The use and reporting of multiple imputation in medical research – a review

● A. Mackinnon

Centre for Youth Mental Health, University of Melbourne, Parkville, Victoria, Australia

Abstract. Mackinnon A (Centre for Youth Mental Health, University of Melbourne, Parkville, Victoria, Australia) The use and reporting of multiple imputation in medical research – a review. *J Intern Med* 2010; **268**: 586–593.

Background. Multiple imputation (MI) is an advanced, principled method of dealing with missing data in statistical analyses, a common problem in medical research. This paper sought to document the use of MI in general medical journals and to evaluate the information provided to readers about the application of the procedure in studies.

Methods. Research articles using MI in analyses published in *JAMA, New England Journal of Medicine, BMJ* and the *Lancet* were identified using full text searches from the earliest date each journal offered such searches until the end of 2008. Ninety-nine articles were found. Studies were classified according to their design.

Results. Multiple imputation was used in 49 RCTs and 50 other types of studies. A third of the articles (n = 33) reported no details of the procedure used. In a third of these (n = 11), it was not possible to infer the approach used from references cited or software used. The nature of the imputation model was rarely reported. MI was frequently used as a secondary analysis (n = 40) either to justify reporting a simpler approach or as a form of sensitivity analysis.

Conclusions. Whilst still relatively uncommon, the use of MI has risen substantially, particularly in trials. MI is rarely adequately reported, leading to doubt about its appropriateness in some cases. This gives rise to uncertainty about conclusions reached and poses a barrier to attempts to replicate analyses. Guidelines for the reporting of MI should be developed.

Keywords: biostatistics, missing data handling, multiple imputation, randomized controlled trials, reporting standards.

Background

Missing data are a ubiquitous problem in nearly all fields of medical research. Complete information from participants or individual items of data is frequently unavailable because of attrition in longitudinal studies and due to a wide variety of other reasons in these and other studies.

Established, routinely used strategies for analysing data with missing observations can be divided into two approaches: (i) using only complete cases and (ii) the filling-in, or imputation, of the missing values. Complete case analysis (CCA) may reduce power substantially, particularly in multiple variable analyses where even low rates of missingness in individual variables may compound to high rates of casewise missingness. More importantly, individuals with complete data may be unrepresentative of the whole sample, leading to biased results. Many

methods exist for imputing missing observations. These include regression-based models, substitution of values from similar, complete cases (hotdeck imputation) and imputations made with recourse to a principle such as conservative estimation of an outcome, as is supposedly achieved by the last observation carried forward approach. All these methods treat imputed values as 'real' data leading, at best, to inflated estimates of precision and possibly bias. Thus, single imputation is rarely defensible.

Multiple imputation (MI) overcomes objections to single imputation by making repeated draws from a model of the distribution of variables that have missing observations, to create a number of complete datasets. These multiple complete datasets are then analysed in parallel. Variation in outcomes between datasets reflects uncertainty because of the imputation process. This uncertainty may be formally incor-

porated into estimates of precision and into tests of significance [1].

Multiple imputation is not new – being introduced in the 1970s [2] – but its application in medicine has been rare. This is likely to be because it is a complex, multistep procedure. It involves assessing the suitability of the dataset for imputation, development of an appropriate imputation model, assessment of the integrity of imputed values, followed by undertaking analyses in each imputed dataset and the combination of results from each analysis. Increasingly, the routine aspects of MI have been incorporated into, and automated in, major statistical packages. A notable addition to this was the recent inclusion of MI in SPSS [3], a package favoured by many researchers who analyse data without professional statistical assistance; it is now possible to use MI by completing dialog boxes, without the need to specify the proce-

dure in detail. Similar procedures are available in other statistical packages.

Increased accessibility brings with it the risk of misapplication of the procedure and misinterpretation of its results. Whilst this is the case with any statistical technique, MI itself is complex and involves making many choices and decisions (even if these are implemented only as defaults by a computer program). Further, MI is essentially a means to the end of conducting a conventional statistical analysis to answer substantive research questions. As such, little space is often devoted to methods used for handling missing data.

Table 1 shows key aspects of data and the imputation process that are required to assess and interpret analyses obtained using MI. These were derived from the work of authorities in the field

Table 1 *Details necessary to interpret analyses carried out using multiple imputation*

	Details to be reported
Missingness	Rates of missing data for each variable imputed
	Casewise missingness rates
	Basis of missingness – absence of participant vs missing measurement on available participant
	Pattern of missingness – dropout *vs* intermittent
	Causes of missing observations or measurements, if known.
	Design factors related to missingness, if appropriate (e.g. planned missingness)
Assumptions	Assumptions underlying MI [usually missing at random (MAR)]
	Assumptions of the particular method used for imputation (e.g. distribution of variables)
	The plausibility of the MAR assumption
Imputation model	Variables imputed and whether they are dependent (outcome) or independent (predictor or covariate) variables
	Variables used to impute missing values
	Overall form of model/procedure (linear, logistic, matching etc.)
	Relationship between imputation and analysis model
Imputation procedure	Details of the software used, including specific procedure and version
	Parameters, as appropriate, including number of imputations Order of imputation, if appropriate
	Steps taken to ensure and to assess imputation integrity, e.g. burn in iterations, tests of autocorrelation, examination of the imputed values
	Method of combining results from individual datasets, particularly if standard methods may not be or are not appropriate
Interpretation	If more than one method of analysing the data is presented, the status of each approach should be explained, with the rationale for preferring one approach over the other(s) being explicit.

[1, 4, 5], a variety of didactic guides [5, 6] and from research into MI [7, 8]. As implemented in software packages and routinely applied, MI assumes that observations are missing at random (MAR). This allows that missingness may be related to observed values but not to the missing observations themselves [4]. Variables related to missingness may include outcomes, covariates and other variables of no substantive interest to the study. Except in quite limited circumstances where fixed covariates predict missingness [9], CCA requires the much more restrictive assumption that missingness is occurring completely at random (MCAR) to produce unbiased results.

It is possible to test whether data are MCAR [10], but should this not be the case, differentiating data that are MAR from observations not missing at random (NMAR) would require access to the missing observations. Thus, formal assessment of the MAR assumption is inherently impossible; it cannot be assessed by examining the pattern of missingness or observed values. Specific MI procedures make additional assumptions about variables, most frequently that they are normally distributed in the case of continuous measures. MI also involves specifying an imputation model. This determines the values imputed. The model must reflect the nature of the analyses planned. Whilst it may include terms and variables not included in the analysis with a view to improving imputation accuracy, it must not exclude variables and terms such as interactions that are to be included in the analysis [6, 7]. Just as with analyses themselves, different issues may arise in imputing missing observations for later missing values, where information about earlier values is available in longitudinal data compared to imputing values collected contemporaneously, such as covariates and predictors in cross-sectional studies.

The current review of research articles using MI sought to determine how the technique is currently implemented and how its use is reported. This was motivated by the observation that at least some papers in medical journals using MI implemented the technique and interpreted its results in ways that were questionable or provided too little information to appraise the application of MI. Characterizing the use of MI in leading general medical journals will assist readers of research and those reporting it, by identifying gaps between the information required and that provided in articles. Furthermore, it will highlight common issues, problems and pitfalls in the use of MI.

Table 2 *Number of articles using multiple imputation by year and type of study*

Journal (Earliest full text searching)	RCTs		Other studies	
	Before 2005	2005 –2008	Before 2005	2005 –2008
BMJ (1994)	2	8	3	7
JAMA (1998)	9	19	9	8
Lancet (1823)[a]	1	5	4	7
New England Journal of Medicine (1993)	1	4	7	5
Total	13	36	23	27

[a]Earliest Multiple imputation (MI) publication 1994.

Methods

The review sought to identify all reports of research in four major medical journals – *JAMA, New England Journal of Medicine, BMJ* and the *Lancet* – in which MI was used in analyses. Full text searches for the phrase 'multiple imputation' were undertaken on each journal's website. Journals were searched from their earliest searchable full text articles (see Table 2) until the end of 2008. Articles were then hand searched to verify their use of MI in analyses reported. Where ambiguity remained, authors were contacted to confirm their use of MI. Where referred to in the article and available from the authors or the publisher, supplementary files were considered as part of the report. Information from primary reports was also consulted for secondary studies, e.g. a report of the cost effectiveness of a treatment might refer to details in the original report of the RCT.

Reports were classified according to the nature of the study – whether it was retrospective, cross-sectional, prospective or a trial. Details of missingness, the imputation procedure and its documentation, and the reporting of, and use to which analyses using MI were put were recorded. In the absence of a second rater, each paper was assessed by the author on two occasions separated by at least a month.

In addition to these data, limited information was collected from papers published in 2009 to enable trends in MI use to be extended to more recent times.

Although information was extracted systematically, this review is of a complete population of papers and

is essentially descriptive. Therefore, inferential statistics are not reported.

Results

Ninety-nine papers were identified as using MI in analyses reported. Figure 1 shows a substantial increase in MI use from 2005 to 2008, notably for RCTs. There was wide variation in MI use between journals (see Table 2). Approximately half the papers in which MI was used (n = 49) were RCTs. Other applications included prospective studies, two of which were uncontrolled trials (n = 21), retrospective studies, including case-control studies and analyses of public and existing datasets (n = 24), and a small number of cross-sectional reports (n = 5). Information provided about the use of MI was extremely variable and frequently minimal, with over a quarter of articles (n = 28) giving no details at all. Table 3 shows the number of articles reporting details relevant to the imputation procedure for RCTs and the other types of studies.

Papers published in 2009 (content not analysed in this review) demonstrate an ongoing, accelerated increase in the use of MI, particularly in trials. Pubmed (pubmed.gov) was used to count all RCTs published in each journal in 2009. A total of 305 such papers were identified, with 27 (9%) of these using MI in some way. There was substantial variation between journals, ranging from 5% (5 of 104) in *New England Journal of Medicine*, through 7% (6 of 87) and 8% (5 of 63) in the *Lancet* and *BMJ*, respectively, to 22% (11 of 51) in *JAMA*.
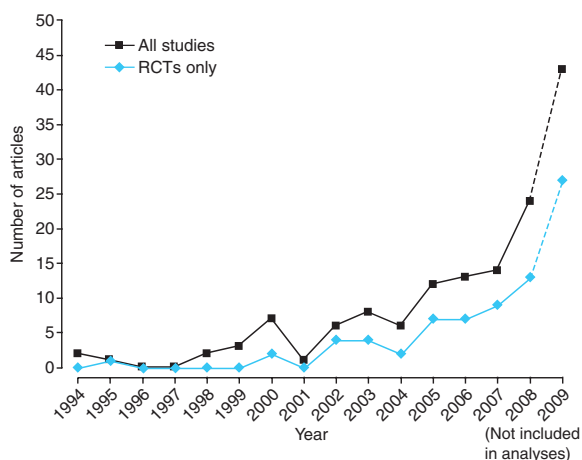
Fig. 1 *Research articles in BJM, JAMA, Lancet and NEJM using multiple imputation.*

Extent and nature of missingness

All but two RCTs included a participant flow diagram as required by the CONSORT statement [11]. Of the 50 other types of studies, nine gave no information about rates of missingness. Where reported or inferable, variable-wise missingness rates ranged from very low (<1%) to very high (above 70%). Casewise missingness rates were only sporadically reported but were often high (over 50%). Number of variables imputed (where stated) was not related to completeness of reporting – even when only one variable was imputed, it was common to give no information about the percentage of missing observations.

Only 16 studies mentioned any assumptions on which MI generally or the particular imputation method used is based. In two cases, this was partially incorrect (see Table 3).

Variables imputed

The nature of the variables imputed differed substantially according to the study design (see Table 3). Outcome measures (dependent variables) were reported as being imputed in all but 4 RCTs. Dependent variables were also imputed in half of the prospective studies (11/21) but in few nonprospective designs (7/29). In nine papers, it was not stated or possible to infer whether outcomes were imputed, and in many (n = 20), the number of dependent measures imputed was not stated or inferable. Where stated, most work involved only one or two outcomes (39 of 63 studies that reported imputing dependent variables).

Explicit imputation of independent variables (predictors or covariates) was rare in RCTs (one trial), but relatively common in other prospective and nonprospective designs (see Table 3). However, in nearly 20% of studies, it was not stated or possible to infer whether independent variables were imputed. In just over one third of the articles in which predictors were imputed (12/34), it was not stated and not possible to infer which or how many variables were involved. Where the number of variables imputed was discernable, most studies (14/22) imputed just one or two.

Imputation details

A third of the studies (n = 33) gave no details of the imputation methods or procedures. In two thirds of these papers (n = 22), other information was given in the article that allowed an inference to be made about the procedure likely to have been used. In most cases,

Table 3 *Details of missingness and imputation procedures reported in articles using multiple imputation (n = 99)*

	RCTs (n = 49)	Prospective studies (n = 21)[a]	Other studies (n = 29)[b]	Total
Missingness				
No details provided	0	4	5	9
Explicit statement of missingness	14	11	19	44
'CONSORT' Participant flow diagram	47	0	na	47
MAR assumption stated?				
No	41	17	25	83
Incorrect statement	0	2	0	2
Yes	8	2	4	14
Variables imputed				
Dependent/outcome variables imputed?				
No	0	7	20	27
Yes	45	11	7	63
Not stated	4	3	2	9
Independent variables imputed?				
No	35	6	5	46
Yes	1	12	21	34
Not stated	13	3	3	19
Imputation details				
No details provided	16	7	10	33
Imputation method stated	23	13	17	53
Model type reported	11	7	11	29
Predictors used in imputation specified	21	10	15	46
Additional variables used specified	16	5	8	29
Number of imputations stated	14	14	14	42
Software used stated?				
Package only	22	10	12	44
Procedure	13	6	9	28
Not stated	14	5	8	27
Analysis status				
Primary	21	15	23	59
Secondary	13	5	5	23
Secondary – sensitivity analysis	15	1	1	17

MAR, missing at random.
[a]Comprises two non-RCT trials and 19 other prospective studies.
[b]Comprises five cross-sectional, five case–control and 19 other retrospective studies.

this was limited to the software package used, but sometimes a routine or reference was given. For the remaining 11 articles, no details were given or could be assumed. Only half (n = 46) of the articles gave details of which variables were used as predictors in the imputation. Fewer (n = 29) stated whether additional variables were used other than those used in subsequent analyses.

Reports of explicit use of imputation approaches that accommodated analytic methods employed were rare. Eight papers used methods of imputation

specifically developed for the project or the problem concerned and reported this in detail or referred to an article covering the method. Other papers cited articles concerning specific imputation methods, but omitted details of its implementation with their data. Three RCTs reported imputing missing data separately in each arm of the trial. A further two RCTs mentioned the use of treatment arm as a variable in the imputation model, one article including it, the other specifically excluding it as a 'conservative measure.'

Very few papers reported any assessment of the integrity of imputations. Individual articles reported distributional suitability of variables to be imputed, the stability of the imputation process and the plausibility of the imputed values, but this was rare.

Analysis status

Fifty-nine papers used MI in the primary analyses; those which addressed the principal aims of the research. The remaining forty papers used MI in secondary analyses aimed at examining the impact of missing data on their conclusions. The primary analyses in these papers used a different approach, usually CCA. In just over half (21) of such papers, the MI results were not reported and were usually described as yielding identical or comparable results to CCA.

Just under half of the studies that undertook secondary MI analyses (17 out of 40) described them as sensitivity analyses. Some unique approaches to reporting MI results were found. These included reporting the results of each imputed dataset or reporting only results where both CCA and MI gave significant outcomes [12]. Reporting MI results in preference to CCA was very rare [13].

Discussion

The use of MI in medical research is increasing, notably in the analysis of RCTs. Currently, readers of research reporting analyses of multiply imputed data must take much on trust. Most papers fail to include critical details of the data imputed or the nature of the imputation process. Thus, it is difficult for any reader to assess the appropriateness of the analyses reported and validity of conclusions reached.

Missingness

Whilst a CONSORT style flow diagram is informative, it may not include all required information, as individuals may have missing measurements even though they were still participating in the trial. The absence of reporting standards for other types of studies often means that problems in the interpretation of results are increased. It is surprising that many of the articles surveyed did not include clear details of the extent or patterns of missingness, as the use of MI suggests that the authors believed that missingness was a problem. All articles using MI should provide a concise summary of both variable-wise and casewise missing rates. Patterns of missingness should also be described. Where this is complex, a summary table of the frequency of each pattern of missingness should be made available (see Table S1). Where a large proportion of data has been imputed, the method and assumptions underlying the procedure are likely to be much more critical than when only a small proportion of observations are involved. The plausibility of the MAR assumption may be able to be judged if information about predictors and the pattern of missingness is presented.

Implementation and imputation model

Failure to provide adequate details of the imputation procedure is perhaps the greatest deficiency of the articles reviewed. In many papers, which provided no explicit information about imputation procedures, details might be inferred from the software used, assuming the implementation of standard procedures and program defaults. This leads to the conclusion that most imputation has been carried out using models involving only linear combinations of predictors. In contrast, analysis of trial data almost invariably involves testing an interaction or differential effect of treatment arm over occasions of measurement. Interactions, in the form of moderators of risks or combined exposures are also important in epidemiological studies. There was little indication that the relationship between the imputation and analytic models was recognized in the papers reviewed. Three RCTs [14–16] reported using an approach that would accommodate the possibility of an interaction. All of these trials imputed data separately in each arm of the trial.

The robustness or sensitivity of MI to imputation models that differ from the analytic model used is dependent on a range of factors, including the amount of missing data, degree of mismatch and whether the analytic model is true. Schafer [4] has described the relationship between the imputation and analytic models as a forgiving one. However, the

dominant use of MI is in RCTs where this cannot be assumed to be the case. Allison [6] provides a striking demonstration of not allowing for an interaction in the imputation model, resulting in failure to detect a significant interaction effect in subsequent modelling. Appropriate imputation models and methods are required to ensure that parameter estimates are unbiased and that standard errors are optimized [6, 7].

The choice of variables used has been shown to be critical [8] to the quality of imputation. Further, the use of additional variables, which are not of direct interest to the analyses in the imputation model as predictors to improve imputation, is an advantage of MI. Failure to document the variables used prejudices the ability to replicate findings. Also crucial to the use of MI is consideration of the suitability of the imputation procedure for the variables being imputed and the integrity of the imputations. Readers cannot know whether these factors have been overlooked by authors or simply not reported.

In the context of the reporting of analyses using MI, the form of the imputation model and the variables included should be stated clearly as should its relationship with the analytic model(s).

Secondary and sensitivity analyses

The logic behind the use of MI as a test of the robustness of complete case analysis or as a sensitivity analysis is flawed. CCA makes the strict assumption that data are MCAR whereas MI requires the more liberal assumption that missingness is at random, conditional on observed data. Emerging research suggests the superiority of MI [17, 18]. Beyond its familiarity to authors and readers, there is little basis for preferring a technique that requires more restrictive assumptions over one with more relaxed assumptions. Thus, *a priori*, results from properly undertaken MI should be preferred and be reported. Agreement between CCA and MI analyses seems to be taken as an indication that data are MCAR. This is false reassurance. It is quite possible for a mechanism to operate that produces nonignorable missingness, yet results in comparable outcomes in these methods. Sensitivity analysis involves exploring the effect of plausible but unobserved missingness mechanisms – for instance, assuming that cases who drop out or with missing data differ from observed cases in a substantial way, which cannot be accounted for by what is known about them. Robustness of the analyses to such mechanisms can increase confidence in the conclusions reached. Any analysis assuming MAR does not do this and is, at best, only the mildest of sensitivity analyses.

Quality of the articles reviewed

The focus of this survey has been on shortcomings of reporting the use of MI. There were, however, articles which gave full details of the MI procedure as outlined in Table 1, either in the body of the article itself or in supplementary documentation. This demonstrates that full reporting of MI within the constraints of journal articles is possible. It is also important that concern expressed about the reporting of MI analyses should not be taken as criticism of the research or the analyses themselves. Many of the articles reviewed were innovative and rigorous in their dealing with missing data. In using MI, the authors of these studies recognized the limitations and potential biases of common approaches to missingness. The principal argument of the current review is not that MI is invariably poorly undertaken, but that the level of reporting makes this difficult to judge.

Conclusions

Multiple imputation is a powerful method for dealing with missing observations, a pervasive problem in medical research. Its incorporation into software packages and its increasing use in analyses is welcome. However, inappropriate or sub-optimal use of MI can lead to biased results and erroneous or unsafe conclusions. Current reporting of MI is inadequate for readers to assess its application in most studies which have used it. This is highlighted in applications where there is debate or uncertainty regarding the best methods of implementing MI. Failure to report details of techniques used may cast doubt over the results obtained [19]. In effect, when MI is used, normal standards of scientific reporting should be applied to this step in the data analysis: MI should not be perceived as a routine background procedure. Given the differing uses to which MI is put, it is unlikely that a single set of reporting standards could be developed for MI. There are specific requirements for trials and other major types of studies. Where it is not possible to include all the information in the body of an article, supplementary material should be held and made available by the publisher. The current survey reinforces the recent paper by Sterne and colleagues [20], which introduces MI and also highlights similar pitfalls in its application to those identified here.

It is also important that users of MI and readers of research using it develop a level of literacy about the technique, particularly about issues surrounding missingness mechanisms, the assumptions and requirements of MI, and imputation models and methods. This will ensure that MI does not create more problems than it solves.

Conflict of interest statement

No conflict of interest to declare.

Author's contributions

The author undertook all aspects of the research reported in this paper.

References

1 Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley, 1987.
2 Rubin DB. Multiple imputation after 18 + years. *J Am Stat Assoc* 1996; **91:** 473–89.
3 SPSS Inc. *SPSS Statistics for Windows 17.0.* Chicago, IL: SPSS Inc, 2008.
4 Schafer JL. *Analysis of incomplete multivariate data.* London: Chapman & Hall, 1997.
5 Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods* 2002; **7:** 147–77.
6 Allison PD. *Missing Data. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136.* Thousand Oaks, CA: Sage, 2001.
7 Meng X-L. Multiple-Imputation inferences with uncongenial sources of input. *Stat Sci* 1994; **9:** 538–58.
8 Moons KGM, Donders RART, Stijnenc T, Harrell FE Jr. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* 2006; **59:** 1092–101.
9 Little RJA. Modeling the drop-out mechanism in repeated measures study. *J Am Stat Assoc* 1995; **90:** 1112–21.
10 Little RJA. A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc* 1988; **83:** 125–34.
11 Moher D, Schulz KF, Altman D, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001; **285:** 1987–91.
12 Melhuish E, Belsky J, Leyland AH, Barnes J. Effects of fully-established sure start local programmes on 3-year-old children and their families living in England: a quasi-experimental observational study. *The Lancet* 2008; **372:** 1641–7.
13 Smith GCS, Wood AM, Pell JP, Dobbie R. Sudden infant death syndrome and complications in other pregnancies. *The Lancet* 2005; **366:** 2107–11.
14 Miranda J, Chung JY, Green BL *et al.* Treating depression in predominantly low-income young minority women: a randomized controlled trial. *JAMA* 2003; **290:** 57–65.
15 Snape MD, Perrett KP, Ford KJ *et al.* Immunogenicity of a tetravalent meningococcal glycoconjugate vaccine in infants: a randomized controlled trial. *JAMA* 2008; **299:** 173–84.
16 Unutzer J, Katon W, Callahan CM *et al.* Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA* 2002; **288:** 2836–45.
17 van der Heijden GJMG, Donders ART, Stijnene T, Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol* 2006; **59:** 1102–9.
18 Janssen KJM, Donders ART, Harrell FE Jr *et al.* Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol* 2010; **63:** 721–7.
19 Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol* 2009; **9:** 57.
20 Sterne JAC, White IR, Carlin JB *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; **338:** b2393.

Correspondence: Andrew Mackinnon, Centre for Youth Mental Health, University of Melbourne, Locked Bag 10, Parkville, Victoria, 3052, Australia.
(fax: +61 3 9342 2941; e-mail: Andrew.Mackinnon@unimelb.edu.au).

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1. Example of a compact display of patterns of missingness in an RCT with three covariates and three occasions of measurement of outcome.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.∎