Social Science & Medicine 209 (2018) 169–173

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

The treatment of incomplete data: Reporting, analysis, reproducibility, and replicability

Yulia Sidi, Ofer Harel*

Department of Statistics, University of Connecticut, 215 Glenbrook Road, Unit 4120, Storrs, CT 06269-4120, United States

ARTICLE INFO

Keywords:
Missing data
Incomplete data
Multiple imputation
Reproducibility
Replicability

ABSTRACT

Proper analysis and reporting of incomplete data continues to be a challenging task for practitioners from various research areas. Recently Nguyen, Strazdins, Nicholson and Cooklin (NSNC; 2018) evaluated the impact of complete case analysis and multiple imputation in studies of parental employment and health. Their work joins interdisciplinary efforts to educate and motivate scientists across the research community to use principled statistical methods when analyzing incomplete data. Although we fully support and encourage work in parallel to NSNC's, we also think that further actions should be taken by the research community to improve current practices. In this commentary, we discuss some aspects and misconceptions related to analysis of incomplete data, in particular multiple imputation. In our view, the missing data problem is part of a larger problem of research reproducibility and replicability today. Thus, we believe that improving analysis and reporting of incomplete data will make reproducibility and replicability efforts easier. We also provide a brief checklist of recommendations which could be used by members of the scientific community, including practitioners, journal editors, and reviewers to set higher publication standards.

1. Introduction

A recent study by Nguyen, Strazdins, Nicholson and Cooklin (Nguyen et al., 2018; hereafter referred as NSNC) discusses the importance of properly handling missing data in studies of parental employment and health. The authors provide an excellent overview of the impact of missing data in this research area. Moreover, they compare, using a case study, two commonly used missing data techniques: complete case analysis (CCA) and multiple imputation (MI). Given the fact that CCA is the most common approach in such studies, their example signifies the implications associated with poorly handling incomplete data.

Although the missing data problem is not new, it continues to be overlooked in many research settings (Bell et al., 2014; Eekhout et al., 2012; Harel et al., 2012; Harel and Boyko, 2013; Karahalios et al., 2012; Little et al., 2012; Perkins et al., 2018; Peugh and Enders, 2004; Powney et al., 2014; Sullivan et al., 2017; Wood et al., 2004). For example, Harel et al. (2012) demonstrated that out of 57 HIV-prevention randomized trials with biological outcomes published between 2005 and 2010 in refereed journals, none mentioned missing data assumptions in their analyses, 74% performed a CCA; most seriously, only 12% are expected to report unbiased results. Eekhout et al. (2012) performed a systematic review of 262 studies published in 2010 in the three leading epidemiology journals that used questionnaires: 85% of these articles had no mention of the missingness assumption at all, and 81% used CCA. Masconi et al. (2015) considered 48 prevalent diabetes risk studies published between 1997 and 2014, where they found that 62.5% of the reported articles offered no information in regard to missing data. Nicholson et al. (2017) evaluated 541 papers related to attrition in developmental psychology, published between 2009 and 2012, to assess whether the *Publication Manual* of the American Psychological Association (2010; APA), which recommended reporting, assessment, and appropriate handling of missing data, had any effect in practice. They found that the Manual did not alter improvement in this area; only 18.3% of the articles they reviewed discussed missing data mechanisms. The common goal of the studies above, as in NSNC's, was to underline the importance of appropriate handling and reporting of incomplete data. Moreover, the variety of mentioned research fields shows that overlooking the missing data problem is not specific for a particular research area. We believe this also has a great impact on reproducibility and replicability in research.

Reproducibility, which is the ability to compute the same result, could be compromised when a published manuscript doesn't share analyzed data or programming code. In contrast, *replicability* refers to a chance of obtaining consistent results by independent studies having similar design and research questions. In general, the latter is a more

* Corresponding author.
E-mail address: ofer.harel@uconn.edu (O. Harel).

https://doi.org/10.1016/j.socscimed.2018.05.037
Received 8 March 2018; Received in revised form 11 May 2018; Accepted 19 May 2018
Available online 21 May 2018
0277-9536/

severe problem, as it puts the research community's credibility in question (Leek and Peng, 2015). The chances of achieving high replicability can be lowered by the lack of scientific reasoning for the analysis assumptions used in a previous research.

Although there are ongoing efforts in the statistical community to increase reproducibility and replicability (Leek and Peng, 2015), it needs to be extended to all areas of science. We believe that this goal could be achieved if scientists, journal editors, and reviewers set higher publication standards. Improving current practices along with more rigorous publication requirements would make reproducibility straight forward and would more importantly increase chances of replicability.

We consider the recent paper by NSNC a great addition to the interdisciplinary effort to emphasize the importance of considering the complications that arise from incomplete data in social sciences and medical research. In this commentary, we expand on some aspects of reporting and handling of incomplete data, and share our thoughts about the currently available practices mentioned by NSNC, specifically when MI is used in the statistical analysis.

2. Important aspects regarding the reporting and handling of incomplete data

2.1. Missingness mechanism - stating the assumptions

Before diving into methodological details, let us first present an artificial example that will be referred to throughout this commentary. Suppose we are interested in determining whether mental health status is associated with physical activity (high/low) and the mental health status is assessed by a questionnaire with possible outcome scores of 0–20, with lower scores representing a better mental health. We further assume that the data are collected in one wave and while the physical activity is recorded for all the subjects in the study, mental health scores are missing for 25% of the participants. In order to analyze such a dataset with incomplete data, a researcher needs to determine plausible statistical assumptions before performing any statistical analysis. These assumptions, which are embedded in any statistical method, must be explicitly noted and justified, in particular when related to missing data.

Following a general notation, let Y_{com} denote a complete dataset we aim to collect, that is, physical activity and mental health status scores. Y_{com} could be conceptually partitioned into observed Y_{obs} and missing Y_{mis} parts, while in practice we see only values for Y_{obs}. Also, let's define θ as parameter of interest (e.g., mean, regression coefficient, or odds ratio) for which we use Y_{obs} to estimate it. Further suppose that R is a matrix of the same dimension as Y_{com}, which consists of 1s where the data values are missing, and 0 otherwise. In our dataset, we would have a missing values indicator for *Mental Status* score as it is the only variable with missing data. In context of the above example, the data structure appears in Table 1.

In general, we would like to infer about θ from $P\ (Y_{com}|\theta)$; however, because the database is incomplete, a joint model of the data (Y_{com}) and missing data mechanism (R) needs to be considered. Consider the following joint model:

$$P\ (Y_{com}, R; \theta, \varphi) = P\ (Y_{com}|\theta)P\ (R|Y_{com}, \varphi)P\ (\theta, \varphi), \tag{1}$$

where φ is a nuisance parameter which characterizes the distribution of R. Due to the missing values, we are unable to summarize information in $P\ (Y_{com}; \theta,)$ (e.g., regression or ANOVA of the complete data) and need to evaluate θ from Equation (1) instead. Of course, it is a much more complicated situation.

The missing data mechanisms could be specified as: missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR) (Rubin, 1976; Little and Rubin. 2014). MCAR indicates that missing data mechanism, R, neither depends on the data Y_{com} we tried to collect, nor on any other information outside the study. In the context of our example, if some records of the mental health scores were deleted by mistake due to a technical problem it will imply that MCAR could be considered as the missing data mechanism. MAR indicates that R depends only on the observed information (Y_{obs}). In our example, this assumption suggests that only exercise status is responsible for the missing information in the mental health scores. Finally, MNAR indicates that R may depend on information that is not available to us, which either is missing due to incomplete variable(s) we are collecting or is missing due to other factors outside the study. MNAR in our example could imply that people with worse mental health status refused to answer this questionnaire. As can be seen, different reasons imply different missing data mechanisms, which consequently lead to different assumptions being made in the statistical analysis.

Certainly in practice it is hard to know the underline reasons (mechanism) that cause missing data, and while some of the missing data mechanism assumptions are testable (MCAR) (Little, 1988), others are not (MAR, MNAR) (Molenberghs et al., 2008). In particular, it is impossible to distinguish between MAR and MNAR structure, with the observed data alone. Yet, this problem makes the assumptions choice argument even more important. Thus, practitioners are encouraged to clearly specify the assumptions they use in the analysis, as well as to justify them in the context of the specific problem they study.

Table 1
Dataset structure for mental health vs. physical activity study.

Y_{obs}- Observed values in the data		R-missing values indicator	
Physical activity	Mental status score	Physical activity	Mental status score
Yes	7	0	0
Yes	?	0	1
No	15	0	0
…	…	…	…
No	13	0	0
No	?	0	1

Note. The character "?" implies a missing datum.

2.2. Ignorability — commonly confused with missing at random

Many researchers confuse ignorability with MAR, mostly because MAR is a needed component (and mostly argued) for ignorability. Yet, the two concepts, while related, differ and should be understood by those dealing with incomplete data. While missing data assumptions are necessary for proper analysis of incomplete data, these are not sufficient for ignorability. Ignorability plays a central role in the analysis of incomplete data, and is usually incorporated as default option for multiple imputation procedures in many statistical software programs (e.g., PROC MI in SAS (SAS Institute Inc, 2011), norm package in R (Novo and Schafer, 2013), and the suite of MI commands in STATA (StataCorp, 2013). As Little and Rubin (2014) described, *ignorability* consists of two assumptions: (a) MAR and (b) distinctness (or *a priori* independence in Bayesian framework) in parameters of the data model (θ) and missing data mechanism (φ). *Distinctness* can be thought of as meaning that a change in one parameter will not influence the other. For example, the mean difference in the mental scores between people with high and low physical activity (θ) is not related to the proportion of study participants who couldn't complete the mental health questionnaire due to time constraints (φ).

Consequently, non-ignorability could be attributed to either MNAR or non-distinctness (or both). Although, non-distinctness is more likely to appear in longitudinal settings, where observations are collected for the same individuals repeatedly over time (or clustered data in general), it can still lead to inefficiency in other types of studies (Little and Rubin, 2014). As was recently evaluated through a thorough simulation study conducted by Yucel (2017), non-distinctness can cause serious reductions in the coverage rates when evaluated in relation to MI. Thus, it would be more difficult to find statistically significant results that truly exist, in the presence of non-distinctness even in scenarios where MAR holds.

Due to the importance of ignorability and its impact on analysis and results, practitioners are encouraged to report, in addition to missing data mechanism, whether the distinctness of the above specified parameters is assumed. Mathematical details, showing how ignorability affects analysis of missing data, appear in Appendix A.

2.3. Multiple imputations - some misconceptions

Because MI is a powerful technique, it has become very popular in the last two decades (Carpenter and Kenward, 2012; Harel and Zhou, 2007; Mackinnon, 2010; Rezvan et al., 2015; Schafer and Graham, 2002). It was originally developed to analyze complex surveys (Rubin, 2004), but has shown to be efficient in various frameworks (Rubin and Schenker, 1991; Sterne et al., 2009). MI consists of three main steps: Imputing incomplete dataset multiple (m) times by replacing the missing data in each replicate with plausible values drawn from an imputation model, analyzing each completed dataset separately using standard statistical methods, and then combining the results utilizing Rubin's (2004) rules (See Schafer (1999) and Sterne et al. (2009) for detailed guidelines for MI implementation.). Furthermore, the assumptions mentioned above are essential for the imputation step of this process. As discussed earlier, most of the standard statistical software is programmed to implement MI based on the ignorability assumption as a default option. If the ignorability assumption is reasonable, the result of MI should be unbiased and efficient, as long as MI follows a reasonable implementation. However, performing MI under ignorability assumption, when the true missing data mechanism is non-ignorable, may lead to biased results as it may not capture the true missing data mechanism (Yucel, 2017). In this case, it is possible to impute the data under the non-ignorable model, which will solve this difficulty.

An additional point we would like to emphasize refers to results obtained from MI. According to NSNC, MI results were reported as more precise with increasing sample size and narrower confidence intervals (CIs) when compared with CCA. We agree that MI helps to estimate parameters of interest unbiasedly. However, MI generated CIs are not always shorter than those obtained from CCA, which is due to the fact that MI procedure not only increases sample size through imputation, but also inserts variability associated with missing data mechanism. Therefore, the trade-off between increasing variability (widens CI) and sample size (shortens CI) depends on the amount of the information added by MI. Furthermore, the efficiency of MI is affected by amount of missing data, the number of imputations used (Rubin, 2004; Harel, 2007), and by inclusion of auxiliary variables in the imputation model (Collins et al., 2001). To illustrate an example in which MI widens CIs, we performed a small simulation for the previously mentioned example (code available at: https://github.com/yuliasidi/CCA-vs-MI). The mental health status scores were assumed to follow a normal distribution with means of 10 and 13 for high and low physical activity groups respectively, while the variance was assumed to be three for both groups. MCAR, MAR, and MNAR were imposed so that 25% of the data was missing, the three resulting datasets were then analyzed using CCA or MI using data augmentation method, assuming ignorability and using five imputations. Confidence intervals (95%) and their length were calculated for the difference in the mental health scores between the exercise groups. In Fig. 1, the 95% CIs obtained from MI were wider than those from CCA for all scenarios. We do not argue against MI, but want users to understand what to expect when using it.

Another common misconception is that MI only works for ignorable models. While most commonly applied MI approaches rely on the ignorability assumption, it is still possible (and not very complicated, either) to perform MI for non-ignorable models (Little and Rubin, 2014; Carpenter and Kenward, 2012). In particular, the imputation stage is the only stage in the MI process that will change and it will be based on non-ignorable models (e.g., pattern-mixture models (Little, 1993), selection models (Heckman, 1976; Amemiya, 1984), or shared parameter models (Daniels and Hogan, 2008)), while the rest of the procedure (analysis of the imputed data and combining results) will follow exactly in the same manner of ignorable models.

2.4. Sensitivity analysis

Sensitivity analysis is widely used in many fields of research. The aim of this type of analysis is to evaluate how robust the results are when the model and consequently assumptions related to it change (Saltelli et al., 2000; Little et al., 2012). Thus, if the results are not altered much by different models, we are more confident about their plausibility. Still, if a small change in the model assumption causes a dramatic change in the results, it should raise a red flag and lead to a further evaluation of the statistical model and its underlying data. Performing sensitivity analysis within the MI framework has been emphasized in a number of guidelines (Burzykowski et al., 2010; Little et al., 2012; Sterne et al., 2009) and reviews (Rezvan et al., 2015; Manly and Wells, 2015), and was also recommended in the recent National Research Council (NRC, 2011; Little et al., 2012). When MI is used for primary analysis, different set of assumptions for MI could be evaluated (MAR/MNAR, ignorable/non-ignorable). Also, different missing data techniques can be used for sensitivity analyses, such as: inverse probability weighting (IPW) (Little and Rubin, 2014; Horvitz and Thompson, 1952; Robins et al., 1994; Robins and Rotnitzky, 1995), maximum likelihood (ML) (Little and Rubin, 2014), pattern-mixture models (Little, 1993), selection models (Heckman, 1976; Amemiya, 1984), and shared parameter models (Daniels and Hogan, 2008).

Recently, three companion papers by Harel et al. (2018), Perkins et al. (2018), and Sun et al. (2018) illustrated how two completely different approaches, namely MI and IPW provide very similar results when applied to three datasets, in which MCAR, MAR, and MNAR were imposed but masked from the teams analyzing them. Moreover, unlike CCA, both MI and IPW approaches were consistent with the results obtained from the full data-analysis for MAR scenario. This joint work not only demonstrates advantages of principal methods over CCA, but also shows how robustness of the results could be assessed by utilizing models based on different assumptions.

In terms of different sets of assumptions for MI, both the degree of ignorability and of missing mechanism uncertainty could be evaluated though different imputation models. As Daniels and Hogan (2008) proposed, to assess departures from MAR mechanism, the missing data mechanism could be categorized and evaluated as: MAR with no uncertainty, MAR with uncertainty, MNAR with no uncertainty, and MNAR with uncertainty. Furthermore, uncertainly of missing data mechanism could be incorporated into the imputation model as well (Siddique et al., 2012, 2014). Siddique et al. (2012) proposed a nested imputation procedure, in which imputation model uncertainty is accounted for by the following four steps: 1) distribution of the imputation models, consisting of ignorable and non-ignorable imputation models is specified 2) M imputation models are randomly drawn from that distribution, and N multiple imputations of the incomplete dataset are generated within each model, 3) standard statistical analysis is applied for each imputed dataset, 4) results of the analysis in step 4 are combined using nested multiple imputation rules (Shen, 2000; Harel, 2009). Such methods provide a comprehensive evaluation of the imputation model used by MI.

3. Conclusion

Similarly to NSNC, there are growing efforts in different areas of research to inform practitioners about consequences of ignoring the missing data problem, while presenting methods that could help to deal with missing data problems. And while these efforts are essential, unfortunately these are not sufficient in minimizing the gap between

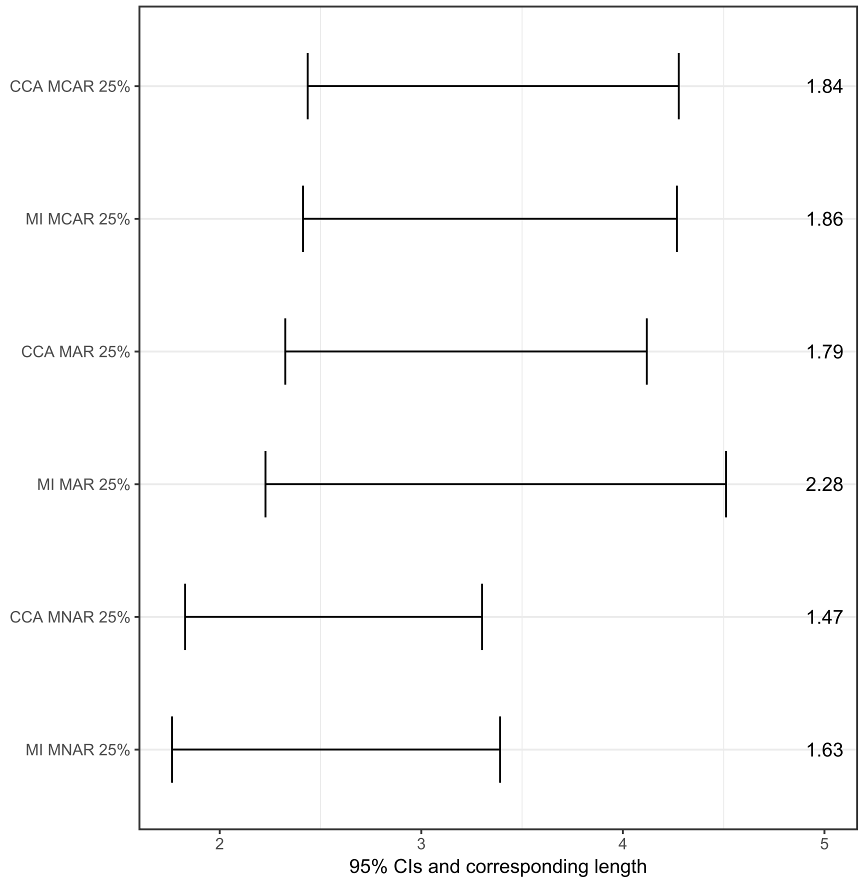

Fig. 1. CCA and MI results for simulated MCAR, MAR and MNAR dataset.

recommendations they supply and further practices. Thus, we believe that in order to trigger significant improvement in reporting and handling of incomplete data, scholars, journal editors, and reviewers should pursue higher standards as implied by this following checklist:

1) Carefully report the degree of the missing data.
2) Specify statistical analysis assumptions, in particular assumptions related to missing data (missing mechanism, ignorability).
3) Justify these assumptions in the context of the specific study.
4) Properly report results obtained from statistical analysis, especially from principled methods such as MI (e.g., reporting the rates of missing information, a measure quantifies the amount of information lost due to the missing values).
5) Perform sensitivity analysis to assess robustness of the assumptions used in the primary analysis. And,
6) Provide code along with data (if not confidential) used in the statistical analyses.

We are convinced that following these steps will foster reproducibility and replicability, and benefit the research community and all those who rely on accurate results.

Appendix A

When using MI, the general imputation model should be based on the predictive distribution of Y_{mis} given observed data Y_{obs} and missingness indicator R, i.e.

P $(Y_{mis}|Y_{obs}, R)$, while integrating out the unknown parameters θ and φ:

$$P(Y_{mis}|Y_{obs}, R) = \iint \frac{P(Y_{com}, R; \theta, \phi)}{P(Y_{obs}, R)} d\theta d\phi \tag{2}$$

replacing the numerator by Equation (1), Equation (2) becomes:

$$\frac{1}{P(Y_{obs}, R)} \iint P(Y_{com}|\theta) P(R|Y_{com}, \phi) P(\theta, \phi) d\theta d\phi \tag{3}$$

MAR and distinctness could be specified by Equation (4), and (5) respectively:

$$\mathrm{P}\ (\mathrm{R}|\mathrm{Y}_{\mathrm{com}}, \varphi) = \mathrm{P}\ (\mathrm{R}|\mathrm{Y}_{\mathrm{obs}}, \varphi), \tag{4}$$

$$P\ (\theta, \varphi) = P\ (\theta) P\ (\varphi). \tag{5}$$

Substituting Equation (4), and (5) into Equation (3) results in:

$$\mathrm{P}\ (\mathrm{Y}_{\mathrm{mis}}|\mathrm{Y}_{\mathrm{obs}}, \mathrm{R}) = \mathrm{P}\ (\mathrm{Y}_{\mathrm{mis}}|\mathrm{Y}_{\mathrm{obs}}). \tag{6}$$

Notice that instead of imputing from $P\,(Y_{mis}|Y_{obs},\,\mathrm{R})$ one now needs to impute from $P\,(Y_{mis}|Y_{obs})$ and the missingness process R is not needed any more (ignored).

In other words, Equation (6) implies that when ignorability holds (i.e., both MAR and distinctness hold), missing values could be predicted from the observed data without taking into account distribution of R. This however will not hold when either MAR or distinctness is violated. If MAR doesn't hold, then $P\,(R|Y_{com},\,\varphi)$ will have to be modeled, and if distinctness doesn't hold then the joint distribution $P\,(\theta,\,\varphi)$ will need to be modeled and correspondingly implemented into the imputation procedure. Similar formulations can be shown for maximum likelihood and Bayesian representations.

References

Amemiya, T., 1984. Tobit models: a survey. J. Econom. 24 (1–2), 3–61.

American Psychological Association, 2010. Publication Manual of the American Psychological Association, sixth ed. American Psychological Association, Washington, DC.

Bell, M.L., Fiero, M., Horton, N.J., Hsu, C.-H., 2014. Handling missing data in rcts; a review of the top medical journals. BMC Med. Res. Meth. 14 (1), 118.

Burzykowski, T., Carpenter, J., Coens, C., Evans, D., France, L., Kenward, M., Lane, P., Matcham, J., Morgan, D., Phillips, A., et al., 2010. Missing data: discussion points from the psi missing data expert group. Pharmaceut. Stat. 9 (4), 288–297.

Carpenter, J., Kenward, M., 2012. Multiple Imputation and its Application. John Wiley & Sons.

Collins, L.M., Schafer, J.L., Kam, C.-M., 2001. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol. Meth. 6 (4), 330.

Daniels, M.J., Hogan, J.W., 2008. Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis. CRC Press.

Eekhout, I., de Boer, R.M., Twisk, J.W., de Vet, H.C., Heymans, M.W., 2012. Missing data: a systematic review of how they are reported and handled. Epidemiology 23 (5), 729–732.

Harel, O., 2007. Inferences on missing information under multiple imputation and two-stage multiple imputation. Stat. Meth. 4 (1), 75–89.

Harel, O., 2009. Strategies for Data Analysis with Two Types of Missing Values. Lap Lambert Academic Publishing GmbH KG.

Harel, O., Boyko, J., 2013. Mi??ing data: should we c?re? Am. J. Publ. Health 103 (2), 200–201. https://www.ncbi.nlm.nih.gov/pubmed/23237157.

Harel, O., Mitchell, E.M., Perkins, N.J., Cole, S.R., Tchetgen Tchetgen, E.J., Sun, B., Schisterman, E.F., 2018. Multiple imputation for incomplete data in epidemiologic studies. Am. J. Epidemiol. 187 (3), 576–584.

Harel, O., Pellowski, J., Kalichman, S., 2012. Are we missing the importance of missing values in hiv prevention randomized clinical trials? review and recommendations. AIDS Behav. 16 (6), 1382–1393.

Harel, O., Zhou, X.-H., 2007. Multiple imputation: review of theory, implementation and software. Stat. Med. 26 (16), 3057–3077.

Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann. Econ. Soc. Meas. 5 (4), 475–492.

Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. 47 (260), 663–685.

Karahalios, A., Baglietto, L., Carlin, J.B., English, D.R., Simpson, J.A., 2012. A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. BMC Med. Res. Meth. 12 (1), 96.

Leek, J.T., Peng, R.D., 2015. Opinion: reproducible research can still be wrong: adopting a prevention approach. Proc. Natl. Acad. Sci. Unit. States Am. 112 (6), 1645–1646.

Little, R.J., 1988. A test of missing completely at random for multivariate data with missing values. J. Am. Stat. Assoc. 83 (404), 1198–1202.

Little, R.J., 1993. Pattern-mixture models for multivariate incomplete data. J. Am. Stat. Assoc. 88 (421), 125–134.

Little, R.J., D'agostino, R., Cohen, M.L., Dickersin, K., Emerson, S.S., Farrar, J.T., Frangakis, C., Hogan, J.W., Molenberghs, G., Murphy, S.A., et al., 2012. The prevention and treatment of missing data in clinical trials. N. Engl. J. Med. 367 (14), 1355–1360.

Little, R.J., Rubin, D.B., 2014. Statistical Analysis with Missing Data, vol. 333 John Wiley & Sons.

Mackinnon, A., 2010. The use and reporting of multiple imputation in medical research–a review. J. Intern. Med. 268 (6), 586–593.

Manly, C.A., Wells, R.S., 2015. Reporting the use of multiple imputation for missing data in higher education research. Res. High. Educ. 56 (4), 397–409.

Masconi, K.L., Matsha, T.E., Echouffo-Tcheugui, J.B., Erasmus, R.T., Kengne, A.P., 2015. Reporting and handling of missing data in predictive research for prevalent undiagnosed type 2 diabetes mellitus: a systematic review. EPMA J. 6 (1), 7.

Molenberghs, G., Beunckens, C., Sotto, C., Kenward, M.G., 2008. Every missingness not at random model has a missingness at random counterpart with equal fit. J. Roy. Stat. Soc. B 70 (2), 371–388.

Nguyen, C.D., Strazdins, L., Nicholson, J.M., Cooklin, A.R., 2018. Impact of missing data strategies in studies of parental employment and health: missing items, missing waves, and missing mothers. Soc. Sci. Med in this issue.

Nicholson, J.S., Deboeck, P.R., Howard, W., 2017. Attrition in developmental psychology: a review of modern missing data reporting and practices. Int. J. Behav. Dev. 41 (1), 143–153.

Novo, A., Schafer, J., 2013. Norm: Analysis of Multivariate Normal Datasets with Missing Values. R Package Version 1.0-9.5. https://CRAN.R-project.org/package=norm.

NRC, 2011. The Prevention and Treatment of Missing Data in Clinical Trials. National Academies Press.

Perkins, N.J., Cole, S.R., Harel, O., Tchetgen Tchetgen, E.J., Sun, B., Mitchell, E.M., Schisterman, E.F., 2018. Principled approaches to missing data in epidemiologic studies. Am. J. Epidemiol. 187 (3), 568–575.

Peugh, J.L., Enders, C.K., 2004. Missing data in educational research: a review of reporting practices and suggestions for improvement. Rev. Educ. Res. 74 (4), 525–556.

Powney, M., Williamson, P., Kirkham, J., Kolamunnage-Dona, R., 2014. A review of the handling of missing longitudinal outcome data in clinical trials. Trials 15 (1), 237.

Rezvan, P.H., Lee, K.J., Simpson, J.A., 2015. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. BMC Med. Res. Meth. 15 (1), 30.

Robins, J.M., Rotnitzky, A., 1995. Semiparametric efficiency in multivariate regression models with missing data. J. Am. Stat. Assoc. 90 (429), 122–129.

Robins, J.M., Rotnitzky, A., Zhao, L.P., 1994. Estimation of regression coefficients when some regressors are not always observed. J. Am. Stat. Assoc. 89 (427), 846–866.

Rubin, D.B., 1976. Inference and missing data. Biometrika 63 (3), 581–592.

Rubin, D.B., 2004. Multiple Imputation for Nonresponse in Surveys, vol. 81 John Wiley & Sons.

Rubin, D.B., Schenker, N., 1991. Multiple imputation in health-are databases: an overview and some applications. Stat. Med. 10 (4), 585–598.

Saltelli, A., Chan, K., Scott, E.M., et al., 2000. Sensitivity Analysis, vol. 1 Wiley New York.

SAS Institute Inc, 2011. *SAS/STATO®* 9.3 User's Guide. SAS Institute Inc, Cary, NC.

Schafer, J.L., 1999. Multiple imputation: a primer. Stat. Meth. Med. Res. 8 (1), 3–15.

Schafer, J.L., Graham, J.W., 2002. Missing data: our view of the state of the art. Psychol. Meth. 7 (2), 147.

Shen, Z., 2000. Nested Multiple Imputations. PhD thesis. Harvard University.

Siddique, J., Harel, O., Crespi, C.M., 2012. Addressing missing data mechanism uncertainty using multiple-model multiple imputation: application to a longitudinal clinical trial. Ann. Appl. Stat. 6 (4), 1814.

Siddique, J., Harel, O., Crespi, C.M., Hedeker, D., 2014. Binary variable multiple-model multiple imputation to address missing data mechanism uncertainty: application to a smoking cessation trial. Stat. Med. 33 (17), 3013–3028.

StataCorp, 2013. *Stata: Release 13.* Statistical Software. StataCorp LP, College Station, TX.

Sterne, J.A., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., Wood, A.M., Carpenter, J.R., 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 338, b2393.

Sullivan, T.R., Yelland, L.N., Lee, K.J., Ryan, P., Salter, A.B., 2017. Treatment of missing data in follow-up studies of randomised controlled trials: a systematic review of the literature. Clin. Trials 14 (4), 387–395.

Sun, B., Perkins, N.J., Cole, S.R., Harel, O., Mitchell, E., Schisterman, E.F., Tchetgen Tchetgen, E.J., 2018. Inverse probability weighted estimation for monotone and nonmonotone missing data. Am. J. Epidemiol. 187 (3), 585–591.

Wood, A.M., White, I.R., Thompson, S.G., 2004. Are missing outcome data adequately handled? a review of published randomized controlled trials in major medical journals. Clin. Trials 1 (4), 368–376.

Yucel, R., 2017. Impact of the non-distinctness and non-ignorability on the inference by multiple imputation in multivariate multilevel data: a simulation assessment. J. Stat. Comput. Simulat. 87 (9), 1813–1826.