</td>
<td>6 (9)</td>
</tr>
<tr>
<td>Open trials</td>
<td>4 (6)</td>
</tr>
<tr>
<td>No report of masking</td>
<td>26 (38)</td>
</tr>
<tr>
<td>Intention-to-treat principle</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Adequate baseline measures of group comparability</td>
<td>40 (59)</td>
</tr>
<tr>
<td>Included key word randomized in title</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>

* In double masking, both participants and researchers in a trial were unaware of the treatments given or received, whereas in single masking, only participants were unaware of the treatment received.

Validity is the degree to which a result is likely to be free of bias, and so, true. *Internal validity* expresses the extent to which the observed effects are true for the study sample, while *external validity* (or generalizability) is the extent to which the observed results reflect what is true for the general population. The reporting of RCTs should allow an assessment of their internal and external validity, especially as these trials contribute strong primary evidence for the effects of health care.

There is now consensus on the formats for this reporting in the face of empirical evidence that poorly and incompletely reported trials are associated with greater bias and an exaggeration of treatment effects, and there are continuing efforts to improve the quality of reporting of RCTs. Our results showed that the reporting of RCTs in *Clinical and Experimental Dermatology* was generally not complete and could be improved. Almost all the reports of trials did not describe the methods of generating random sequences or of concealing (protecting) random allocation schedules. Sample sizes were generally small, making the trials prone to type II errors (falsely claiming no treatment effect when a clinically useful effect might still exist), and the studies did not report sample size and power calculations based on a main a priori hypothesis. Almost two fifths of the studies did not report masking and were presumably not adequately masked. Almost 90% excluded (or appeared to have excluded) some randomized participants from their analyses. These problems suggest that many of these studies may have exaggerated treatment effects.

Our results agree with other surveys that described inadequate execution and/or reporting of RCTs.

The issues highlighted may represent incomplete reporting, methodological and design flaws, or both. In one research report, in which clarification from the authors was sought, the results suggested that inadequate reporting was a reflection of inadequate methods. Thus, deficient reporting of data in RCTs can be used as proxy for their defective design and execution.

The importance of the concealment or protection of the allocation schedule in RCTs is probably not as widely appreciated as that of generating a random sequence for allocation to comparison groups. Yet, it is an integral part of the randomization process, and the bias introduced from failure to conceal the allocation schedule can be more serious than that from defective random sequence generation. Therefore, for complete reporting of the randomization process, it is desirable to specify the steps taken to protect the allocation code.
The close similarity between the sizes of intervention groups within trials (Figure) that used simple randomization was greater than that attributable to chance alone. Based on the expected distribution of difference in sample sizes according to the total trial size, none of the 38 trials were outside the boundary, where we expected 50% of the trials to be. This observation suggests that in some of the randomized trials allocation to groups may not have been strictly carried out following a random, unpredictable sequence, as previously described and discussed, and it may be that trialists erroneously thought that a more uniform distribution of sizes of groups conferred greater credibility on their trials.

Moreover, the comparison sample sizes were generally small, such that most have enough power (ie, the ability to detect a difference in treatment outcomes, if there is a real one) to detect only greater than a doubling of relative treatment effects for categorical outcomes. The methodological flaw of small sample sizes and insufficient statistical power is not peculiar to skin trials. Reviewing 71 meta-analysis. One recommendation of the CONSORT published evidence available for a systematic review or including all published trials and, consequently, making all the pending on hand searching. It is the only sure way of finding all published trials by 60% compared with electronic searching, validating the effects that the Cochrane Collaboration is expending on hand searching. It is the only sure way of finding all published trials and, consequently, making all the published evidence available for a systematic review or meta-analysis. One recommendation of the CONSORT statement is that reports of RCTs include some form of the word randomized in the title (only 1 of 68 did in this survey). The universal adoption of this style by dermatology journals will considerably improve the electronic searching yield of studies and obviate the need for prospective hand searching. Use of the approaches described herein will facilitate the emergence of high-quality evidence in systematic reviews and meta-analyses summarizing the effects of health care interventions.

Accepted for publication December 19, 1999.

This study was supported in part by the United Kingdom National Health Service Research and Development Programme, London, England.

Corresponding author: Kayode Adetugbo, PhD, 15 Dolphin Rd, Northolt, UB5 6UQ, Middlesex, England (e-mail: adetugbo@hotmail.com).

REFERENCES


