# BMJ

# RESEARCH

# Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8

OPEN ACCESS

N M Ivers *family physician*[1], M Taljaard *scientist*[2], S Dixon *postdoctoral fellow*[3], C Bennett *research coordinator*[2], A McRae *research director*[4], J Taleban *PhD candidate*[5], Z Skea *research fellow*[6], J C Brehaut *scientist*[2], R F Boruch *professor of education and statistics*[7], M P Eccles *professor of clinical effectiveness*[8], J M Grimshaw *senior scientist*[2], C Weijer *professor of philosophy and medicine*[9], M Zwarenstein *senior scientist*[10], A Donner *professor of epidemiology and biostatistics*[11]

[1]Women's College Hospital, 76 Grenville Street, Toronto, ON, Canada M5S 1B2; [2]Ottawa Hospital Research Institute, Clinical Epidemiology Program, Ottawa Hospital, Civic Campus, 1053 Carling Avenue, Ottawa, ON; [3]Department of Epidemiology and Biostatistics, Schulich School of Medicine and Density, University of Western Ontario, London, ON; [4]Division of Emergency Medicine, University of Calgary, Foothills Medical Centre, 1403-29 Street NW, Calgary, AB; [5]Department of Epidemiology and Biostatistics, University of Western Ontario, London, ON; [6]Health Services Research Unit, Health Sciences Building, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK; [7]Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104, USA; [8]Institute of Health and Society, Newcastle University, Baddiley-Clark Building, Newcastle upon Tyne; [9]Rotman Institute of Philosophy, Department of Philosophy, University of Western Ontario, London, ON; [10]Institute for Clinical Evaluative Sciences and Sunnybrook Research Institute, G106 - 2075 Bayview Avenue, Toronto, ON; [11]Schulich School of Medicine and Dentistry, Kresge Building, University of Western Ontario, London, ON

## Abstract

**Objective** To assess the impact of the 2004 extension of the CONSORT guidelines on the reporting and methodological quality of cluster randomised trials.

**Design** Methodological review of 300 randomly sampled cluster randomised trials. Two reviewers independently abstracted 14 criteria related to quality of reporting and four methodological criteria specific to cluster randomised trials. We compared manuscripts published before CONSORT (2000-4) with those published after CONSORT (2005-8). We also investigated differences by journal impact factor, type of journal, and trial setting.

**Data sources** A validated Medline search strategy.

**Eligibility criteria for selecting studies** Cluster randomised trials published in English language journals, 2000-8.

**Results** There were significant improvements in five of 14 reporting criteria: identification as cluster randomised; justification for cluster randomisation; reporting whether outcome assessments were blind; reporting the number of clusters randomised; and reporting the number of clusters lost to follow-up. No significant improvements were found in adherence to methodological criteria. Trials conducted in clinical rather than non-clinical settings and studies published in medical journals with higher impact factor or general medical journals were more likely to adhere to recommended reporting and methodological criteria overall, but there was no evidence that improvements after publication of the CONSORT extension for cluster trials were more likely in trials conducted in clinical settings nor in trials published in either general medical journals or in higher impact factor journals.

**Conclusion** The quality of reporting of cluster randomised trials improved in only a few aspects since the publication of the extension of CONSORT for cluster randomised trials, and no improvements at all were observed in essential methodological features. Overall, the adherence to reporting and methodological guidelines for cluster randomised trials remains suboptimal, and further efforts are needed to improve both reporting and methodology.

## Introduction

In recent years, increasing attention has been paid to the importance of good reporting practices as they relate to the potential utility of a manuscript.[1] The CONSORT (consolidated standards of reporting trials) statement, originally published in 1996 and updated in 2001 and 2010, provides authors and editors with a checklist for a minimum set of recommendations for

reporting the trial design, analysis, and results.[2] Although certain inadequacies remain common, the quality of reporting of randomised controlled trials in medical journals seems to be improving over time.[3] A recent systematic review indicated that the CONSORT statement has played an important role in this progression.[4]

Unfortunately, reviews of published cluster randomised trials (see box 1) have repeatedly found important shortcomings in their methodological conduct and reporting.[5-18] For example, a review of 152 cluster randomised trials published 1997-2000 found that most of them did not adhere to recommended methods for cluster randomised trials.[8] To help address this problem, an extension for the original CONSORT guideline, specifically addressing the unique methodological features of cluster randomised trials, was published in 2004.[19] In this extension, the authors altered the recommendations for 15 of 22 items on the original CONSORT checklist to emphasise the additional requirements for adequate methodological conduct and reporting of cluster randomised trials. Encouragingly, a review of 34 primary care trials published in seven major medical journals during the two years after the extension found that most trials properly accounted for clustering in the sample size and in the analysis.[7] Nevertheless, that review still found that cluster randomised trials often had suboptimal reporting to the extent that both internal and external validity were uncertain.[7] For example, blinding of participants was not reported clearly in 33% and blinding of outcome assessors was not reported clearly in 38%. Such factors are included in the CONSORT checklist for a good reason: lack of blinding can lead to substantially inflated estimates of effect,[20 21] making clear reporting essential for interpretation. Although blinding of participants is often impossible in cluster randomised trials, especially in those evaluating interventions to change behaviour, we believe that inability to blind trial participants should not be invoked as an excuse for poor reporting.

Given that cluster trials are increasingly common but carry unique risks for bias, adequate reporting is even more important. The CONSORT extension for cluster randomised trials might have had a positive impact on reporting; our group has recently shown that more reports of cluster randomised trials now mention the clustered nature of the trial in the title or abstract, or both.[22] We performed a secondary analysis of data originally abstracted to investigate the unique ethical issues that arise as a result of randomising groups rather than individuals.[23] Using data from a random sample of published cluster randomised trials from 2000-8, we examined trends in the reporting quality of these trials. In addition to investigating whether there was an improvement in reporting of certain items recommended by the CONSORT extension, we assessed whether there were improvements in essential methodological requirements for cluster randomised trials. To do so, we made a distinction between reporting in the manuscript (such as presence of a sample size calculation) and proper methodological conduct (such as accounting for the intracluster correlation in that calculation). Finally, we examined whether trends in trial reporting and methods varied according to characteristics of the study or journal.

## Methods

### Search strategy and article selection

We used a previously published electronic search strategy (box 2) to identify reports of cluster randomised trials in health research, published in English language journals from 2000 to 2008.[22] As described in more detail elsewhere,[22 23] the search strategy was derived and validated with an ideal set of cluster randomised trials identified from manual examination of a large sample of health journals, as well as an independent sample of cluster randomised trials included in previously published reviews. The sensitivity of the search strategy against the ideal set of trials, defined as the proportion of cluster randomised trials that are retrieved by the search, was 90.1%.

Reports identified by the search strategy were sorted in random order with a computer generated random number, and two reviewers (MT and CB) screened titles and abstracts of reports (as well as full text when necessary) to identify cluster randomised trials that met our eligibility criteria. Both reviewers initially screened reports to assess agreement in the identification of eligible trials. After reaching satisfactory agreement (defined as $\kappa \geq 0.85$), reports were screened independently until the target sample size of 300 trials was reached (fig 1). An article was included if it was clearly the main report of a cluster randomised trial. We excluded studies identified by the trial authors as "pilot" or "feasibility" studies, trial protocols, trials randomising households or dyads of different individuals, short communications or conference proceedings, trials with further allocation of individuals within clusters, those using quasi-randomised designs, and studies that reported only baseline findings or secondary analyses of trials. We considered a study as a "secondary analysis" if it was identified as such by the study authors, referenced the main publication elsewhere, or presented only secondary outcomes.

### Data abstraction

The research team developed and pilot tested the data abstraction instrument used for the larger project examining ethical issues in cluster randomised trials. This was then applied to a sample of 21 cluster randomised trials to calibrate reviewers. Six reviewers (MT, AMcR, CB, SD, JT, ZS) independently abstracted these 21 trials. Differences were identified and resolved by discussion. The rest of the trials were then abstracted independently by rotating pairs of reviewers. After each set of 20 trials had been abstracted, discrepancies were reviewed within the pair and resolved by consensus. If differences could not be resolved, one reviewer (MT) was the arbitrator.

### Outcomes

Similar to the approach taken in a previous review,[8] we considered criteria relating to both reporting and methodological quality of the trials. Reporting quality was assessed on the basis of the presence or absence of a subset of criteria in the CONSORT extension to cluster randomised trials; methodological quality was assessed on the basis of four methodological requirements specific to the conduct of such trials. These are related to the criteria described in the CONSORT extension checklist, but they have been abstracted to assess appropriateness of trial conduct rather than simply trial reporting. Table 1 compares the criteria in the CONSORT checklist and the variables assessed as outcomes in the present study.

### Reporting criteria

Although there are 22 items listed in the CONSORT reporting checklist for cluster randomised trials, some are difficult to abstract in a standardised fashion and others can be broken down into multiple variables. As part of the larger project investigating ethical issues in cluster randomised trials, we abstracted 14 CONSORT related reporting variables. The choice not to abstract all CONSORT criteria represented a compromise

**Box 1 Brief description of cluster randomised trials**

- Cluster randomised trials differ from classic (individual level) randomised controlled trials in that the unit of randomisation includes a group (or cluster) of patients—such as a medical practice, hospital, or entire community—rather than an individual patient
- Cluster randomised trials are often done for pragmatic purposes (such as in public health trials where the intervention is directed at the whole community) or to avoid contamination of the treatment arm (such as in health services trials where patients in the intervention group share a healthcare provider)
- Individuals nested within a cluster might be more similar than individuals from other clusters; this "intracluster correlation" must be accounted for in the design and analysis
- Failing to account for the intracluster correlation in the calculation of the sample size can lead to an underpowered trial, and failing to account for it during the analysis can lead to spuriously significant results
- Numerous other challenges separate cluster randomised trials from individual level trials (for example, loss to follow-up of clusters can substantially reduce power compared with loss of individuals)

**Box 2: Medline search strategy to identify cluster randomised trials**

1. randomized controlled trial.pt.
2. animals/
3. humans/
4. 2 NOT (2 AND 3)
5. 1 NOT 4
6. cluster$ adj2 randomi$.tw.
7. ((communit$ adj2 intervention$) OR (communit$ adj2 randomi$)).tw.
8. group$ randomi$.tw.
9. 6 OR 7 OR 8
10. intervention?.tw.
11. cluster analysis/
12. health promotion/
13. program evaluation/
14. health education/
15. 10 OR 11 OR 12 OR 13 OR 14
16. 9 OR 15
17. 16 AND 5

between comprehensiveness and feasibility, given the large sample size involved. For every trial report, we classed each of the following criteria as "reported" or "not reported":

- Clear identification of cluster randomised in the title or abstract of the report
- Explicit provision of a rationale or justification for using a clustered design (such as avoidance of contamination)
- Reporting of clearly defined primary outcome measures
- Presentation of calculation of sample size
- Identification of who enrolled participants in the trial (excluding trials with no enrolment of participants—for example, trials using data from secondary sources only)
- Reporting of the blinding of participants
- Reporting of the blinding of administrators or outcome assessors, or both
- Presentation of a clearly defined approach to analysis
- Reporting of the number of clusters randomised to each arm
- Reporting of the number of clusters that withdrew
- Reporting of the number of clusters that were lost to follow-up

- Reporting of the size of clusters in each arm
- Reporting of the number of individuals lost to follow-up
- Reporting of an estimated intracluster correlation (excluding trials using a pair matched design or those where the analysis was at the cluster level).

## Methodological criteria

We abstracted four criteria related to the appropriate conduct of a cluster randomised trial:

- Whether or not the sample size calculation (if reported) accounted for clustering.[24] A trial was classified as meeting the sample size requirement if the sample size calculation was presented and clearly accounted for clustering (such as by using the intracluster correlation, coefficient of variation, or cluster level summary statistics).
- Whether or not the analysis accounted for clustering.[24] A trial was classified as meeting the analysis requirement if the method of analysis was reported and was clearly appropriate for the clustered design (such as by adjusting for the intracluster correlation, using a mixed effects regression analysis, or using cluster level summary statistics).

- Whether any attempt was made beyond simple (unrestricted) randomisation to attain balance at baseline—cluster randomised trials have a greater risk of chance imbalances at baseline compared with trials randomising individual patients because of the limited number of clusters that can feasibly be randomised in any one trial. Restricted randomisation (using stratification, pair matching, or minimisation) to limit the chance of baseline imbalances is therefore recommended.[19]

- As in a previous review,[8] we abstracted whether the number of clusters randomised per arm was greater than four as trials randomising fewer than four clusters per arm might be severely limited in their statistical power.[25] Unlike each of the variables above, this criterion was not explicitly recommended in the CONSORT extension for cluster trials.

## Study and journal characteristics

We assessed the study setting as well as the type of journal and its impact factor for each manuscript. To distinguish between trials conducted in clinical settings or non-clinical settings we assessed the unit of allocation. We classified the study setting as "clinical" when the unit of allocation was a healthcare provider, teams of healthcare providers, or healthcare organisations (such as a primary care practice or group of practices, hospital or hospital wards, nursing home) or if the trial was conducted in a healthcare organisation; the remainder of the trials (such as those randomising schools or classrooms; residential areas; worksites; and sports teams, clubs, churches, or other social groups) were classified as "non-clinical." We obtained journal impact factors from journal citation reports (ISI Web of Science, 2009). When a journal's ranking was unavailable, we used the impact ranking of the open access SMImago journal and country rank database, if available.[26] This ranking is calculated with a similar formula and is strongly correlated with the journal citation impact factor.[27] We used journal citation reports from ISI Web of Knowledge to identify general medical journals according to those classified as "medicine, general and internal."

## Analysis

Results were summarised with frequencies and percentages for categorical variables and medians and interquartile ranges for continuous or ordinal variables. For our primary objective of determining whether there has been an improvement in reporting and methodological quality over time, we compared the proportions of manuscripts meeting the recommended criteria published before CONSORT (2000-4) with those published after CONSORT (2005-6 and 2007-8) using Cochran-Armitage tests for trend. As seen in figure 2⇊, the number of citations of the extension increased linearly over this timeframe, and we therefore chose two cut-off points to account for the expected gradual dissemination of the guidelines over time. To quantify the magnitude of change in adherence over time, we calculated the absolute change in the proportion of trials meeting each recommendation from before to after CONSORT, together with 95% asymptotic confidence intervals (or exact confidence intervals in the case of small expected frequencies). We also created a summary score for the 14 reporting criteria, representing the proportion of items adhered to in the report. Differences in the summary score over time were analysed with one way analysis of variation (ANOVA) with publication year as a three level categorical variable. For our secondary objective (to investigate variations in adherence to reporting and methodological criteria according to study or journal

characteristics) we repeated the above analyses after classifying trial setting as clinical versus non-clinical and after classifying journals as higher (above the median) versus lower impact factors and as general medical journals versus other. All analyses were carried out with SAS v.9.2 with a level of significance set at $\alpha=0.05$.

## Results

Table 2 shows characteristics of the 300 randomly selected trials included in the review⇊. Trials were published in 150 different journals; 103 (34%) were published in general medical journals. Journal impact factors ranged from 0.45 to 50; the median impact factor was 2.9. The median impact factor of the general medical journals was 9.2.

Table 3⇊ shows the percentage of trials that met each of the reporting criteria before and after CONSORT, together with the tests for trend and confidence intervals for absolute change in adherence. Five of the 14 reporting criteria showed a significant trend for improvement: reporting on loss of clusters to follow-up (P=0.01); identification as "cluster randomised" in title or abstract (P=0.038); providing a justification for the clustered design (P=0.038); reporting whether or not outcome assessors had been blinded (P=0.019); and reporting of the number of clusters randomised (P=0.035). Among these criteria, the absolute improvement in adherence ranged from 6.6% (95% confidence interval −1.1% to 14.3%) for reporting of number of clusters randomised to 13.9% (3.1% to 24.7%) for reporting whether outcome assessors were blinded. Notably, there was no improvement in the proportion of trials clearly identifying a primary outcome, with fewer than half of trials overall meeting this important criterion; moreover, about half of trials overall failed to report a sample size calculation and this does not seem to have improved over time. Based on the summary score, there was minimal improvement in overall reporting, with papers from 2000-4 reporting a mean of 60% of criteria, those from 2005-6 reporting 62%, and those from 2007-8 reporting 66% (P=0.09 for trend).

We found no trend over time in the methodological criteria that we chose to abstract. Overall, 56% of trials used restricted randomisation, 70% accounted for clustering in analysis, 60% of those presenting sample size calculations accounted for clustering in the design, and 86% allocated more than four clusters per arm.

Tables 4, 5, and 6 show the results of our secondary analyses, which examined the role of study and publication characteristics on improvement in quality of reporting and methodological conduct before and after the publication of the CONSORT extension for cluster randomised trials. In particular, table 4 shows that general medical journals performed at a higher standard overall for nearly every criterion with the exception of justification for clustering⇊. When considering the change from before to after CONSORT, the general medical journals showed significant improvement in only one criterion: reporting of number of clusters lost to follow-up (absolute improvement 17%, 3% to 31%). No significant improvements in any of the methodological criteria were observed in trials published in either the general medical or other journals. Based on the summary score, trials published in general medical journals reported a mean of 66.5% of criteria before and a mean of 71.6% of criteria after the CONSORT extension was published (absolute improvement 5%, −1.7% to 12%), while trials published in other journals reported a mean of 56% before and 60% after (absolute improvement 4.3%, −1.1% to 9.7%).

Table 5⇓ shows similar findings: higher impact journals tended to score better in most reporting and methodological criteria. Higher impact factor journals showed significant improvements in two of 14 reporting criteria and in one of four methodological criteria, while lower impact factor journals improved in two reporting criteria. Based on the summary score, trials published in higher impact factor journals reported a mean of 66.0% of criteria before and a mean of 68.3% of criteria after the CONSORT extension was published (absolute improvement 2.2%, −3.6% to 8.3%), while trials published in lower impact factor journals reported a mean of 54.4% before and 59.2% after (absolute improvement 4.8%, −1.4% to 11.0%).

The results in table 6⇓ follow the same pattern: trials conducted in clinical settings tended to meet more of the reporting and methodological criteria than trials conducted in non-clinical settings. Trials from clinical settings showed significant improvements in one reporting criterion and one methodological criterion, while the trials from non-clinical settings did not show significant improvements in any of the criteria. Based on the summary score, trials conducted in clinical settings reported a mean of 63.3% of criteria before and a mean of 67.8% of criteria after the CONSORT extension was published (absolute improvement 4.6%, −1.1% to 10.3%), while trials published in non-clinical settings reported a mean of 55.3% before and 58.2% after (absolute improvement 2.9%, −3.7% to 9.4%).

## Discussion

### Statement of principal findings

There have been significant improvements over time in only five of 14 CONSORT related reporting criteria for cluster randomised trials. In particular, reporting whether outcome assessments were blind, as well as the number of clusters randomised, withdrawn, and lost to follow-up per arm has considerable implications for assessment of internal validity.[28] Although the absolute improvements were small, the trends represent an important accomplishment because through improved clarity of reporting, readers can make better judgments regarding the risks of bias for any given study.[1] While this progress is welcome, the pace remains slow, and there remains considerable room for further improvement. The findings from our review suggest that future updates of the CONSORT extension for cluster trials should be accompanied by additional interventions to help investigators and editors improve the standards of these trials.

Of the five reporting criteria that had a greater than 75% adherence rate, four can easily be summarised in a well prepared cluster-patient flow diagram or in a table summarising baseline data for cluster and individual level variables, or both. Moreover, of the five criteria that showed significant improvements over time, two related to flow of clusters-patients (reporting of numbers of clusters randomised and lost to follow-up). Therefore, encouraging the use of such flow diagrams (as recommended by the CONSORT extension) seems to have been successful in improving reporting of enrolment and losses of patients and clusters. Even for trials conducted in clinical settings and published in journals with higher impact factors, however, there remains a great need for further improvements in other reporting criteria, possibly through the development of similar devices within manuscripts that facilitate the communication of key information. Studies conducted in clinical rather than non-clinical settings and published in general medical and other higher impact journals were more likely to meet the criteria for both reporting and methodological conduct, but the standards of these studies failed to improve significantly in

many areas after the CONSORT extension was published. More stringent editorial policies might be required to bring about substantial improvement. Although journal editorial policies that promote CONSORT are associated with improved reporting,[29] a recent survey of 165 high impact journals found that only 3% refer to the extension to cluster trials in the online instructions for authors.[30]

Some CONSORT related variables might be poorly reported because they are deemed unnecessary by investigators or editors. For example, justification for conducting a cluster level randomisation might be considered unnecessary in cases when the intervention itself is at the level of the cluster. In a post hoc analysis in which we excluded trials with interventions solely at the cluster level (n=99), however, the proportion reporting a justification remained low (33%). We agree with the authors of CONSORT that explicit justification is important because cluster randomised trials involve unique methodological challenges that require special attention during the design and analysis.[19] For example, the intracluster correlation must be accounted for in both the sample size calculation and analysis; failure to clearly report whether this has been done leads to questions regarding the validity of the findings. Unfortunately, we found no evidence of significant improvement over time in four key methodological criteria. Indeed, the methodological quality of cluster randomised trial reports after publication of the CONSORT extension remains disappointingly poor; this is especially true for trials published in specialty (non-general) medicine journals.

### Comparison with literature

Many previous authors have noted problems with the design, analysis, and reporting of cluster randomised trials. For example, the reporting of an intracluster correlation represents an important contribution to the literature, allowing future studies to plan for adequate power. Only 18% of the 300 manuscripts that we reviewed, however, reported an intracluster correlation (or 16% after we excluded trials with a pair matched design or those with analysis at the cluster level). This is higher than previous estimates of 4%[8] and 8%.[5] Our finding that only 31% of manuscripts provided a clear justification for using a cluster randomised design also seems positive in light of the previous finding from a review of 152 cluster randomised trials published from 1997 to 2000, which found that only 14% justified the use of a clustered design.[8] In contrast, we found that only 50% reported clearly whether administrators or participants were blinded, and only 38% reported whether outcome assessors were blinded, whereas a previous review of 34 cluster randomised trials in primary care published in 2004 or 2005 found that 67% and 62% clearly reported whether participants or outcome assessors were blinded, respectively.[7] Nevertheless, our findings indicate that the reporting of blinding of outcome assessment could be improving over time. This is encouraging as blinding plays an essential role in any assessment of internal validity or risk of bias, or both. Intriguingly, reporting of blinding might be superior in cluster randomised trials than in individually randomised trials; a recent review of 144 trials from 55 high impact factor (median 7.67) journals found that only 25% adequately reported blinding.[31]

About 40% of the trials in our review that reported a sample size calculation failed to account for clustering, while 30% failed to account for clustering in the analysis. Improvement over time in these crucial aspects was suggested by a previous review that assessed 18 trials[12] published from 1983 to 2003 and was also observed in an (unpublished) overview of methodological reviews that assessed the quality of cluster randomised trials published from 1973 to 2008.[32] Still, inappropriate analytical

methods that inflate the risk of type 1 and type 2 error remain common in cluster randomised trials in various clinical specialties.[5-8][13-15]

We found that only 86% of trials allocated a minimum of four clusters to each arm of the study. Past reviews of 152 cluster randomised trials in primary care and 60 trials in public health found that 91%[8] and 92%[6] of trials, respectively, met this criterion, but a review of 75 cluster randomised trials in cancer care found that only 67% had sufficient numbers of clusters per arm.[5] This might reflect the challenge of recruiting clusters rather than individuals, but the ability to make valid inferences regarding intervention effects in studies with few clusters is severely compromised. When further recruitment is impossible, investigators should consider other design options.[33] The reviews focusing on public health and on cancer trials each found 5% of studies allocated only one cluster per arm. In our study, 10 trials (3%) had only one cluster per arm. Investigators should recognise that in trials with only one cluster per arm, the intervention effect is completely confounded by the cluster effect.

Low numbers of clusters increase the likelihood of cluster level differences between study arms. Furthermore, clustering itself can make it difficult to achieve balance across arms for individual level variables when only simple randomisation is used to allocate clusters.[34] In our review, 41% of the trials used simple randomisation rather than an allocation technique more likely to achieve balance at baseline. Previous reviews found that trials used simple randomisation rather than stratification or matching techniques 40%[5] to 46%[8] of the time. The CONSORT extension for cluster randomised trials warns against the risk of baseline imbalance from use of simple randomisation,[19] and, although matching could overcome this problem, it leads to difficulties in estimating intracluster correlations and can complicate analyses.[10] Although we did not actually assess for imbalance at baseline, one previous review found that three of 36 cluster randomised trials published in the *BMJ*, *Lancet*, and *New England Journal of Medicine* from 1997 to 2002 had evidence for cluster imbalance, while in 15 the adequacy of balance was deemed unclear.[13] This indicates the need for investigators (and editors) to strongly consider the value of allocation techniques other than simple randomisation to achieve baseline balance in cluster randomised trials.[35]

## Strengths and weaknesses of the study

Any study that does not consider the entire population could be susceptible to selection bias; it is plausible that different conclusions would be found by looking at a different sample. We mitigated this risk in our study by abstracting what is (to our knowledge) the largest ever sample of cluster randomised trials for a methodological review and the first to use a search strategy that aimed to produce a sample of cluster trials representative of all Medline publications. This search strategy had a sensitivity of 90%, meaning that the random sample used in this study was representative of most cluster randomised trials in health research published in Medline from 2000 to 2008. If the 10% of cluster randomised trials not identified by our search strategy were systematically different with respect to reporting or methodological quality, however, this could bias our results. In particular, as trials that are not clearly identified as "cluster randomised" in titles and abstracts of reports might also be less likely to adhere to recommended methodological and reporting standards, our results might overestimate the proportion of trials adhering to these standards. Although the sample was large relative to previous methodological reviews, it was determined by the objectives of the larger study focusing on ethical issues

in cluster trials and we were therefore not specifically powered to detect small improvements in reporting. In particular, assuming the most conservative estimate of adherence of 50% before CONSORT, improvements to 60% and 70% adherence after CONSORT (2005-6 and 2007-8, respectively) would have been required to allow 80% probability of detecting a significant trend. Naturally, similar limitations extend to the secondary analyses involving subgroups defined by study or journal characteristics.

We used a unique summary score for adherence to 14 CONSORT based reporting variables relevant to cluster randomised trials. We believe this is helpful because it provides a sense of overall quality of reporting, but we do not intend to suggest by providing this score that each criterion is of equal importance. Although similar summary scores have been used before in other reviews of reporting standards,[4] readers should note this caveat in addition to the other known limitations of quality scales.[36] Furthermore, we acknowledge the risk of spurious findings associated with multiple testing in our analyses and recommend that significant results be interpreted cautiously. Finally, there are numerous criteria that we did not abstract in this review that could have important implications with respect to internal or external validity, such as baseline imbalances for clusters and participants, as well as risks of identification and recruitment bias. The extension to CONSORT for cluster randomised trials might have resulted in improvements in reporting for criteria other than those that we abstracted. It is also likely that the extension had a greater impact in journals that actively endorsed (and explicitly enforced) the guideline. Unfortunately, we were unable to evaluate this because of the difficulty in determining when CONSORT extensions were endorsed by each journal or how they were enforced.

## Implications

Although we observed some improvement over time related to the publication of the CONSORT extension, we found an ongoing need for attention to the proper reporting and methodological conduct in cluster randomised trials. As the extension to CONSORT for cluster randomised trials was published in a general medical journal, investigators conducting trials in non-clinical settings or publishing in other journal categories (such as public health journals) might not have been aware of the extension, thus leading to the relatively poorer performance observed in this group. The most recent CONSORT guidelines for individual level randomised controlled trials were published in a wide range of clinical journals. Future updates for the CONSORT extension to cluster trials should consider a similar (or even broader) approach.

The slow uptake of guidelines into practice is a problem that extends beyond the clinic and into the realm of publication. It seems likely that more than the publication of the CONSORT guideline is required to assist editors and investigators in proper conduct and reporting of cluster randomised trials. Improved collaboration with guideline developers and new tools to support the efforts of editors and investigators should be made available to support the publication of more transparent and higher quality cluster randomised trial manuscripts.

> **What is already known on this topic**
>
> Previous methodological reviews have noted poor reporting of cluster randomised trials and have also found that the methods used often fail to properly account for the clustered nature of the data
>
> An extension to the consolidated standards of reporting trials (CONSORT) specific for cluster randomised trials was published in 2004
>
> **What this study adds**
>
> Reporting and methodological conduct of cluster randomised trials remains suboptimal
>
> The extension to CONSORT for cluster randomised trials led to some improvements in reporting but not in methodological conduct
>
> More than guidelines/checklists are needed to support investigators and editors in publishing cluster randomised trials that provide adequate information to properly assess internal and external validity

1    Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med* 2010;8:24.
2    Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:e1-37.
3    Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010;340:c723.
4    Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006;185:263-7.
5    Murray DM, Pals SL, Blitstein JL, Alfano CM, Lehman J. Design and analysis of group-randomized trials in cancer: a review of current practices. *J Natl Cancer Inst* 2008;100:483-91.
6    Varnell SP, Murray DM, Janega JB, Blitstein JL. Design and analysis of group-randomized trials: a review of recent practices. *Am J Public Health* 2004;94:393-9.
7    Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ* 2008;336:876-80.
8    Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials* 2004;1:80-90.
9    Hahn S, Puffer S, Torgerson DJ, Watson J. Methodological bias in cluster randomised trials. *BMC Med Res Methodol* 2005;5:10.
10   Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health* 2004;94:416-22.
11   Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979-1989. *Int J Epidemiol* 1990;19:795-800.
12   Bland JM. Cluster randomised trials in the medical literature: two bibliometric surveys. *BMC Med Res Methodol* 2004;4:21.
13   Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ* 2003;327:785-9.
14   Handlos LN, Chakraborty H, Sen PK. Evaluation of cluster-randomized trials on maternal and child health research in developing countries. *Trop Med Int Health* 2009;14:947-56.
15   Bowater RJ, Abdelmalik SM, Lilford RJ. The methodological quality of cluster randomised controlled trials for managing tropical parasitic disease: a review of trials published from 1998 to 2007. *Trans R Soc Trop Med Hyg* 2009;103:429-36.
16   Isaakidis P, Ioannidis JP. Evaluation of cluster randomized controlled trials in sub-Saharan Africa. *Am J Epidemiol* 2003;158:921-6.
17   Simpson JM, Klar N, Donnor A. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993. *Am J Public Health* 1995;85:1378-83.
18   Chuang JH, Hripcsak G, Jenders RA. Considering clustering: a methodological review of clinical decision support system studies. *Proc AMIA Symp* 2000:146-50.
19   Campbell MK, Elbourne DR, Altman DG, CONSORT group. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328:702-8.
20   Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358-61.
21   Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
22   Taljaard M, McGowan J, Grimshaw JM, Brehaut JC, McRae A, Eccles MP, et al. Electronic search strategies to identify reports of cluster randomized trials in MEDLINE: low precision will improve with adherence to reporting standards. *BMC Med Res Methodol* 2010;10:15.
23   Taljaard M, McRae AD, Weijer C, Bennett C, Dixon S, Taleban J, et al. Inadequate reporting of research ethics review and informed consent in cluster randomised trials: review of random sample of published trials. *BMJ* 2011;342:d2496.
24   Donner A, Klar N. Design and analysis of cluster randomization trials in health research. Oxford University Press, 2000.
25   Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol* 1994;140:279-301.
26   SJR—SCImago Journal and Country Rank. Science analysis. 2011. www.scimagojr.com.
27   Falagas ME, Kouranos VD, Arencibia-Jorge R, Karageorgopoulos DE. Comparison of SCImago journal rank indicator with journal impact factor. *FASEB J* 2008;22:2623-8.
28   Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions. Version 5.0.2. Cochrane Collaboration, 2009.
29   Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the consolidated standards of reporting trials (CONSORT) checklist. *Control Clin Trials* 2002;23:380-8.
30   Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT statement by high impact factor medical journals: a survey of journal editors and journal "instructions to authors." *Trials* 2008;9:20.
31   Reveiz L, Cortes-Jofre M, Asenjo Lobos C, Nicita G, Ciapponi A, Garcia-Dieguez M, et al. Influence of trial registration on reporting quality of randomized trials: study from highest ranked journals. *J Clin Epidemiol* 2010;63:1216-22.
32   Rotondi R. Selected topics in the meta-analysis of cluster randomized trials: aspects of evidence-synthesis and experimental design [PhD thesis]. University of Western Ontario, 2010.
33   Shadish WR. Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin Company, 2002.
34   Glynn RJ, Brookhart MA, Stedman M, Avorn J, Solomon DH. Design of cluster-randomized trials of quality improvement interventions aimed at medical care providers. *Med Care* 2007;45(10 suppl 2):S38-43.
35   Raab GM, Butcher I. Balance in cluster randomized trials. *Stat Med* 2001;20:351-65.
36   Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;16:62-73.

Cite this as: *BMJ* 2011;343:d5886

# Tables

**Table 1| Comparison of recommendations in extension to CONSORT for cluster randomised trials and variables abstracted for present study**

| CONSORT criterion | Criterion included in present study? | Description of variable abstracted for this review |
|---|---|---|
| 1) Specify that allocation was based on clusters | Yes (reporting) | Clearly identified as cluster randomised trial in title or abstract |
| 2) Rationale for using cluster design | Yes (reporting) | Justification provided for using clustered design |
| 3) Eligibility criteria for participants and clusters | No | — |
| 5) Interventions intended for individual level, cluster level, or both | No | — |
| 5) Specific objectives for individual level, cluster level, or both | No | — |
| 6) Report outcome measures for individual level, cluster level, or both | Yes (reporting) | Primary outcome identified clearly |
| 7) How total sample size was determined including method of calculation, No of clusters, cluster size, coefficient of intracluster correlation | Yes (reporting) | Sample size calculation presented |
| | Yes (methodology) | Accounted for clustering in sample size |
| 8) Method used to generate random allocation sequence | Yes (methodology) | Used stratification/matching/minimisation |
| 9) Method used to implement random allocation sequence | No | — |
| 10) Who generated allocation sequence and enrolled and assigned participants | Yes (reporting) | Identified who enrolled patients |
| 11) Whether participants, those administering interventions, and those assessing outcomes were blinded to group assignment | Yes (reporting) | Reported on blinding of outcome assessors; reported on blinding of participants/administrators |
| 12) Statistical methods used to compare groups for primary outcome(s) indicating how clustering was taken into account | Yes (reporting) | Reported methods of analysis |
| | Yes (methodology) | Accounted for clustering in analysis |
| 13) Flow of clusters and individual participants through each stage | Yes (reporting) | Reported No of clusters randomised; reported No of individuals lost to follow-up; reported No of clusters lost to follow-up; reported No of clusters withdrew; reported size of clusters in each arm |
| 14) Dates defining periods of recruitment and follow up | No | — |
| 15) Baseline information for each group for individual and cluster levels | No | — |
| 16) No of clusters and participants in each group included in each analysis and whether analysis was by intention to treat | Yes (reporting) | Similar to 13 above |
| | Yes (methodology) | Allocated minimum of four clusters per arm |
| 17) For each outcome, summary of results for each group for individual or cluster level, and coefficient of intracluster correlation. | Yes (reporting) | Reported intracluster correlation |
| 18) Address multiplicity by reporting any other analyses performed | No | — |
| 19) All important adverse events or side effects in each intervention group | No | — |
| 20) Interpretation of results (internal validity) | No | — |
| 21) Generalisability (external validity) | No | — |
| 22) Interpretation in context of current evidence | No | — |

**Table 2| Characteristics of 300 cluster randomised trials included in review of studies for compliance with CONSORT extension. Figures are numbers (percentage) of trials unless stated otherwise**

| Characteristic | Data |
|---|---|
| Publication year: | |
| 2000-4 | 139 (46) |
| 2005-6 | 93 (31) |
| 2007-8 | 68 (23) |
| Journal impact factor (n=294): | |
| Median (IQR) | 2.9 (2.1-5.1) |
| Range | 0.45-50.0 |
| Country of study recruitment: | |
| USA | 114 (38) |
| UK or Ireland | 50 (17) |
| Canada | 16 (5) |
| Australia or New Zealand | 16 (5) |
| Other | 104 (35) |
| Clinical setting (unit of allocation): | |
| Medical practices or clinics | 81 (27) |
| Individual health professionals | 41 (14) |
| Hospitals, hospital units, hospital wards | 25 (8) |
| Nursing homes or wards | 16 (5) |
| Other (such as postal codes of family practices) | 6 (2) |
| Non-clinical setting (unit of allocation): | |
| Schools or classrooms | 66 (22) |
| Residential areas (such as villages, districts, housing units) | 39 (13) |
| Worksites | 16 (5) |
| Sports teams, clubs, churches, other social groups | 10 (3) |
| No of clusters randomised (n=285): | |
| Median (IQR) | 21.0 (12-52) |
| Range | 2-605 |
| Average cluster size (n=271): | |
| Median (IQR) | 33.9 (12.5-88.5) |
| Range | 1.7-122 855 |
| No of participants per arm (n=290): | |
| Median (IQR) | 329 (143-866) |
| Range | 20-614 275 |

IQR=interquartile range.

**Table 3| Adherence (number (percentage)) to standard criteria for reporting and methodology for cluster randomised trials, overall and before and after publication of CONSORT extension for cluster randomised trials**

| | Overall (n=300) | Before (2000-4, n=139) | After (2005-6, n=93) | After (2007-8, n=68) | P value for trend | Absolute change in % adherence (before to after) (95% CI) |
|---|---|---|---|---|---|---|
| **Criteria related to quality of reporting** | | | | | | |
| Clearly identified as clustered in title or abstract | 145 (48) | 59 (43) | 47 (50.5) | 39 (57) | 0.038 | 11.0 (–0.3 to 22.2) |
| Justification provided for using cluster design | 94 (31) | 37 (27) | 29 (31.2) | 28 (41) | 0.038 | 8.8 (–1.6 to 19.2) |
| Reported on blinding of outcome assessors | 113 (38) | 42 (30) | 40 (43.0) | 31 (46) | 0.019 | 13.9 (3.1 to 24.7) |
| Reported on blinding of participants/administrators | 151 (50) | 72 (52) | 50 (53.8) | 29 (43) | 0.292 | –2.7 (–14.1 to 8.6) |
| Primary outcome identified clearly | 141 (47) | 62 (45) | 43 (46.2) | 36 (53) | 0.284 | 4.5 (–6.9 to 15.8) |
| Sample size calculation presented | 164 (55) | 77 (55) | 44 (47.3) | 43 (63) | 0.483 | –1.4 (–12.7 to 9.9) |
| Identified who enrolled participants* | 134 (58) | 59 (60) | 41 (55.4) | 34 (57) | 0.674 | –3.6 (–16.4 to 9.2) |
| Reported No of clusters randomised | 261 (87) | 116 (84) | 81 (87.1) | 64 (94) | 0.035 | 6.6 (–1.1 to 14.3) |
| Reported No of clusters lost to follow-up | 235 (78) | 99 (71) | 78 (83.9) | 58 (85) | 0.010 | 13.3 (3.9 to 22.6) |
| Reported No of clusters that withdrew | 256 (85) | 115 (83) | 78 (83.9) | 63 (93) | 0.078 | 4.8 (–3.3 to 12.9) |
| Reported size of clusters in each arm | 262 (87) | 117 (84) | 82 (88.2) | 63 (93) | 0.081 | 5.9 (–1.7 to 13.5) |
| Reported No of individuals lost to follow-up | 228 (76) | 109 (78) | 65 (69.9) | 54 (79) | 0.860 | –4.5 (–14.1 to 5.1) |
| Reported methods of analysis | 281 (94) | 128 (92) | 89 (95.7) | 64 (94) | 0.456 | 2.9 (–2.7 to 8.6) |
| Reported intracluster correlation† | 35 (16) | 21 (22) | 4 (6.2) | 10 (18) | 0.323 | –10.2 (–20.3 to 0) |
| Mean (SD) summary score | 61.8 (19.2) | 59.9 (19.6) | 61.5 (19.9) | 66.1 (16.9) | 0.092 | 3.5 (–0.8 to 7.8) |
| **Criteria relating to methodological quality** | | | | | | |
| Used restricted randomisation | 167 (56) | 72 (52) | 55 (59.1) | 40 (59) | 0.272 | 7.2 (–4.1 to 18.5) |
| Allocated minimum of four clusters per arm‡ | 244 (86) | 111 (85) | 76 (85.4) | 57 (86) | 0.867 | 0.4 (–7.8 to 8.6) |
| Accounted for clustering in sample size§ | 100 (61) | 51 (66) | 21 (47.7) | 28 (65) | 0.662 | –9.9 (–24.8 to 4.9) |
| Accounted for clustering in analysis | 209 (70) | 100 (72) | 63 (67.7) | 46 (68) | 0.474 | –4.2 (–14.6 to 6.2) |

*Excludes 67 trials with no participant enrolment.

†Excludes 84 trials with pair matched designs or primary analysis at cluster level.

‡Excludes 15 studies with unclear number of clusters.

§Excludes 136 trials with no sample size calculation presented.

**Table 4| Change in adherence (number (percentage)) to standard criteria for reporting and methodology for cluster randomised trials by journal type, before and after publication of CONSORT extension for cluster randomised trials**

| | General medical journals | | | Other journals | | |
|---|---|---|---|---|---|---|
| | Before (2000-4, n=55) | After (2005-8, n=48) | Absolute change in adherence (before to after) (95% CI) | Before (2000-4, n=84) | After (2005-8, n=113) | Absolute change in adherence (before to after) (95% CI) |
| **Criteria related to quality of reporting** | | | | | | |
| Clearly identified as clustered in title or abstract | 29 (53) | 33 (69) | 16.0 (−2.6 to 34.6) | 30 (36) | 53 (47) | 11.2 (−2.6 to 25.0) |
| Justification provided for using cluster design | 13 (24) | 16 (33) | 9.7 (−7.7 to 27.1) | 24 (29) | 41 (36) | 7.7 (−5.4 to 20.8) |
| Reported on blinding of outcome assessors | 22 (40) | 27 (56) | 16.3 (−2.8 to 35.3) | 20 (24) | 44 (39) | 15.1 (2.3 to 27.9) |
| Reported on blinding of participants/administrators | 34 (62) | 30 (63) | 0.7 (−18.1 to 19.5) | 38 (45) | 49 (43) | −1.9 (−15.9 to 12.2) |
| Primary outcome identified clearly | 33 (60) | 34 (71) | 10.8 (−7.4 to 29.1) | 29 (35) | 45 (40) | 5.3 (−8.3 to 18.9) |
| Sample size calculation presented | 36 (66) | 34 (71) | 5.4 (−12.6 to 23.4) | 41 (49) | 53 (47) | −1.9 (−16.0 to 12.2) |
| Identified who enrolled participants* | 26 (74) | 24 (65) | −9.4 (−30.6 to 11.7) | 33 (52) | 51 (53) | 1.0 (−14.8 to 16.8) |
| Reported No of clusters randomised | 47 (86) | 44 (93) | 6.2 (−6.0 to 18.4) | 69 (82) | 101 (89) | 7.2 (−2.7 to 17.2) |
| Reported No of clusters lost to follow-up | 41 (75) | 44 (93) | 17.1 (3.2 to 31.0) | 58 (69) | 92 (81) | 12.4 (0.2 to 24.6) |
| Reported No of clusters that withdrew | 47 (86) | 44 (93) | 6.2 (−6.0 to 18.4) | 68 (81) | 97 (86) | 4.9 (−5.7 to 15.5) |
| Reported size of clusters in each arm | 48 (87) | 44 (93) | 4.4 (−7.4 to 16.2) | 69 (82) | 101 (89) | 7.2 (−2.7 to 17.2) |
| Reported No of individuals lost to follow-up | 47 (86) | 38 (79) | −6.3 (−21.2 to 8.5) | 62 (74) | 81 (72) | −2.1 (−14.7 to 10.4) |
| Reported methods of analysis | 51 (93) | 47 (98) | 5.2 (−2.8 to 13.2) | 77 (92) | 106 (94) | 2.1 (−5.3 to 9.5) |
| Reported intracluster correlation† | 10 (30) | 5 (14) | −16.0 (−38.7 to 7.7) | 11 (18) | 9 (11) | −6.9 (−23.0 to 9.5) |
| Mean (SD) summary score | 66.5 (18.1) | 71.6 (16.7) | 5.1 (−1.7 to 12.0) | 55.7 (19.5) | 60.0 (18.6) | 4.3 (−1.1 to 9.7) |
| **Criteria relating to methodological quality** | | | | | | |
| Used restricted randomisation | 36 (66) | 32 (67) | 1.2 (−17.1 to 19.5) | 36 (43) | 63 (56) | 12.9 (−1.1 to 26.9) |
| Allocated minimum of four clusters per arm‡ | 47 (89) | 44 (94) | 4.9 (−6.1 to 16.0) | 64 (83) | 89 (82) | −0.7 (−11.7 to 10.3) |
| Accounted for clustering in sample size§ | 30 (55) | 24 (50) | −4.6 (−23.9 to 14.8) | 21 (25) | 25 (22) | −2.9 (−14.9 to 9.1) |
| Accounted for clustering in analysis | 46 (84) | 42 (88) | 3.9 (−9.7 to 17.4) | 54 (64) | 67 (59) | −5.0 (−18.7 to 8.7) |

*Excludes 67 trials with no participant enrolment.

†Excludes 84 trials with pair matched designs or primary analysis at cluster level.

‡Excludes 15 studies with unclear number of clusters.

§Excludes 136 trials with no sample size calculation presented.

**Table 5| Change in adherence (number (percentage)) to standard criteria for reporting and methodology for cluster randomised trials by journal impact factor\*, before and after publication of CONSORT extension for cluster randomised trials**

| | Higher impact factor journals | | | Lower impact factor journals | | |
|---|---|---|---|---|---|---|
| | Before (2000-4, n=71) | After (2005-8, n=78) | Absolute change in adherence (before to after) (95% CI) | Before (2000-4, n=63) | After (2005-8, n=82) | Absolute change in adherence (before to after) (95% CI) |
| **Criteria related to quality of reporting** | | | | | | |
| Clearly identified as clustered in title or abstract | 40 (56) | 49 (63) | 6.5 (–9.3 to 22.2) | 19 (30) | 37 (45) | 15.0 (–0.7 to 30.6) |
| Justification provided for using cluster design | 18 (25) | 33 (42) | 17.0 (2.0 to 31.9) | 17 (27) | 24 (29) | 2.3 (–12.5 to 17.0) |
| Reported on blinding of outcome assessors | 24 (34) | 43 (55) | 21.3 (5.7 to 36.9) | 18 (29) | 27 (33) | 4.4 (–10.7 to 19.5) |
| Reported on blinding of participants/administrators | 40 (56) | 38 (49) | –7.6 (–23.6 to 8.4) | 30 (48) | 41 (50) | 2.4 (–14.0 to 18.8) |
| Primary outcome identified clearly | 40 (56) | 48 (62) | 5.2 (–10.6 to 21.0) | 20 (32) | 31 (38) | 6.1 (–9.5 to 21.6) |
| Sample size calculation presented | 45 (63) | 52 (67) | 3.3 (–12.0 to 18.6) | 32 (51) | 35 (43) | –8.1 (–24.5 to 8.2) |
| Identified who enrolled participants† | 33 (62) | 37 (58) | –4.5 (–22.3 to 13.3) | 24 (57) | 38 (55) | –2.1 (–21.1 to 17.0) |
| Reported No of clusters randomised | 65 (92) | 70 (90) | –1.8 (–11.1 to 7.5) | 48 (76) | 74 (90) | 14.1 (1.7 to 26.4) |
| Reported No of clusters lost to follow-up | 56 (79) | 67 (86) | 7.0 (–5.2 to 19.3) | 40 (64) | 68 (83) | 19.4 (5.0 to 33.9) |
| Reported No of clusters that withdrew | 64 (90) | 70 (90) | –0.4 (–10.1 to 9.3) | 48 (76) | 70 (85) | 9.2 (–3.8 to 22.2) |
| Reported size of clusters in each arm | 64 (90) | 71 (91) | 0.9 (–8.5 to 10.3) | 49 (78) | 73 (89) | 11.3 (–1.1 to 23.5) |
| Reported No of individuals lost to follow-up | 59 (83) | 61 (78) | –4.9 (–17.5 to 7.8) | 46 (73) | 58 (71) | –2.3 (–17.0 to 12.5) |
| Reported methods of analysis | 66 (93) | 77 (99) | 5.8 (–0.7 to 12.2) | 57 (91) | 76 (93) | 2.2 (–7.0 to 11.4) |
| Reported intracluster correlation‡ | 14 (28) | 6 (10) | –17.7 (–35.6 to 1.3) | 7 (17) | 8 (13) | –3.6 (–23.0 to 16.0) |
| Mean (SD) summary score | 66.0 (18.7) | 68.3 (18.0) | 2.2 (–3.6 to 8.3) | 54.4 (19.1) | 59.2 (18.3) | 4.8 (–1.4 to 11.0) |
| **Criteria relating to methodological quality** | | | | | | |
| Used restricted randomisation | 37 (52) | 45 (8) | 5.6 (–10.4 to 21.6) | 33 (52) | 50 (61) | 8.6 (–7.6 to 24.8) |
| Allocated minimum of four clusters per arm§ | 59 (84) | 71 (96) | 11.7 (2.0 to 21.3) | 48 (86) | 62 (78) | –8.2 (–21.2 to 4.7) |
| Accounted for clustering in sample size¶ | 31 (69) | 34 (65) | –3.5 (–22.2 to 15.2) | 20 (63) | 15 (43) | –19.6 (–43.1 to 3.8) |
| Accounted for clustering in analysis | 58 (82) | 65 (83) | 1.6 (–10.6 to 13.9) | 38 (60) | 44 (54) | –6.7 (–22.9 to 9.5) |

\*Excludes six trials with no impact factor.

†Excludes 67 trials with no participant enrolment.

‡Excludes 84 trials with pair matched designs or primary analysis at cluster level.

§Excludes 15 studies with unclear number of clusters.

¶Excludes 136 trials with no sample size calculation presented.

**Table 6| Change in adherence (number (percentage)) to standard criteria for reporting and methodology for cluster randomised trials by trial setting, before and after publication of CONSORT extension for cluster randomised trials**

| | Clinical settings | | | Non-clinical settings | | |
|---|---|---|---|---|---|---|
| | Before (2000-4, n=81) | After (2005-8, n=88) | Absolute change in adherence (before to after) (95% CI) | Before (2000-4, n=58) | After (2005-8, n=73) | Absolute change in adherence (before to after) (95% CI) |
| **Criteria related to quality of reporting** | | | | | | |
| Clearly identified as clustered in title or abstract | 41 (51) | 54 (61) | 10.8 (–4.2 to 25.7) | 18 (31) | 32 (44) | 12.8 (–3.7 to 29.3) |
| Justification provided for using cluster design | 24 (30) | 33 (38) | 7.9 (–6.3 to 22.1) | 13 (22) | 24 (33) | 10.5 (–4.8 to 25.7) |
| Reported on blinding of outcome assessors | 29 (36) | 44 (50) | 14.2 (–0.6 to 29.0) | 13 (22) | 27 (37) | 14.6 (–0.9 to 30.0) |
| Reported on blinding of participants/administrators | 50 (62) | 52 (59) | –2.6 (–17.4 to 12.1) | 22 (38) | 27 (37) | –0.9 (–17.6 to 15.8) |
| Primary outcome identified clearly | 38 (47) | 51 (58) | 11.0 (–3.9 to 26.0) | 24 (41) | 28 (38) | –3.0 (–19.9 to 13.9) |
| Sample size calculation presented | 58 (72) | 58 (66) | –5.7 (–19.6 to 8.3) | 19 (33) | 29 (40) | 7.0 (–9.5 to 23.5) |
| Identified who enrolled participants* | 33 (69) | 46 (67) | –2.1 (–19.3 to 15.1) | 26 (51) | 29 (45) | –6.4 (–24.7 to 11.9) |
| Reported No of clusters randomised | 66 (82) | 77 (88) | 6.0 (–4.9 to 16.9) | 50 (86) | 68 (93) | 6.9 (–3.7 to 17.5) |
| Reported No of clusters lost to follow-up | 54 (67) | 75 (85) | 18.6 (5.9 to 31.2) | 45 (78) | 61 (84) | 6.0 (–7.7 to 19.7) |
| Reported No of clusters that withdrew | 65 (80) | 75 (85) | 5.0 (–6.4 to 16.4) | 50 (86) | 66 (90) | 4.2 (–7.0 to 15.4) |
| Reported size of clusters in each arm | 71 (88) | 82 (93) | 5.5 (–3.4 to 14.4) | 46 (79) | 63 (86) | 7.0 (–6.1 to 20.1) |
| Reported No of individuals lost to follow-up | 68 (84) | 73 (83) | –1.0 (–12.2 to 10.2) | 41 (71) | 46 (63) | –7.7 (–23.8 to 8.4) |
| Reported methods of analysis | 72 (88.9) | 82 (93) | 4.3 (–4.3 to 12.9) | 56 (97) | 71 (97) | 0.7 (–5.3 to 6.7) |
| Reported intracluster correlation† | 13 (21) | 12 (16) | –5.1 (–21.8 to 11.9) | 8 (23) | 2 (4) | –18.5 (–39.1 to 3.7) |
| Mean (SD) summary score | 63.3 (19.7) | 67.8 (17.7) | 4.6 (–1.1 to 10.3) | 55.3 (18.8) | 58.2 (18.7) | 2.9 (–3.7 to 9.4) |
| **Criteria relating to methodological quality** | | | | | | |
| Used restricted randomisation | 38 (47) | 58 (66) | 19.0 (4.3 to 33.7) | 34 (59) | 37 (51) | –7.9 (–25.0 to 9.2) |
| Allocated minimum of four clusters per arm | 63 (84) | 72 (87) | 2.8 (–8.3 to 13.8) | 48 (87) | 61 (85) | –2.6 (–14.7 to 9.6) |
| Accounted for clustering in sample size‡ | 37 (64) | 31 (54) | –10.3 (–28.2 to 7.5) | 14 (74) | 18 (62) | –11.6 (–38.2 to 14.9) |
| Accounted for clustering in analysis | 62 (77) | 66 (75) | –1.5 (–14.5 to 11.4) | 38 (66) | 43 (59) | –6.6 (-23.3 to 10.0) |

*Excludes 67 trials with no participant enrolment.

†Excludes 84 trials with pair matched designs or primary analysis at cluster level.

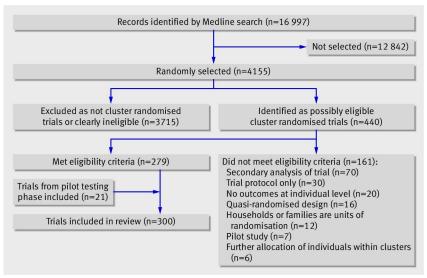‡Excludes 136 trials with no sample size calculation presented.

# Figures



Records identified by Medline search (n=16 997)

Not selected (n=12 842)

Randomly selected (n=4155)

Excluded as not cluster randomised trials or clearly ineligible (n=3715)

Identified as possibly eligible cluster randomised trials (n=440)

Met eligibility criteria (n=279)

Did not meet eligibility criteria (n=161):
Secondary analysis of trial (n=70)
Trial protocol only (n=30)
No outcomes at individual level (n=20)
Quasi-randomised design (n=16)
Households or families are units of randomisation (n=12)
Pilot study (n=7)
Further allocation of individuals within clusters (n=6)

Trials from pilot testing phase included (n=21)

Trials included in review (n=300)

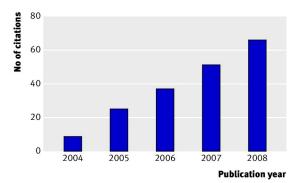**Fig 1** Identification of sample of 300 cluster randomised trials



**Fig 2** Number of citations (assessed with SCOPUS) of CONSORT extension for cluster randomised trials by year of publication