MISSING DATA: IS IT ALL THE SAME?

Sian Lydersen, Norwegian University of Science and Technology, Department of mental health, Trondheim, Norway

Introduction: In most studies, there are ‘holes’ in the data set, such that data are missing partly or completely for some of the subjects. This results in reduced statistical power. More seriously, missing data may cause biased results, since data are usually not missing completely at random.

Missing data mechanisms: In which way, and to what extent, does the probability that data are missing, depend on observed and/ or unobserved data values? This is called the missing data mechanism, and is important for an appropriate choice of analysis method. Three types of missing data mechanisms have been defined, see for example (Sterne et al. 2009).

MCAR: Missing completely at random. There are no systematic differences between the missing values and the observed values. The probability that values are missing does not depend on any of the data, observed or unobserved.

MAR: Missing at random. The probability that values are missing may depend on observed data values, but not on unobserved data values. For example, some variables are missing more frequently for patients with higher age. Data are missing at random conditionally on observed data.

MNAR: Missing not at random. Even after the observed data are taken into account, systematic differences remain between the missing values and the observed values: The probability that values are missing depends on unobserved data values as well. It is possible to distinguish between MCAR and MAR by inspecting the data at hand. But it is never possible to determine whether data are MNAR from your data.

Methods for handling missing data

Some commonly used methods, from the less to the more complex ones, are listed below, with an indication of when they give unbiased estimates:

- Complete case analysis (disregarding cases with partially missing data) (MCAR)
- Single imputation methods (sometimes under MAR but underestimates uncertainty)
- Multiple imputation (MAR)
- Full information maximum likelihood (MAR)
- Linear mixed models in longitudinal studies (MAR)

In longitudinal studies, last observation carried forward (LOCF) is a simple method for imputing missing values. But it is not unbiased under any sensible assumptions, and should not be used (Lydersen 2019).

Reporting: In general, report the amount of missing data in the different variables, and how this was handled in the analysis. This is recommended in the Consort Guidelines (Moher et al. 2010) as well as the STROBE Statement (STROBE 2014), see also (Lydersen 2015).

REFERENCES:


Disclosure of Interests: None declared


CON SESSION: PREPARE YOUR DATA COLLECTION

WHAT DO WE KNOW – WHAT SHALL WE DO – WHAT DO WE TELL

Timothy R. Radstake, University Medical Center Utrecht, Netherlands

There is an increasing need for patients to be involved in research as well as matters that can influence their own clinical journey. One aspect of science that links optimal clinical care is the increased use of biomarkers for so-called personalized medicine approaches.