*Journal of Statistical Software*

*MMMMMM YYYY, Volume VV, Issue II.* [*doi: 10.18637/jss.v000.i00*](https://doi.org/10.18637/jss.v000.i00)

**Imputation of Incomplete Multilevel Data with** mice

### **Hanne Oberman**

Utrecht University

### **Johanna Mu~noz Avila**

University Medical Center Utrecht

### **Valentijn de Jong**

University Medical Center Utrecht

### **Gerko Vink**

Utrecht University

### **Thomas Debray**

University Medical Center Utrecht

**Abstract**

This is a tutorial paper on imputing incomplete multilevel data with mice. Footnotes in the current version show work in progress/under construction. The last section is not part of the manuscript, but purely for reminders.

*Keywords*: missing data, multilevel, clustering, mice, R.

# **Introduction**

Contemporary decision making is driven increasingly often and sometimes even entirely dictated by data. Although data relevant for decision makers were initially collected using small and well-designed studies, there has been a growing need to address more complex questions at a wider scale. This not only requires to collect larger amounts of data, but also from more diverse settings and populations.

For example, [introduce IPDMA here; eg. HIV studies Johanna is referring to]. Another example is the use of large databases with information from thousands or even millions of individuals from multiple cities, regions or even countries. Finally, a third situation arises when data are collected from multicenter studies [maybe refer to the schools/classes example here]

For example, students may be clustered in classes in psychometrics research, or patients may be clustered in studies in individual patient data meta-analyses.

A common characteristic in aforementioned examples is the presence of clustering. [briefly explain what is clustering?]

Ignoring the clustered structure of such multilevel data can be harmful to the statistical inferences and bias results. ([Hox,](#_bookmark14) [Moerbeek, and van de Schoot](#_bookmark14) [2017](#_bookmark14)). These relations can and should be taken into account when developing analysis models for multilevel data for the simple reason that groups of observations share some common variance. The variability due to clustering is often mea- sured by means of the intraclass coefficient (ICC, [Fisher 1925](#_bookmark11)). The ICC can be seen as the percentage of variance that can be attributed to the cluster-level, where a high ICC would

[need to explain/introduce in the introduction::

* When/whyy do we need to account for clustering in the analysis of clustered data? [you already highlight this issue briefly above]
* Why is the presence of clustering relevant when considering multipe imputation of missing data ? e.g. distinction between systematically and sporadically missing data. But also: mechanisms of missing data (e.g. MCAR, MAR, MNAR) may differ between clusters. But also: relation between observed data may differ between clusters
* When/ why should we avoid using traditional imputation methods? => congeniality issues

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | cluster |  | X1 |  | X2 |  | X3 |
| 1 | 1 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| 2 | 1 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| 3 | 2 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| 4 | 2 |  |  |  |  |  |  |
|  | ... |  | ... |  | ... |  | ... |
| n | N |  |  |  |  |  |  |

Figure 1: Missingness in multilevel data

indicate that a lot of variability is due to the cluster structure.

Multilevel models typically accommodate for this variability by including a separate group mean for each cluster. In addition to random intercepts, multilevel models can also include random effects and hetero- geneous residual error variances across clusters (see e.g. [Gelman and Hill](#_bookmark12) [2006](#_bookmark12), [Hox et al.](#_bookmark14) ([2017](#_bookmark14)) and [de Jong, Moons, Eijkemans, Riley, and Debray](#_bookmark8) ([2021](#_bookmark8))). There are many names for models that take clustering into account. Some popular examples are ‘multilevel models’, ‘hierarchical models’, ‘mixed effect models’ and ‘random effect models’.

### **Missingness in multilevel data**

The process of analyzing multilevel data is further complicated when not all data entries are observed. Just as with single level data, missingness may occur at the unit level. But with multiple levels of data comes the potential for clustered missingness. Incomplete multi- level data can be categorized into two general patterns: systematic missingness and sporadic missingness ([Resche-Rigon, White, Bartlett, Peters, and Thompson 2013](#_bookmark19)). Systematic miss- ingness implies that one or more variables are never observed in a certain cluster. With sporadic missingness there may be observed data for some but not all units in a cluster ([Van Buuren 2018](#_bookmark21); [Jolani 2018](#_bookmark15)). We have visualized this difference in Figure 1, which shows an *n p* set **X** = *X*1*, . . . , Xp*, with *n* units distributed over *N* clusters, and *p* = 3 columns. Column *X*1 is completely observed, column *X*2 is systematically missing, and column *X*3 is sporadically missing. To analyze these incomplete data, we have to take the nature of the missingness and the cluster structure into account. For example, the sporadic missingness in *X*3 could be easily amended if this would be a cluster-level variable (and thus constant within clusters). We could then just extrapolate the true (but missing) value of *X*3 for unit

×

1 from unit 2, and the value for unit 4 from unit 3. If *X*3 would instead be a unit-level variable (which may vary within clusters), we could not just recover the unobserved ‘truth’, but need some kind of missing data method or discard the incomplete units altogether (i.e., list-wise deletion/complete case analysis). And, with the systematic missingness in *X*2, it is impossible to fit a multilevel model since we cannot estimate the intercept of cluster 2. We would have to exclude this cluster from our analyses entirely to obtain any results. Obviously, excluding observations is not a desirable workflow.

Ignoring the missingness in analyses can be extremely harmful to inferences. Complete case analysis can introduce bias in statistical inferences and lowers statistical power. Instead, the missingness should be accommodated before or within the analysis of scientific interest. Especially the former is very generic and popular. Imputing (i.e., filling in) the missing values separates the missing data problem from the scientific problem: missing data are replaced by plausible values whereafter the completed data is analyzed as if it were completely observed. The R package mice has become the de-facto standard for imputation by chained equations, which iteratively solves the missingness on a variable-by-variable basis. mice is known to yield valid inferences under many different missing data circumstances ([Van Buuren](#_bookmark21) [2018](#_bookmark21)). In this paper, we will discuss how to use mice in the context of multilevel data.

### **Aim of this paper**

This papers serves as a tutorial for imputing incomplete multilevel data with mice. We provide practical guidelines and code snippets for different missing data situations, including missing not at random (MNAR) mechanisms (where the probability to be missing depends on unrecorded information, making the missingness non-ignorable, [Rubin 1976](#_bookmark20); [Meng 1994](#_bookmark16)). For reasons of brevity, we focus on imputation by chained equations wit mice exclusively1. Other useful resources for the analysis of incomplete multilevel data include the R packages mitml, miceadds, and mdmb, and empirical work by [Audigier, White, Jolani, Debray, Quartagno,](#_bookmark7) [Carpenter, van Buuren, and Resche-Rigon](#_bookmark7) ([2018](#_bookmark7)) and [Grund, Lüdtke, and Robitzsch](#_bookmark13) ([2018](#_bookmark13)). Please note that this tutorial paper assumes a basic level of knowledge on multilevel models.2 Assumed knowledge also includes the use of the ‘piping operator’, %>%, adopted from the magrittr package, and the lme4 notation for multilevel models.3

We illustrate how to impute incomplete multilevel data by means of three case studies:

* + - popmis from the mice package (simulated data on perceived popularity, *n* = 2*,* 000 pupils across *N* = 100 schools);
    - hiv from the GJRM package (simulated data on HIV diagnoses, *n* = 6*,* 416 patients across *N* = 9 regions);
    - impact from the metamisc package (empirical data on traumatic brain injuries, *n* = 11*,* 022 patients across *N* = 15 studies).

For each of these datasets, we will discuss the nature of the missingness, choose one or more imputation models and evaluate the imputed data, but we will also highlight one specific

1Note that the alternative, joint modeling imputation for multilevel data or jomo [Quartagno, Grund, and](#_bookmark17) [Carpenter](#_bookmark17) ([2019](#_bookmark17)), has been implemented in mice as well but is outside the scope of this tutorial.

2Note to self: We’re providing an overview of implementations. It’s up-to the reader to decide which

multilevel strategy suits their data. We won’t go into detail for the different methods (and equations). This paper is just a software tutorial, so we’ll keep it practical.

3Add environment info, seed and version number(s) somewhere!

aspect of the imputation workflow. With the popmis data, we show how (not) to develop an imputation model. With the hiv data we focus on extending the imputation model to include Heckman-type selection-inclusion methods. And with the impact data we provide a example of multivariate missingness in real-world data. Together, this should give enough scaffolding for applied researchers who are faced with incomplete multilevel data.

REVIEW UNTIL THIS POINT SVP

# **How (not) to impute (Case Study I: Popularity)**

In this section we’ll go over the different steps involved with imputing incomplete multilevel data. The data we’re using is the popmis dataset from the mice package. This is a simulated dataset with pupils (*n* = 2000) clustered within schools (*N* = 100). In this tutorial we’ll use the following variables:

* + - school, school identification number (clustering variable);
    - popular, pupil popularity (self-rating between 0 and 10; unit-level);
    - sex, pupil sex (0=boy, 1=girl; unit-level);
    - texp, teacher experience (in years; cluster-level).

The analysis model corresponding to this dataset is multilevel regression with random inter- cepts, random slopes and a cross-level interaction. The outcome variable is popular, which is predicted from the unit-level variable sex and the cluster-level variable texp. The regression equation4 and lme4 notation for this model are

popular*ij* = *γ*00 + *γ*10 sex*ij* + *γ*01 texp*j* + *γ*11 texp*j* × sex*ij* + *u*0*j* + *u*1*j* sex*ij* + *eij*

popular ∼ 1 + sex + texp + sex:texp + (1 + sex | school)

Since the data is simulated and the missingness is induced, we can compare our inferences after imputation to the true complete data. The data is created in such a way that the clustering variable school explains quite some variance in the outcome variable popular. We express this using the intraclass correlation, ICC = 0.58. We’ll evaluate the ICC after each missing data strategy, and compare the estimated fixed effects:

Estimate with 95% CI

|  |  |  |
| --- | --- | --- |
| 1 | (Intercept) | 3.314 [ 2.998, 3.629] |
| 2 | sex | 1.330 [ 1.069, 1.590] |
| 3 | texp | 0.110 [ 0.090, 0.130] |
| 4 | sex:texp | -0.034 [-0.051, -0.017] |

4add the ‘level notation’ (Bryk and Raudenbush, 1992) and/or matrix notation (‘linear mixed effects model’; Laird and Ware, 1982) too?

## Missing data pattern

Total number of missing entries: 848

schol sex texp poplr

Number of missing entries per pattern

1152 0

Pattern frequency

848 1

0 0 0 848

Number of missing entries per variable

Note. schol = school, poplr = popular

Figure 2: Missing data pattern in the popularity data

#### Incomplete data

Load the data into the environment and select the relevant variables:

*R> popmis <- mice::popmis[, c("school", "popular", "sex", "texp")]*

The missingness is univariate and sporadic, which is illustrated in the missing data pattern in Figure [2](#_bookmark4). The ICC in the incomplete data is 0.56.

To develop the best imputation model for the incomplete variable popular, we need to know whether the missingness depends on the observed values of other variables. We’ll highlight only one other variable to illustrate, but ideally one would inspect all relations. The questions we’ll ask are: ‘Does the missing data of pupil popularity (popular) depend on observed teacher popularity (texp)?’. This can be evaluated statistically, but visual inspection usually suffices. We’ll make a histogram of texp separately for the pupils with known popularity and missing popularity.

In Figure [3](#_bookmark5) we see that **[update this part]** the distribution for the missing popular is further to the right than the distribution for observed popular. This would indicate a right-tailed MAR missingness. (In fact, this is exactly what happens, because the missingness in these data was created manually.) We’ve made it observable by examining the relations between the missingness in popular and the observed data in texp.

#### Complete case analysis (not recommended)

Complete case analysis ignores the observations with missingness altogether, which may even introduce bias in MCAR situations.

Cluster 1

Cluster 2

0.04

0.02

0.00

Cluster 4

density

0.04

Cluster 3

0.02

0.00

12.5 15.0 17.5 20.0 22.5 12.5 15.0 17.5 20.0 22.5

texp

popular observed missing

Figure 3: Conditional distributions in the popularity data

#### Imputation ignoring the cluster variable (not recommended)

The first imputation model that we’ll use is likely to be invalid. In this model, we ignore the multilevel structure of the data, despite the high ICCs. And assumes exchangeability between units. This is purely to illustrate the effects of ignoring the clustering in our imputation effort.

We’ll use predictive mean matching to impute the continuous variables and logistic regression to impute the binary variable sex. We do not use the observation identifier pupil or cluster identifier school as predictors to impute other variables.

*R> # dry run to get imputation parameters R> ini <- mice(popmis, maxit = 0)*

*R>*

*R> # extract predictor matrix and adjust R> pred <- ini$pred*

*R> pred[, c("school")] <- 0 R>*

*R> # impute the data, ignoring the cluster structure*

*R> imp\_ignored <- mice(popmis, maxit = 1, pred = pred, print = FALSE)*

8



6

popular

Imputed Observed



4



2

0 1 2 3 4 5

Imputation

As the original ICCs show, 100% of the variance in texp can be attributed to the clustering variable school. This tells us that the multilevel structure of the data should be taken into account. If we don’t, we’ll end up with incorrect imputations, biasing the effect of the clusters towards zero.

We can also observe that the teacher experience increases slightly after imputation. This is due to the MNAR missingness in texp. Higher values for texp have a larger probability to be missing. This may not a problem, however, if at least one pupil in each school has teacher experience recorded, we can deductively impute the correct (i.e. true) value for every pupil in the school.

#### Imputation with the cluster variable as predictor (not recommended)

We’ll now use school as a predictor to impute all other variables. This is still not recom- mended practice, since it only works under certain circumstances and results may be biased ([Drechsler 2015](#_bookmark9); [Enders, Mistler, and Keller 2016](#_bookmark10)). But at least, it includes some multilevel aspect. This method is also called ‘fixed cluster imputation’, and uses N-1 indicator variables representing allocation of N clusters as a fixed factor in the model ([Reiter, Raghunathan, and](#_bookmark18) [Kinney 2006](#_bookmark18); [Enders et al. 2016](#_bookmark10)). Colloquially, this is ‘multilevel imputation for dummies’.

Add: doesn’t work with syst missing (only sporadically). There’s some pro’s and con’s. May not differ much if the number of clusters is low.

The more the random effects are of interest, the more you need ml models.

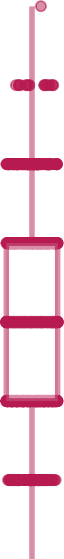
*R> # adjust the predictor matrix R> pred <- ini$pred*

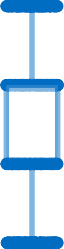
*R> # pred[, "pupil"] <- 0 R>*

*R> # impute the data, cluster as predictor*

*R> imp\_predictor <- mice(popmis, maxit = 1, pred = pred, print = FALSE)*

8



6

popular

Imputed Observed



4



2

0 1 2 3 4 5

Imputation

Now, we can clearly see that the imputed values of texp are higher than the observed values, which is in line with right-tailed MAR.

The ICCs are way more in line with the ICCs in the incomplete data. But this is a quick and dirty way of imputing multilevel data. We should be using a multilevel model.

#### Imputation with random effects

With 2l.norm we impute the outcome with a multilevel model assuming random slopes for each variable in the imputation model and homogeneous within-cluster variance.

“Van Buuren (2011) considered the homoscedastic linear mixed model as invalid for imputing incomplete predictors, and investigated only the 2l.norm method, which allows for heteroge- neous error variances” ([Van Buuren 2018](#_bookmark21)).

*R> pred <- ini$pred*

*R> pred["popular", ] <- c(-2, 2, 2, 2)*

*R> #-2 for the cluster variable, 2 for random effects R> meth <- ini$meth*

*R> meth <- c("", "2l.norm", "", "")*

*R> # meth <- c("", "2l.pmm", "", "") R> imp\_norm\_2l <-*

*+ mice(*

*+ popmis %>% mutate(school = as.integer(school)),*

*+ pred = pred,*

*+ meth = meth,*

*+ maxit = 1,*

*+ print = FALSE*

*+ )*

Warning: Removed 848 rows containing non-finite values (stat\_boxplot). Warning: Removed 848 rows containing missing values (geom\_point).

9



Imputed



popular

6

Observed

3

0 1 2 3 4 5

Imputation

#### Imputation with random effects and heterogeneity

This method assumes random slopes for each variable in the imputation model. In contrast to 2l.norm this method allows a cluster-specific residual error variance.

*R> pred["popular", ] <- c(-2, 2, 1, 2)*

*R> meth <- c("", "2l.pan", "", "") R> imp\_pan\_2l <-*

*+ mice(*

*+ popmis %>% mutate(school = as.integer(school)),*

*+ pred = pred,*

*+ meth = meth,*

*+ maxit = 1,*

*+ print = FALSE*

*+ )*

Warning: Removed 848 rows containing non-finite values (stat\_boxplot).

Warning: Removed 848 rows containing missing values (geom\_point).



7.5

5.0

popular



2.5

Imputed Observed

0 1 2 3 4 5



Imputation

# **How to handle non-random selection (Case study II: HIV)**

Data are simulated and included in the GJRM package. We will use the following variables:

* + - region Cluster variable,
    - hiv HIV diagnosis (0=no, 1=yes),
    - age Age of the patient,
    - marital Marital status,
    - condom Condom use during last intercourse,
    - smoke Smoker (levels; inclusion restriction variable).

The imputation of these date is based on the toy example from [IPDMA Heckman Github](https://github.com/johamunoz/Heckman-IPDMA/blob/main/Toy_example.R) [repo](https://github.com/johamunoz/Heckman-IPDMA/blob/main/Toy_example.R).

## Missing data pattern

Total number of missing entries: 1318

regin age martl condm smoke hiv

Number of missing entries per pattern

5098 0

Pattern frequency

1318 1

0 0 0 0 0 1318

Number of missing entries per variable

Note. regin = region, martl = marital, condm = condom

From the missing data pattern we see that we can set maxit to 1, since there is only one variable with missingness.

The inclusion restriction variable should be a predictor of the the actual value of the variable of interest, but not of missingness indicator for the variable of interest. In this case, the data were simulated to adhere to this requirement. Namely, *βsmoke* = -0.064, 95% CI [-0.256, 0.126] for the analysis model (formula = hiv ~ .), and *βsmoke* = -0.265, 95% CI [-0.422,

-0.11] for the selection model (formula = is.na(hiv) ~ .). This means the assumptions

for the Heckman-type selection model are met.

# **How to handle multivariate missingness (Case study III: IMPACT)**

impact is traumatic brain injury data with patients, *n* = 11022, clustered in studies, *N* = 15. With the following 11 variables:

* + - name Name of the study,
    - type Type of study (RCT: randomized controlled trial, OBS: observational cohort),
    - age Age of the patient,
    - motor\_score Glasgow Coma Scale motor score,
    - pupil Pupillary reactivity,
    - ct Marshall Computerized Tomography classification,
    - hypox Hypoxia (0=no, 1=yes),
    - hypots Hypotension (0=no, 1=yes),
    - tsah Traumatic subarachnoid hemorrhage (0=no, 1=yes),
    - edh Epidural hematoma (0=no, 1=yes),
    - mort 6-month mortality (0=alive, 1=dead).

The analysis model for this dataset is a prediction model with mort as the outcome.5 The data is already imputed (Steyerberg et al, 2008), so we’ve induced missingness again based on the missingness in the original data.

Missing data pattern

Total number of missing entries: 11870

3808

937

681

238

724

121

290

81

687

309

217

32

321

39

220

20

240

259

225

221

40

37

31

38

6

Pattern frequency

4

5

3

2

1

1

614

64

44

22

51

18

20

21

45

23

15

4

11

10

14

5

25

35

23

22

22

9

18

21

7

8

6

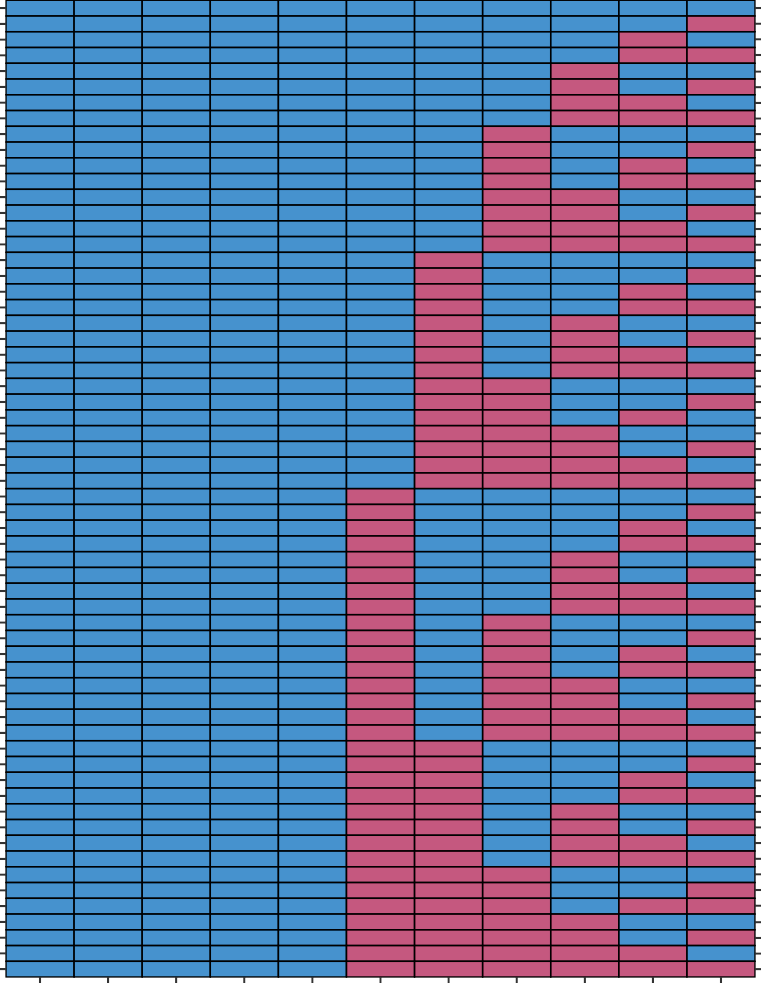
1

2

1

3

name type age mtr\_s mort pupil hypts tsah edh hypox ct

0

1

1

2

1

2

2

3

1

2

2

3

2

3

3

4

1

2

2

Number of missing entries per pattern

3

2

3

3

4

2

3

3

3

4

4

5

1

2

2

3

2

3

3

4

2

3

3

4

3

4

4

5

2

3

3

4

3

4

4

5

3

4

5

4

5

5

6

0 0 0 0 0 1184 1316 2022 2196 2540 2612

Number of missing entries per variable

Note. mtr\_s = motor\_score, hypts = hypots



5Look at analysis model, maybe copy from GREAT data example e.g., adjusted prognostic effect of ct on unfortunate outcomes. Add something about systematically missing data here.

## Missing data pattern

Total number of missing entries: 215

916

98

62

Pattern frequency

5

14

2

1

2

9

5

3

1

nametype age mtr\_smort ct hypts edh tsah pupilhypox

0

Number of missing entries per pattern

1

1

2

1

2

2

3

1

1

1

2

0 0 0 0 0 4 5 9 19 70 108

Number of missing entries per variable

Note. mtr\_s = motor\_score, hypts = hypots

## Missing data pattern

Total number of missing entries: 414

362

363

Pattern frequency

7

11

3

3

1

5

1

nametype age mtr\_s edh morthypoxpupil tsahhypts ct

0

Number of missing entries per pattern

1

1

2

1

2

1

2

2

0 0 0 0 0 0 1 6 6 18 383

Number of missing entries per variable

Note. mtr\_s = motor\_score, hypts = hypots

# **Discussion**

* + - JOMO in mice -> on the side for now
    - Additional levels of clustering
    - More complex data types: timeseries and polynomial relationship in the clustering.

# **Think about**

* + - Adding some kind of help function to mice that suggests a suitable predictor matrix to the user, given a certain analysis model.
    - Adding a multilevel\_ampute() wrapper function in mice.
    - Exporting mids objects to other packages like lme4 or coxme?
    - Adding a ICC=0 dataset to show that even if there is no clustering it doesn’t hurt.
    - env dump in repo

# **References**

Audigier V, White IR, Jolani S, Debray TPA, Quartagno M, Carpenter J, van Buuren S, Resche-Rigon M (2018). “Multiple Imputation for Multilevel Data with Continuous and Binary Variables.” Statistical Science, **33**(2), 160–183. ISSN 0883-4237, 2168-8745. [doi:](https://doi.org/10.1214/18-STS646)

[10.1214/18-STS646](https://doi.org/10.1214/18-STS646). 1702.00971.

de Jong VMT, Moons KGM, Eijkemans MJC, Riley RD, Debray TPA (2021). “Developing More Generalizable Prediction Models from Pooled Studies and Large Clustered Data Sets.” Statistics in Medicine, **40**(15), 3533–3559. ISSN 1097-0258. [doi:10.1002/sim.8981](https://doi.org/10.1002/sim.8981).

Drechsler J (2015). “Multiple Imputation of Multilevel Missing Data—Rigor Versus Sim- plicity.” Journal of Educational and Behavioral Statistics, **40**(1), 69–95. ISSN 1076-9986. [doi:10.3102/1076998614563393](https://doi.org/10.3102/1076998614563393).

Enders CK, Mistler SA, Keller BT (2016). “Multilevel Multiple Imputation: A Review and Evaluation of Joint Modeling and Chained Equations Imputation.” Psychological Methods, **21**(2), 222–240. ISSN 1939-1463. [doi:10.1037/met0000063](https://doi.org/10.1037/met0000063).

Fisher RA (1925). “Statistical Methods for Research Workers.” In S Kotz, NL John- son (eds.), Breakthroughs in Statistics: Methodology and Distribution, Springer Se- ries in Statistics, pp. 66–70. Springer, New York, NY. ISBN 978-1-4612-4380-9. [doi:](https://doi.org/10.1007/978-1-4612-4380-9_6) [10.1007/978-1-4612-4380-9\_6](https://doi.org/10.1007/978-1-4612-4380-9_6).

Gelman A, Hill J (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press. ISBN 978-1-139-46093-4.

Grund S, Lüdtke O, Robitzsch A (2018). “Multiple Imputation of Missing Data for Multilevel Models: Simulations and Recommendations.” Organizational Research Methods, **21**(1), 111–149. ISSN 1094-4281. [doi:10.1177/1094428117703686](https://doi.org/10.1177/1094428117703686).

Hox JJ, Moerbeek M, van de Schoot R (2017). Multilevel Analysis: Techniques and Applications, Third Edition. Routledge. ISBN 978-1-317-30868-3.

Jolani S (2018). “Hierarchical Imputation of Systematically and Sporadically Missing Data: An Approximate Bayesian Approach Using Chained Equations.” Biometrical Journal. Biometrische Zeitschrift, **60**(2), 333–351. ISSN 1521-4036. [doi:10.1002/bimj.201600220](https://doi.org/10.1002/bimj.201600220).

Meng XL (1994). “Multiple-Imputation Inferences with Uncongenial Sources of Input.” Statistical Science, **9**(4), 538–558. ISSN 0883-4237, 2168-8745. [doi:10.1214/ss/](https://doi.org/10.1214/ss/1177010269)

[1177010269](https://doi.org/10.1214/ss/1177010269).

Quartagno M, Grund S, Carpenter J (2019). “Jomo: A Flexible Package for Two-level Joint Modelling Multiple Imputation.” The R Journal, **11**(2), 205–228. ISSN 2073-4859.

Reiter JP, Raghunathan T, Kinney SK (2006). “The Importance of Modeling the Sampling Design in Multiple Imputation for Missing Data.” undefined.

Resche-Rigon M, White IR, Bartlett JW, Peters SAE, Thompson SG (2013). “Multiple Im- putation for Handling Systematically Missing Confounders in Meta-Analysis of Individual Participant Data.” Statistics in medicine, **32**(28), 4890–4905. ISSN 1097-0258 0277-6715. [doi:10.1002/sim.5894](https://doi.org/10.1002/sim.5894).

Rubin DB (1976). “Inference and Missing Data.” Biometrika, **63**(3), 581–592. [doi:10.2307/](https://doi.org/10.2307/2335739) [2335739](https://doi.org/10.2307/2335739).

Van Buuren S (2018). Flexible Imputation of Missing Data. Chapman and Hall/CRC.

### **Affiliation:**

Hanne Oberman Utrecht University Padualaan 14

3584 CH Utrecht

E-mail: [h.i.oberman@uu.nl](mailto:h.i.oberman@uu.nl)

URL: <https://hanneoberman.github.io/>

*Journal of Statistical Software* <http://www.jstatsoft.org/>

published by the Foundation for Open Access Statistics <http://www.foastat.org/>

MMMMMM YYYY, Volume VV, Issue II *Submitted:* yyyy-mm-dd

[doi:10.18637/jss.v000.i00](https://doi.org/10.18637/jss.v000.i00) *Accepted:* yyyy-mm-dd