

RESEARCH METHODS & REPORTING

Statistics Notes: Missing outcomes in randomised trials

Andrew J Vickers *attending research methodologist*¹, Douglas G Altman *professor of statistics in medicine*²

¹Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA; ²Centre for Statistics in Medicine, University of Oxford, Oxford OX2 6UD, UK

In most randomised trials, some patients fail to provide data for study endpoints.¹ We have previously described the analysis of a trial of acupuncture versus sham acupuncture for the treatment of shoulder pain.² All 52 randomised patients provided baseline data on pain and range of motion, but only 45 returned for follow-up testing. The statistical question is how to handle those seven patients with missing data. The most straightforward approach is simply to ignore the seven patients and do what is known as an “available case analysis” (often confusingly known as “complete case analysis”). As not all randomised patients are included in the analysis, this leads to reduced statistical power.¹

A method that attempts to include all randomised patients is “last observation carried forward,” in which the last measurement obtained from the patient is used for all data points that were subsequently missed. This method is attractive because it is simple, but it has little else to recommend it. Substituting a missing data point with a value is known as “imputation,”³ and the data analyst needs a clear rationale for the type of imputation used. That a patient’s responses would remain the same after drop-out is generally implausible. This is most obvious in chronic degenerative diseases. For instance, cognitive function scores decrease over time in dementia, so last observation carried forward gives overoptimistic scores for patients who drop out (figure 1). If a treatment was associated with toxicity, and this led to earlier drop-out than in the control group, the method would give results biased in favour of the experimental arm.³ By contrast, shoulder pain generally gets better over time, either because treatment is effective or because of the placebo effect and regression to the mean.⁴ In the randomised trial, patients in the control group improved by a mean of 9.8 points out of 100 from baseline to post-treatment follow-up, whereas patients who received acupuncture improved by 21.5 points. So assuming that patients lost to follow-up experienced precisely zero change in pain scores makes little sense. Last observation carried forward may also underestimate the standard deviation of the endpoint, especially in cases in which the last observation is the baseline, leading to confidence intervals that are too narrow.

A more sophisticated approach to missing data is known as multiple imputation, which uses a regression model to predict missing values.⁵ In randomised trials, the strongest predictors of future outcome are often the scores provided by the patient so far, but other variables can be included. To avoid underestimating the width of the confidence interval, multiple imputation involves a form of random sampling. For a given patient with a missing outcome, regression is used to predict the mean value of the missing outcome for similar patients and also the variability around the mean; a value is then selected at random from this distribution. The results from several imputations (hence “multiple”) are combined using a method known as “Rubin’s rules.”^{5 6} Multiple imputation is widely believed to be the preferred approach to missing data, not just for randomised trials.⁷ It is computationally complex, however, and needs to be implemented by special software, such as the “ice” command in Stata (see www.multiple-imputation.com).

The table 1 shows the results of the shoulder pain study analysed by each method. The estimates for available case and multiple imputation do not differ much, although multiple imputation has a slightly narrower confidence interval. Last observation carried forward appears to be biased—it underestimates the effects of acupuncture—and gives a confidence interval that is too narrow.

Multiple imputation works best when good predictors of outcome are available. In the shoulder pain example, baseline score was only moderately correlated with follow-up score ($r=0.4$). Had outcome been assessed halfway through treatment, this measure would have been more highly correlated with post-treatment score, markedly improving the properties of the multiple imputation.

Multiple imputation has several important strengths, but it does not adjust for the sort of bias created if patients were less likely to return for follow-up if they were in a lot of pain; this is an inherent limitation to missing data analysis. We cannot know whether patients’ pain levels affect the chance that they will complete a pain questionnaire because, obviously enough, we do not have the pain scores of non-respondents.

Sometimes simple common sense is more important than complex statistics. In the shoulder pain trial, three of the seven drop-outs were in the acupuncture group and four were controls, so it seems implausible that their omission had materially affected the results of the trial. If drop-out rates were very different between the two arms of a trial, that may raise concerns about bias. Above all, analysis of missing data teaches us the importance of avoiding missing data in the first place: an informed guess, even using a technique as sophisticated as multiple imputation, is still a guess.

Contributors: AJV and DGA jointly wrote and agreed the text.

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; not externally peer reviewed.

- 1 Altman DG, Bland JM. Missing data. *BMJ* 2007;334:424.
- 2 Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. *BMJ* 2001;323:1123-4.
- 3 Molnar FJ, Man-Son-Hing M, Hutton B, Fergusson DA. Have last-observation-carried-forward analyses caused us to favour more toxic dementia therapies over less toxic alternatives? A systematic review. *Open Med* 2009;3:e31-50.
- 4 Bland JM, Altman DG. Regression towards the mean. *BMJ* 1994;308:1499.
- 5 Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.
- 6 White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;30:377-99.
- 7 White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ* 2011;342:d40.

Cite this as: *BMJ* 2013;346:f3438

© BMJ Publishing Group Ltd 2013

Table

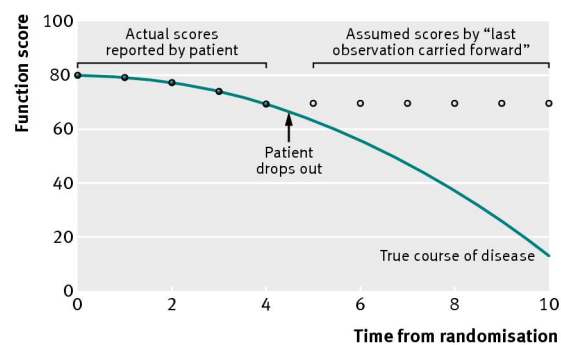
Table 1| Analysis of shoulder pain trial with three statistical methods

Analysis	Effect of acupuncture—difference in points (95% CI)	Standard error	P value
Available case	14.2 (5.1 to 23.4)	4.53	0.003
Last observation carried forward*	12.6 (3.9 to 21.3)	4.33	0.005
Multiple imputation†	14.3 (5.4 to 23.3)	4.45	0.002

*Missing final value replaced by baseline value (n=7).

†Using baseline score and treatment group.

Figure



Function scores over time for patient with chronic degenerative disease