

"Missing Data in Clinical Studies"

Geert Molenberghs and Michael G. Kenward, 2007, 504 pages, John Wiley & Sons, \$110.00

Review by Norman M. Goldfarb

"Missing Data in Clinical Studies" provides sophisticated, yet practical, methods for modeling datasets with missing data. These methods minimize bias and maximize information from the data that are available. Of course, the missing data are still missing, so study results are not as definitive as they would be with complete data.

Almost all clinical studies generate data that are incomplete: Subjects miss visits, lose interest, or expire.

Investigators neglect to record values or record invalid data. Lab

reports and case report form pages mysteriously disappear. The simple methods for handling missing data are to discard subjects without reasonably complete data (complete case analysis) or to assume that the missing values are the same as previously recorded values (last observation carried forward). Both of these methods require unrealistic assumptions that are likely to distort the results. For example, if a subject drops out because of nausea (unknown to us), the complete case analysis method requires us to pretend he never existed. If a subject misses a visit because she has been hospitalized for chest pains (unknown to us), the last observation carried forward method forces us to pretend that everything continued normally.

This book has been selected for
[The First Clinical Research Bookshelf](#)
Essential reading for clinical research professionals

In 1976, Donald B. Rubin made the key insight in the analysis of missing data. He classified missing data into three types:

- **Missing Completely at Random (MCAR)**. Data that may be missing because of observed covariates (e.g., age or dosage), but not because of unobserved covariates, observed outcomes, or unobserved outcomes.
- **Missing at Random (MAR)**. Data that may be missing because of observed outcomes and also possibly because of observed covariates, but not because of unobserved covariates or outcomes.
- **Missing Not at Random (MNAR)**. Data that may be missing because of unobserved measurements and also possibly because of observed measurements.

Depending on the classification, study results can be corrected to a greater or lesser extent. For example, if the study drug worked well with 70% of female subjects, but 10% of female subjects dropped out of the study because of childcare obligations (MCAR), it can be assumed that a 70% of the drop-outs experienced positive effects from the study drug.

The book consists of 25 chapters in seven sections:

- Preliminaries
- Classical Techniques and the Need for Modeling
- Missing at Random
- Ignorability
- Missing Not at Random

- Sensitivity Analysis
- Case Studies

Seventeen examples demonstrate how to use the methods presented in the text.

The book is available in bookstores.

Reviewer

Norman M. Goldfarb is Managing Director of First Clinical Research LLC, a provider of clinical research best practices information, consulting and training services. Contact him at 1.650.465.0119 or ngoldfarb@firstclinical.com.