

## Missing data in clinical research: an integrated approach

Loes Hollestein, PhD.

Erasmus MC University Medical Center, Rotterdam, The Netherlands

Netherlands Comprehensive Cancer Center (IKNL), Utrecht, The Netherlands.

James Carpenter, DPhil

London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding author:

Loes Hollestein

PO Box 2040, 3000 CA Rotterdam, the Netherlands

M: [l.hollestein@erasmusmc.nl](mailto:l.hollestein@erasmusmc.nl)

T: 0031107035886

Wordcount: 1683 [max 2000]

The clinical study with no missing data has yet to be conducted – and never will be! Yet, despite its ubiquity, missing data and the issues they raise are still too often brushed under the carpet or inappropriately handled<sup>1</sup>. Despite the ready availability of software<sup>2</sup>, this situation has changed surprisingly slowly over the last decade<sup>3,4</sup>.

Fortunately, the careful handling and reporting of missing data in the STOP GAP trial, reported in this issue, provides a ready model for researchers<sup>5</sup>. STOP GAP, a multicenter observer blind randomized controlled trial, compared the cost-effectiveness of ciclosporin to prednisolone-initiated treatment for Pyoderma Gangrenosum over a period of 24 weeks. Data on health care resource use was collected at 8 and 24 weeks by clinic visits, telephone, trial drug logs and patient diaries. Following the CONSORT recommendations, their 'Table 1' includes the completeness of data, reporting of the number of participants per group for all analyses. The primary analysis (referred to as the base-case analysis by the authors) included multiple imputation to account for missing data, assuming data were *missing at random* (explained below).

Prevention is always better than cure, so we first review how the authors of the STOP GAP trial designed to minimize the extent of missing data, before discussing the primary and sensitivity analysis. This leads to some suggestions for good practice.

### Design to avoid missing data

At the design stage, consideration of likely reasons for, and extent of, missing data - preferably informed by previous studies in the target population – is crucial. We highlight three areas<sup>6</sup>:

- Choice of outcome measure

To minimize participant burden, outcomes need to be carefully selected. The Core Outcome Set Initiative (COUSIN)<sup>7</sup> which identifies the minimal set of core outcomes relevant to patients, carers and health decision makers, is an excellent starting point. Following these guidelines

ensures that a common set of outcomes are available across trials, facilitating meta-analysis<sup>7</sup>. Core outcome sets are being developed for a range of dermatological diseases<sup>8</sup>. There is no core outcome set for Pyoderma Gangerosum, but the patients were involved in the choice of outcome measure to ensure relevance to the target population and avoid unnecessary burden. Moreover, the authors recognized that a lengthy follow-up leads to an increasing proportion of missing values and therefore chose a primary endpoint at 6 weeks of follow-up (velocity of healing) which showed to be a good surrogate for subsequent healing<sup>9</sup>.

- Strategies to minimize drop out

Practical steps to reduce the burden on participants are likely to pay dividends. These include using data which do not require a clinic visit (e.g. routinely collected health care data), reducing the number of visits, and allowing a relatively large time window for each assessment<sup>6</sup>. In the STOP GAP trial the time window was relatively large (16 weeks).

- Imaginative strategies for maintaining follow-up

A skilled trial management team is at the heart of this. Maintaining follow-up is easier if patients give consent for follow-up separate to consent to treatment, so they can be followed up after any treatment withdrawal. A strategy of enhanced contact attempts (e.g. phone) for a selection of non-responders may also help. Participants of the STOP GAP trial were called by telephone when clinical visits were missed.

With any trial there is a tension between internal and external validity, and this applies to missing data too. We may target recruitment at patients who a-priori we believe are more likely to remain in the study; by doing so we enhance our internal validity, possibly at the expense of external validity.

### **Analysis of incomplete data**

The consequence of missing data is a loss of information to test our hypothesis (which may not be a simple factor of the proportion of observations missing). In our analysis, we can buy some of this information back with assumptions about the distribution of the missing values.

We assume that the data are not missing completely randomly. Beyond this, a natural starting point is to consider strata of patients who are in the same treatment group, and have similar baseline and early follow up data. If we assume that such patients have the same distribution of their final endpoint *regardless of whether we were able to observe it* then we are essentially assuming the data are *missing at random*<sup>10</sup>.

When some patients are missing later follow-up data, we can exploit this assumption to reduce the bias and increase the precision of the estimated treatment effect<sup>11</sup>. We can either do this using an appropriate longitudinal model for all the observed data, or equivalently we can use a technique called *multiple imputation*<sup>10</sup>, which is now implemented in a range of software packages.

In analyzing the STOP GAP trial, the authors rightly first explored whether the probability of missing outcome data depended on certain baseline variables, and whether these baseline variables predicted the missing values, finding they did both. Thus, data are not missing completely at random, so restricting the analysis to the subset of patients with no missing data (i.e. the completers) will

generally result in bias and loss of information. Anticipating this, the authors base-case analysis assumed values were missing at random, and used multiple imputation.

Multiple imputation generates a number of ‘completed’ data sets in which missing values are imputed under the missing at random assumption. Roughly speaking, in each imputed data set, each patient’s missing final follow-up data is replaced by a draw from a pool of patients from the same treatment group who have closely matching baseline and early follow-up data. Since we can never know the missing values, but only the ‘pool’ (i.e. statistical distribution) they belong to, a single imputed data set is not sufficient. Instead, we create multiple imputed data sets, and in each one each patient’s missing values are replaced by a draw from their appropriate ‘pool’. Then, we estimate the treatment effect from each imputed dataset, before combining the results using a set of simple rules (Rubin’s rules) to arrive at the final treatment estimate, p-value and confidence interval <sup>12</sup>.

The ‘multiple’ aspect of this process is key: in general, single imputation methods, such as last observation carried forward, generally leading to biased estimates and discard the uncertainty of imputation, leading to underestimated standard errors and misleading p-values<sup>13,14</sup>.

### **Sensitivity analysis of incomplete data**

Sensitivity analysis explores the robustness of our scientific conclusions by doing different, contextually plausible, analysis assumptions. In the context of missing data, we explore the robustness of our conclusions to departures from the missing at random assumption.

Recalling the discussion of missing at random above, it follows that departures from this assumption occur if a patient with missing outcome data has outcome values that are systematically worse or better than what we would predict from a pool of patients from the same treatment group with broadly similar baseline and early follow-up data. In statistical terminology, departures from missing at random are called *missing not at random*.

The authors of the STOP GAP trial compared the results of their base-case multiple imputation based analysis with a complete case sensitivity analysis. The results of both analyses are compatible, although the multiple imputation analysis is more precise (i.e. narrower confidence intervals). Both the multiple imputation and complete case analyses adjust for the same covariates that are predictive of the potentially missing outcomes. The similarity of the resulting point estimates suggests that, given these covariates, the inclusion of outcome data during follow-up has relatively little effect on the predicted final outcomes. This reason for this is that the STOP GAP base-case and sensitivity analysis both rest on a broadly similar missing at random assumption.

To explore departures from missing at random, we need the imputed values to lie *outside* the missing at random ‘pools’. It would have been interesting if the authors had explored missing not at random scenarios. This can be achieved by *changing* the imputed values to represent the departure from missing at random. Once this is done, we proceed as before – analyzing each imputed data set and combining the results using Rubin’s rules. The changes need to be contextually appropriate.

### Last but not least: reporting of incomplete data

To improve statistical reporting, the BJD recommends all authors to follow the SAMPL guidelines<sup>15</sup>. These state that all supplementary analyses, such as imputation of missing data, should be described in the methods. Study-specific reporting guidelines also provide guidance for reporting of incomplete data. The STROBE statement for observational studies explicitly states that authors should address how missing data was handled in the analysis and that the number of participants with missing data for each variable of interest should be indicated<sup>15</sup>. The CONSORT guidelines state that the number of missing data should be reported by treatment arm<sup>15</sup>.

To this we would add the following: the assumptions about missing data the analyses rest on should be accessibly reported, in particular the assumption about missing data that was specified in the protocol for the primary analysis. Further, when different assumptions lead to different scientific conclusions, the authors need to clearly communicate which assumptions are most contextually plausible.

The authors of the STOP GAP trial broadly complied with this, although as we noted above the extent that their sensitivity analysis explores departures from the missing at random assumption could be more fully discussed.

### Conclusion: design holds the key!

Prevention is better than cure, and to minimize the loss of information due to missing data it is important to consider the issues carefully in the design stage. Mechanisms of missing data can frequently be anticipated before the start of the study. Variables, which may be predictable for both the chance of data being missed and the missing value, should be included in the planned set of measurements. A practical solution is then to pre-specify their use as part of a multiple imputation analysis assuming data are missing at random. Such an analysis will often be the primary analysis. Relatively simple multiple imputation-based sensitivity analysis can also usefully be pre-specified<sup>16,17</sup>. Finally, following the reporting guidelines will not only increase readers' confidence in your scientific conclusions, but also build on your success in addressing the missing data issues!

### References

- 1 Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005; **365**: 1159-62.
- 2 Sterne JA, White IR, Carlin JB *et al*. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; **338**: b2393.
- 3 Bell ML, Fiero M, Horton NJ *et al*. Handling missing data in RCTs; a review of the top medical journals. *BMC Med Res Methodol* 2014; **14**: 118.
- 4 Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials* 2004; **1**: 368-76.
- 5 Mason JM, Thomas KS, Ormerod AD *et al*. Ciclosporin compared to prednisolone therapy for patients with pyoderma gangrenosum: cost-effectiveness analysis of the STOP GAP trial. *Br J Dermatol* 2017.
- 6 National Research Council. The prevention and treatment of missing data in clinical trials. *The National Academies Press* 2010.

- 7 Cochrane Skin Group - Core Outcome Set Initiative (CSG-COUSIN), <https://www.uniklinikum-dresden.de/de/das-klinikum/universitaetscentren/zegv/cousin/>, Last accessed: 18 April 2017
- 8 The Core Outcome Measures in Effectiveness Trials (COMET) initiative <http://www.comet-initiative.org/>, Last accessed: 18 April 2017
- 9 Craig FF, Thomas KS, Mitchell EJ *et al.* UK Dermatology Clinical Trials Network's STOP GAP trial (a multicentre trial of prednisolone versus ciclosporin for pyoderma gangrenosum): protocol for a randomised controlled trial. *Trials* 2012; **13**: 51.
- 10 Donders AR, van der Heijden GJ, Stijnen T *et al.* Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006; **59**: 1087-91.
- 11 Carpenter J, Kenward M. *Multiple Imputation and its Application*. Chichester: Wiley. 2013.
- 12 Rubin D. Inference and missing data. *Biometrika* 1976; **63**: 581-92.
- 13 Streiner DL. Missing data and the trouble with LOCF. *Evid Based Ment Health* 2008; **11**: 3-5.
- 14 Ware JH. Interpreting incomplete data in studies of diet and weight loss. *N Engl J Med* 2003; **348**: 2136-7.
- 15 Enhancing the QUALity and Transparency Of health Research (EQUATOR), <http://www.equator-network.org/>, Last accessed: 31 May 2017
- 16 Mason AJ, Gomes M, Grieve R *et al.* Development of a practical approach to expert elicitation for randomised controlled trials with missing health outcomes: Application to the IMPROVE Trial. *Clinical Trials* 2017; **Accepted for publication**.
- 17 White IR, Carpenter J, Evans S *et al.* Eliciting and using expert opinions about dropout bias in randomized controlled trials. *Clin Trials* 2007; **4**: 125-39.