



Published in final edited form as:

Clin Trials. 2012 August ; 9(4): 396–407. doi:10.1177/1740774512450098.

Including all individuals is not enough: lessons for intention-to-treat analysis

Ian R. White, M.Sc.,

MRC Biostatistics Unit, Cambridge, UK

James Carpenter, D.Phil., and

London School of Hygiene and Tropical Medicine, UK

Nicholas J. Horton, Sc.D.

Smith College, Northampton, USA

Abstract

Background—Intention-to-treat analysis requires all randomised individuals to be included in the analysis in the groups to which they were randomised. However, there is confusion about how intention-to-treat analysis should be performed in the presence of missing outcome data.

Purpose—To explain, justify and illustrate an intention-to-treat analysis strategy for randomised trials with incomplete outcome data.

Methods—We consider several methods of analysis and compare their underlying assumptions, plausibility, and numbers of individuals included. We illustrate the intention-to-treat analysis strategy using data from the UK700 trial in the management of severe mental illness.

Results—Depending on the assumptions made about the missing data, some methods of analysis that include all randomised individuals may be less valid than methods that do not include all randomised individuals. Further, some methods of analysis that include all randomised individuals are essentially equivalent to methods that do not include all randomised individuals.

Limitations—This work assumes that the aim of analysis is to obtain an accurate estimate of the difference in outcome between randomised groups, not to obtain a conservative estimate with bias against the experimental intervention.

Conclusions—Clinical trials should employ an intention-to-treat analysis strategy, comprising a design that attempts to follow up all randomised individuals, a main analysis that is valid under a stated plausible assumption about the missing data, and sensitivity analyses which include all randomised individuals in order to explore the impact of departures from the assumption underlying the main analysis. Following this strategy recognises the extra uncertainty arising from missing outcomes and increases the incentive for researchers to minimise the extent of missing data.

Keywords

Intention-to-treat analysis; missing data; sensitivity analysis; mixed models; last observation carried forward; analysis of covariance; multiple imputation; clinical trials

1 Introduction

Intention-to-treat (ITT) analysis is essential in avoiding bias in the analysis of randomised trials [1]. The ITT principle states that all individuals randomised in a clinical trial should be included in the analysis, in the groups to which they were randomised, regardless of any departures from randomised treatment. By following this principle, data analysts preserve the benefit of randomisation in creating treatment groups that do not differ systematically on any factors except those assigned in the trial, whereas not following the ITT principle risks introducing selection bias.

One implication of the ITT principle is that investigators should aim to collect outcome data on all randomised individuals. It is essential to maximise the extent of outcome data collected by careful trial design, including appropriate eligibility criteria, attention to the burden of data collection on participants, and by energetic measures to remain in contact with participants and regain contact with lost participants. Further information is given by [2–4].

Despite investigators' best efforts, missing outcome data are common. From a statistical perspective, any analysis of a clinical trial with incomplete outcome data makes untestable assumptions. For example, it is often assumed that the data are missing at random (MAR), which means that missing data are equal in distribution to observed data, conditional on other variables included in the analysis [5]. Some analyses may make the stronger assumption that the data are missing completely at random (MCAR), which means that missing data are unconditionally equal in distribution to observed data. It is essential that the assumptions made are transparent and plausible, based on knowledge of the trial and the subject matter area.

A recent report by the Committee on National Statistics (CNSTAT) for the US National Academy of Sciences clarifies many of the design and analysis issues [4]. In particular, it stresses the importance of careful pre-specification of the causal estimands of primary interest (Recommendation 1); choosing designs that minimise treatment withdrawal (Recommendation 2); pre-specification of statistical methods and their assumptions in a way that can be understood by clinicians (Recommendation 9); and collecting ancillary data that are associated with reasons for missing values, and/or intensively following up a sample of non-respondents (Recommendation 15). It describes analysis methods for trials with incomplete data, focusing on methods that assume MAR (chapter 4). It then argues forcefully for analyses that explore the sensitivity of the results to departures from MAR (Recommendation 16), and extensively describes how such sensitivity analyses could be performed (chapter 5), although methodology for sensitivity analysis is noted as requiring more statistical research (Recommendation 20).

However, there is confusion about how the ITT principle should be applied in the presence of missing outcome data. A strict view would hold that no analysis with missing outcome data can be described as ITT, but such an unattainable standard is unhelpful. The explanatory paper to the 2001 revision of the CONSORT statement suggested acceptance of an analysis of observed data: “Although those participants [who drop out] cannot be included in the analysis, it is customary still to refer to analysis of all available participants as an intention-to-treat analysis” [6]. On the other hand, Hollis and Campbell argued that “Complete case analysis, which was the approach used in most trials, violates the principle of intention to treat” [7]. Often, ITT is taken to require imputation: the European Medicines Agency wrote “The statistical analysis of a clinical trial generally requires the imputation of values to those data that have not been recorded...” [8], and Altman wrote “No analysis option is ideal here; there is, in effect, a choice between omitting participants without final

outcome data or estimating (imputing) the missing outcome data” [9]. In new advice, the European Medicines Agency takes a more relaxed view: “Full set analysis generally requires the imputation of values or modelling for the unrecorded data” [10], and the 2010 CONSORT checklist no longer includes the “widely misused” phrase “intention to treat analysis” [11], and instead separately asks whether the analysis was by original assigned groups and what numbers were included in the analyses.

To resolve this confusion, we recently proposed a four-point *ITT analysis strategy* for trials with incomplete outcome data [12]:

1. Attempt to follow up all randomised individuals, even if they withdraw from allocated treatment.
2. Perform a main analysis that is valid under a plausible assumption about the missing data and uses all observed data
3. Perform sensitivity analyses to explore the impact of departures from the assumption made in the main analysis.
4. Account for all randomised individuals, at least in the sensitivity analyses.

The aim of this paper is to detail the rationale underpinning this strategy and to illustrate its application. We assume that interest lies in testing and estimating the effect of treatment assignment on clinical outcomes over all randomised individuals: this is the ‘ITT estimand’ or the ‘ITT treatment effect’ and is usually the most clinically- and policy-relevant estimand in large-scale randomised trials. We do not consider: other possible estimands discussed in the CNSTAT report, relating to subsets who adhere to treatment; estimating the causal effect of treatment itself, although this may be a useful ancillary analysis [13]; or estimating the effect of treatment assignment on a composite outcome that includes missingness as one component (common in HIV trials aimed at comparing HIV RNA levels for antiretroviral drugs, where missing values are taken as failures [14], but difficult to interpret clinically).

Throughout the paper, we consider an outcome either measured at just one time point, or measured repeatedly where interest lies mainly in the treatment effect at the last time. Our arguments would apply equally when interest lies in an average outcome such as the area under the curve. We mainly discuss quantitative outcomes, and consider other outcome types in the discussion. Our focus is on missing values in the outcome, although in our example we also deal with missing values of baseline variables.

The paper is organised as follows. In Section 2, we describe various commonly used methods of analysis, and their underlying assumptions. Section 3 details the rationale underpinning the four points of the ITT analysis strategy. Section 4 shows why the ITT analysis strategy does not require all randomised individuals to be included in the main analysis. Section 5 uses the UK700 trial in mental health to exemplify the ITT analysis strategy, and then to illustrate the argument in Section 4. We conclude with a discussion in Section 6.

2 Methods and assumptions

In this section, we discuss various methods of analysis, noting whether they include all randomised individuals and elucidating their underlying assumptions.

2.1 Last observation carried forward

Last observation carried forward (LOCF) replaces missing final outcomes by the last observed outcome (which could be the baseline value of the outcome). While it is widely

used [15], and attractive because it usually allows all individuals to be included in the analysis, it has been widely criticised [16–21].

The assumption underlying LOCF is often mis-stated. When the analysis is an unadjusted comparison of means or proportions, LOCF is unbiased if, in each randomised group, the mean of the unobserved values of the final outcome equals (in expectation) the mean of the last observed outcomes in the individuals who drop out. We call this the LOCF assumption. When the analysis is covariate-adjusted, the LOCF assumption is conditional on covariates. LOCF does not require the data to be MCAR, although some authors claim that it does [17, 18]: MCAR would instead require the missing data to be equal to the observed data in expectation at the final time point [5].

If the LOCF assumption is false, bias in LOCF analyses can arise in various ways. If there is a treatment effect at intermediate times but not at the final time, then carrying forward intermediate values can artifactually create a treatment effect at the final time. If unobserved outcomes improve over time, then LOCF tends to favour treatment groups with less drop-out, while if unobserved outcomes deteriorate over time, then LOCF tends to favour treatment groups with more drop-out.

LOCF validly estimates weighted averages of subgroup-specific means at different time points [22], but the weights may differ between randomised groups, so this parameter lacks clinical interest and causal interpretation [19]. LOCF is also sometimes defended as being conservative: for conditions that tend to improve over time, it is indeed likely to be conservative for arm-specific mean outcomes, but its bias for the estimated treatment effect is not necessarily in a conservative direction [20]. An appropriate justification of LOCF should argue that average *unobserved* outcomes within each randomised group do not change over time; we have never seen such a justification. Instead, analysts more commonly attempt to justify LOCF by the stability over time of *observed* outcomes, which is not a sufficient argument [21].

2.2 Missing=failure

In some clinical areas, it is common to assume that missing values represent failures. This is only possible when the outcome is categorical (usually binary): for example, in smoking cessation studies (e.g. [23]). ‘Missing=failure’ is the same as LOCF when the outcome is a measure of improvement observed at just one time point.

Like LOCF, ‘missing=failure’ makes it easy to include all randomised individuals in the analysis. However, the underlying assumption needs to be carefully justified. In particular, ‘missing=failure’ logically implies that every success is actually observed, often a rather implausible assumption. If the assumption is false then ‘missing=failure’, like LOCF, gives conservative estimates of outcomes within randomised groups, but not necessarily a conservative estimate of the difference between groups, especially if the amounts of missing data, or their reasons, differ between randomised groups.

2.3 Complete-case analysis

In a trial with outcome measured at one time point, a complete-case analysis typically involves a simple outcome comparison between groups or an analysis of covariance in which the outcome is regressed on randomised group, adjusted for baseline variables. These analyses are valid under the assumption that response is MAR given randomised group or given randomised group and baseline, respectively, and can be viewed as likelihood-based.

In a trial with outcome measured repeatedly, a complete-case analysis would typically exclude any individual whose outcome was not observed at the final follow-up time.

Excluding individuals whose outcome is observed at intermediate follow-up times is clearly inappropriate. However, in a survey of 35 such trials, 17 used a complete-case analysis [15]. Likelihood-based analysis of all observed data is preferable.

2.4 Likelihood-based methods

A likelihood-based analysis fits a suitable statistical model to all the observed data. Often this would be a linear mixed model [24]. Likelihood-based analyses (including Bayesian analyses) implicitly assume that the data are MAR, unless the missing data mechanism is explicitly modelled. In the case of a trial with outcome measured repeatedly, this means that missing data are equal in distribution to observed data, conditional on the baseline and follow-up variables included in the analysis [5]. With non-monotone missing data patterns, the MAR assumption can be hard to interpret [25].

2.5 Multiple imputation

Multiple imputation (MI) is a broadly applicable technique for handling missing data [26, 27]. MI is usually able to include all randomised individuals in the analysis. Briefly, missing data are imputed more than once, in a way that reflects the uncertainty about the missing values. In Rubin's formulation, each imputed data set is analysed by standard methods, and the point estimates and standard errors are combined to provide inferences that reflect the uncertainty about the missing values [26]. Standard implementations of MI [28–32] assume MAR, although in principle MI may be performed under other missing data mechanisms. Other formulations of MI may provide more accurate standard errors in some less-standard settings, but are not available in standard software [33].

Many MI analyses can be viewed as computationally convenient approximations to likelihood-based analyses based on the observed data [34]. For example, if the variables used in imputing the missing data correspond to the variables in the analysis model and a (multivariate) Normal assumption is made in both analyses, then a MI analysis approximates a likelihood-based analysis. The quality of the approximation is determined by the Monte Carlo error inherent in MI analysis, which decreases as the number of imputations increases [35].

In some cases, an MI procedure can be improved by including in the imputation model 'auxiliary variables' that are not in the analysis model [36, Chapter 4]: auxiliary variables in a randomised trial might be secondary outcomes or compliance summaries. MI then produces estimates of the treatment effect that are genuinely different from a likelihood-based analysis, by incorporating information on individuals with missing outcome but observed values of auxiliary variables. However, in our experience, the contribution to such an analysis of individuals missing the outcome of interest is moderate unless correlations between the outcome and one or more auxiliary variables are substantial [37].

2.6 Illustration

Three different assumptions are explored in Figure 1, which depicts mean outcomes in one arm of a randomised trial. Higher outcomes are assumed to be worse. Individuals with complete data (the solid line) start with mean outcome 10 and improve by a mean of 2 units at time 1, with this improvement sustained at time 2. Individuals who drop out after time 1 started with a better mean outcome and also had a mean improvement of 2 units at time 1. A LOCF analysis (depicted in the left-hand panel) assumes that this mean improvement was sustained up to time 2. An analysis based on MCAR (such as a complete-case analysis) assumes that individuals who drop out after time 1 are similar to completers at time 2, which in this example corresponds to their improvement being transient (middle panel). An analysis based on MAR (such as a likelihood-based analysis) assumes that the missing

outcomes at time 2 can be predicted using the relationship in completers between outcomes at the three times. Suppose that this relationship is $E[Y_2 | Y_0, Y_1] = \alpha + \beta Y_1$ with $\beta = 0.5$, where (Y_0, Y_1, Y_2) are the three outcomes. The observed mean difference between completers and dropouts at time 1 is 2 units, so the MAR assumption implies a mean difference of $2 \times \beta = 1$ unit at time 2 (right-hand panel).

3 ITT analysis strategy

Having discussed common analyses and their assumptions, we now discuss the rationale for the four-point ITT analysis strategy.

3.1 Attempt to follow up all randomised individuals, even if they withdraw from allocated treatment

This point refers to the design of trials in which patients may withdraw from their allocated treatment during the trial. Some trials do not attempt to follow up patients after treatment withdrawal. This has four serious disadvantages. First, it is contrary to the spirit of the ITT principle. Second, it means that all observed data are ‘on-treatment’ and so standard analyses based on observed data attempt to estimate an ‘on-treatment’ effect and not the ITT treatment effect [36]. (If the ‘on-treatment’ effect is really of interest then a different approach to the design and analysis may be appropriate [4].) Third, it often makes MAR less plausible, because individuals who stop trial treatment are often more highly selected than those who are simply lost to follow-up. Fourth, even if it introduces no bias, it can reduce the power of the trial if treatment effects are long-lasting [38].

We therefore believe that no primary analysis of such a trial should be described as ITT. Instead, trials should attempt to follow up all randomised individuals, including those who withdraw from treatment (an ‘ITT design’) [39, 40]. Individuals who have withdrawn from trial treatment tend to be harder to follow up, but if at least some data are collected, then analysis based on MAR can allow for treatment withdrawals and can attempt to estimate the ITT treatment effect [36, 41].

3.2 Perform a main analysis that is valid under a plausible assumption about the missing data and uses all observed data

This point emphasises the importance of assumptions. Any trial report should state the assumption made about the missing data in the main analysis – for example, MAR or the LOCF assumption – and give reasons why the assumption is plausible [42].

We require the inclusion of all observed outcome data. Analyses that exclude some observed outcome data would not be acceptable without strong rationale such as doubt over the integrity of the data. In particular, complete-case analysis of repeated measures data (Section 2.3) would not be consistent with the ITT analysis strategy.

The controversial point here is that we do not require the inclusion of all randomised individuals in the main analysis – that is, we do not require inclusion of individuals with no outcome measures – because the validity of an analysis is determined by whether its assumptions are correct: a valid estimate of the ITT estimand is consistent with the ITT principle. Thus, analyses of all observed data (such as mixed models) should be acceptable if the MAR assumption is reasonably plausible in the clinical context. Harmful consequences of requiring the inclusion of all randomised individuals in the main analysis are given in Section 4. Of course, analyses that *do* include all randomised individuals are acceptable if they make a plausible assumption: for example, in a smoking cessation trial, it might be plausible to assume that all individuals with missing outcomes are still smoking.

3.3 Perform sensitivity analyses to explore the impact of departures from the assumption made in the main analysis

All analyses with missing data make untestable assumptions, so it is always important to perform sensitivity analyses exploring the impact of departure from the assumptions [43, 44]. Appropriate sensitivity analyses should address departures from the assumptions that are relevant for the estimand at hand in an accessible way.

Unfortunately, many sensitivity analyses used in practice are inappropriate. For example, in one survey, the most common form of sensitivity analysis was LOCF when the primary analysis adopted a complete-case analysis [15]. Agreement between the results of LOCF and complete-case analysis is not necessarily reassuring because the assumptions underlying the two methods could both be wrong, and so both results could be biased. Figure 2 illustrates this problem. Although LOCF and MCAR impute the missing values at time 2 in different ways, they both impute the same mean value, 7. It would be wrong to derive reassurance from this agreement. In fact, an MAR analysis would impute a different value, 6.

Instead, a ‘principled’ sensitivity analyses should move smoothly away from the assumptions underpinning the primary analysis, in a way that is clinically plausible and accessible to those interpreting and using the study results. Kenward *et al.* describe the procedure thus: “It is necessary to properly parameterise the set of models considered by means of one or more continuous parameters and then to consider all or at least a range of models along such a continuum” [45]. For example, one might define a parameter δ equal to the difference between the mean of the observed data and the mean of the unobserved data, adjusted for other observed variables. Under an MAR analysis, δ is assumed to be zero. A sensitivity analysis would consider plausible alternative values of δ . It is important to consider the possibility that δ differs across randomised groups: for example, missing data after a psychological intervention may be further from MAR than after no intervention [46]. This idea underlies computational [47–49] and graphical [50, 51] approaches to sensitivity analysis.

We have focussed here on sensitivity analyses to the untestable assumptions about the missing data; it is also important to verify testable assumptions, such as a Normality assumption for outcome data, or the way baseline covariates are entered in the model [4], although estimated treatment effects are usually far more robust to departures from testable than untestable assumptions.

3.4 Account for all randomised individuals, at least in the sensitivity analyses

A key feature of a principled sensitivity analysis as described above is that all individuals must be included in the analysis. For example, if $\delta \neq 0$ so that missing values differ from observed values, then a complete-case analysis is no longer acceptable. Thus, although an analysis based on MAR need not include all randomised individuals, analyses assuming departures from MAR must include them.

This point provides a key link with previous conceptions of intention-to-treat analysis: inclusion of all randomised individuals is important, but the place of that inclusion is in the sensitivity analysis.

4 Harmful consequences of requiring inclusion of all randomised individuals in the main analysis

4.1 Implausible assumptions

We compare LOCF and likelihood-based analyses (noting that complete-case analysis of a trial with outcome measured at one time point is effectively a likelihood-based method). The different assumptions underlying these methods were described in Section 3. The methods also differ in which randomised individuals are included: if the baseline observation is complete then LOCF includes all individuals in the analysis, but likelihood-based methods exclude individuals who provide no post-baseline outcome data.

The MAR assumption is often seen as a natural starting point for analysis [17, 52]. A stronger belief in MAR than other assumptions led Molenberghs *et al.* to write, “A likelihood based ignorable analysis should be seen as a proper way to accommodate information on a patient with postrandomization outcomes, even when such a patient's profile is incomplete” and “This fact, in conjunction with the use of treatment allocation as randomized rather than as received, shows that [a mixed model analysis] is fully consistent with ITT” [17]. These authors do not explain what they mean by ITT, but seem to be arguing that if an analysis is suitable, it must conform to ITT.

We avoid blanket statements about the plausibility of particular assumptions: this must instead be determined in each trial using subject-matter knowledge. However, in some trials, MAR is more plausible than the LOCF assumption. In such trials, an MAR-based analysis excluding individuals who provide no post-baseline outcome data would be preferable to an LOCF analysis including them. Thus requiring inclusion of all randomised individuals in the main analysis would invite analysts to adopt a less plausible assumption.

4.2 Unnecessary complexity

We now describe two situations where simple analyses that do not include all randomised individuals are approximately equivalent to, and make the same assumption as, more complex analyses that do include all randomised individuals.

First, when a likelihood-based analysis is used, baseline values of the outcome can be included either as a covariate or as an outcome [53, 54]. For example, a trial with outcome measured at baseline and one follow-up time can use either an analysis of covariance (ANCOVA) or a mixed model with baseline and follow-up as a bivariate outcome. These methods give identical point estimates and very similar standard errors when the baseline is complete [53]. They also give very similar results when the baseline is incomplete, provided a suitably modified ANCOVA avoids dropping individuals with missing baselines [55, 56]. However, the analysis using baseline as an outcome includes all individuals with the outcome observed at baseline or follow-up, whereas the analysis using baseline as a covariate includes only those individuals with the outcome observed at follow-up.

Second, MI may be a computationally convenient alternative to a likelihood-based analysis (Section 2.5), and it typically includes all individuals in the analysis; but in the absence of strong auxiliary variables, MI may be inferior to likelihood-based analysis unless the number of imputations is large enough to minimise Monte Carlo error. However, authors' desire to include all randomised individuals in the analysis favours MI: for example, in a trial of a web-based self-help intervention for problem drinkers, the authors claimed “We then performed intention-to-treat analysis, using multiple imputation to deal with loss to follow-up” [57].

In both cases, requiring all randomised individuals to be included in the analysis would invite analysts to adopt an unnecessarily complex analysis, with consequent greater opportunity for human error.

5 Case study: the UK700 trial

We use the UK700 trial to illustrate both the ITT analysis strategy and the arguments of Section 4. This trial compared intensive case management with standard case management for 708 people with severe mental illness living in the community [58]. We here consider two outcomes: psychopathology score (CPRS) and satisfaction with services (SAT), which were measured in interviews at baseline, year 1 and year 2. A third outcome, days in hospital for mental health reasons (HOS), was recorded at baseline and year 2 from hospital notes and therefore had few missing values. Key variables are summarised in Table 1.

Missing data in CPRS occurred mainly when individuals did not attend interviews at years 1 and 2. Missing data in SAT occurred additionally because the variable was not included in early versions of the baseline interview and because some interviews were incomplete (SAT, unlike CPRS, came near the end of the interview). Missing data patterns are summarised in Table 2.

5.1 ITT analysis strategy for UK700

For point 1, attempts were made to follow up all randomised individuals.

For point 2, we need a plausible assumption for a main analysis. In this mental health setting, individuals with missing values may have worse psychopathology and greater dissatisfaction than observed individuals, and their psychopathology and dissatisfaction may have worsened over time. Thus neither the LOCF assumption nor MAR seems entirely satisfactory. The published analysis was based on an MAR assumption, and we follow that here, recognising that sensitivity analysis to departures from MAR will be essential. We can make the MAR assumption more plausible (and possibly gain precision) by including in the analyses the third outcome, HOS, which was more completely observed than CPRS and SAT: this will be done in a sensitivity analysis. The assumption underlying LOCF cannot be amended to account for HOS.

The chosen main analysis is therefore a mixed model for CPRS or SAT, using all the observed data at all three time points, and adjusted for trial centre and the baseline value of the outcome variable. Covariate effects varied by year (that is, the model included interactions between year and covariates); treatment effects also varied by year but were absent at baseline. Outcome covariance matrices were unstructured but equal across arms. The estimated intervention effect (95% confidence interval) was -0.39 units (-2.40 to +1.62) on CPRS and -0.35 (-1.15 to +0.45) on SAT.

For the SAT analysis, 101 randomised individuals have missing baseline values but one or more observed outcomes. Suitable methods for including these individuals in the analysis can be surprisingly simple, because the role of baseline covariates in randomised trials is only to increase power and not to remove confounding [55]. Thus mean imputation methods, which are inappropriate for missing covariates in non-randomised studies [59], are appropriate for missing baseline values in randomised trials, provided that the imputed values respect the independence of baseline values and randomised group [55]. In the analysis described above, missing baseline values of SAT were imputed by the centre-specific mean of the observed baseline values.

For point 3, we require sensitivity analyses exploring the impact of departures from the MAR assumption: we illustrate them for the CPRS outcome. The main analysis assumed that $\delta = 0$, as defined in Section 3.3. Positive values of δ indicate that missing individuals have worse psychopathology than observed individuals, which seems the likely direction of departure from MAR in a mental health context. Let f_1 and f_0 be the fractions of individuals with missing outcome at the final time in the intervention and control arm respectively: in the UK700 data, $f_1 = 0.12$ and $f_0 = 0.20$. The sensitivity analysis is done by adding a quantity Δ to the treatment effect estimated under the MAR assumption, where $\Delta = f_1 \delta$ if data depart from MAR in the intervention arm only, $\Delta = -f_0 \delta$ if data depart from MAR in the control arm only, and $\Delta = (f_1 - f_0) \delta$ if data depart from MAR in the same way in both arms. We allow δ to take values from 0 to 10: since the standard deviation of CPRS is 14 (Table 1), this represents a fairly wide range. We make the approximation that the standard error of the parameter estimate is unaffected by the sensitivity analysis: other work, the subject of a future report, shows that this approximation works well over a wide range of δ . More generally, we could allow δ to take values δ_0 and δ_1 in the intervention and control arm respectively, so that $\Delta = f_1 \delta_1 - f_0 \delta_0$. A fuller treatment of sensitivity analysis, including expert elicitation of the range of values for δ , is given in [60].

Figure 3 shows how the estimated intervention effect varies in the sensitivity analyses. Departures from MAR have more impact in the control arm than in the intervention arm, because $f_0 > f_1$. Under MAR, the trial showed no significant benefit of intervention; for this conclusion to be changed would require missing CPRS values to average some 8 points (more than half a standard deviation) more than the observed values in the control arm only, which seems relatively implausible.

A second sensitivity analysis used MI with auxiliary variables to make better use of the observed data and to make the MAR assumption somewhat more plausible. The auxiliary variables were the baseline and follow-up values of the other two outcomes (HOS and CPRS for SAT; HOS and SAT for CPRS). MI was implemented by the MICE algorithm [28,61,62]. Monte Carlo error was reduced by using 1000 imputed data sets [63]. The estimated intervention effect (95% confidence interval) was -0.43 units (-2.43 to +1.58) on CPRS and -0.40 (-1.20 to +0.39) on SAT, which show much less difference from the main results than does the sensitivity analysis in Figure 3.

For point 4, all randomised individuals are included in this set of analyses, because each missing individual contributes to one of the quantities f_1 or f_0 .

5.2 Comparison of different analyses

We performed further mixed model, LOCF and MI analyses to illustrate the arguments of Section 4.

We first consider results ignoring the year 1 data (Table 3, top part), thus illustrating results for a trial with outcome measured at one time point. ANCOVA with mean-imputed missing baselines and a mixed model with baseline as outcome give almost identical results for CPRS and very similar results for SAT; greater differences are expected for SAT since it has more missing values at baseline. MI, using a basic imputation model including all variables from the analysis model, gave results very similar to the other methods. However, these three analyses that give similar results, and rest on similar assumptions, include very different numbers of individuals in the analysis.

We next consider analyses using data from all three time points (Table 3, lower part). The LOCF estimate differs substantially from all the other estimates for CPRS and has smaller standard error than the other methods, because its implicit assumption allows more

information to be drawn from individuals with missing data: this suggests that greater caution needs to be attached to LOCF analyses. Mixed model analysis of available cases gives very similar estimates whether baseline is included as a covariate or as an outcome. MI using a basic imputation model agrees closely with mixed model analysis of available cases, and MI using the extended imputation model of Section 5.1 shows small changes as noted there. Again, methods based on the same missing data assumption – mixed models on available cases, whether with baseline as outcome or covariate, and MI using the basic imputation model – give very similar answers, as theory suggests, despite including very different numbers of individuals.

These results illustrate that the choice of assumption matters far more than how many individuals are included in the analysis.

6 Discussion

We believe that excessive focus on including all individuals in the analysis of randomised trials with missing outcomes can lead to a choice of analysis that rests on implausible or unnecessarily complex imputation. In the ITT analysis strategy, we have therefore proposed that the main focus in choosing the analysis should be the plausibility of its assumptions, while inclusion of all randomised individuals is a requirement only for sensitivity analyses.

Our approach has been to obtain the best possible estimate of the intervention effect. Some analyses, particularly LOCF, are popular because they are believed to be conservative, but this is misguided [20]. It is hard to be sure that an analysis is conservative without attempting to compare it with an unbiased estimate of the intervention effect. We believe that conservatism is best achieved by attempting unbiased estimation but appropriately allowing for the uncertainty due to the missing data [46].

We have discussed incomplete quantitative outcomes. Our proposal for an ITT analysis strategy applies equally well with other outcomes. However, some different modelling issues arise. For trials with repeatedly measured incomplete binary outcomes, when interest lies in a treatment effect on the log odds scale, complications arise because of the differences between ‘population-averaged’ and ‘subject-specific’ approaches [64]. The goal of ITT analysis is usually a population-averaged odds ratio, which can be directly estimated by generalised estimating equations and multiple imputation, but not by mixed models, which directly estimate the subject-specific odds ratio. For trials with time-to-event outcomes, the missing data are the censored outcomes, and in practice the plausible assumption about the missing data is nearly always that censoring is non-informative (similar to an MAR assumption). Methods for sensitivity analysis to informative censoring are not well developed.

Our considerations have led us to propose a framework for ITT analysis with missing data that complements and extends the CNSTAT report [4]. We believe that if trialists follow this framework then there is scope for considerable improvement in the appropriateness, consistency, and reporting of ITT analyses when outcomes are missing. However, the best approach to missing data is always to design and conduct the trial to maximise data collection [2, 3]. A careful ITT analysis strategy, and in particular an appropriate sensitivity analysis, recognises the increase in uncertainty that arises from missing outcomes, and therefore increases the incentive for researchers to maximise their data completeness.

Acknowledgments

Grant support: Medical Research Council Unit Programme number U105260558 (IRW), ESRC Research Fellowship RES-063-27-0257 (JC), NIH grant 5R01MH054693-11 (NJH).

References

1. Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *International Journal of Epidemiology*. 1992; 21:837–841. [PubMed: 1468842]
2. McKnight, P.; McKnight, KM.; Sidani, S.; Figueredo, AJ. *Missing data: A gentle introduction*. The Guilford Press; New York: 2007.
3. PSI Missing Data Expert Group. *Missing data: Discussion points from the PSI missing data expert group*. *Pharmaceutical Statistics*. 2009;10.1002/pst.391
4. National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press; Washington, DC: 2010. Panel on Handling Missing Data in Clinical Trials Committee on National Statistics, Division of Behavioral and Social Sciences and Education.
5. Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. 2nd. Wiley; Hoboken, N. J.: 2002.
6. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine*. 2001; 134:663–694. [PubMed: 11304107]
7. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Medical Journal*. 1999; 319:670–4. [PubMed: 10480822]
8. Committee for Proprietary Medicinal Products (CPMP). Points to consider on missing data. 2001. <http://www.emea.europa.eu/pdfs/human/ewp/177699EN.pdf>
9. Altman D. Missing outcomes in randomized trials: addressing the dilemma. *Open Medicine*. 2009; 3(2):e51. [PubMed: 19946393]
10. European Medicines Agency. *Guideline on missing data in confirmatory clinical trials*. 2010. <http://www.ema.europa.eu/ema/pages/includes/document/open-document.jsp?webContentId=WC500096793>
11. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *British Medical Journal*. 2010; 340:c332. [PubMed: 20332509]
12. White IR, Horton N, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *British Medical Journal*. 2011; 342:d40. [PubMed: 21300711]
13. White IR. Uses and limitations of randomization-based efficacy estimators. *Statistical Methods in Medical Research*. 2005; 14:327–347. [PubMed: 16178136]
14. Center for Drug Evaluation and Research (CDER). [accessed 18 Oct 2010] *Guidance for industry: Antiretroviral drugs using plasma HIV RNA measurements – clinical considerations for accelerated and traditional approval*. 2002. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatory-Information/Guidances/UCM070968.pdf>
15. Wood A, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials*. 2004; 1:368–376. [PubMed: 16279275]
16. Kenward MG, Molenberghs G. Last observation carried forward: A crystal ball? *Journal of Biopharmaceutical Statistics*. 2009; 19:872–888. [PubMed: 20183449]
17. Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, Carroll RJ. Analyzing incomplete longitudinal clinical trial data. *Biostatistics*. 2004; 5:445–464. [PubMed: 15208205]
18. Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharmaceutical Statistics*. 2008; 7:93–106. [PubMed: 17351897]
19. Carpenter J, Kenward M, Evans S, White I. 'Last observation carryforward and last observation analysis' by J. Shao and B. Zhong (*Statistics in Medicine* 2003; 22: 2429-2441). *Statistics in Medicine*. 2004; 23:3241–2. [PubMed: 15449330]
20. Siddiqui O, Hung HM, O'Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *Journal of Biopharmaceutical Statistics*. 2009; 19:227–246. [PubMed: 19212876]
21. Heyting A, Tolboom JTBM, Essers JGA. Statistical handling of dropouts in longitudinal clinical trials. *Statistics in Medicine*. 1992; 11:2043–2061. [PubMed: 1293667]
22. Shao J, Zhong B. Last observation carry-forward and last observation analysis. *Statistics in Medicine*. 2003; 22:2429–2441. [PubMed: 12872300]

23. Sutton S, Gilbert H. Effectiveness of individually tailored smoking cessation advice letters as an adjunct to telephone counselling and generic self-help materials: randomized controlled trial. *Addiction*. 2007; 102:994–1000. [PubMed: 17523995]
24. Verbeke, G.; Molenberghs, G., editors. *Linear Mixed Models in Practice*. Springer-Verlag; New York: 1997.
25. Potthoff RF, Tudor GE, Pieper KS, Hasselblad V. Can one assess whether missing data are missing at random in medical studies? *Statistical Methods in Medical Research*. 2006; 15:213–34. [PubMed: 16768297]
26. Rubin, DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons; New York: 1987.
27. van Buuren, S. *Flexible Imputation of Missing Data*. CRC press; 2012.
28. Royston P. Multiple imputation of missing values. *Stata Journal*. 2004; 4:227–241.
29. Novo AA, Schafer JL. R package version 1.0-9.2. Package NORM. 2010
30. van Buuren, S.; Oudshoorn, CGM. *Multivariate Imputation by Chained Equations: MICE V1.0 User's manual*. Leiden: TNO Preventie en Gezondheid; 2000. TNO Report PG/VGZ/00.038. Available at <http://www.multiple-imputation.com/>
31. SAS Institute Inc.. *SAS/STAT 9.1 User's Guide*. SAS Institute Inc.; Cary, NC: 2004. chapter 46.
32. StataCorp. *Stata Multiple-Imputation Reference Manual*. Stata Press; College Station, TX: 2009.
33. Robins J, Wang N. Inference for imputation estimators. *Biometrika*. 2000; 87(1):113–124.
34. Meng XL. Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*. 1994; 9:538–558.
35. Royston P, Carlin JB, White IR. Multiple imputation of missing values: new features for mim. *Stata Journal*. 2009; 9:252–264.
36. Carpenter, JR.; Kenward, MG. *Missing Data in Clinical Trials — a Practical Guide*. Birmingham: National Institute for Health Research; 2008. Publication RM03/JH17/MK Available at http://www.pcpoh.bham.ac.uk/publichealth/methodology/projects/RM03_JH17_MK.shtml
37. White IR, Moodie E, Thompson SG, Croudace T. A modelling strategy for the analysis of clinical trials with partly missing longitudinal data. *International Journal of Methods in Psychiatric Research*. 2003; 12:139–150. [PubMed: 12953141]
38. Lachin JM. Statistical considerations in the intent-to-treat principle. *Controlled Clinical Trials*. 2000; 21:167–189. [PubMed: 10822117]
39. Lavori PW. Clinical trials in psychiatry – should protocol deviation censor patient data? *Neuropsychopharmacology*. 1992; 6:39–48. [PubMed: 1571068]
40. Lavori P, Dawson R. Designing for intent-to-treat. *Drug Information Journal*. 2001; 35:1079–1086.
41. Little R, Yau L. Intent-to-treat analysis for longitudinal studies with dropout. *Biometrics*. 1996; 52:1324–1333. [PubMed: 8962456]
42. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal*. 2009; 338:b2393. [PubMed: 19564179]
43. Sabin CA, Lepri AC, Phillips AN. A practical guide to applying the intention-to-treat principle to clinical trials in HIV infection. *HIV clinical trials*. 2000; 1:31–38. [PubMed: 11590497]
44. Shih WJ. Problems in dealing with missing data and informative censoring in clinical trials. *Current Controlled Trials in Cardiovascular Medicine*. 2002; 3:4. [PubMed: 11985778]
45. Kenward MG, Goetghebeur EJT, Molenberghs G. Sensitivity analysis for incomplete categorical tables. *Statistical Modelling*. 2001; 1:31–48.
46. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomised controlled trials. *Clinical Trials*. 2007; 4:125–139. [PubMed: 17456512]
47. Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*. 1998; 93:1321–1339.
48. Daniels MJ, Hogan JW. Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics*. 2000; 56:1241–1248. [PubMed: 11129486]
49. Molenberghs G, Kenward MG, Goetghebeur E. Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Journal of the Royal Statistical Society (C)*. 2001; 50:15–29.

50. Matts JP, Launer CA, Nelson ET, Miller C, Dain B. A graphical assessment of the potential impact of losses to follow-up on the validity of study results. *Statistics in Medicine*. 1997; 16:1943–1954. [PubMed: 9304765]
51. Hollis S. A graphical sensitivity analysis for clinical trials with non-ignorable missing binary outcome. *Statistics in Medicine*. 2002; 21:3823–3834. [PubMed: 12483769]
52. Rubin DB, Stern HS, Vehovar V. Handling “don't know” survey responses: The case of the Slovenian plebiscite. *Journal of the American Statistical Association*. 1995; 90:822–828.
53. Liang KY, Zeger SL. Longitudinal data analysis of continuous and discrete responses for pre-post designs. *Sankhya: The Indian Journal of Statistics, Series B*. 2000; 62:134–148.
54. Liu GF, Lu K, Mogg R, Mallick M, Mehrotra DV. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? *Statistics in Medicine*. 2009; 28:2509–2530. [PubMed: 19610129]
55. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomised trials. *Statistics in Medicine*. 2005; 24:993–1007. [PubMed: 15570623]
56. Kenward M, White IR, Horton N, Carpenter J. Re: Liu et al, Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? (*Statistics in Medicine* 2009; 28: 2509 - 2530). *Statistics in Medicine*. 2010; 29:1455–6. [PubMed: 20527014]
57. Riper H, Kramer J, Smit F, Conijn B, Schippers G, Cuijpers P. Web-based self-help for problem drinkers: a pragmatic randomized trial. *Addiction*. 2008; 103:218–227. [PubMed: 18199300]
58. Burns T, Creed F, Fahy T, Thompson S, Tyrer P, White I, for the UK700 trial group. Intensive versus standard case management for severe psychotic illness: a randomised trial. *Lancet*. 1999; 353:2185–89. [PubMed: 10392982]
59. Greenland S, Finkle W. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *American Journal of Epidemiology*. 1995; 142:1255–1264. [PubMed: 7503045]
60. White IR, Kalaitzaki R, Thompson SG. Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an Internet-based alcohol trial. *Statistics in Medicine*. 2011; 30:3192–3207. [PubMed: 21948462]
61. Royston P. Multiple imputation of missing values: update. *Stata Journal*. 2005; 5:188–201.
62. Royston P. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. *Stata Journal*. 2009; 9:466–477.
63. Wood A, White I, Hillsdon M, Carpenter J. Comparison of imputation and modelling methods in the analysis of a physical activity trial with missing outcomes. *International Journal of Epidemiology*. 2005; 34:89–99. [PubMed: 15333619]
64. Zeger S, Liang K, Albert P. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988; 44:1049–1060. [PubMed: 3233245]

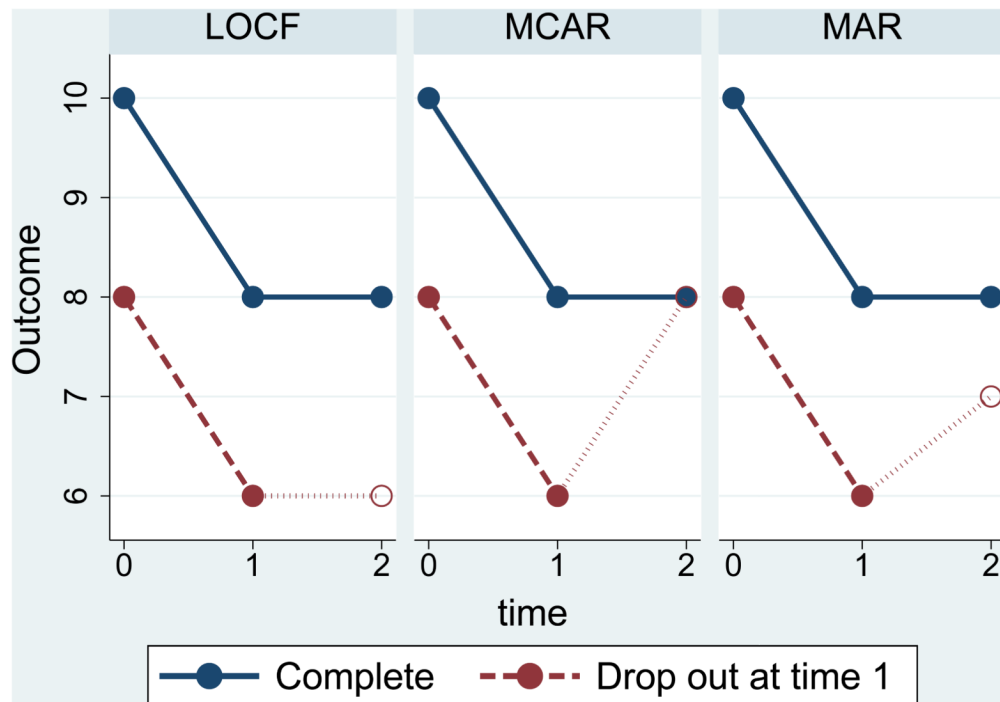


Figure 1. Mean profiles in one arm of a hypothetical randomised trial for individuals who have complete data (solid line) and those who drop out at time 1 (dashed line), illustrating three possible assumptions for missing data at time 2 (dotted line).

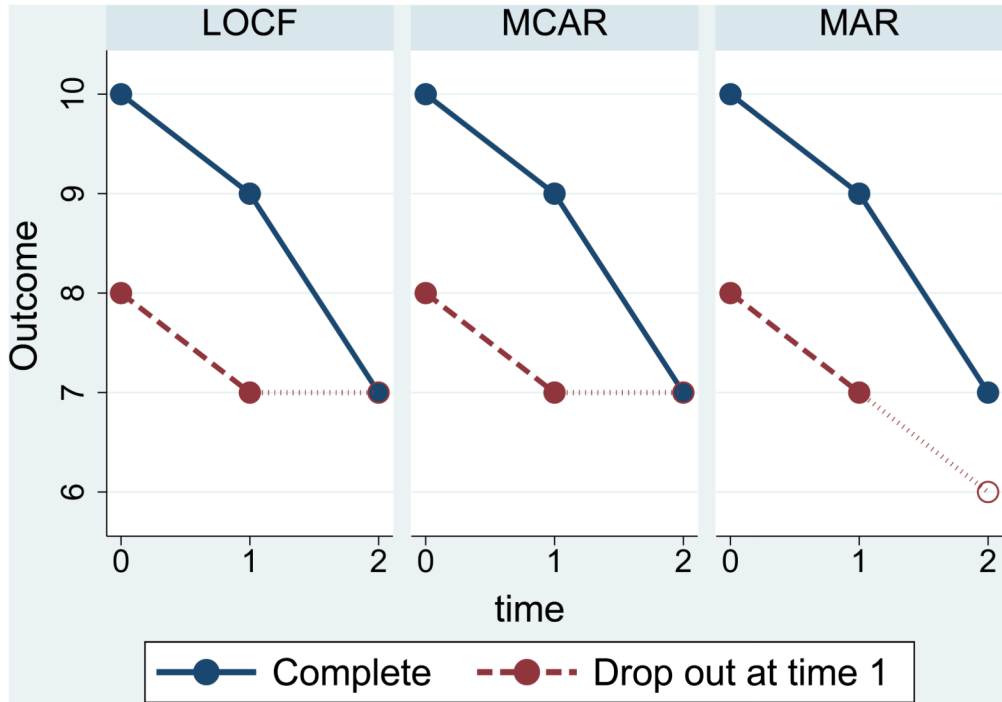


Figure 2. Mean profiles in one arm of a hypothetical randomised trial for individuals who have complete data (solid line) and those who drop out at time 1 (dashed line), showing how LOCF and MCAR can agree without necessarily being correct.

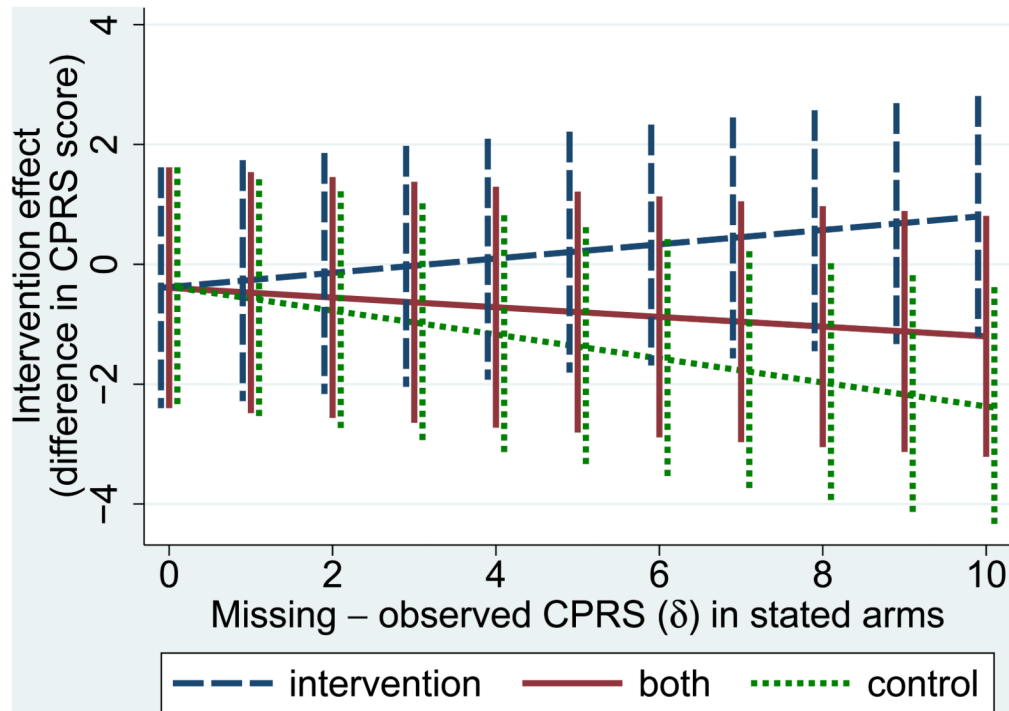


Figure 3. UK700 trial: sensitivity analysis for departures from MAR. The basic analysis is the mixed model with CPRS as outcome and baseline CPRS as covariate using available cases. The parameter δ is the difference between missing and observed CPRS, adjusted for baseline CPRS, in one or both arms. Vertical bars are 95% confidence intervals.

Table 1
UK700 trial: data summary

Variable	% missing	Values	Mean	SD
Centre	0%	0/1/2/3		
Randomised group (standard/intensive)	0%	0/1		
CPRS: psychopathology score				
baseline	0.4%		18.8	12.7
year 1	16.2%		17.2	13.1
year 2	16.0%		18.3	13.8
SAT: satisfaction score ¹				
baseline	19.4%		18.9	4.8
year 1	27.8%		17.2	4.7
year 2	30.8%		16.9	4.8
HOS: days in hospital for mental health reasons over past 2 years				
baseline	0.1%		108.9	112.6
year 2	4.1%		73.3	117.8

¹Higher values denote lower satisfaction.

\$watermark-text

\$watermark-text

\$watermark-text

Table 2

UK700 trial: missing data patterns. ‘O’ denotes an observed value, ‘.’ a missing value. HOS was only observed at years 0 and 2.

Baseline	Year 1	Year 2	CPRS	SAT	HOS
0	0	0	546	355	0
0	0	.	45	70	0
0	.	0	49	55	679
0	.	.	65	91	28
.	0	0	0	65	0
.	0	.	2	21	0
.	.	0	0	15	0
.	.	.	1	36	1

Table 3

UK 700 trial: estimated intervention effect on CPRS and SAT at year 2, adjusted for trial centre and baseline, by various analyses which assume the data are missing at random but include different numbers of the 708 individuals randomised. Results marked * were used in the ITT analysis strategy (Section 5.1).

Method	Assumption			CPRS			SAT		
	Indivs ¹	Obs ²	Estimate	Std. Err	Indivs ¹	Obs ²	Estimate	Std. Err	
<i>Analyses using baseline and year 2 data only</i>									
Analysis of covariance ³	MAR	595	595	-0.3898	1.0348	490	490	-0.3851	0.4139
Mixed model, baseline as outcome	MAR	705	595	-0.3898	1.0304	651	490	-0.3757	0.4139
Multiple imputation ⁴	MAR	708	595	-0.3834	1.0369	708	490	-0.3800	0.4159
<i>Analyses using baseline and year 1 and 2 data</i>									
LOCF	LOCF	707	642	-0.7176	0.9115	672	581	-0.3189	0.3437
Mixed model, baseline as covariate ³	MAR	642	1188	-0.3916*	1.0256*	581	1001	-0.3502*	0.4092*
Mixed model, baseline as outcome	MAR	707	1188	-0.3992	1.0215	672	1001	-0.3452	0.4094
Multiple imputation ⁴									
Basic model	MAR	708	1188	-0.4027	1.0268	708	1001	-0.3444	0.4087
Extended model	MAR ⁵	708	1188	-0.4282*	1.0232*	708	1001	-0.4025*	0.4050*

¹Number of individuals included in the analysis.

²Number of post-randomisation observations included in the analysis.

³Mean imputation used for missing baseline values.

⁴Monte Carlo error [35] is approximately 0.01 for point estimates and 0.002 for standard errors.

⁵MAR conditional on all variables included in the extended model.