Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines

Karla Díaz-Ordaz, Michael G Kenward, Abie Cohen, Claire L Coleman and Sandra Eldridge

Clin Trials published online 5 June 2014
DOI: 10.1177/1740774514537136

Review

CLINICAL TRIALS

Clinical Trials
1–11

DOI: 10.1177/1740774514537136
ctj.sagepub.com
SAGE

Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines

Karla Díaz-Ordaz[1], Michael G Kenward[1], Abie Cohen[2], Claire L Coleman[2] and Sandra Eldridge[2]

Abstract

Background: Missing data are a potential source of bias, and their handling in the statistical analysis can have an important impact on both the likelihood and degree of such bias. Inadequate handling of the missing data may also result in invalid variance estimation. The handling of missing values is more complex in cluster randomised trials, but there are no reviews of practice in this field.

Objectives: A systematic review of published trials was conducted to examine how missing data are reported and handled in cluster randomised trials.

Methods: We systematically identified cluster randomised trials, published in English in 2011, using the National Library of Medicine (MEDLINE) via PubMed. Non-randomised and pilot/feasibility trials were excluded, as were reports of secondary analyses, interim analyses, and economic evaluations and those where no data were at the individual level. We extracted information on missing data and the statistical methods used to deal with them from a random sample of the identified studies.

Results: We included 132 trials. There was evidence of missing data in 95 (72%). Only 32 trials reported handling missing data, 22 of them using a variety of single imputation techniques, 8 using multiple imputation without accommodating the clustering and 2 stating that their likelihood-based complete case analysis accounted for missing values because the data were assumed Missing-at-Random.

Limitations: The results presented in this study are based on a large random sample of cluster randomised trials published in 2011, identified in electronic searches and therefore possibly missing some trials, most likely of poorer quality. Also, our results are based on information in the main publication for each trial. These reports may omit some important information on the presence of, and reasons for, missing data and on the statistical methods used to handle them. Our extraction methods, based on published reports, could not distinguish between missing data in outcomes and missing data in covariates. This distinction may be important in determining the assumptions about the missing data mechanism necessary for complete case analyses to be valid.

Conclusions: Missing data are present in the majority of cluster randomised trials. However, they are poorly reported, and most authors give little consideration to the assumptions under which their analysis will be valid. The majority of the methods currently used are valid under very strong assumptions about the missing data, whose plausibility is rarely discussed in the corresponding reports. This may have important consequences for the validity of inferences in some trials. Methods which result in valid inferences under general Missing-at-Random assumptions are available and should be made more accessible.

Keywords

Cluster randomised trials, missing data, multiple imputation

Introduction

In cluster randomised trials, the unit of randomisation is a group or cluster of individuals, for example, a hospital, rather than the individuals themselves. It is well known that there is greater homogeneity of individuals in clusters than in the population at large,[1] and that

[1]Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK
[2]Centre for Primary Care and Public Health, Queen Mary University of London, London, UK

Corresponding author:
Karla Díaz-Ordaz, Department of Medical Statistics, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK.
Email: karla.diaz-ordaz@lshtm.ac.uk

ignoring this homogeneity within clusters in the analyses can result in an overestimation of precision at the group level, potentially leading to spuriously significant results and inappropriate inferences.

Another important problem that compromises the validity of the trial results is that of missing data. Because entire clusters may withdraw altogether or fail to collect certain data from all individuals at follow-up visits, cluster randomised trials may sometimes have a high proportion of missing values. In addition, the probability that individuals are lost to follow-up or miss their follow-up appointment may depend on cluster as well as individual level characteristics.

In order to draw valid inferences in the presence of missing data, it is necessary to make assumptions about the statistical behaviour of the unobserved data. Broadly, data are said to be *Missing-Completely-at-Random* if the probability of a missing value is the same for all individuals in the trial irrespective of their covariates or outcomes. This is likely to be the case if, for example, questionnaires got lost in the post. Data are said to be *Missing-at-Random* if, having allowed for possible dependence on observed data, there is no additional dependence of the probability of missingness on unobserved data. For example, drop out in a longitudinal study may depend on previously observed outcomes of the participant but, given these, not on future, unobserved values. A likelihood-based analysis using only the observed data can be valid under Missing-at-Random if the analysis adjusts for those variables associated with the probability of not being observed. Finally, if the missing data are not Missing-Completely-at-Random or Missing-at-Random, the data are said to be *Missing-Not-at-Random*: the probability of a missing observation depends on unobserved data, even after allowing for dependence on all observed data. This could arise, for example, in measurements of health status if patients withdraw for reasons directly associated with poor health that are not captured by observed measurements.

Depending on the particular association between missingness and the data, ignoring the mechanism by which data became missing may result in invalid inferences. In practice, this mechanism is rarely known and cannot be assessed from the data under analysis. Hence, missing data usually introduce some degree of unavoidable ambiguity into a statistical analysis.

Simple methods may be valid in some circumstances. For example, an analysis restricted to those subjects with complete sets of data, a so-called Complete Case Analysis, will be valid if the probability of being a complete case is independent of the outcome given the covariates in the analysis model, that is, if missingness is driven by the explanatory variables, and given these, not the outcome.[2] However, Complete Case Analysis may be unacceptably inefficient if more than a small proportion of individuals is omitted.[3] More generally, other ad hoc methods, such as single imputation, will not be valid, as without suitable modification to obtain valid estimates of precision, they do not produce analyses that properly capture the precision of estimators, are often incompatible with the assumptions underlying the analysis model and rely on assumptions that are not always explicitly stated and whose plausibility is rarely discussed.

A principled tool for handling missing data is multiple imputation.[4,5] Missing data are imputed using the *imputation model*, producing several completed data sets. Each set is then analysed separately using the original analysis model, and the resulting parameter estimates and associated measures of precision are combined using formulae derived by Rubin,[4] to produce the multiple imputation estimators and their variances. Under certain broad conditions, inferences based on these will be valid, with the uncertainty due to the missing data properly accounted for.

Multiple imputation can increase precision, by incorporating information from individuals with partly observed data, and remove bias that might be present from a Complete Case Analysis under Missing-at-Random mechanisms. In the special case, when only outcomes are missing, it may well be that multiple imputation has no advantages over a conventional likelihood analysis,[2] although it offers the possibility of including in the imputation model variables which are predictive of missingness and outcome, but which are not wanted in the analysis model, so-called *auxiliary variables*. This makes the assumption that the data are Missing-at-Random more plausible.

As for individually randomised trials, in cluster randomised trials, a Complete Case Analysis may be valid under the assumptions already described. Single imputation methods can be carried out as well, although they suffer from the same inadequacies. To ensure valid inferences, a multiple imputation model needs to be *multilevel*, that is, reflect the within-cluster dependence in this hierarchical design.[6] The use of a multilevel imputation model may or may not affect consistency of the parameter estimators, but can be critical for proper estimation of their precision.[7] Box 1 summarises the most common methods aimed to handle missing data, whether hierarchical or not. This list is by no means exhaustive.

A review of cluster randomised trials was undertaken to assess the quality of handling and reporting missing data, in particular, establishing the prevalence of missing data and describing the current handling of this potentially serious problem. We propose some guidelines for handling missing values in the analysis of cluster randomised trials, and for reporting this appropriately.

Methods

Inclusion and exclusion criteria

All cluster randomised trials published in English during 2011 (including online date of publication) were

Box 1. Categories of the most common techniques used to handle missing data.

Category	Definition and assumptions for validity
A – Complete case analysis	Excludes subjects with missing data. In general, only valid under Missing-Completely-at-Random, except when the missingness is independent of the outcome given the covariates, for example, if all missing outcome data are dependent on baseline covariates. Inverse probability weighting can be used to make a valid complete case analysis under Missing-at-Random (see, for example, Little and Rubin,[8] Chapter 3).
B – Single imputation (for continuous data)	Missing values are filled in with a single value (the choice of this value may vary, see below). The variance is underestimated, as the completed data set is treated as if it was the observed data set, not accounting for the imputation.
Last observation carried forward for longitudinal studies	Imputes missing values with the individual's last observation. The underlying assumption for validity is that the missing value is statistically exchangeable with previous measurements, implying a specific and implausible Missing-Not-at-Random mechanism.[9]
Mean imputation and regression imputation	Missing values are either assumed to be the marginal mean of the observed values or predicted from the individuals' observed data, using a model based on observed individuals. For cluster randomised trials, there are two choices for mean imputation: using all the data pooled across clusters in each intervention group, or using the data from each cluster only (within-cluster mean imputation).[10] Regression imputation should account for the hierarchical nature of the data. These methods can provide unbiased estimators under certain Missing-at-Random mechanisms and models, but uncorrected measures of precision will be incorrect.
C – Single imputation (for binary data); best/worst case (or assuming zero for binary missing data)	Imputes all missing values with the best or worst case value respectively. For example, it assumes that individuals have missing data because their health has improved, or alternatively because of poor health. This is rarely a plausible assumption, but is typically used as part of a sensitivity analysis.
D – Multiple imputation	A method of 'filling' in the missing values multiple times, by drawing from an appropriate distribution. Multiple imputation treats missing data as an explicit source of random variability and incorporates this uncertainty explicitly, by accounting for the between-imputation variability.
Single-level multiple imputation	The imputation model ignores clustering.
Fixed-effects multiple imputation	The imputation model includes a dummy variable for each cluster (see, for example, Graham[11]).
Random-effects multiple imputation	The imputation model accounts for clustering through a random effect (see, for example, Díaz-Ordaz et al.[12]).

considered for inclusion. Quasi-experimental cluster designs; studies reported as 'pilot', 'feasibility' or 'preliminary' studies; studies in which no data at the individual level were collected; and reports only of cost-effectiveness were excluded. No studies were excluded on the basis of quality because our aim was to provide a description of quality. Secondary reports of trials were excluded.

Data sources and search methods

PubMed was searched for relevant reports in June 2012, using a search strategy previously used by our research team and which compares well with other cluster randomised trial searches.[13] The full electronic search strategy is given in Box A in the Supplemental File.

Sifting and validation

Two researchers examined titles and abstracts of trials identified from electronic searches and screened them for possible inclusion in the study. Full texts were obtained for those reports definitely or possibly satisfying the inclusion criteria. For validation, a third researcher carried out the same process independently on a random 10% selection. This same researcher acted as an arbitrator, when there was disagreement whether a trial should be included.

Piloting

Two researchers piloted the full inclusion criteria and extraction process on one-third of the reports identified by electronic searches. The definitions used for the inclusion criteria and extraction, for example, the

criteria for establishing the occurrence of cluster and individual missing data, were refined, and this was recorded in the protocol, which is available from the corresponding author. The reports used for the pilot were then re-extracted according to the updated criteria.

Data extraction

After sifting and piloting, we randomly ordered the reports identified electronically, excluding the one-third used in the pilot, and extracted from this list until we had included twice the number of reports accepted in the pilot. The total sample is thus formed of those reports identified in the pilot, representing one-third of the sample, plus those reports accepted in random-order from the two-thirds of the total of possible studies. Written guidance, agreed in advance of any extractions, was used when extracting data. Data were recorded onto a spreadsheet. Discrepancies were resolved by discussion.

Data were extracted on journal, country in which the trial was carried out and one primary outcome per report, defined as that specified by the authors or, if not specified, the outcome used in sample size calculations. If no primary outcome was specified and no sample size calculation was reported, the first outcome presented in the abstract was considered primary. Information was collected on the principal analysis, defined as the main analysis presented for the primary outcome. Descriptions of attrition, loss to follow-up and missing values were also recorded.

Data on total number of clusters and individuals randomised and analysed were extracted, and the difference between those randomised and analysed was used as an indication for the occurrence of missing data. To provide complete information on reasons for cluster drop out, we identified clusters that had been recruited and randomised to a study but did not manage to recruit individuals, separately from those randomised clusters in which individuals were recruited but no data from the cluster were available. Losing clusters in either way can cause bias. We also recorded whether there was mention of missing data, and any quote with regard to the use of imputation techniques.

For those trials reporting using an imputation technique, one researcher extracted information on whether *sensitivity analyses* were done with respect to the missing data assumptions, that is, if principal analysis assumed Missing-at-Random, was another analysis, assuming some other missing data mechanism, performed on the same outcome to assess the robustness of the results? If sensitivity analyses were reported, the results were examined to see whether the method for handling missing data made a difference to the study conclusions, for example, a substantial change in estimates, measures of precision and inferences drawn, and whether this was recorded. When multiple imputation was reported, we also extracted the number of imputations used and whether the variables included in the imputation model were reported.

Analysis

The prevalence of missing data is reported based on a combination of reported differences in the total number of clusters and individuals randomised and analysed, and whether missing data were mentioned in the report. The interquartile range of the proportion of missing individual data is also reported, for trials for which this proportion could be calculated from reported data.

The methods used for dealing with missing data in analyses were categorised according to Box 1. If weighting methods were used in the Complete Case Analysis, this was recorded. The number (%) of articles using each missing data approach is reported.

Results

We identified 526 published reports via our electronic search, of which 188 were excluded on the basis of titles or abstracts as definitely not meeting our inclusion criteria. Of the remaining 338 reports, a randomly selected one-third (112 reports) was used for piloting, from which 44 trials were included. From the 226 remaining reports, we examined 164 and accepted a further 88 trials. In all, 62 full reports were not examined and thus effectively excluded at random. A further 144 reports were excluded after assessing their full-text report, as they did not meet our eligibility criteria. See Figure 1 for the flow diagram. The full list of 132 reports which met our eligibility criteria and are included in this study is reported in the Supplemental File.

In the screening phase, that is, screening title and abstract of each report for inclusion, the two extractors had an agreement of 93%, the third researcher taking final decision on the remaining 35 studies. During the full-text extraction, agreement after discussion was 97%, with the arbitrator being called to look at only eight studies.

Trial characteristics

Over half of the included trials were conducted in Europe (50 trials, 38%) or North America (36 trials, 27%). The number of clusters randomised ranged from 4 to 596 and the number of individuals included in the analysis from 39 to 59,666. Clustering was accounted for in the analyses of 70% (92/132) of the trials, while an extra 10 (8%) trials adjusted only the standard error for clustering. In further 11 (8%) trials, it was unclear whether clustering had been accounted for and remainder did not adjust for clustering.

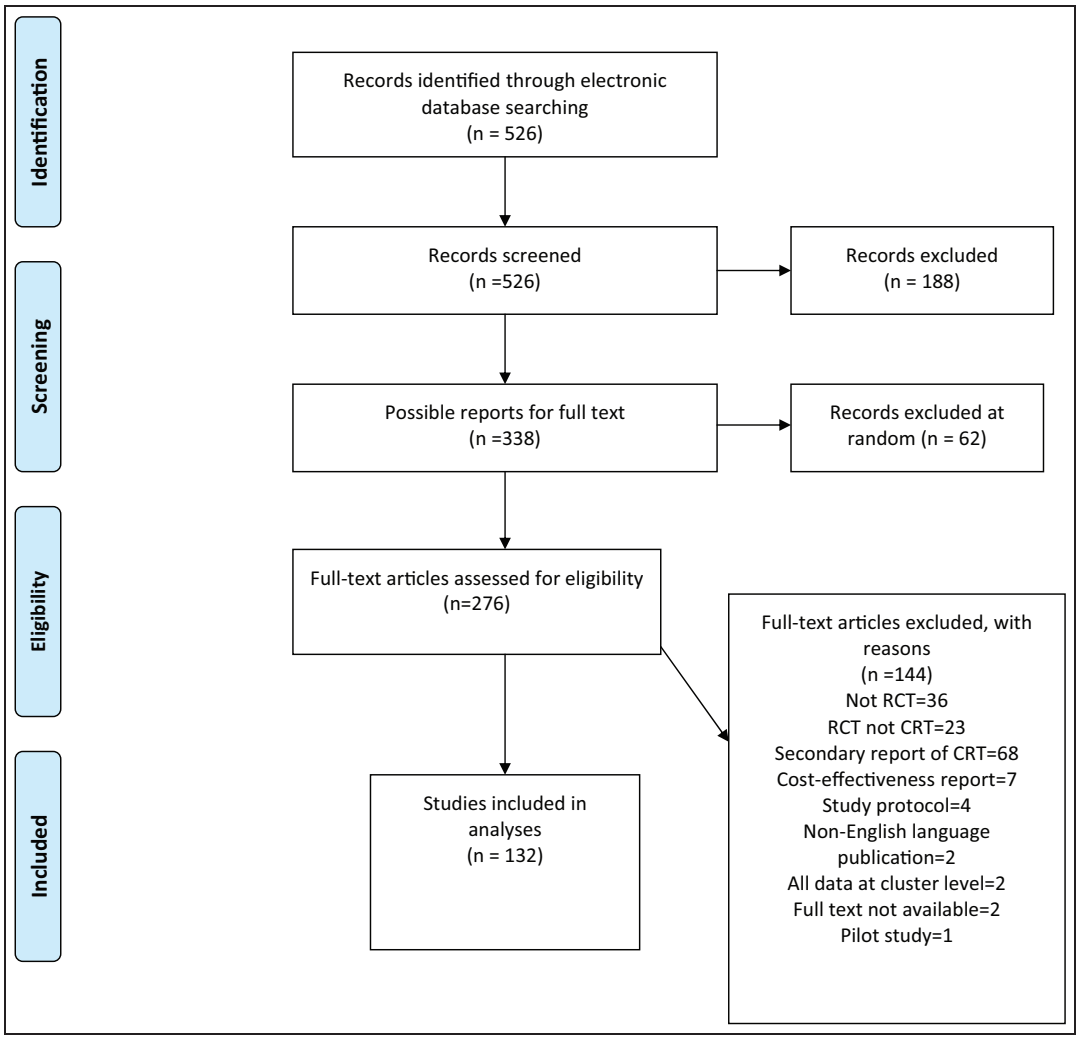

Figure 1. Flow diagram of the identification process for the sample of 132 cluster randomised trials included in this review.

Whole clusters were dropped from analyses in 24 (18%) trials. This is based on the number of clusters randomised and those included in the principal analyses, but excludes three trials (2%) where clusters were randomised but failed to recruit any individual participants. Reports were not clear enough for us to assess cluster attrition in 12 trials (9%). Among those studies with missing clusters (24 trials), the median percentage of missing clusters was 10% (interquartile range: 5%–29%). Six trials lost over 30% of the clusters randomised.

In all, 63 trials (48%) had missing individual data. For a further 41 studies (31%), missing data at the individual level could not be verified from the published report. The proportion of missing individual data among those 63 studies ranged from 1% to 47%, with an interquartile range 5%–23%, and median 13% missing individual data. Together with the information that trials reported concerning the imputation of data, it was estimated that only 37 studies (28%) had complete data.

From the 95 trials with missing data (which could be missing outcome or covariates or both), only 32 trials (34%) reported how they handled missing data, including two studies that used likelihood-based analysis explicitly stating the fact that they believe data were Missing-at-Random, and hence they would obtain unbiased results.

The most popular method for explicitly handling the missing data was Last Observation Carried Forward, used in 11 studies (12%); investigators used a variety of single imputation techniques in nine studies and multiple imputation in eight (6%) (Table 1).

The use of multiple imputation

Of the eight trials in which multiple imputation was used, none used an imputation model that accounted for clustering, despite the fact that six accounted for clustering in the analysis model. Only two reported the number of imputations performed; this was five in both cases. None specified the variables used in the

Table 1. Frequency of methods used to handle the missing data.[a]

Imputation technique	Number of reports (N = 95)	Percentage
A – Complete case[b]	65	66
B – Ad hoc single imputation	16	
Last observation carried forward	11	13
Linear extrapolation	2	2
Regression imputation	1	1
Mean imputation	2	2
C – Ad hoc single imputation for binary data	4	
Best/worst case	1	1
Assumed zeros	3	4
D – Multiple imputation	8	6
Single-level multiple imputation	8	6
Fixed-effects multiple imputation	0	0
Random-effects multiple imputation	0	0
E – Unclear	2	2

[a]Based on studies which acknowledged having missing data or where numbers analysed and randomised were not equal.
[b]Of which 28 (30%) used a likelihood-based hierarchical model, which may result in valid inferences for certain missing data mechanism. Only two studies discussed these assumptions explicitly.

imputation models. One of the trials using multiple imputation did not use appropriate rules to combine estimates obtained from multiply imputed data sets.[14] Multiple imputation was used as principal analysis in five studies, while it was used as a sensitivity analysis in a further three (Table 2).

Sensitivity analyses

A sensitivity analysis was reported in only 10 trials (11%). In all, 23 trials used the imputed data, whether single or multiple for the principal analysis, and a further 7 studies used Complete Case as the principal analysis, with some imputation technique as a sensitivity analysis. Of those using imputation as a principal analysis, only three subsequently reported having performed a sensitivity analysis and all used Complete Case Analysis. Six trials reported that sensitivity analyses either resulted in a substantial change of treatment estimate or that results changed from being statistically significant to non-significant or vice versa (Table 2).

Discussion

Main findings

More than two-thirds of the trials in this review had missing data (in either outcome or covariates), but only about one-third of those explicitly accounted for missing values in analyses, the rest used Complete Case

Table 2. Reporting of sensitivity analyses (10 studies).

Study	Principal analysis	Sensitivity analysis	Results from sensitivity analysis reported?	Changes to conclusion
Aburto et al.[15]	Complete case	Last observation carried forward	Not tabulated, stated results to be similar	No
Bodin and Strandberg[16]	Complete case	Multiple imputation	Tabulated	No
		Best case scenario	Tabulated	No
		Worst case	Tabulated	Inference is changed
Caria et al.[17]	Complete case	Last observation carried forward	Not tabulated, stated results to be similar	No
		Best case scenario	Not tabulated, stated results to be similar	No
		Worst case scenario	Yes	Inference is changed
Ezendam et al.[18]	Complete case	Baseline/last observation carried forward	Not tabulated. The text reported only results when significant changes occurred	Inference is changed for some secondary outcomes
Meyer et al.[19]	Complete case	Multiple imputation	Tabulated	No
Slade et al.[20]	Complete case	Regression imputation	Tabulated	No
Taft et al.[21]	Complete case	Multiple imputation	Tabulated	Inference is changed
Au et al.[22]	Multiple imputation	Complete case	In text	Inference is changed
Jorgensen et al.[23]	Last observation carried forward	Complete case	Tabulated	Inference is changed
Tylleskar et al.[24]	Recorded missing as non-events	Complete case	Tabulated (in web-appendix)	Important changes to point estimates (>1 SD)

SD: standard deviation.

Analysis without discussion of the plausibility of the assumptions required for these analyses to be valid.

In trials in which imputation was used to handle missing values, methods were inadequate in every case, either because they are generally inappropriate, for example, Last Observation Carried Forward and single-value imputation, or because they are not suitable for use in cluster randomised trials, for example, multiple imputation which ignores the hierarchical nature of the data.

The analyses based on single imputation methods (22 studies) will have over-estimated the precision in the data and are based on assumptions which are difficult to understand and rarely plausible. At least half of the studies using single imputation used a likelihood-based analysis model, which could have yielded consistent estimators based on completers under Missing-Completely-at-Random, which is a more transparent assumption, even if fairly strong.

Sensitivity analyses were rarely performed, and when they were, the results of the sensitivity analyses were not always reported. In particular, the different assumptions about the missing data used and the changes that these differences made to the inference or the magnitude of the treatment effect were not discussed at any length in the reports or web-appendices.

Interpretation. Complete Case Analyses may result in valid inferences when missingness is driven by the explanatory variables included in the analysis model, and given these, not the outcome, but this is difficult to ascertain from the trial data, and even more so from the information included in each trial report. Moreover, if the analysis model does not use likelihood-based estimation, in general, these Complete Case Analyses will only be valid under the restrictive Missing-Completely-at-Random assumption.

While Last Observation Carried Forward may seem initially an attractive method for longitudinal data because it is easy to implement, it has been shown that its validity depends on implausible assumptions and a very specific Missing-Not-at-Random mechanism.[9,25] In addition, single-value imputations, without appropriate modification, underestimate the variance because in these analyses, the imputed data are analysed as if they were 'real' data points.[26] The use of worst/best case imputation,[27] while intended as a sensitivity analysis, is limited to binary data, and typically leads to a very wide range of inferences. Arguably, it is better to use *controlled imputation methods* for sensitivity analyses (Mallinckrodt,[28] Section 10.5).

Sensitivity analyses are intended to assess the robustness of trial findings to untestable assumptions made about the missing data mechanism, and, as such, they should be based on assumptions that deviate in a plausible way from those in the principal analysis. The plausibility of these assumptions is trial-specific. Simplistic ad hoc analyses, such as Last Observation Carried Forward and mean imputation, typically rest on implausible assumptions that are incompatible with the assumptions implicit in the model underlying the analysis.

The use of multiple imputation under Missing-at-Random as a sensitivity analysis of a likelihood-based model can be justified if the multiple imputation is bringing extra information that the analysis model was not, for example, by using auxiliary variables. For this reason, it is very important that the variables used in the imputation model are reported,[29] but none of the studies in our sample reported this information.

Ignoring clustering in the multiple imputation can lead to an increase in Type I errors,[10] while including cluster as a fixed effect in the imputation model, as recommended in White et al.,[30] produces imputations with excessive between-cluster variability, especially in circumstances with small cluster sizes and low intra-class correlation coefficient.[31]

The correct form of multiple imputation for cluster randomised trials uses a multilevel imputation model. Appropriate multilevel multiple imputation routines are now implemented in a variety of commercial and freeware software packages: PAN package[32] in R, the mi macro[7] which operates within MLwiN and can handle up to four hierarchical levels and the REALCOM-IMPUTE software[33] which unlike the other options can handle cluster-level variables with missing data. Although there exist now user-contributed packages in SAS (MMI_IMPUTE macro[34]) and Stata (program REALCOM Impute, which allows data to be exported to REALCOM), we recognise that there is a need for multilevel multiple imputation packages to be embedded in the most commonly used statistical software packages, SAS and Stata,[35] in order to increase the accessibility of this method.

Where whole clusters provide no data, imputation of cluster and individual level data in preparation for analysis would not add any new information, as values would be imputed from the distribution predicted by the imputation model. If, however, we have some information about these clusters, for example, cluster-level covariates measured at the time of cluster recruitment, multilevel multiple imputation may help reduce the bias introduced by cluster non-response. Whether multiple imputation is used or not, we encourage investigators to consider carefully what missing mechanisms might be operating and to include a discussion of possible biases introduced in their discussions.

Strengths and limitations

We have used a rigorous search and data extraction procedure, including double extraction. Inclusion criteria were wide, so our sample should be representative

of recently published cluster randomised trials. However, as with reviews of this nature, the assessment of quality was based only on information in trial reports. It is possible that trial teams had conducted sensitivity analyses to assess the robustness of their conclusions to the missing data, but not reported it in the published version. Space limitations in journals may sometimes preclude the reporting of factors relevant to assess the appropriateness of the methods used to handle the missing data, although this may change with increasing online publishing and the availability of web-appendices.

In addition, because we based our judgement on the evidence of missing data partly on the differences between randomised and analysed participants as reported on the published articles, it was not possible to determine whether the missing data were in the outcome or the covariates; this distinction is important as a Complete Case Analysis may be appropriate under certain Missing-at-Random or even Missing-Not-at-Random mechanisms, for example, a Complete Case Analysis that uses maximum likelihood could be appropriate under Missing-at-Random, if only outcome is missing.

Box 2. Guidelines for analysing and reporting cluster randomised trials with missing data.

Recommendation	Rationale
Report the number of clusters and individuals lost to follow-up, as well as numbers of missing values for each variable of interest.	This is likely to be important when deciding what analysis to use, and designing a relevant sensitivity analysis. For a cluster randomised trial, the variables associated with the missing data mechanism may be at the cluster or at the individual level.
Collect and report information about reasons for losses to follow-up and other missing values.	Knowledge about the reasons for missing data may help when deciding which auxiliary variables to use in an imputation model, as well as the plausibility of the Missing-at-Random assumption.
Justify the choice of principal analysis, the missing data mechanism assumed, and the plausibility of these assumptions.	The validity of the analysis is determined by the correctness and plausibility of the underlying assumptions.
Perform and report sensitivity analyses	It is important to explore the robustness of the trial results to departures from the missing data assumption made in the primary analysis.
For analyses where the treatment estimate of interest is an intention-to-treat	
The principal analysis should involve at least all the observed outcome data.[41]	Likelihood analyses of all observed data, such as mixed models, are acceptable if the Missing-at-Random assumption is reasonably plausible.
Include all randomised individuals in a sensitivity analysis.	Analyses that exclude data from randomised participants are not acceptable without strong assumptions about the missing data (Missing-Completely-at-Random), and are not consistent with the strict intention-to-treat analysis strategy, so it is important to explore the impact of those individuals on the trial results.
For analyses based on multiple imputation	
Use a method that accounts for clustering adequately, that is, multilevel multiple imputation.	The validity of the results when using multiple imputation depends on modelling the distribution of each variable with missing values, in terms of the observed data appropriately, including the dependence structure of the data. This is crucial to obtain correct estimates for the standard errors.
Report number of imputations and software used.	The number of imputed data sets has an impact on the variability from the imputation process. Knowing which software was used is important too, in order to understand the assumptions under which the results obtained are valid, as some assume normality for all continuous variables.
Report the details of the imputation model, variables included and how non-normally distributed and categorical variables were dealt with.	The imputation model must include all variables in the analysis model. Adding auxiliary variables helps making the Missing-at-Random assumption more plausible. Including non-normally distributed variables in the imputation model may introduce bias if the software used assumes normality and no measures had been taken to improve normality.

Comparison with previous research

Two previous reviews[36,37] of the quality of conduct and reporting of cluster randomised trials found that 39% and 70% of the trials reported on loss to follow-up for clusters. The cluster attrition percentage median was 0% (interquartile range: 0%–6%) in Eldridge et al.[36] and 0% (interquartile range: 0%–5%) in Eldridge et al.[37] Our findings suggest that there has been some improvement in cluster follow-up rates, as the cluster attrition percentage median is 0% (interquartile range: 0%–0%).

A previous review looked at how missing outcome data were reported and handled in individually randomised clinical trials.[38] The review included 71 trials published in 2001, of which 63 reported missing outcome data (89%). Among these, 92% used Complete Case Analysis, 8% Last Observation Carried Forward and one study used multiple imputation.

Another review considered cost-effectiveness analyses using randomised clinical trial data published post-2003.[39] There, the inclusion criterion was any mention of missing data. Among the 88 cost-effectiveness analyses, 31% used Complete Case Analysis, 2% used Last Observation Carried Forward and 9% used multiple imputation, the remainder using other ad hoc methods.

There is some indication from our review that the proportion of authors reporting handling missing values and using multiple imputation in recent cluster randomised trials is greater than the proportion doing so in less recent individually randomised trials. It is plausible, particularly given the increased interest in the area of missing values in recent years, that this reflects a rising number of authors wanting to account for missing values in cluster randomised trials. Unfortunately, within our review, no authors used multilevel multiple imputation methods. This may have important consequences for inference in some trials.

Inadequate methods to handle missing data may be frequently used because researchers believe that a valid intention-to-treat analysis necessarily requires imputed data.[18,23,40] In fact, analyses that do not include all individuals may have greater validity than those that do;[41,42] so imputation methods should be used with care.

Conclusion

Our review shows that most of the studies with missing data use Complete Case Analysis, and that while this is valid under Missing-Completely-at-Random assumptions, the use of appropriate multilevel methods to obtain valid inferences under the more general Missing-at-Random assumption is rare in cluster randomised trials, and there is a need to redress this given the potential for invalid inferences.

In Box 2, we make recommendations for analysing cluster randomised trials with missing data. These are similar to general guidelines for individually randomised trials,[29,30,43] but we emphasise the importance of collecting and using information both at the cluster and individual level, and the need to account for clustering in all analyses, including those used to handle missing data.

It is important to understand that the validity of the results when using multiple imputation depends on obtaining an appropriate estimate for the standard errors, for which it is crucial to use a method that correctly models the dependence structure of the data. Appropriate methods to handle missing values in cluster randomised trials under the Missing-at-Random assumption are available. To promote their use, methodologists should publish good examples of the effects of using and not using such methods based on either empirical or simulated trial data.

Authors' contributions

K.DO. conceived and designed the study and wrote the protocol. S.E. and M.G.K. contributed to the design, the protocol and the manuscript. A.C. and C.L.C. undertook independent data extraction. K.DO. extracted additional data, settled disagreements among those extracting data, conducted all the analyses and took primary responsibility writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Jonathan Bartlett for his useful comments on a draft of this article.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Funding

NIHR Research Methods fellowship and internships (K.DO., A.C. and C.L.C.).

References

1. Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol* 1978; 108: 100–102.
2. White IR and Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat Med* 2010; 29(28): 2920–2931.
3. Vach W and Blettner M. Missing data in epidemiologic studies. In: P Armitage and T Colton (eds) *Encyclopedia of biostatistics*, vol. 2. Chichester: John Wiley & Sons, Ltd, 2005, pp. 1255–1276.
4. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York: Wiley, 1987.
5. Carpenter J and Kenward M. *Multiple imputation and its application*. New York: Wiley, 2013.
6. Schafer AL. Multiple imputation with PAN. In: L Collins and A Sayer (eds) *New methods for the analysis of change*. Washington, DC: American Psychological Association, 2001, pp. 355–377.

7. Carpenter J and Goldstein H. Multiple imputation using MLwiN. *Multilevel Model Newsl* 2004; 16(2): 9–18.
8. Little RJ and Rubin DB. *Statistical analysis with missing data*. New York: Wiley, 2002.
9. Kenward MG and Molenberghs G. Last observation carried forward: a crystal ball? *J Biopharm Stat* 2009; 19(5): 872–888.
10. Taljaard M, Donner A and Klar N. Imputation strategies for missing continuous outcomes in cluster randomized trials. *Biom J* 2008; 50(3): 329–345.
11. Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol* 2009; 60: 549–576.
12. Díaz-Ordaz K, Kenward MG and Grieve R. Handling missing values in cost effectiveness analyses that use data from cluster randomized trials. *J R Stat Soc Ser A Stat Soc* 2013; 177: 457–474.
13. Taljaard M, McGowan J, Grimshaw J, 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(1): 15.
14. Jackson C, Cheater FM, Harrison W, et al. Randomised cluster trial to support informed parental decision-making for the MMR vaccine. *BMC Public Health* 2011; 11: 475.
15. Aburto NJ, Fulton JE, Safdie M, et al. Effect of a school-based intervention on physical activity: cluster-randomized trial. *Med Sci Sports Exerc* 2011; 43(10): 1898–1906.
16. Bodin MC and Strandberg AK. The Orebro prevention programme revisited: a cluster-randomized effectiveness trial of programme effects on youth drinking. *Addiction* 2011; 106(12): 2134–2143.
17. Caria MP, Faggiano F, Bellocco R, et al. Effects of a school-based prevention program on European adolescents' patterns of alcohol use. *J Adolesc Health* 2011; 48(2): 182–188.
18. Ezendam NP, Brug J and Oenema A. Evaluation of the Web-based computer-tailored FATaintPHAT intervention to promote energy balance among adolescents: results from a school cluster randomized trial. *Arch Pediatr Adolesc Med* 2012; 166(3): 248–255.
19. Meyer C, Ulbricht S, Gross B, et al. Adoption, reach and effectiveness of computer-based, practitioner delivered and combined smoking interventions in general medical practices: a three-arm cluster randomized trial. *Drug Alcohol Depend* 2012; 121(1–2): 124–132.
20. Slade GD, Bailie RS, Roberts-Thomson K, et al. Effect of health promotion and fluoride varnish on dental caries among Australian Aboriginal children: results from a community-randomized controlled trial. *Community Dent Oral Epidemiol* 2011; 39(1): 29–43.
21. Taft AJ, Small R, Hegarty KL, et al. Mothers' AdvocateS In the Community (MOSAIC) – non-professional mentor support to reduce intimate partner violence and depression in mothers: a cluster randomised trial in primary care. *BMC Public Health* 2011; 11: 178.
22. Au DH, Udris EM, Engelberg RA, et al. A randomized trial to improve communication about end-of-life care among patients with COPD. *Chest* 2012; 141(3): 726–735.
23. Jorgensen MB, Ektor-Andersen J, Sjogaard G, et al. A randomised controlled trial among cleaners – effects on strength, balance and kinesiophobia. *BMC Public Health* 2011; 11: 776.
24. Tylleskar T, Jackson D, Meda N, et al. Exclusive breastfeeding promotion by peer counsellors in sub-Saharan Africa (PROMISE-EBF): a cluster-randomised trial. *Lancet* 2011; 378(9789): 420–427.
25. Molnar FJ, Hutton B and Fergusson D. Does analysis using 'last observation carried forward' introduce bias in dementia research? *CMAJ* 2008; 179(8): 751–753.
26. Allison PD. Multiple imputation for missing data: a cautionary tale. *Sociol Methods Res* 2000; 28(3): 301–309.
27. Lachin JM. Worst-rank score analysis with informatively missing observations in clinical trials. *Control Clin Trials* 1999; 20(5): 408–422.
28. Mallinckrodt CH. *Preventing and treating missing data in longitudinal clinical trials: a practical guide*: Cambridge: Cambridge University Press, 2013.
29. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; 338: b2393.
30. White IR, Royston P and Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011; 30(4): 377–399.
31. Andridge RR. Quantifying the impact of fixed effects modeling of clusters in multiple imputation for cluster randomized trials. *Biom J* 2011; 53(1): 57–74.
32. Schafer JL and Yucel RM. Computational strategies for multivariate linear mixed-effects models with missing values. *J Comput Graph Stat* 2002; 11(2): 437–457.
33. Carpenter JR, Goldstein H and Kenward MG. REALCOM-IMPUTE software for multilevel multiple imputation with mixed response types. *J Stat Softw* 2011; 45(5): 1–14.
34. Mistler SA. A SAS macro for applying multiple imputation to multilevel data. In: *Proceedings of the SAS global forum: 2013*, San Francisco, CA, 28 April–1 May 2013, contributed paper (Statistics and Data Analysis) 438–2013.
35. Dembe A, Partridge J and Geist L. Statistical software applications used in health services research: analysis of published studies in the US. *BMC Health Serv Res* 2011; 11(1): 252.
36. Eldridge SM, Ashby D, Feder GS, et al. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials* 2004; 1(1): 80–90.
37. Eldridge S, Ashby D, Bennett C, et al. Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ* 2008; 336(7649): 876–880.
38. Wood AM, White IR and Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials* 2004; 1(4): 368–376.
39. Noble SM, Hollingworth W and Tilling K. Missing data in trial-based cost-effectiveness analysis: the current state of play. *Health Econ* 2012; 21(2): 187–200.
40. Jaglal SB, Donescu OS, Bansod V, et al. Impact of a centralized osteoporosis coordinator on post-fracture osteoporosis management: a cluster randomized trial. *Osteoporos Int* 2012; 23(1): 87–95.

41. White IR, Carpenter J and Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. *Clin Trials* 2012; 9(4): 396–407.
42. White IR, Horton NJ, Carpenter J, et al. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ* 2011; 342: d40.
43. Carpenter J and Kenward MG. *Missing data in clinical trials – a practical guide*. National Health Service Coordinating Centre for Research Methodology, 2007.