Master Thesis: Multilevel Multivariate Imputation by Chained Equations through Bayesian Additive Regression Trees

Methodology and Statistics for the Behavioural, Biomedical and Social Sciences

Heleen Brüggen

Utrecht University

Word count:	2496
Candidate Journal:	Computational Statistics & Data Analysis
FETC Case Number:	23-1778
Supervisors:	
MSc. T. Volker	Utrecht University
Dr. G. Vink	Utrecht University
MSc. H. Oberman	Utrecht University

Contents

1 Introduction

Incomplete data is a common challenge in many fields of research. Frequently used ad hoc strategies to deal with missing data, such as complete case analysis or mean imputation often lead to erroneous inferences in realistic situations. These strategies don't consider the multivariate nature of the data. Missingness can depend on observed data or even unobserved data, leading to biased estimates and inaccurate variance estimates when using one of these ad hoc strategies (Austin et al., 2021; Enders, 2017; Kang, 2013; Little and Rubin, 2002; van Buuren, 2018). Multiple imputation (MI; Rubin 1987) is considered an effective method for dealing with incomplete data supported by much methodological research (Audigier et al., 2018; Austin et al., 2021; Burgette and Reiter, 2010; Enders, 2017; Grund et al., 2021; Hughes et al., 2014; Little and Rubin, 2002; Mistler and Enders, 2017; Van Buuren, 2007; van Buuren, 2018).

MI allows us to separate the missing data problem from the analysis problem (Audigier et al., 2018; Austin et al., 2021; Burgette and Reiter, 2010; Enders, 2017; Grund et al., 2021; Hughes et al., 2014; Little and Rubin, 2002; Mistler and Enders, 2017; Van Buuren, 2007; van Buuren, 2018). MI is used to impute each missing value in the dataset more than once given the observed data, considering necessary variation associated with the missingness problem. The multiply imputed datasets are analyzed, and the corresponding inferences are pooled according to Rubin's rules (Austin et al., 2021; Carpenter and Kenward, 2013; Rubin, 1987; van Buuren, 2018). However, specifying the imputation models, the models used to impute the missing data, can be challenging. The concept of congeniality dictates that the imputation models should be at least as general as the analysis model and preferably all-encompassing (Bartlett et al., 2015; Enders et al., 2018a; Grund et al., 2016, 2018b; Little and Rubin, 2002; Meng, 1994). Otherwise, it will not capture every aspect of the data and the analysis model estimates may be biased. So, when the complexity of data increases, specifying the imputation models becomes more difficult (Grund et al., 2018b; van Buuren, 2018).

Congeniality-issues become more pronounced when MI is used in a multilevel data context (Audigier et al., 2018; Dong and Mitani, 2023; Enders et al., 2020, 2018a,b, 2016; Grund et al., 2016, 2018a,b, 2021; Lüdtke et al., 2017; Mistler and Enders, 2017; Quartagno and Carpenter, 2022; Resche-Rigon and White, 2018; Taljaard et al., 2008; van Buuren, 2018). Multilevel data is hierarchically structured, where, for example, students are nested within schools (Hox and Roberts, 2011; Hox et al., 2017). When analyzing multilevel data, the hierarchical structure should be considered. Ignoring the hierarchical structure will underestimate the intra-class correlation (ICC; Hox and Roberts 2011; Lüdtke et al. 2017; Taljaard et al. 2008; van Buuren 2018), which can be interpreted as the proportion of the total variance at level-2 (Gulliford et al., 2005; Hox and Roberts, 2011; Shieh, 2012). This can be done using multilevel models (MLM; Hox and Roberts 2011; Hox et al. 2017; Lüdtke et al. 2017). MLMs can contain both level-1, and level-2 variables, relating to the individual and class respectively, random intercepts, random slopes, and cross-level interactions (Hox and Roberts, 2011; Hox et al., 2017). Typically, the complexity of the multilevel analysis model is built step-wise with non-linearities, meaning the analysis model is not determined beforehand (Hox and Roberts, 2011; Hox et al., 2017). Thus, including the hierarchical structure, along with the complicated non-linearities from cross-level interactions in imputation models can be quite challenging (Burgette and Reiter, 2010; Hox and Roberts, 2011; van Buuren, 2018) and a very complex model might not converge (van Buuren, 2018).

A popular and flexible implementation of MI is fully conditional specification (FCS), otherwise known as chained equations(Audigier et al., 2018; Burgette and Reiter, 2010; Grund et al., 2018a; Van Buuren, 2007). FCS iteratively imputes each incomplete variable conditional on complete and previously imputed variables (Enders et al., 2018a,b, 2016; Grund et al., 2018a; Hughes et al., 2014; Mistler and Enders, 2017; van Buuren, 2018). In a multilevel context, FCS employs univariate linear mixed models to account for the hierarchical structure (Enders et al., 2018a; Mistler and Enders, 2017; Resche-Rigon and White, 2018). Furthermore, FCS can be used to impute non-linearities, such as cross-level interactions, by using 'passive imputation' or defining a separate imputation model for the non-linearities (Grund et al., 2018b; van Buuren, 2018). Still, imputation models including cross-level interaction or non-linear terms in FCS are very complicated (Grund et al., 2018b, 2021) and, thus, researchers' focus has predominantly been on the inclusion of random intercepts and slopes, but not of cross-level interactions (Enders et al., 2020, 2018a,b, 2016; Grund et al., 2016, 2018a).

Using non-parametric tree-based models might solve this problem because these models do not assume a specific data distribution and, thus, implicitly model non-linear relationships and interactions between the predictor variables, and handle continuous and categorical variables simultaneously (Breiman et al., 1984; Burgette and Reiter, 2010; Chipman et al., 2010; Hill et al., 2020; James et al., 2021; Lin and Luo, 2019; Salditt et al., 2023). In a single-level imputation context, the use of tree-based, non-parametric models like regression trees, random forests, or Bayesian Additive Regression Trees (BART) simplified imputation models and performed better than parametric methods: the imputations showed better confidence interval coverage of the population parameters, lower variance and lower bias, especially in non-linear and interactive contexts (Burgette and Reiter, 2010; Silva and Gutman, 2022; Xu et al., 2016). Waljee et al. (2013) also found lower imputation error when imputing with a random forest algorithm compared to multivariate imputation by chained equations (`MICE`), K-nearest neighbors (KNN) and mean imputation.

BART models have been implemented in a multilevel prediction context. However, multilevel-BART models (M-BART) have predominantly been implemented with only random intercepts (Chen, 2020; Tan et al., 2016; Wagner et al., 2020; Wundervald et al., 2022). In a prediction context, Wagner et al. (2020) have found that this random intercept M-BART model provided better predictions with a lower mean squared error (MSE) compared to a parametric MLM, Tan et al. (2016) found higher area under the curve (AUC) values compared to a singel-level BART model and linear random intercept model, and Chen (2020) found better predictions and better coverage of the estimates compared to parametric models and a single-level BART model. Other researchers modeled the random intercept as an extra split on each terminal node and found a lower MSE compared to a standard BART model and parametric MLMs (Wundervald et al., 2022). Dorie et al. (2022) developed a multilevel BART model that included random intercepts and random slopes by modeling the random parts with Stan (Lee et al., 2017) and the fixed parts with BART. Their results showed that their algorithm `stan4bart` showed better coverage of the population values and lower root mean squared error (RMSE) compared to BART models with varying intercept, BART models ignoring the multilevel structure, bayesian causal forests, and parametric MLMs.

Despite these promising findings, M-BART models have yet to be implemented in a multilevel multiple imputation context. Thus, my thesis research question will be: *Can multivariate imputation by chained equations through a multilevel bayesian additive regression trees model improve the bias, variance, and coverage of the estimates in a multilevel context compared to current practices?* Given the success of non-parametric models in single-level MI, I anticipate that employing M-BART models in a multilevel missing data context will reduce bias, accurately model variance, and improve estimate coverage compared to conventional implementations of multilevel MI, single-level MI, and complete case analysis in the R-package `MICE` (Buuren and Groothuis-Oudshoorn, 2011).

The research report's sections will cover theoretical background, methods for evaluating M-BART models, preliminary results, and discussion of next steps.

2 Method

2.1 Theoretical background

2.1.1 Bayesian Additive Regression Trees (BART)

BART is a sum-of-trees model proposed by Chipman et al. (2010) that has regression trees as its building blocks (Chipman et al., 2010; Hill et al., 2020; James et al., 2021). Regression trees divide the data into subgroups by recursively splitting the data into binary subgroups based on the predictors minimizing variability within the subgroups (Hastie, 2017; James et al., 2021; Salditt et al., 2023). Recursive binary partitioning of the predictor space doesn't assume a specific data form, making this a non-parametric model (Hastie, 2017; James et al., 2021; Salditt et al., 2023) and allows regression trees to model non-linearities well and automatically (Burgette and Reiter, 2010; Hill et al., 2020). Chipman et al. (2010) define the BART model as:

$$f(\mathbf{x}) = \sum_{k=1}^{m} g(\mathbf{x}; T_k, M_k), \tag{1}$$

where $f(\mathbf{x})$ is the overall fit of the model: the sum of m regression trees, $\mathbf{x}$ are the predictor variables, T_k is the k$^{\text{th}}$ tree and M_k is the collection of leaf parameters within the k$^{\text{th}}$ tree (Chipman et al., 2010; Hill et al., 2020; James et al., 2021). The data are assumed to arise from a model with additive normally distributed errors: $Y = \sum_{k=1}^{m} g(\mathbf{x}; T_k, M_k) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2)$. Next to the sum-of-trees model, BART also includes a regularization prior that constrains the size and fit of each tree so that each contributes only a small part to prevent overfitting (Chipman et al., 2010; Hill et al., 2020; James et al., 2021). BARTs are estimated using the Bayesian back-fitting Markov Chain Monte Carlo (MCMC) algorithm. It updates individual trees, considering the remaining trees, their associated parameters, and the residual standard deviation (σ). It fits a new tree to the partial residuals, r_i, treating them as the data, by perturbing the tree from the previous iteration. Perturbations entail either *growing*, *pruning*, or *changing* a tree. *Growing* means adding additional splits, *pruning* removes splits, and *changing* changes decision rules. The algorithm stops after the specified number of iterations. The partial residuals are defined as:

$$r_i = y_i - \sum_{k'<k} \hat{f}_{k'}^{b}(x_i) - \sum_{k'>k} \hat{f}_{k'}^{b-1}(x_i), \text{with } i = 1, \dots, N \tag{2}$$

where $\hat{f}_k^b(x_i)$ is the prediction of the k^{th} tree in the b^{th} iteration for person i and sample size N.

2.1.2 Multilevel-BART (M-BART)

Chen (2020); Wagner et al. (2020) and Tan et al. (2016) define a M-BART model including a random intercept building on the work of Lin and Luo (2019). The M-BART algorithm breaks down the observed variable into fixed and random components. The fixed components are modeled by BART and the random components are modeled by a linear mixed effects model (Chen, 2020; Tan et al., 2016; Wagner et al., 2020). The BART model (1) can be extended to include a random intercept by:

$$f(\mathbf{x}) = \sum_{k=1}^{m} g(\mathbf{x}; T_k, M_k) + \alpha_j, \tag{3}$$

where, now, $f(\mathbf{x})$ is the overall fit of the model incorporating random intercept α_j for cluster j.

2.2 Simulation study

2.2.1 Data generating mechanism

The population data-generating mechanism was based on the following MLM:

$$y_{ij} = \beta_{0j} + \sum_{k=1}^{7} \beta_{kj} X_{kij} + \epsilon_{ij}, \qquad X_{kij} \sim \mathcal{MVN}(0, \mathbf{\Sigma}_x), \;\; \epsilon_{ij} \sim \mathcal{N}(0, 25), \tag{4a}$$

$$\beta_{0j} = \gamma_{00} + \sum_{q=1}^{2} \gamma_{0q} Z_{qj} + \upsilon_{0j}, \tag{4b}$$

$$\beta_{kj} = \gamma_{k0} + \sum_{q=1}^{2} \gamma_{kq} Z_{qj} + \upsilon_{kj}, \qquad Z_{qj} \sim \mathcal{MVN}(0, \mathbf{\Sigma}_z), \;\; \boldsymbol{\upsilon}_j \sim \mathcal{MVN}(0, \mathbf{T}), \tag{4c}$$

$$\gamma_{00} = 10, \boldsymbol{\gamma}_{0q} = \begin{pmatrix} .5 \\ .5 \end{pmatrix}, \boldsymbol{\gamma}_{k0} = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \gamma_{50} \\ \gamma_{60} \\ \gamma_{70} \end{pmatrix}, \boldsymbol{\gamma}_{kq} = \begin{pmatrix} .35 & 0 \\ .35 & 0 \\ 0 & .35 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{4d}$$

$$\mathbf{\Sigma}_x = \begin{pmatrix} 6.25 & & & & & & \\ 2.25 & 9 & & & & & \\ 1.5 & 1.8 & 4 & & & & \\ 2.25 & 3.06 & 2.04 & 11.56 & & & \\ 1.5 & 1.8 & 1.2 & 2.04 & 4 & & \\ 1.125 & 1.35 & 0.9 & 1.53 & .9 & 2.25 & \\ 3.3 & 3.96 & 2.64 & 4.488 & 2.64 & 1.98 & 19.36 \end{pmatrix}, \mathbf{\Sigma}_z = \begin{pmatrix} 1 & \\ .48 & 2.56 \end{pmatrix}, \tag{4e}$$

$$\mathbf{T} = \begin{pmatrix} t_{00} & & & & & & & \\ .3 & 1 & & & & & & \\ .3 & .3 & 1 & & & & & \\ .3 & .3 & .3 & 1 & & & & \\ 0 & 0 & 0 & 0 & 0 & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{4f}$$

where y_{ij} is a continuous level-1 outcome variable for person i in group j and X_{kij} are k continuous level-1 variables and Z_{qj} are q continuous level-2 variables. They are multivariate normally distributed with means of 0 and variance-covariance matrix $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_z$ respectively. Equation 4e shows the variance-covriance matrices of the level-1 and level-2 variables. The covariances between the variables were calculated as such that the correlation between the variables was .3, aligned with Cohen's (1990) medium effect size benchmark. The residuals are normally distributed with a mean of 0 and a variance of 25. The random intercept β_{0j} is determined by the overall intercept γ_{00}, the q group-level effects $\gamma_{0q} Z_{qj}$ and the group-level random residuals υ_{0j}. The overall intercept γ_{00} was set to 10 and the group-level effects γ_{01} and γ_{02} to .5. The k regression coefficients β_{kj} for the continuous variables X_{kij} depend on the intercepts γ_{k0}, the cross-level interactions $\gamma_{kq} Z_{qj}$, and the random slopes υ_{kj}. The k intercepts, or within-group effect sizes, γ_{kj} were varied in the simulations, the cross-level interactions γ_{11}, γ_{21}, and γ_{32} were set to .35. The random slopes are multivariate normally distributed with a mean of 0 and a variance-covariance matrix $\mathbf{T}$ shown in equation 4f. Again, the covariances were calculated to yield a correlation of .3. The variance of υ_{0j}, the group-level random residuals t_{00}, were scaled such that the specified ICC values as in table 1 was obtained. The following formula was used to calculate υ_{0j} following the variance decomposition from Rights and Sterba (2019):

$$\texttt{ICC} = \frac{\boldsymbol{\gamma}^{b\prime} \boldsymbol{\phi}^b \boldsymbol{\gamma}^b + \tau_{00}}{\boldsymbol{\gamma}^{w\prime} \boldsymbol{\phi}^w \boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime} \boldsymbol{\phi}^b \boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}, \tag{5}$$

where γ^b and γ^w are the level-1 and level-2 fixed effects, ϕ^b and ϕ^w are the variance-covariance matrices of a vector with 1, for the intercept, and all level-2 predictors and all cluster-mean-centered level-1 predictors respectively, τ_{00} is the variance of the random intercept, $\mathbf{T}$ is the variance-covariance matrix of the random intercept and slopes, $\mathbf{\Sigma}$ is the variance-covariance matrix of a vector containing 1, for the intercept, and the level-1 variables, and σ^2 is the residual variance. The value for τ_{00} was calculated using the function `uniroot` in R (R Core Team, 2023).

2.2.2 Simulation design

Table 1 shows the variations considered in the simulation study. They are realistic in practice and/or previously proposed (Enders et al., 2020, 2018b; Grund et al., 2018b; Gulliford et al., 1999; Hox et al., 2017; Murray and Blitstein, 2003). For each combination of varying parameters, 1000 datasets were simulated. 6 different imputation methods were compared:

1. complete case analysis,
2. conventional single-level imputation,
3. conventional multilevel imputation,
4. single-level BART imputation,
5. multilevel BART imputation accounting for random intercepts (Chen, 2020; Tan et al., 2016; Wagner et al., 2020; Wundervald et al., 2022),
6. multilevel BART imputation accounting for random effects and cross-level interactions (Dorie et al., 2022).

They were compared on the pooled estimates after fitting the analysis model in equations 4a, 4b, and 4c to the imputed datasets. The analysis models were fitted using the R-package `lme4` (Bates et al., 2015) and the estimates were pooled together using the R-package `mice` (Buuren and Groothuis-Oudshoorn, 2011).

The second and third methods were implemented with the R-package `mice`. The conventional single-level imputation were implemented with the imputation method `pmm` (predictive mean matching) from the `mice`-package and the conventional multilevel imputation was supplemented with the imputation method `pmm.2lonly` from the `mice`-package for the level-2 variables.

Table 1: Simulation design

Parameter	Values
Number of clusters (j)	30, 50
Within-cluster sample size (n_j)	15, 35, 50
Intraclass Correlation (ICC)	0, .3
Missing data mechanism	MAR, MCAR
Amount of missingness	0%, 25%, 50%
Within-group effect size (γ)	.2, .5

The fourth, single-level BART, fifth, random intercept BART and sixth method, multilevel BART methods were implemented by writing functions in R (R Core Team, 2023) for the package `MICE`. The functions `bart` and `rbart_vi` from the `dbarts` package were used for the single-level and random intercept BART imputation methods. The function `stan4bart` from the package `stan4bart` was used for the multilevel BART imputation method accounting for random effects and cross-level interactions. The functions were written such that they can be used as imputation methods in the `mice` package.

2.2.3 Missing data generation

As can be seen in table 1, the missing data mechanism was either Missing At Random (MAR) or Missing Completely At Random (MCAR) and either 0%, 25%, or 50% of the data was missing. The missing data was generated using the function `ampute` from the package `mice`.

For both the MCAR and MAR mechanism, there were patterns of missingness defined with missing values for 1 to 5 missing values per case. These patterns had the same relative frequency of occurence in the data sets. For the MAR mechanism, the weighted sum of scores on the observed variables was used to predict the probability of missingness for a case. The weights of the variables *x4* and *z1* were set to 2 and 1.5 respectively when they remained observed in a specific pattern, while the weights of the other variables that remained observed in a specific pattern are set to 1. The type of missingess was set to *"RIGHT"* meaning that cases with a higher weighted sum of scores had a higher probability of becoming incomplete. So, this means that cases with higher values on *x4* and *z1* were more likely to become incomplete.

2.2.4 Evaluation

The estimated from the analysis models were evaluated in terms of relative bias and Mean Squared Error (MSE) and coverage of 95% confidence intervals (Morris et al., 2019):

$$\texttt{Bias} = \frac{1}{n_{\text{sim}}} \sum_{t=1}^{n_{\text{sim}}} (\hat{\theta}_t - \theta), \tag{6a}$$

$$\texttt{MSE} = \frac{1}{n_{\text{sim}}} \sum_{t=1}^{n_{\text{sim}}} (\hat{\theta}_t - \theta)^2, \tag{6b}$$

$$\texttt{Coverage} = \Pr(\hat{\theta}_{\text{low,i}} \leq \theta \leq \hat{\theta}_{\text{upp,i}}) = \frac{1}{n_{\text{sim}}} \sum_{t=1}^{n_{\text{sim}}} 1(\hat{\theta}_{\text{low,i}} \leq \theta \leq \hat{\theta}_{\text{upp,i}}), \tag{6c}$$

where $\hat{\theta}_t$ is the estimated parameter in simulation t, θ is the true value, and n_{sim} is the number of simulated datasets. The lower and upper bounds of the 95% confidence intervals are denoted as $\hat{\theta}_{\text{low,i}}$ and $\hat{\theta}_{\text{upp,i}}$ respectively. The coverage is the proportion of the 95% confidence intervals that contain the true value.

3 Results

4 Discussion

5 Conclusion

6 Appendix

Listing 1: Imputation function for single-level BART

```
mice.impute.bart <- function(y, ry, x, wy = NULL, use.matcher = FALSE, donors = 5L, ...) {
    install.on.demand("dbarts", ...)
    if (is.null(wy)) {
        wy <- !ry
    }

    # Parameter estimates
    fit <- dbarts::bart(x, y, keeptrees = TRUE, verbose = FALSE)

    yhatobs <- fitted(fit, type = "ev", sample = "train")[ry]
    yhatmis <- fitted(fit, type = "ev", sample = "train")[wy]

    # Find donors
    if (use.matcher) {
        idx <- matcher(yhatobs, yhatmis, k = donors)
    } else {
        idx <- matchindex(yhatobs, yhatmis, donors)
    }

    return(y[ry][idx])
}
```

Listing 2: Imputation function for random intercept BART

```
    mice.impute.2l.rbart <- function(y, ry, x, wy = NULL, type, use.matcher = FALSE, donors = 5L, ...) {
    install.on.demand("dbarts", ...)
    if (is.null(wy)) {
        wy <- !ry
    }

    clust <- names(type[type == -2])
    effects <- names(type[type != -2])
    X <- x[, effects, drop = FALSE]

    model <- paste0(
        "y ~ ", paste0(colnames(X), collapse = " + ")
    )

    fit <- dbarts::rbart_vi(formula = formula(model), group.by = clust, data = data.frame(y, x), verbose = FALSE, ...)

    yhatobs <- fitted(fit, type = "ev", sample = "train")[ry]
    yhatmis <- fitted(fit, type = "ev", sample = "train")[wy]

    # Find donors
    if (use.matcher) {
        idx <- matcher(yhatobs, yhatmis, k = donors)
    } else {
        idx <- matchindex(yhatobs, yhatmis, donors)
    }

    return(y[ry][idx])
}
```

Listing 3: Imputation function for multilevel BART with random effects and cross-level interactions

```
    mice.impute.2l.bart <- function(y, ry, x, wy = NULL, type, intercept = TRUE, use.matcher = FALSE, donors = 5L, ...) {
    install.on.demand("stan4bart", ...)
    if (is.null(wy)) {
        wy <- !ry
    }

    if (intercept) {
        x <- cbind(1, as.matrix(x))
        type <- c(2, type)
```

```r
        names(type)[1] <- colnames(x)[1] <- "(Intercept)"
    }

    clust <- names(type[type == -2])
    rande <- names(type[type == 2])
    fixe <- names(type[type > 0])

    lev <- unique(x[, clust])

    X <- x[, fixe, drop = FALSE]
    Z <- x[, rande, drop = FALSE]
    xobs <- x[ry, , drop = FALSE]
    yobs <- y[ry]
    Xobs <- X[ry, , drop = FALSE]
    Zobs <- Z[ry, , drop = FALSE]

    # create formula
    fr <- ifelse(length(rande) > 1,
        paste0("+ (1 +", paste(rande[-1L], collapse = "+")),
        " + (1 "
    )
    randmodel <- paste0(
        "y ~ bart(", paste0(fixe[-1L], collapse = " + "), ")",
        fr, "| ", clust, ")"
    )

    fit <- eval(parse(text = paste("stan4bart::stan4bart(", randmodel,
    ", data = data.frame(y, x),
        verbose = -1,
        ...
    )", collapse = "")))

    yhatobs <- fitted(fit, type = "ev", sample = "train")[ry]
    yhatmis <- fitted(fit, type = "ev", sample = "train")[wy]

    # Find donors
    if (use.matcher) {
        idx <- matcher(yhatobs, yhatmis, k = donors)
    } else {
        idx <- matchindex(yhatobs, yhatmis, donors)
    }

    return(y[ry][idx])
}
```

References

Audigier, V., White, I. R., Jolani, S., Debray, T. P. A., Quartagno, M., Carpenter, J., Van Buuren, S., and Resche-Rigon, M. (2018). Multiple Imputation for Multilevel Data with Continuous and Binary Variables. *Statistical Science*, 33(2).

Austin, P. C., White, I. R., Lee, D. S., and Van Buuren, S. (2021). Missing Data in Clinical Research: A Tutorial on Multiple Imputation. *Canadian Journal of Cardiology*, 37(9):1322–1331.

Bartlett, J. W., Seaman, S. R., White, I. R., Carpenter, J. R., and for the Alzheimer's Disease Neuroimaging Initiative* (2015). Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical Methods in Medical Research*, 24(4):462–487.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, 67(1).

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification And Regression Trees*. Routledge, 1 edition.

Burgette, L. F. and Reiter, J. P. (2010). Multiple Imputation for Missing Data via Sequential Regression Trees. *American Journal of Epidemiology*, 172(9):1070–1076.

Buuren, S. V. and Groothuis-Oudshoorn, K. (2011). **Mice** : Multivariate Imputation by Chained Equations in *R*. *Journal of Statistical Software*, 45(3).

Carpenter, J. R. and Kenward, M. G. (2013). *Multiple Imputation and Its Application*. Wiley, 1 edition.

Chen, S. (2020). *A New Multilevel Bayesian Nonparametric Algorithm and Its Application in Causal Inference*. PhD thesis, Texas A&M University.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1).

Cohen, J. (1990). Statistical power analysis for the behavioral sciences. *Computers, Environment and Urban Systems*, 14(1):71.

Dong, M. and Mitani, A. (2023). Multiple imputation methods for missing multilevel ordinal outcomes. *BMC Medical Research Methodology*, 23(1):112.

Dorie, V., Perrett, G., Hill, J. L., and Goodrich, B. (2022). Stan and BART for Causal Inference: Estimating Heterogeneous Treatment Effects Using the Power of Stan and the Flexibility of Machine Learning. *Entropy*, 24(12):1782.

Enders, C. K. (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. *Behaviour Research and Therapy*, 98:4–18.

Enders, C. K., Du, H., and Keller, B. T. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms. *Psychological Methods*, 25(1):88–112.

Enders, C. K., Hayes, T., and Du, H. (2018a). A Comparison of Multilevel Imputation Schemes for Random Coefficient Models: Fully Conditional Specification and Joint Model Imputation with Random Covariance Matrices. *Multivariate Behavioral Research*, 53(5):695–713.

Enders, C. K., Keller, B. T., and Levy, R. (2018b). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods*, 23(2):298–317.

Enders, C. K., Mistler, S. A., and Keller, B. T. (2016). Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychological Methods*, 21(2):222–240.

Grund, S., Lüdtke, O., and Robitzsch, A. (2016). Multiple imputation of missing covariate values in multilevel models with random slopes: A cautionary note. *Behavior Research Methods*, 48(2):640–649.

Grund, S., Lüdtke, O., and Robitzsch, A. (2018a). Multiple Imputation of Missing Data at Level 2: A Comparison of Fully Conditional and Joint Modeling in Multilevel Designs. *Journal of Educational and Behavioral Statistics*, 43(3):316–353.

Grund, S., Lüdtke, O., and Robitzsch, A. (2018b). Multiple Imputation of Missing Data for Multilevel Models: Simulations and Recommendations. *Organizational Research Methods*, 21(1):111–149.

Grund, S., Lüdtke, O., and Robitzsch, A. (2021). Multiple imputation of missing data in multilevel models with the R package mdmb: A flexible sequential modeling approach. *Behavior Research Methods*, 53(6):2631–2649.

Gulliford, M., Adams, G., Ukoumunne, O., Latinovic, R., Chinn, S., and Campbell, M. (2005). Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *Journal of Clinical Epidemiology*, 58(3):246–251.

Gulliford, M. C., Ukoumunne, O. C., and Chinn, S. (1999). Components of Variance and Intraclass Correlations for the Design of Community-based Surveys and Intervention Studies: Data from the Health Survey for England 1994. *American Journal of Epidemiology*, 149(9):876–883.

Hastie, T. J., editor (2017). *Statistical Models in S*. Routledge, 1st edition.

Hill, J., Linero, A., and Murray, J. (2020). Bayesian Additive Regression Trees: A Review and Look Forward. *Annual Review of Statistics and Its Application*, 7(1):251–278.

Hox, J. and Roberts, J. K., editors (2011). *Handbook of Advanced Multilevel Analysis*. Routledge, 0 edition.

Hox, J. J., Moerbeek, M., and Van De Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications*. Routledge, Third edition. | New York, NY : Routledge, 2017. |, 3 edition.

Hughes, R. A., White, I. R., Seaman, S. R., Carpenter, J. R., Tilling, K., and Sterne, J. A. (2014). Joint modelling rationale for chained equations. *BMC Medical Research Methodology*, 14(1):28.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer US, New York, NY.

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, 64(5):402.

Lee, D., Carpenter, B., Li, P., Morris, M., Betancourt, M., Maverickg, Brubaker, M., Trangucci, R., Inacio, M., Kucukelbir, A., Buildbot, S., Bgoodri, Seantalts, Arnold, J., Tran, D., Hoffman, M., Margossian, C., Modrák, M., Adler, A., Sakrejda, K., Stukalov, A., Lawrence, M., Goedman, R. J., Van Horn, K. S., Vehtari, A., Gabry, J., Casallas, J. S., and Bales, B. (2017). Stan-dev/stan: V2.17.1. Zenodo.

Lin, S. and Luo, W. (2019). A New Multilevel CART Algorithm for Multilevel Data with Binary Outcomes. *Multivariate Behavioral Research*, 54(4):578–592.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics. Wiley, 1 edition.

Lüdtke, O., Robitzsch, A., and Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, 22(1):141–165.

Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical science*, pages 538–558.

Mistler, S. A. and Enders, C. K. (2017). A Comparison of Joint Model and Fully Conditional Specification Imputation for Multilevel Missing Data. *Journal of Educational and Behavioral Statistics*, 42(4):432–466.

Morris, T. P., White, I. R., and Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11):2074–2102.

Murray, D. M. and Blitstein, J. L. (2003). Methods To Reduce The Impact Of Intraclass Correlation In Group-Randomized Trials. *Evaluation Review*, 27(1):79–103.

Quartagno, M. and Carpenter, J. R. (2022). Substantive model compatible multilevel multiple imputation: A joint modeling approach. *Statistics in Medicine*, 41(25):5000–5015.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Resche-Rigon, M. and White, I. R. (2018). Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Statistical Methods in Medical Research*, 27(6):1634–1649.

Rights, J. D. and Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, 24(3):309–338.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* Wiley, New York.

Salditt, M., Humberg, S., and Nestler, S. (2023). Gradient Tree Boosting for Hierarchical Data. *Multivariate Behavioral Research*, pages 1–27.

Shieh, G. (2012). A comparison of two indices for the intraclass correlation coefficient. *Behavior Research Methods*, 44(4):1212–1223.

Silva, G. C. and Gutman, R. (2022). Multiple imputation procedures for estimating causal effects with multiple treatments with application to the comparison of healthcare providers. *Statistics in Medicine*, 41(1):208–226.

Taljaard, M., Donner, A., and Klar, N. (2008). Imputation Strategies for Missing Continuous Outcomes in Cluster Randomized Trials. *Biometrical Journal*, 50(3):329–345.

Tan, Y. V., Flannagan, C. A. C., and Elliott, M. R. (2016). Predicting human-driving behavior to help driverless vehicles drive: Random intercept Bayesian Additive Regression Trees.

Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, 16(3):219–242.

van Buuren, S. (2018). *Flexible Imputation of Missing Data.* Chapman & Hall/CRC Interdisciplinary Statistics Series. CRC Press, Taylor & Francis Group, Boca Raton London New York, second edition edition.

Wagner, J., West, B. T., Elliott, M. R., and Coffey, S. (2020). Comparing the Ability of Regression Modeling and Bayesian Additive Regression Trees to Predict Costs in a Responsive Survey Design Context. *Journal of Official Statistics*, 36(4):907–931.

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Zhu, J., and Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*, 3(8):e002847.

Wundervald, B., Parnell, A., and Domijan, K. (2022). Hierarchical Embedded Bayesian Additive Regression Trees.

Xu, D., Daniels, M. J., and Winterstein, A. G. (2016). Sequential BART for imputation of missing covariates. *Biostatistics*, 17(3):589–602.