The problem of missing values – under- or overestimated? Some theses

Workshop “Missing Data”

Köln, 2004-12-03

summarized by

Gerhard Nehmiz

Boehringer Ingelheim Pharma GmbH & Co KG

Under- or overestimated?

Boehringer Ingelheim

It depends!

Under- or overestimated?

(1)

If only very few data are missing, it can be argued that a detailed analysis is not necessary.

What means “very few”, is problem-dependent. It can be valuable to look at single cases with missing values, in order to detect problems.

If the data are observed in short time distances from each other, the problems are usually smaller than if the measurements are sparse.

Not only the missingness itself is the problem but also imbalanced missingness between observation groups (e.g. treatments).

All this does not mean that no prespecification is to be made. The prespecification can however be that no special investigations will be made.

Under- or overestimated?

Boehringer Ingelheim

(2)

Deviations from the MAR assumption are common but their relevance is difficult to assess.

- It is worth-while to define influence measures, and to develop graphical diagnostic displays.

See e.g. Verbeke G et al.: Sensitivity Analysis for Nonrandom Dropout: A Local Influence Approach. Biometrics 2001; 57 (1): 7-14.

The effect and/or its SE could be shown in dependence from indicators for differential missingness like the IMOR (see presentation of I White)

Under- or overestimated?

(3)
Methodology is most advanced for quantitative (normally-distributed) data. Except for the GLMM, not much is known for ordinal, binary and (interval-)censored data.

- It is worth-while to continue the investigation of dependent censoring, competing risks, ...

Under- or overestimated?

Boehringer Ingelheim

(4)

Post-hoc comparisons of baseline values / covariates between completed and discontinued subjects are not meaningful.

- It is necessary to investigate the prognosis for completion, dependent on early values.

The pattern-mixture model does not contradict to this thesis. It predicts the outcome of the experiment (e.g. the treatment difference), taking into account differential missingness.

Under- or overestimated?

(5)

More methodology for analysis of intermittent missing data is needed, in addition to methods for drop-out data.

Under- or overestimated?

Boehringer Ingelheim

(6)

Analysis of, and extrapolation from, the data should use any available external (medical, ...) knowledge.

- This is implicitly Bayesian.

The external contribution is in practice often qualitative, e.g. exclusion of certain effects or influence factors.

Subject-matter knowledge is necessary if, in a special context, different data-analytic methods lead to divergent results.