



EDITORIAL

Missing data in clinical research.

DOI: 10.17126/joralres.2016.047

Clinical research, especially clinical trials, often have a large proportion of missing data. One of the problems frequently observed is that patients have non-compliance with the treatment protocol, *i.e.* they do not follow the instructions of the researcher/clinician. This fact generates problems with the statistical analysis because they can bias the outcomes of the research.¹

In many clinical trials, probably most of them, missing data is almost inevitable. In recent years this problem has been intensively discussed in the scientific literature, showing the complexity and difficulties to complete the statistical analysis when data are incomplete.

A group of international experts established a set of general principles included in methodological guidelines to ensure the quality of results of clinical trials. In general, all guidelines/principles focus on designing strategies in order to minimize possible missing data. Another strategy is to enhance the statistical analysis through the use of all the information obtained from the subjects, *i.e.* other variables with no missing data, or using sensitivity analysis.¹

Missing data is more common than we think. Several authors have identified a rate ranging from 15% to 20% in educational research, and from 36% to 48% in psychological research.^{2,3} Statistical literature shows an extensive description and application of methods to handle missing data.

Some researchers address this problem excluding patients who did not comply with the protocol. This method called *per-protocol-analysis* corresponds to the statistical analysis of the patients who received randomized assigned intervention.

The *intention-to-treat* (ITT) analysis is another methodological strategy to handle missing data, mainly in clinical trials. This methodology considers the analysis of all individuals enrolled in the trial, according to their allocation, even if they have not complied with the assigned

protocol. ITT analysis has two basic rules: analyze according to original allocation and include all randomized subjects in the analysis. Both rules generally try to prevent selection and drop-out bias.⁴ Thus, the goal to balance known and unknown prognostic factors through the randomization is accomplished.

The ITT analysis has many advantages, especially to reveal the actual clinical setting, because it considers that some patients may not comply with the protocol established in the clinical trial, giving an unbiased estimate of the treatment effect. The bias would occur if the non-compliance or drop-out subjects were eliminated from the final analysis, because this could create larger differences in the outcome of each group. Another advantage is that ITT analysis preserves the original sample size, so the type I error is minimized, allowing a greater generalization of the results.

However, the ITT analysis has disadvantages such as the underestimation of the magnitude of the effect. This underestimation is caused by the subjects who do not comply with the protocol. In the case of studies assessing adverse effects, the ITT analysis may not be the most appropriate.

Missing data in the ITT also corresponds to the removal of patients due to adverse effects related to interventions tested in the clinical studies. These data can be classified as missing data completely at random (MCAR), missing at random data (MAR) or missing data not at random (MNAR) data. The latter type of data occurs when the missing data depends on unobserved values, so it can cause a bias in the study outcomes.

The problem generated by missing data in completely randomized clinical trials can be addressed by imputing data through statistical estimations. The two imputation methods commonly used are last observation carried forward (LOCF) or baseline observation carried forward (BOCF). LOCF uses the last observed value in place of



the missing outcome, whereas BOCF uses the baseline value to replace the missing value. Both methods have been widely criticized because they introduce a bias. A newer method is multiple imputations. This involves the estimation of plausible values based on a Bayesian approach⁴ in order to replace the missing values.

Despite the fact that there are many methods and approaches to manage missing data, the first recommen-

dation is to prevent missing data. Therefore, in clinical research the variables of interest should be planned before the study and rigorously measured over the data collection.

MABEL BRUNOTTO. DDS MSc PhD.

Facultad de Odontología,
Universidad Nacional de Córdoba, Argentina.

REFERENCES.

1. Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs; a review of the top medical journals. *BMC Med Res Methodol.* 2014;14:118.
2. Enders CK. A Primer on Maximum Likelihood Algorithms Available for Use With Missing Data. *Struct Equation Modeling.* 2001;8(1):128–41.
3. Peugh JL, Enders CK. Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement. *Rev Educ Res.* 2004;74(4):525–56.
4. Fiero M, Huang S, Bell ML. Statistical analysis and handling of missing data in cluster randomised trials: protocol for a systematic review. *BMJ Open.* 2015;5:e007378.