

Missing Data

CCTS Biostatistics Core

February 2019

Missing data is common in health research and must be dealt with before and during analysis. Missing data is a topic of active research. Many publications discuss the causes and patterns of missing data, as well as statistical and practical approaches. [Some resources are listed below.](#)

Multiple Imputation

Multiple imputation is a common way to deal with missing data.

Imputation means filling in a missing data point with a value. In the past, it was common to fill in “holes” in the data with the mean of a variable’s available values. This approach, called *single imputation*, leads to underestimation of a variable’s variance, since one presumed value is standing in for many different unknown values. It can lead to erroneous inference because too many statistical tests return significant results.

Multiple imputation (MI) uses regression on other variables in the data set to generate plausible values where data are missing. It also makes sure there’s some variance in the predicted plausible values. The process includes the following steps:

1. An MI program creates, say, 30 completed data sets, each one including a different dose of random noise
2. The analyst then runs her analysis (e.g., linear mixed model for repeated measures) on each one of the 30 data sets separately
3. The results of each of those 30 analyses are averaged. There’s a special formula to make sure the standard errors of estimates take into account the extra uncertainty

Most major statistical programs can assist with the generation of multiply imputed data sets and the combination of repeated analyses. Two common software options are [PROC MI in SAS](#) and [the mice package in R](#).

Resources

This page includes publications and tools that our consultants have found useful. For more information on this topic, including advice about how to apply it in your research, consider [scheduling a consultation with a biostatistician](#).

While we hope this resource list serves as a helpful starting point for other researchers, we provide no guarantee of its comprehensiveness or of the accuracy or reliability of the works cited. If you have concerns or suggestions to improve this page, please [contact us](#).

Understanding Missing Data

Akl, E. A., M. Briel, J. J. You, et al. (2012). “Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review”. In: *BMJ* 344.may18 1, p. e2809–e2809. ISSN: 1756-1833. DOI: 10.1136/bmj.e2809. <http://dx.doi.org/10.1136/bmj.e2809>.

- Allison, Paul (2002). *Missing Data*. SAGE Publications, Inc. ISBN: 9781412985079. DOI: 10.4135/9781412985079. <http://dx.doi.org/10.4135/9781412985079>.
- Alosh, Mohamed (2009). "The Impact of Missing Data in a Generalized Integer-Valued Autoregression Model for Count Data". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 1039–1054. ISSN: 1520-5711. DOI: 10.1080/10543400903242787. <http://dx.doi.org/10.1080/10543400903242787>.
- Angrist, Joshua D. and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. ISBN: 9780691120348. DOI: 10.2307/j.ctvc4j72. <http://dx.doi.org/10.2307/j.ctvc4j72>.
- Bell, Melanie L, Mallorie Fiero, Nicholas J Horton, et al. (2014). "Handling missing data in RCTs; a review of the top medical journals". In: *BMC Medical Research Methodology* 14.1. ISSN: 1471-2288. DOI: 10.1186/1471-2288-14-118. <http://dx.doi.org/10.1186/1471-2288-14-118>.
- Beunckens, Caroline, Geert Molenberghs, Herbert Thijs, et al. (2007). "Incomplete hierarchical data". In: *Statistical Methods in Medical Research* 16.5, p. 457–492. ISSN: 1477-0334. DOI: 10.1177/0962280206075310. <http://dx.doi.org/10.1177/0962280206075310>.
- Bhaumik, Dulal K., Anindya Roy, Subhash Aryal, et al. (2008). "Sample Size Determination for Studies with Repeated Continuous Outcomes". In: *Psychiatric Annals* 38.12. ISSN: 1938-2456. DOI: 10.3928/00485713-20081201-01. <http://dx.doi.org/10.3928/00485713-20081201-01>.
- Birmingham, Jolene and Garrett M. Fitzmaurice (2002). "A Pattern-Mixture Model for Longitudinal Binary Responses with Nonignorable Nonresponse". In: *Biometrics* 58.4, p. 989–996. ISSN: 0006-341X. DOI: 10.1111/j.0006-341x.2002.00989.x. <http://dx.doi.org/10.1111/j.0006-341x.2002.00989.x>.
- Chen, Qixuan, Andrew Gelman, Melissa Tracy, et al. (2015). "Incorporating the sampling design in weighting adjustments for panel attrition". In: *Statistics in Medicine* 34.28, p. 3637–3647. ISSN: 1097-0258. DOI: 10.1002/sim.6618. <http://dx.doi.org/10.1002/sim.6618>.
- Daniels, Michael J. and Joseph W. Hogan (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman and Hall/CRC. ISBN: 9780429145704. DOI: 10.1201/9781420011180. <http://dx.doi.org/10.1201/9781420011180>.
- Demirtas, Hakan and Joseph L. Schafer (2003). "On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out". In: *Statistics in Medicine* 22.16, p. 2553–2575. ISSN: 1097-0258. DOI: 10.1002/sim.1475. <http://dx.doi.org/10.1002/sim.1475>.
- DeSouza, Cynthia M., Anna T. R. Legedza, and Abdul J. Sankoh (2009). "An Overview of Practical Approaches for Handling Missing Data in Clinical Trials". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 1055–1073. ISSN: 1520-5711. DOI: 10.1080/10543400903242795. <http://dx.doi.org/10.1080/10543400903242795>.
- Díaz-Ordaz, Karla, Michael G Kenward, Abie Cohen, et al. (2014). "Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines". In: *Clinical Trials* 11.5, p. 590–600. ISSN: 1740-7753. DOI: 10.1177/1740774514537136. <http://dx.doi.org/10.1177/1740774514537136>.
- Fiero, Mallorie H., Shuang Huang, Eyal Oren, et al. (2016). "Statistical analysis and handling of missing data in cluster randomized trials: a systematic review". In: *Trials* 17.1. ISSN: 1745-6215. DOI: 10.1186/s13063-016-1201-z. <http://dx.doi.org/10.1186/s13063-016-1201-z>.
- Fitzmaurice, Garrett M., Nan M. Laird, and James H. Ware (2011). *Applied Longitudinal Analysis*. Wiley. ISBN: 9781119513469. DOI: 10.1002/9781119513469. <http://dx.doi.org/10.1002/9781119513469>.
- Flyer, Paul and Joseph Hirman (2009). "Missing Data in Confirmatory Clinical Trials". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 969–979. ISSN: 1520-5711. DOI: 10.1080/10543400903242746. <http://dx.doi.org/10.1080/10543400903242746>.
- Gibbons, Robert D. (2008). "Design and Analysis of Longitudinal Studies". In: *Psychiatric Annals* 38.12. ISSN: 1938-2456. DOI: 10.3928/00485713-20081201-03. <http://dx.doi.org/10.3928/00485713-20081201-03>.

Giusta, Caterina and Roderick J. A. Little “An Analysis of nonignorable nonresponse to income in a survey with rotating panel design”. In: *Journal of Official Statistics* 27.2, pp. 211-229.

Graham, John W. (2009). “Missing Data Analysis: Making It Work in the Real World”. In: *Annual Review of Psychology* 60.1, p. 549–576. ISSN: 1545-2085. DOI: 10.1146/annurev.psych.58.110405.085530. <http://dx.doi.org/10.1146/annurev.psych.58.110405.085530>.

Graham, John W., Scott M. Hofer, Stewart I. Donaldson, et al. (1997). “Analysis with missing data in prevention research.” In: *The science of prevention: Methodological advances from alcohol and substance abuse research*. American Psychological Association, p. 325–366. ISBN: 1557984395. DOI: 10.1037/10222-010. <http://dx.doi.org/10.1037/10222-010>.

Hedeker, Donald and Robert D. Gibbons (1997). “Application of random-effects pattern-mixture models for missing data in longitudinal studies.” In: *Psychological Methods* 2.1, p. 64–78. ISSN: 1082-989X. DOI: 10.1037/1082-989x.2.1.64. <http://dx.doi.org/10.1037/1082-989X.2.1.64>.

Hedeker, Donald and Robert D. Gibbons (2006). *Longitudinal Data Analysis*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9780470036488. DOI: 10.1002/0470036486. <http://dx.doi.org/10.1002/0470036486>.

Hedeker, Donald and Jennifer S. Rose (2000). “The Natural History of Smoking: A Pattern–Mixture Random-Effects Regression Model”. In: *Multivariate Applications in Substance Use Research: New Methods for New Questions*. Ed. by Jennifer S. Rose, Laurie Chassin, Clark C. Presson and Steven J. Sherman. Hillsdale, NJ: Psychology Press, pp. 79-112. ISBN: 9781410604217. DOI: 10.4324/9781410604217. <http://dx.doi.org/10.4324/9781410604217>.

Herbert, Robert D., Jessica Kasza, and Kari Bø (2018). “Analysis of randomised trials with long-term follow-up”. In: *BMC Medical Research Methodology* 18.1. ISSN: 1471-2288. DOI: 10.1186/s12874-018-0499-5. <http://dx.doi.org/10.1186/s12874-018-0499-5>.

Horton, Nicholas J and Ken P Kleinman (2007). “Much Ado About Nothing: A Comparison of Missing Data Methods and Software to Fit Incomplete Data Regression Models”. In: *The American Statistician* 61.1, p. 79–90. ISSN: 1537-2731. DOI: 10.1198/000313007x172556. <http://dx.doi.org/10.1198/000313007X172556>.

Ibrahim, Joseph G. and Geert Molenberghs (2009). “Missing data methods in longitudinal studies: a review”. In: *TEST* 18.1, p. 1–43. ISSN: 1863-8260. DOI: 10.1007/s11749-009-0138-x. <http://dx.doi.org/10.1007/s11749-009-0138-x>.

Ibrahim, Joseph G, Ming-Hui Chen, Stuart R Lipsitz, et al. (2005). “Missing-Data Methods for Generalized Linear Models: A Comparative Review”. In: *Journal of the American Statistical Association* 100.469, p. 332–346. ISSN: 1537-274X. DOI: 10.1198/016214504000001844. <http://dx.doi.org/10.1198/016214504000001844>.

Kolamunnage-Dona, Ruwanthi, Colin Powell, and Paula Ruth Williamson (2016). “Modelling variable dropout in randomised controlled trials with longitudinal outcomes: application to the MAGNETIC study”. In: *Trials* 17.1. ISSN: 1745-6215. DOI: 10.1186/s13063-016-1342-0. <http://dx.doi.org/10.1186/s13063-016-1342-0>.

Kong, Fanhui, Yeh-Fong Chen, and Kun Jin (2009). “A Bias Correction in Testing Treatment Efficacy Under Informative Dropout in Clinical Trials”. In: *Journal of Biopharmaceutical Statistics* 19.6, p. 980–1000. ISSN: 1520-5711. DOI: 10.1080/10543400903242753. <http://dx.doi.org/10.1080/10543400903242753>.

Lavori, Philip W., C. Hendricks Brown, Naihua Duan, et al. (2008). “Missing Data in Longitudinal Clinical Trials Part A: Design and Conceptual Issues”. In: *Psychiatric Annals* 38.12. ISSN: 1938-2456. DOI: 10.3928/00485713-20081201-04. <http://dx.doi.org/10.3928/00485713-20081201-04>.

Little, Roderick J. A. (1988). “A Test of Missing Completely at Random for Multivariate Data with Missing Values”. In: *Journal of the American Statistical Association* 83.404, p. 1198–1202. ISSN: 1537-274X. DOI: 10.1080/01621459.1988.10478722. <http://dx.doi.org/10.1080/01621459.1988.10478722>.

Little, Roderick J. A. (1993). “Pattern-Mixture Models for Multivariate Incomplete Data”. In: *Journal of the American Statistical Association* 88.421, p. 125. ISSN: 0162-1459. DOI: 10.2307/2290705. <http://dx.doi.org/10.2307/2290705>.

- Little, Roderick J. A. (1995). "Modeling the Drop-Out Mechanism in Repeated-Measures Studies". In: *Journal of the American Statistical Association* 90.431, p. 1112–1121. ISSN: 1537-274X. DOI: 10.1080/01621459.1995.10476615. <http://dx.doi.org/10.1080/01621459.1995.10476615>.
- Little, Roderick J. A. and Nathaniel Schenker (1995). "Missing Data". In: *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. Springer US, p. 39–75. ISBN: 9781489912923. DOI: 10.1007/978-1-4899-1292-3_2. http://dx.doi.org/10.1007/978-1-4899-1292-3_2.
- Little, Roderick and Donald Rubin (2019). *Statistical Analysis with Missing Data, Third Edition*. Wiley. ISBN: 9781119482260. DOI: 10.1002/9781119482260. <http://dx.doi.org/10.1002/9781119482260>.
- Little, Roderick and Linda Yau (1996). "Intent-to-Treat Analysis for Longitudinal Studies with Drop-Outs". In: *Biometrics* 52.4. SAS code available at <http://sitemaker.umich.edu/rlittle/files/linda.htm>, p. 1324. ISSN: 0006-341X. DOI: 10.2307/2532847. <http://dx.doi.org/10.2307/2532847>.
- Ma, Guoguang, Andrea B. Troxel, and Daniel F. Heitjan (2005). "An index of local sensitivity to nonignorable drop-out in longitudinal modelling". In: *Statistics in Medicine* 24.14, pp. 2129-2150. DOI: <https://doi.org/10.1002/sim.2107>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2107>.
- Marcus, Sue M., Juned Siddique, Thomas R. Ten Have, et al. (2008). "Balancing Treatment Comparisons in Longitudinal Studies". In: *Psychiatric Annals* 38.12. ISSN: 1938-2456. DOI: 10.3928/00485713-20081201-05. <http://dx.doi.org/10.3928/00485713-20081201-05>.
- Mclsaac, Michael and RJ Cook (2017). "Statistical methods for incomplete data: Some results on model misspecification". In: *Statistical Methods in Medical Research* 26.1, pp. 248-267. DOI: 10.1177/0962280214544251. <https://doi.org/10.1177/0962280214544251>.
- Meng, Xiangyi and Nathaniel Schenker (1999). "Maximum likelihood estimation for linear regression models with right censored outcomes and missing predictors". In: *Computational Statistics & Data Analysis* 29.4, p. 471–483. ISSN: 0167-9473. DOI: 10.1016/s0167-9473(98)00074-7. [http://dx.doi.org/10.1016/S0167-9473\(98\)00074-7](http://dx.doi.org/10.1016/S0167-9473(98)00074-7).
- Molenberghs, G. (2004). "Analyzing incomplete longitudinal clinical trial data". In: *Biostatistics* 5.3, p. 445–464. ISSN: 1468-4357. DOI: 10.1093/biostatistics/kxh001. <http://dx.doi.org/10.1093/biostatistics/kxh001>.
- Molenberghs, Geert, Caroline Beunckens, Ivy Jansen, et al. (2014). "Missing Data". In: *Handbook of Epidemiology*. Springer New York, p. 1283–1335. DOI: 10.1007/978-0-387-09834-0_20. http://dx.doi.org/10.1007/978-0-387-09834-0_20.
- Molenberghs, Geert and Michael G. Kenward (2007). *Missing Data in Clinical Studies*. Wiley. ISBN: 9780470510445. DOI: 10.1002/9780470510445. <http://dx.doi.org/10.1002/9780470510445>.
- Molenberghs, Geert and Geert Verbeke (2005). *Models for discrete longitudinal data*. Springer Series in Statistics. Part VI, Chapters 26-32, considers missing data analysis.. New York: Springer Science+Business Media, Inc. DOI: 10.1007/0-387-28980-1. <http://dx.doi.org/10.1007/0-387-28980-1>.
- Molenberghs, Geert, Geert Verbeke, Herbert Thijs, et al. (2001). "Influence analysis to assess sensitivity of the dropout process". In: *Computational Statistics & Data Analysis* 37.1, p. 93–113. ISSN: 0167-9473. DOI: 10.1016/s0167-9473(00)00065-7. [http://dx.doi.org/10.1016/S0167-9473\(00\)00065-7](http://dx.doi.org/10.1016/S0167-9473(00)00065-7).
- Moore, K. L. and M. J. van der Laan (2009). "Increasing Power in Randomized Trials with Right Censored Outcomes Through Covariate Adjustment". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 1099–1131. ISSN: 1520-5711. DOI: 10.1080/10543400903243017. <http://dx.doi.org/10.1080/10543400903243017>.
- Nie, L., H. Chu, Y. Cheng, et al. (2009). "Marginal and Conditional Approaches to Multivariate Variables Subject to Limit of Detection". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 1151–1161. ISSN: 1520-5711. DOI: 10.1080/10543400903243033. <http://dx.doi.org/10.1080/10543400903243033>.
- Powney, Matthew, Paula Williamson, Jamie Kirkham, et al. (2014). "A review of the handling of missing longitudinal outcome data in clinical trials". In: *Trials* 15.1. ISSN: 1745-6215. DOI: 10.1186/1745-6215-15-237. <http://dx.doi.org/10.1186/1745-6215-15-237>.

- Rässler, Susanne, Donald B. Rubin, and Schenker, Nathaniel. (2008). "Incomplete Data". In: *International Handbook of Survey Methodology*. Ed. by Edith D. de Leeuw, Joop Hox and Don Dillman. New York: Routledge. ISBN: 9780805857528. DOI: 10.4324/9780203843123.ch19. <http://dx.doi.org/10.4324/9780203843123.ch19>.
- Raykov, Tenko and George A. Marcoulides (2013). "Identifying Useful Auxiliary Variables for Incomplete Data Analyses: A Note on a Group Difference Examination Approach". In: *Educational and Psychological Measurement* 74.3, p. 537–550. ISSN: 1552-3888. DOI: 10.1177/0013164413511326. <http://dx.doi.org/10.1177/0013164413511326>.
- Resseguier, Noémie, Roch Giorgi, and Xavier Paoletti (2011). "Sensitivity Analysis When Data Are Missing Not-at-random". In: *Epidemiology* 22.2. SensMice" package, p. 282. ISSN: 1044-3983. DOI: 10.1097/ede.0b013e318209dec7. <http://dx.doi.org/10.1097/EDE.0b013e318209dec7>.
- Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao (1995). "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data". In: *Journal of the American Statistical Association* 90.429, p. 106–121. ISSN: 1537-274X. DOI: 10.1080/01621459.1995.10476493. <http://dx.doi.org/10.1080/01621459.1995.10476493>.
- Rombach, Ines, Crispin Jenkinson, Alastair Gray, et al. (2018). "Comparison of statistical approaches for analyzing incomplete longitudinal patient-reported outcome data in randomized controlled trials". In: *Patient Related Outcome Measures* Volume 9, p. 197–209. ISSN: 1179-271X. DOI: 10.2147/prom.s147790. <http://dx.doi.org/10.2147/PROM.S147790>.
- Rombach, Ines, Oliver Rivero-Arias, Alastair M. Gray, et al. (2016). "The current practice of handling and reporting missing outcome data in eight widely used PROMs in RCT publications: a review of the current literature". In: *Quality of Life Research* 25.7, p. 1613–1623. ISSN: 1573-2649. DOI: 10.1007/s11136-015-1206-1. <http://dx.doi.org/10.1007/s11136-015-1206-1>.
- Rothmann, Mark D., Kallappa Koti, Kyung Yul Lee, et al. (2009). "Missing Data in Biologic Oncology Products". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 1074–1084. ISSN: 1520-5711. DOI: 10.1080/10543400903242993. <http://dx.doi.org/10.1080/10543400903242993>.
- Schafer, Joseph L. (1997). *Analysis of incomplete multivariate data*. Vol. 72. Monographs on Statistics and Applied Probability. London: Chapman & Hall.
- Schafer, Joseph L. and John W. Graham (2002). "Missing data: Our view of the state of the art." In: *Psychological Methods* 7.2, p. 147–177. ISSN: 1082-989X. DOI: 10.1037/1082-989x.7.2.147. <http://dx.doi.org/10.1037/1082-989X.7.2.147>.
- Seaman, Shaun R. and Stijn Vansteelandt (2018). "Introduction to Double Robust Methods for Incomplete Data". In: *Statistical Science* 33.2. ISSN: 0883-4237. DOI: 10.1214/18-sts647. <http://dx.doi.org/10.1214/18-STs647>.
- Seaman, Shaun R and Ian R White (2011). "Review of inverse probability weighting for dealing with missing data". In: *Statistical Methods in Medical Research* 22.3, p. 278–295. ISSN: 1477-0334. DOI: 10.1177/0962280210395740. <http://dx.doi.org/10.1177/0962280210395740>.
- Shardell, Michelle and Samer S. El-Kamary (2009). "Sensitivity Analysis of Informatively Coarsened Data Using Pattern Mixture Models". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 1018–1038. ISSN: 1520-5711. DOI: 10.1080/10543400903242779. <http://dx.doi.org/10.1080/10543400903242779>.
- Shen, Shuyi, Caroline Beunckens, Craig Mallinckrodt, et al. (2006). "A Local Influence Sensitivity Analysis for Incomplete Longitudinal Depression Data". In: *Journal of Biopharmaceutical Statistics* 16.3, p. 365–384. ISSN: 1520-5711. DOI: 10.1080/10543400600609510. <http://dx.doi.org/10.1080/10543400600609510>.
- Siddique, Juned, C. Hendricks Brown, Donald Hedeker, et al. (2008). "Missing Data in Longitudinal Trials – Part B, Analytic Issues". In: *Psychiatric Annals* 38.12. ISSN: 1938-2456. DOI: 10.3928/00485713-20081201-09. <http://dx.doi.org/10.3928/00485713-20081201-09>.

- Soon, Guoxing (Greg) (2009). "Editorial: Missing Data—Prevention and Analysis". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 941–944. ISSN: 1520-5711. DOI: 10.1080/10543400903331226. <http://dx.doi.org/10.1080/10543400903331226>.
- Sullivan, Thomas R, Lisa N Yelland, Katherine J Lee, et al. (2017). "Treatment of missing data in follow-up studies of randomised controlled trials: A systematic review of the literature". In: *Clinical Trials* 14.4, p. 387–395. ISSN: 1740-7753. DOI: 10.1177/1740774517703319. <http://dx.doi.org/10.1177/1740774517703319>.
- Ten Have, Thomas R., Sharon-Lise T. Normand, Sue M. Marcus, et al. (2008). "Intent-to-treat vs. Non-intent-to-treat Analyses under Treatment Non-adherence in Mental Health Randomized Trials". In: *Psychiatric Annals* 38.12. ISSN: 1938-2456. DOI: 10.3928/00485713-20081201-10. <http://dx.doi.org/10.3928/00485713-20081201-10>.
- Thijs, Herbert, Geert Molenberghs, and Geert Verbeke (2000). "The Milk Protein Trial: Influence Analysis of the Dropout Process". In: *Biometrical Journal* 42.5, p. 617–646. ISSN: 1521-4036. DOI: 10.1002/1521-4036(200009)42:5<617::aid-bimj617>3.0.co;2-n. [http://dx.doi.org/10.1002/1521-4036\(200009\)42:5%3C617::AID-BIMJ617%3E3.0.CO;2-N](http://dx.doi.org/10.1002/1521-4036(200009)42:5%3C617::AID-BIMJ617%3E3.0.CO;2-N).
- Troxel, Andrea, Guoguang Ma, and Daniel F. Heitjan (2004). "An index of local sensitivity to nonignorability". In: *Statistica Sinica* 14.4. ISNI index, pp. 1221-1237. <http://www.jstor.org/stable/24307229>.
- Verbeke, Geert and Geert Molenberghs (2000). *Linear mixed models for longitudinal data*. Springer Series in Statistics. Chapters 16-21 consider missing data analysis.. New York: Springer-Verlag. DOI: 10.1007/b98969. <http://dx.doi.org/10.1007/b98969>.
- Verbeke, Geert, Geert Molenberghs, and Caroline Beunckens (2008). "Formal and informal model selection with incomplete data". In: *Statistical Science* 23.2, pp. 201-218. DOI: 10.48550/ARXIV.0808.3587. <https://arxiv.org/abs/0808.3587>.
- Verbeke, Geert, Geert Molenberghs, Herbert Thijs, et al. (2001). "Sensitivity Analysis for Nonrandom Dropout: A Local Influence Approach". In: *Biometrics* 57.1, p. 7–14. ISSN: 1541-0420. DOI: 10.1111/j.0006-341x.2001.00007.x. <http://dx.doi.org/10.1111/j.0006-341x.2001.00007.x>.
- Walton, Marc K. (2009). "Addressing and Advancing the Problem of Missing Data". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 945–956. ISSN: 1520-5711. DOI: 10.1080/10543400903238959. <http://dx.doi.org/10.1080/10543400903238959>.
- Wang, Ming-Dauh, Jiajun Liu, Geert Molenberghs, et al. (2018). "An evaluation of the trimmed mean approach in clinical trials with dropout". In: *Pharmaceutical Statistics* 17.3, p. 278–289. ISSN: 1539-1612. DOI: 10.1002/pst.1858. <http://dx.doi.org/10.1002/pst.1858>.
- Wang, Xiaofei, Yougui Wu, and Haibo Zhou (2009). "Outcome- and Auxiliary-Dependent Subsampling and Its Statistical Inference". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 1132–1150. ISSN: 1520-5711. DOI: 10.1080/10543400903243025. <http://dx.doi.org/10.1080/10543400903243025>.
- Wittes, Janet (2009). "Missing Inaction: Preventing Missing Outcome Data in Randomized Clinical Trials". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 957–968. ISSN: 1520-5711. DOI: 10.1080/10543400903239825. <http://dx.doi.org/10.1080/10543400903239825>.
- Wood, Angela M, Ian R White, and Simon G Thompson (2004). "Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals". In: *Clinical Trials* 1.4, p. 368–376. ISSN: 1740-7753. DOI: 10.1191/1740774504cn032oa. <http://dx.doi.org/10.1191/1740774504cn032oa>.
- Xie, Hui (2007). "A local sensitivity analysis approach to longitudinal non-Gaussian data with non-ignorable dropout". In: *Statistics in Medicine* 27.16, p. 3155–3177. ISSN: 1097-0258. DOI: 10.1002/sim.3117. <http://dx.doi.org/10.1002/sim.3117>.
- Xie, Hui (2009). "Bayesian inference from incomplete longitudinal data: A simple method to quantify sensitivity to nonignorable dropout". In: *Statistics in Medicine* 28.22, p. 2725–2747. ISSN: 1097-0258. DOI: 10.1002/sim.3655. <http://dx.doi.org/10.1002/sim.3655>.

- Xie, Hui (2010). "Adjusting for Nonignorable Missingness When Estimating Generalized Additive Models". In: *Biometrical Journal* 52.2, p. 186–200. ISSN: 1521-4036. DOI: 10.1002/bimj.200900202. <http://dx.doi.org/10.1002/bimj.200900202>.
- Xie, Hui (2012). "Analyzing longitudinal clinical trial data with nonignorable missingness and unknown missingness reasons". In: *Computational Statistics & Data Analysis* 56.5, p. 1287–1300. ISSN: 0167-9473. DOI: 10.1016/j.csda.2010.11.021. <http://dx.doi.org/10.1016/j.csda.2010.11.021>.
- Xie, Hui, Weihua Gao, Baodong Xing, et al. (2018). "Measuring the Impact of Nonignorable Missingness Using the R Package isni". In: *Computer Methods and Programs in Biomedicine* 164, p. 207–220. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2018.06.014. <http://dx.doi.org/10.1016/j.cmpb.2018.06.014>.
- Xie, Hui and Daniel F. Heitjan (2009). "Local Sensitivity to Nonignorability: Dependence on the Assumed Dropout Mechanism". In: *Statistics in Biopharmaceutical Research* 1.3, p. 243–257. ISSN: 1946-6315. DOI: 10.1198/sbr.2009.0028. <http://dx.doi.org/10.1198/sbr.2009.0028>.
- Xie, Hui and Yi Qian (2012). "Measuring the impact of nonignorability in panel data with non-monotone non-response". In: *Journal of Applied Econometrics* 27.1, p. 129–159. ISSN: 1099-1255. DOI: 10.1002/jae.1157. <http://dx.doi.org/10.1002/jae.1157>.
- Xie, Hui, Yi Qian, and Leming Qu (2010). *A Semiparametric Approach for Analyzing Nonignorable Missing Data*. DOI: 10.3386/w16270. <http://dx.doi.org/10.3386/w16270>.
- Xu, Shu and Shelley A. Blozis (2011). "Sensitivity Analysis of Mixed Models for Incomplete Longitudinal Data". In: *Journal of Educational and Behavioral Statistics* 36.2, p. 237–256. ISSN: 1935-1054. DOI: 10.3102/1076998610375836. <http://dx.doi.org/10.3102/1076998610375836>.
- Yan, Xu, Shiohjen Lee, and Ning Li (2009). "Missing Data Handling Methods in Medical Device Clinical Trials". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 1085–1098. ISSN: 1520-5711. DOI: 10.1080/10543400903243009. <http://dx.doi.org/10.1080/10543400903243009>.
- Zhang, Hui and Myunghee Cho Paik (2009). "Handling Missing Responses in Generalized Linear Mixed Model Without Specifying Missing Mechanism". In: *Journal of Biopharmaceutical Statistics* 19.6, p. 1001–1017. ISSN: 1520-5711. DOI: 10.1080/10543400903242761. <http://dx.doi.org/10.1080/10543400903242761>.
- Zhu, Hong-Tu and Sik-Yum Lee (2001). "Local Influence for Incomplete Data Models". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 63.1, p. 111–126. ISSN: 1467-9868. DOI: 10.1111/1467-9868.00279. <http://dx.doi.org/10.1111/1467-9868.00279>.

Multiple Imputation

- Ayele, Birhanu Teshome, Ilya Lipkovich, Geert Molenberghs, et al. (2014). "A Multiple-Imputation-Based Approach to Sensitivity Analyses and Effectiveness Assessments in Longitudinal Clinical Trials". In: *Journal of Biopharmaceutical Statistics* 24.2, p. 211–228. ISSN: 1520-5711. DOI: 10.1080/10543406.2013.859148. <http://dx.doi.org/10.1080/10543406.2013.859148>.
- Azur, Melissa J., Elizabeth A. Stuart, Constantine Frangakis, et al. (2011). "Multiple imputation by chained equations: what is it and how does it work?" In: *International Journal of Methods in Psychiatric Research* 20.1, p. 40–49. ISSN: 1557-0657. DOI: 10.1002/mpr.329. <http://dx.doi.org/10.1002/mpr.329>.
- Barnard, John, Donald B. Rubin, and Nathaniel Schenker (2005). *Multiple Imputation Methods*. DOI: 10.1002/0470011815.b2a16040. <http://dx.doi.org/10.1002/0470011815.b2a16040>.
- Berglund, Patricia and Steven G. Heeringa (2014). *Multiple imputation of missing data using SAS*. ISBN: 9781629592039. Cary, NC: SAS Institute, Inc.
- Buuren, Stef van (2007). "Multiple imputation of discrete and continuous data by fully conditional specification". In: *Statistical Methods in Medical Research* 16.3, p. 219–242. ISSN: 1477-0334. DOI: 10.1177/0962280206074463. <http://dx.doi.org/10.1177/0962280206074463>.

- Buuren, Stef van (2018). *Flexible Imputation of Missing Data, Second Edition*. Chapman and Hall/CRC. ISBN: 9780429492259. DOI: 10.1201/9780429492259. <http://dx.doi.org/10.1201/9780429492259>.
- Enders, Craig K., Brian T. Keller, and Roy Levy (2018). "A fully conditional specification approach to multilevel imputation of categorical and continuous variables." In: *Psychological Methods* 23.2, p. 298–317. ISSN: 1082-989X. DOI: 10.1037/met0000148. <http://dx.doi.org/10.1037/met0000148>.
- Galimard, Jacques-Emmanuel, Sylvie Chevret, Emmanuel Curis, et al. (2018). "Heckman imputation models for binary or continuous MNAR outcomes and MAR predictors". In: *BMC Medical Research Methodology* 18.1. ISSN: 1471-2288. DOI: 10.1186/s12874-018-0547-1. <http://dx.doi.org/10.1186/s12874-018-0547-1>.
- Galimard, Jacques-Emmanuel, Sylvie Chevret, Camelia Protopopescu, et al. (2016). "A multiple imputation approach for MNAR mechanisms compatible with Heckman's model". In: *Statistics in Medicine* 35.17, p. 2907–2920. ISSN: 1097-0258. DOI: 10.1002/sim.6902. <http://dx.doi.org/10.1002/sim.6902>.
- Gomes, Manuel, Karla Díaz-Ordaz, Richard Grieve, et al. (2013). "Multiple Imputation Methods for Handling Missing Data in Cost-effectiveness Analyses That Use Data from Hierarchical Studies: An Application to Cluster Randomized Trials". In: *Medical Decision Making* 33.8, p. 1051–1063. ISSN: 1552-681X. DOI: 10.1177/0272989x13492203. <http://dx.doi.org/10.1177/0272989x13492203>.
- Graham, John W. and Joseph L. Schafer (1999). "On the performance of multiple imputation for multivariate data with small sample size". In: *Statistical strategies for small sample research*. Ed. by R. Hoyle. Thousand Oaks, CA: Sage Publications, Inc., pp. 1-29.
- Hayati Rezvan, Panteha, Katherine J Lee, and Julie A Simpson (2015). "The rise of multiple imputation: a review of the reporting and implementation of the method in medical research". In: *BMC Medical Research Methodology* 15.1. ISSN: 1471-2288. DOI: 10.1186/s12874-015-0022-1. <http://dx.doi.org/10.1186/s12874-015-0022-1>.
- Héraud-Bousquet, Vanina, Christine Larsen, James Carpenter, et al. (2012). "Practical considerations for sensitivity analysis after multiple imputation applied to epidemiological studies with incomplete data". In: *BMC Medical Research Methodology* 12.1. ISSN: 1471-2288. DOI: 10.1186/1471-2288-12-73. <http://dx.doi.org/10.1186/1471-2288-12-73>.
- Jakobsen, Janus Christian, Christian Gluud, Jørn Wetterslev, et al. (2017). "When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts". In: *BMC Medical Research Methodology* 17.1. ISSN: 1471-2288. DOI: 10.1186/s12874-017-0442-1. <http://dx.doi.org/10.1186/s12874-017-0442-1>.
- Leacy, Finbarr P., Sian Floyd, Tom A. Yates, et al. (2017). "Analyses of Sensitivity to the Missing-at-Random Assumption Using Multiple Imputation With Delta Adjustment: Application to a Tuberculosis/HIV Prevalence Survey With Incomplete HIV-Status Data". In: *American Journal of Epidemiology*. ISSN: 1476-6256. DOI: 10.1093/aje/kww107. <http://dx.doi.org/10.1093/aje/kww107>.
- Lewis-Beck, Michael, Alan Bryman, and Tim Futing Liao (2004). *The SAGE Encyclopedia of Social Science Research Methods*. Imputation chapter. DOI: 10.4135/9781412950589. <http://dx.doi.org/10.4135/9781412950589>.
- Ma, Jinhui, Noori Akhtar-Danesh, Lisa Dolovich, et al. (2011). "Imputation strategies for missing binary outcomes in cluster randomized trials". In: *BMC Medical Research Methodology* 11.1. ISSN: 1471-2288. DOI: 10.1186/1471-2288-11-18. <http://dx.doi.org/10.1186/1471-2288-11-18>.
- Rubin, Donald B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley. ISBN: 9780470316696. DOI: 10.1002/9780470316696. <http://dx.doi.org/10.1002/9780470316696>.
- Rubin, Donald B. and Nathaniel Schenker (1991). "Multiple imputation in health-care databases: An overview and some applications". In: *Statistics in Medicine* 10.4, p. 585–598. ISSN: 1097-0258. DOI: 10.1002/sim.4780100410. <http://dx.doi.org/10.1002/sim.4780100410>.
- Rubin, Donald B. and Nathaniel Schenker (2005). "Imputation". In: *Encyclopedia of statistical sciences*. Ed. by S. Kotz, C. B. Read and D. L. Banks. New York: John Wiley & Sons, Inc., pp. 336-342. ISBN:

9780471667193. DOI: 10.1002/0471667196.ess0659.pub2. <http://dx.doi.org/10.1002/0471667196.ess0659.pub2>.

SAS Institute Inc. (2015). "The MI Procedure". In: *SAS/STAT 14.1 User's Guide*. Cary, NC. Chap. 75. <https://support.sas.com/documentation/onlinedoc/stat/141/mi.pdf>.

Schafer, Joseph L (1999). "Multiple imputation: a primer". In: *Statistical Methods in Medical Research* 8.1, p. 3–15. ISSN: 1477-0334. DOI: 10.1177/096228029900800102. <http://dx.doi.org/10.1177/096228029900800102>.

Schafer, Joseph L. and Maren K. Olsen (1998). "Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective". In: *Multivariate Behavioral Research* 33.4, p. 545–571. ISSN: 1532-7906. DOI: 10.1207/s15327906mbr3304_5. http://dx.doi.org/10.1207/s15327906mbr3304_5.

Schafer, Joseph L. and Nathaniel Schenker (2000). "Inference with Imputed Conditional Means". In: *Journal of the American Statistical Association* 95.449, p. 144–154. ISSN: 1537-274X. DOI: 10.1080/01621459.2000.10473910. <http://dx.doi.org/10.1080/01621459.2000.10473910>.

Schenker, Nathaniel and Jeremy M.G. Taylor (1996). "Partially parametric techniques for multiple imputation". In: *Computational Statistics & Data Analysis* 22.4, p. 425–446. ISSN: 0167-9473. DOI: 10.1016/0167-9473(95)00057-7. [http://dx.doi.org/10.1016/0167-9473\(95\)00057-7](http://dx.doi.org/10.1016/0167-9473(95)00057-7).

Seaman, Shaun R., Ian R. White, Andrew J. Copas, et al. (2011). "Combining Multiple Imputation and Inverse-Probability Weighting". In: *Biometrics* 68.1, p. 129–137. ISSN: 1541-0420. DOI: 10.1111/j.1541-0420.2011.01666.x. <http://dx.doi.org/10.1111/j.1541-0420.2011.01666.x>.

Siddique, Juned, Ofer Harel, and Catherine M. Crespi (2012). "Addressing missing data mechanism uncertainty using multiple-model multiple imputation: Application to a longitudinal clinical trial". In: *The Annals of Applied Statistics* 6.4. ISSN: 1932-6157. DOI: 10.1214/12-aos555. <http://dx.doi.org/10.1214/12-AOAS555>.

Sterne, J. A C, I. R White, J. B Carlin, et al. (2009). "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls". In: *BMJ* 338.jun29 1, p. b2393–b2393. ISSN: 1468-5833. DOI: 10.1136/bmj.b2393. <http://dx.doi.org/10.1136/bmj.b2393>.

UCLA IDRE Statistical Consulting Group (2021). *Multiple Imputation in SAS Part 1*. https://stats.oarc.ucla.edu/u/sas/seminars/multiple-imputation-in-sas/mi_new_1/.

Van Buuren, S., J. P.L. Brand, C. G.M. Groothuis-Oudshoorn, et al. (2006). "Fully conditional specification in multivariate imputation". In: *Journal of Statistical Computation and Simulation* 76.12, p. 1049–1064. ISSN: 1563-5163. DOI: 10.1080/10629360600810434. <http://dx.doi.org/10.1080/10629360600810434>.

van Buuren, Stef and Karin Groothuis-Oudshoorn (2011). "mice: Multivariate Imputation by Chained Equations in R". In: *Journal of Statistical Software* 45.3, pp. 1-67. DOI: 10.18637/jss.v045.i03.

Wood, A. M., I. R. White, M. Hillson, et al. (2004). "Comparison of imputation and modelling methods in the analysis of a physical activity trial with missing outcomes". In: *International Journal of Epidemiology* 34.1. DOI: 10.1093/ije/dyh297 PMID: 15333619, p. 89–99. DOI: 10.1093/ije/dyh297. <http://dx.doi.org/10.1093/ije/dyh297>.