Differential dropout and bias in randomised controlled trials: when it matters and when it may not

Dropout in randomised controlled trials is common and threatens the validity of results, as completers may differ from people who drop out. Differing dropout rates between treatment arms is sometimes called differential dropout or attrition. Although differential dropout can bias results, it does not always do so. Similarly, equal dropout may or may not lead to biased results. Depending on the type of missingness and the analysis used, one can get a biased estimate of the treatment effect with equal dropout rates and an unbiased estimate with unequal dropout rates. We reinforce this point with data from a randomised controlled trial in patients with renal cancer and a simulation study.

Melanie LBell senior research fellow1, Michael G Kenward professor2, Diane L Fairclough professor3, Nicholas J Horton professor4

1Psycho-Oncology Co-operative Research Group (PoCoG), University of Sydney, Sydney Australia; 2Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK; 3Department of Preventive Medicine and Biometry, University of Colorado at Denver, Denver, CO, USA; 4Department of Mathematics and Statistics, Smith College, Northampton, MA, USA

Introduction

Dropout in longitudinal randomised controlled trials is common and a potential source of bias in terms of evidence based medicine. A review of 71 randomised controlled trials in four top medical journals showed dropout rates of 20% or more in 18% of the trials.1 Similar rates were found in a review of quality of life outcomes.2 In specialist journals, the rates are likely to be higher.

When dropout rates differ between treatment arms, so that fewer patients are followed up in one arm than the other, it is sometimes called “differential dropout” or “differential attrition.” Despite extensive literature on incomplete data methods for randomised controlled trials, guidance from the CONSORT reports, and the National Research Council’s recent report on missing data,3 many applied researchers have misconceptions about how to handle dropout.1 4 6 Two common misunderstandings about differential dropout need to be debunked:

- Myth 1—if dropout rates in longitudinal clinical trials are similar between study arms, bias is not a concern.
- Myth 2—if dropout rates are dissimilar between study arms, the results will necessarily be biased.

Although differential dropout can bias results, equal dropout rates between study arms does not imply that results will be unbiased (myth 1). The inverse is true as well: unequal dropout rates do not mean the results are biased (myth 2). Whether dropout rates between the arms are differential or not can be a red herring; the two key factors are the type of missingness and the statistical analysis. Our aims were to show that for continuous outcome data biased estimates of treatment effects can be obtained when no differential dropout occurs (refute myth 1); unbiased estimates of treatment effects can be obtained when differential dropout occurs (refute myth 2); and when missingness is non-random, the degree of bias depends, in part, on the actual mechanism of dropout. We begin with an example from a real randomised controlled trial and then show the generalisability of our assertions with a computer simulation study (see glossary).

Example

A phase III randomised controlled trial was carried out to compare two treatments in patients with advanced renal cell carcinoma (details discussed elsewhere6). This disease is aggressive, and the treatment is toxic; patients discontinue treatment (one form of dropout) owing to toxicity, disease progression, and death. Quality of life and symptom sub-studies are commonly carried out in clinical trials, as patient reported outcomes are increasingly recognised as important for understanding patients’ experience during treatment and providing support for or against experimental treatments.7 The renal cell trial used a quality of life questionnaire that included questions on physical and
Glossary

Missing completely at random (MCAR)—Outcome data are MCAR if their missingness does not depend on observed covariates or previous outcomes. As a result, no systematic differences exist between dropouts and completers. For example, if a staff member forgets to administer questionnaires, then these missing data are likely to be MCAR.

Missing at random (MAR)—Data following dropout are MAR if the future statistical behaviour of a participant’s outcomes, given his or her past outcomes and covariates, would have been the same whether he/she had dropped out or not. For example, missing post-baseline quality of life data may occur in patients with low baseline quality of life. However, after this information has been taken into account (using baseline quality of life as a covariate in a regression model), no systematic differences exist between dropouts and completers.

Missing not at random (MNAR)—Data following dropout are MNAR if the future statistical behaviour of a participant’s outcomes, given his/her past outcomes and covariates, would not have been the same whether he/she had dropped out or not. This yields systematic differences between dropouts and completers that cannot be adjusted for on the basis of observed quantities. For example, if a patient is too ill to fill in a quality of life questionnaire, the missing data may be MNAR if the current disease state is not captured by the previously observed outcomes.

Missing data mechanism(s)—The underlying reason(s) why data are missing. For example, this might be adverse effects or disease progression.

Multiple imputation—A method of filling in missing data by drawing from a distribution of likely values while properly accounting for the uncertainty associated with doing so. It is called multiple imputation because multiple datasets are created, analysed, and then combined to get the final results.

Single imputation—Also called simple imputation. Missing values are filled in with a single value, such as the mean of a participant’s previous non-missing data, the baseline value, or the last observed value (last observation carried forward). Because this does not account for the imputation, this can cause bias and underestimation of variance.

Computer simulation study—A computationally intensive study that creates many datasets with known characteristics (such as the true treatment effect in a randomised controlled trial) to investigate and illustrate statistical properties (such as bias) of an analytical approach.

Mixed models—A flexible family of regression models that accommodate non-independent data, such as clustered, repeated measures, and longitudinal data. Fitzmaurice, Laird, and Ware provide an excellent and readable text.1

Key concepts

Types of missing data

To investigate the effect of dropout in the example, we need to understand the different types of missing data and how they relate to bias. Rubin defined a taxonomy of missingness, shown in the glossary, which underpins the choice of appropriate analytical approach in a trial with dropout. If patients withdraw from a study for a reason unrelated to their disease or treatment (for example, because they have moved overseas) their data are probably missing completely at random, because systematic differences between them and the patients who remained in the study are unlikely. If patients withdraw from the study for reasons related to their disease or treatment (such as progression or toxicity) the missing data are either missing at random or missing not at random; their quality of life measures would have been, on average, worse than those of patients who remained in the study. One can distinguish between missing at random and missing completely at random by testing, for example, whether baseline quality of life or progression rates differ between dropouts and completers or by using a graphical approach. The figure shows quality of life plotted against time for the renal cell trial, stratified by dropout time. As the trajectories differ substantially by dropout time, the data are not missing completely at random; patients with lower baseline quality of life were more likely to drop out, particularly in the experimental arm. Distinguishing between missing at random and missing not at random from the data under analysis is generally impossible.8 9

Statistical analysis: what works and what does not work

Bias in the estimate of treatment effect caused by dropout can also depend on the analytical approach used. When data are missing completely at random or missing at random, likelihood based mixed models (see glossary) can yield unbiased estimates of treatment effect.10 11 The mixed model for repeated measures is appropriate if participants are assessed at common time points, as is usual in clinical trials. This model allows the outcome to potentially differ by group and time.12 Implementation in a variety of software packages (SAS, Stata, SPSS, and R code are shown in the appendix) is straightforward. If no dropouts occur, the estimate of the difference between treatment and control arms from this mixed model will yield identical results to the simple calculation of a difference in means. Even if dropout differs between the treatment groups, if the data are missing completely at random or missing at random and an appropriate mixed model is used then estimates of the treatment effect will, on average, be unbiased. This is because information from patients with complete data is used to implicitly impute the missing values. (Other methods, such as multiple imputation, Bayesian, and weighted approaches also can give unbiased estimation,9 12 but these are more computationally intensive, may offer no advantage over the likelihood based approach, and are beyond the scope of this paper.) Table 1 shows the lack of bias with mixed models in the renal cell trial. Unfortunately, data from randomised controlled trials are not typically analysed using methods that are valid under missing at random, which can in many circumstances lead to biased estimates.1 4 More commonly, simpler methods such as complete case analysis and single imputation are used, and these are valid only under strong and typically unrealistic assumptions. These methods have been shown to be biased if data are not missing completely at random.8 9 11 Complete case analysis includes only participants with complete data, which can cause bias if those who drop out are sicker. Popular single imputation methods are last observation carried forward and mean imputation. With last observation carried forward, patients with declining health and quality of life who drop out will have their missing quality of life values replaced by their last observed
values, which is likely to overestimate quality of life and yield
biased estimates. Last observation carried forward, because of
its unlikely assumptions about trajectories over time, is generally
biased even when data are missing completely at random.10 11

The common misuse of these methods has potentially serious
consequences.12 In the renal cell trial, these methods were
associated with bias of up to a 74% overestimate and 72%
underestimate of the treatment effect for the complete dataset
(table 1). (Note that we cannot know what the bias is for the
non-augmented real dataset.) This bias translates to a
standardised effect size of 0.4 (0.4 of the baseline standard
deviation). In some cases, the statistical significance changed
as well, which may have implications for decision making.

When data are missing not at random, no method of obtaining
unbiased estimates exists that does not incorporate the
mechanism of non-random missingness, which is nearly always
unknown. Some evidence, however, shows that the use of a
method that is valid under missing at random can provide some
reduction in bias.13 This is the main reason that experts
recommend methods such as mixed modelling as the primary
analysis followed by sensitivity analyses,14 15 in which the
stability of estimates is examined when assumptions about
missingness are varied.3 6 9 11

Returning to the myths, we see from table 1 that biased
estimates can result from equal dropout, particularly if complete
case analysis is used, suggesting that myth 1 may be false. Of
the methods shown, the mixed model gave almost unbiased
treatment estimates, even when the dropout rates were different,
suggesting that myth 2 may also be false. To confirm this, we
did a simulation study.

Simulation study

The simulation study involved creation of 10 000 datasets
representing a two arm randomised controlled trial for quality
of life, with 200 participants, over three time points, in which
the true treatment effect was known. We deleted data to make
equal and unequal dropout scenarios. We created different types
of missingness and different “directions” of dropout (patients
in the treatment group dropped out when they had good
outcomes, and patients in the control group dropped out when
they had poor outcomes). We used two common patterns for
quality of life over time: a linear decline and a temporary
change. We analysed each of these datasets by computing the
difference in mean quality of life at the final time point between
the treatment and control groups by using complete case
analysis, single imputation, and a mixed model. Details are
given in the appendix.

Myth 1—Table 2 (second and fourth columns) shows that equal
dropout rates do not guarantee unbiased results. In the equal
dropout simulation, the single imputation approaches all yielded
biased estimates of the treatment effect, for every type of
dropout, even missing completely at random. Last observation
carried forward, for example, underestimated the treatment
effect by 22% for missing completely at random for both the
same and different direction mechanisms. For missing at
random, last observation carried forward underestimated by
24% for the same direction mechanism but overestimated by
6% for the different direction mechanism.

Myth 2—Table 2 (third and fifth columns) shows that unequal
dropout rates do not imply biased results. The mixed model
yielded unbiased estimates for all missing completely at random
and missing at random scenarios, regardless of equal or unequal
dropout rates and differing mechanisms (directions) for dropout.

Discussion

We have shown that the two myths described in the introduction
are indeed myths: equal dropout rates can still yield biased
estimates of treatment effects, and studies with unequal dropout
rates can be analysed to produce unbiased results. Likelihood
based analyses, such as the mixed model presented here,
represent one approach for valid analysis when the dropout
mechanism is data missing completely at random or missing at
random. A non-random mechanism will typically lead to bias
under any analysis that does not incorporate this (usually
unknown) mechanism. Because of this, if missing not at random
is suspected a sensitivity analyses should be done.3 6 9 11

We have discussed these concepts in the context of quality of life
in oncology trials, but the results are generalisable to any
randomised controlled trial with continuous outcomes and
dropout.

Unfortunately, the consideration of differential dropout in the
literature has omitted a discussion of key elements. In a review
of randomised controlled trials of palliative care, Thomas et al
stated that they “recorded an attrition analysis as performed
only if the authors reported an analysis assessing if the
intervention and control arms were differentially affected by
attrition.”16 Heneghan et al, in considering differential attrition
in randomised controlled trials in diabetes, created a “relative
attrition statistic.”17 Neither of these papers discusses type of
missingness or analyses used for estimation of treatment effect.
The CONSORT statement says: “There should be concern when
the frequency or the causes of dropping out differ between the
intervention groups.”18 Although this is certainly true, it may
be misinterpreted as researchers focus on frequency instead of
cause and are lulled into complacency about unbiased estimation
when dropout rates are equal between trial arms. We do not
wish to give the misleading impression that unequal dropout
rates between arms in a clinical trial is not a problem, but we
wish to stress that it does not always cause bias. An important
reason to consider differential dropout is the acceptability
and uptake of the intervention. Different adverse effect profiles
between the arms, for example, could cause differential dropout.
We recommend, as do others in the missing data field, that a
sensible analytical approach for data with dropout begins with
an assumption of missing at random, includes careful
examination and documentation of the missingness
mechanism(s), and includes sensitivity analyses if missing not
at random data are suspected. With appropriate analyses, such as use of mixed models, bias can be eliminated or reduced.

Data presented are derived (and used with permission) from a trial conducted by Memorial Sloan-Kettering Cancer Center and Eastern Cooperative Oncology Group funded by the National Cancer Institute grant CA-05826.

Contributors: MLB had the original idea, did the analyses, and wrote the first draft. DLF aided in design. All authors revised the manuscript and approved the final version.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Not commissioned; externally peer reviewed.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Not commissioned; externally peer reviewed.


Accepted: 27 November 2012

Cite this as: BMJ 2013;346:e8668

© BMJ Publishing Group Ltd 2013
### Summary points

- Equal dropout rates between treatment arms in a randomised controlled trial do not imply that estimates of treatment effect are unbiased.
- Similarly, unequal dropout rates do not imply that estimates are biased.
- Bias depends on the type of missingness, the analysis method, and the effect that is being estimated.
- Likelihood-based methods such as mixed models can be used to estimate unbiased treatment effects, under assumptions regarding the missingness mechanism(s).

### Tables

**Table 1** Treatment effect estimates and statistical significance at time 4: difference in quality of life between control and experimental arms (QoLcontrol−QoLexperimental) in renal cell trial

<table>
<thead>
<tr>
<th>Dataset and dropout rates (control %:experimental %)</th>
<th>Real data (64:70)</th>
<th>Equal dropout (41:44)</th>
<th>Unequal dropout (25:51)</th>
<th>Unequal dropout (51:25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>Estimate</td>
<td>P value</td>
<td>Estimate (% bias*)</td>
<td>P value</td>
</tr>
<tr>
<td>True difference</td>
<td>?</td>
<td></td>
<td>7.2</td>
<td>0.002</td>
</tr>
<tr>
<td>Mixed model</td>
<td>8.1</td>
<td>0.03</td>
<td>7.2 (0)</td>
<td>0.005</td>
</tr>
<tr>
<td>Complete case†</td>
<td>0.4</td>
<td>0.9</td>
<td>4.5 (−38)</td>
<td>0.1</td>
</tr>
<tr>
<td>Mean imputation‡</td>
<td>5.0</td>
<td>0.008</td>
<td>5.9 (−18)</td>
<td>0.003</td>
</tr>
<tr>
<td>Last observation carried forward§</td>
<td>6.3</td>
<td>0.01</td>
<td>5.1 (−29)</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Complete dataset was created by averaging 50 multiply imputed datasets, then patients were dropped out using missing at random process (see web appendix for more detail); true estimate from this complete dataset is 7.2 (see web appendix for more detail).

*100×(estimate−true)/true, where estimate=difference in mean quality of life at final time point.
†No imputation used.
‡Missing values replaced with mean of patient’s previously observed values.
§Missing values replaced with patient’s last observed value.
Table 2 | Comparison of simple methods and mixed modelling for linear decline pattern

| Missingness and analysis | | | 
|--------------------------|---|---|---|---|
|                          | Same direction† | Different direction† |
| Equal dropout (30%)      | Unequal dropout (20% treatment, 40% control) | Unequal dropout (20% treatment, 40% control) |

| No missing data: | | | 
| Complete case | 0 | 0 | 0 | 0 |
| Mixed model | 0 | 0 | 0 | 0 |

| Missing completely at random: | | | 
| Complete case‡ | −1 | 0 | 0 | 0 |
| Mean imputation§ | −27 | −46 | −27 | −47 |
| Last observation carried forward¶ | −22 | −40 | −22 | −41 |
| Mixed model | 0 | 0 | 0 | 0 |

| Missing at random: | | | 
| Complete case‡ | 0 | 26 | −60 | −61 |
| Mean imputation§ | −28 | −17 | −6 | −24 |
| Last observation carried forward¶ | −24 | −19 | 6 | −10 |
| Mixed model | 0 | 0 | 0 | 0 |

| Missing not at random: | | | 
| Complete case‡ | 0 | 35 | −80 | −82 |
| Mean imputation§ | −26 | −5 | −44 | −61 |
| Last observation carried forward¶ | −25 | −2 | −41 | −57 |
| Mixed model | 0 | −18 | −40 | −41 |

*100×(estimate–true)/true, where estimate=difference in mean quality of life (QoL) at final time point.
†Different direction=patients in treatment group dropped out when they had high QoL, patients in control group dropped out when they had poor QoL; same direction=patients in both group dropped out when QoL scores were low.
‡No imputation used.
§Missing values replaced with mean of patient’s previously observed values.
¶Missing values replaced with patient’s last observed value.
Quality of life (QoL) stratified by treatment arm and dropout time. Possible range of QoL=0-100, with higher values indicating better QoL. If data were missing completely at random (MCAR), the within arm trajectories would be indistinguishable. As patients with lower baseline QoL are more likely drop out, these data are not MCAR.