

Quantile Regression-Based Multiple Imputation of Missing Values — An Evaluation and Application to Corporal Punishment Data

Kristian Kleinke¹, Markus Fritsch², Mark Stemmler³, Jost Reinecke⁴, Friedrich Lösel^{3,5}

[1] Department of Education Studies and Psychology, University of Siegen, Siegen, Germany. [2] School of Business, Economics and Information Systems, University of Passau, Passau, Germany. [3] Institute of Psychology, University of Erlangen-Nurnberg, Erlangen, Germany. [4] Faculty of Sociology, University of Bielefeld, Bielefeld, Germany. [5] Institute of Criminology, University of Cambridge, Cambridge, United Kingdom.

Methodology, 2021, Vol. 17(3), 205–230, <https://doi.org/10.5964/meth.2317>

Received: 2019-12-02 • **Accepted:** 2021-06-16 • **Published (VoR):** 2021-09-30

Corresponding Author: Kristian Kleinke, University of Siegen, Department of Education Studies and Psychology, Adolf-Reichwein-Str. 2a, D-57068, Siegen, Germany. Tel.: +49 27174016906, E-mail: kristian.kleinke@uni-siegen.de

Abstract

Quantile regression (QR) is a valuable tool for data analysis and multiple imputation (MI) of missing values – especially when standard parametric modelling assumptions are violated. Yet, Monte Carlo simulations that systematically evaluate QR-based MI in a variety of different practically relevant settings are still scarce. In this paper, we evaluate the method regarding the imputation of ordinal data and compare the results with other standard and robust imputation methods. We then apply QR-based MI to an empirical dataset, where we seek to identify risk factors for corporal punishment of children by their fathers. We compare the modelling results with previously published findings based on complete cases. Our Monte Carlo results highlight the advantages of QR-based MI over fully parametric imputation models: QR-based MI yields unbiased statistical inferences across large parts of the conditional distribution, when parametric modelling assumptions, such as normal and homoscedastic error terms, are violated. Regarding risk factors for corporal punishment, our MI results support previously published findings based on complete cases. Our empirical results indicate that the identified “missing at random” processes in the investigated dataset are negligible.

Keywords

missing values, multiple imputation, quantile regression, random forest, corporal punishment, parenting behavior



Haupt et al. (2014) point out a gap between modern statistical methods and routine practice in applied psychological research. They argue that robust statistical methods are not widely used, and that analysis of variance (ANOVA) or ordinary least squares (OLS) regression are still employed frequently, even when the underlying parametric modelling assumptions are not met (e.g., in the presence of heteroscedasticity, outliers in the response, or other forms of non-normality like skewness or multi-modality).

This is also the case when imputing missing values in a dataset, where creating multiple imputations (MI) of each missing value based on Rubin's (1987) theory is one of the state-of-the-art approaches (cf. Schafer & Graham, 2002). Most of the currently available MI procedures involve strong parametric assumptions, which are often not met by empirical data in various fields of the social sciences: Haupt et al. (2014) discuss methods to account for outliers in the response and heteroskedasticity in their analysis of risk factors for corporal punishment of children by their fathers; Kleinke, Reinecke, and Weins (2020) analyse highly skewed longitudinal data on juvenile delinquency with missing values and compare modelling results based on normal model-based MI to results based on imputation methods, which rely on more suitable distributional assumptions. Further examples are response times in psychological testing (e.g., Ranger & Wolgast, 2019) – where empirical distributions are typically not normal and often difficult to parameterise – or research on substance abuse and (behavioural) addiction (e.g., Chodkiewicz et al., 2020; Sugiyama et al., 2020), which commonly involves skewed empirical distributions. Though, fully parametric normal model-based MI is still employed (e.g., Wayman, 2002), approaches beyond these methods are often more suitable. An argument frequently made in applied research is that MI is “robust” in general. See the discussion regarding the *self-correcting property of MI* (i.e., masking bias in the model of scientific interest or *analysis model* due to violated assumptions by overestimated standard errors) in Kleinke, Reinecke, Salfrán et al. (2020, Chapter 4.5). However, this is true only to a certain extent. Unfortunately, even imputation methods, which are believed to be robust against mild to moderate violations of parametric assumptions like predictive mean matching (PMM) for example, tend to produce biased inferences, when empirical data are highly skewed (Kleinke, 2017; Yu et al., 2007). Yu et al. (2007) recommend that the use of imputation methods with mis-specified distributional assumptions should be avoided, when the given empirical data deviate too strongly from these assumptions. Alternatives to normal model-based MI are quantile regression (QR)-based MI (Geraci, 2016a, 2016b, 2017) and MI of missing values based on (semi-parametric) generalized additive models for location, scale, and shape (GAMLSS; see e.g., Rigby & Stasinopoulos, 2005; Stasinopoulos & Rigby, 2007; for an implementation in R, see Salfrán & Spiess, 2018).

The two approaches provide flexibility in modelling parametric distributions (GAMLSS-based MI) – location, scale and shape of the distribution, to be more specific – or avoid parametric distributions (QR-based MI) and focus on modelling conditional quantiles of interest directly (Koenker, 2005; Koenker & Hallock, 2001). Despite their

obvious advantages over normal model-based MI, systematic evaluations of QR- and GAMLSS-based MI are still scarce. Notable exceptions are [Bottai and Zhen \(2013\)](#) and [Geraci and McLain \(2018\)](#). The former paper investigated different sample sizes and error distributions and found that QR-based MI was superior to standard imputation methods like PMM. The latter paper examined a QR-based MI approach for bounded data, and found superior performance compared to PMM. Little is known, yet about QR-based MI in the case of discrete data (see the discussion in [Bottai & Zhen, 2013](#), regarding QR-based MI of count data) or quasi-continuous data: In the social sciences, scale scores are often derived from ordinal data (e.g., the mean of Likert scale items). Resulting scale scores are often treated as quasi-continuous and analysed by methods for continuous data.

This paper evaluates QR-based MI in a setting common in the social sciences, where empirical data are frequently not genuinely continuous, normal, and homoscedastic and often contain outliers in the response. Our contribution is two-fold: First, we provide a Monte Carlo simulation in which we compare QR- and GAMLSS-based MI to normal model-based MI and PMM. Second, we replicate and extend the empirical analysis of the risk factors for corporal punishment of children by their fathers in [Haupt et al. \(2014\)](#) and [Fritsch et al. \(2019\)](#) with QR-based MI. In both papers, the results were based on the complete cases only and the response was derived from Likert scale items.

The paper is structured as follows: First, we briefly introduce the theoretical background of MI and QR. We then introduce the QR- and GAMLSS-based MI approaches. In Section 3, we present and discuss the results from our Monte Carlo simulation. In Section 4, we replicate the results in [Haupt et al. \(2014\)](#) and [Fritsch et al. \(2019\)](#) based on QR-based MI and contrast our results with the findings for the complete cases. A general discussion concludes.

Theoretical Background

Multiple Imputation

Multiple imputation based on [Rubin's \(1987\)](#) theory is a three-step process: First, each missing value is imputed multiple times by *plausible* values, so that researchers end up with M complete datasets. Plausible in this context refers to any value that leads to unbiased statistical inferences.¹ Second, these M datasets are analysed separately by any complete data technique such as OLS regression, or QR and yield M sets of modelling results. In the final step, the M sets of modelling results are combined into an overall result based on combination rules such as the large-sample formula given in [Rubin \(1987\)](#) or adaptations for small samples ([Barnard & Rubin, 1999](#)). For a detailed discussions, see

1) See [Rubin \(1996\)](#) for a discussion of criteria that are required to create proper imputations.

Kleinke, Reinecke, Salfrán et al. (2020, Chapter 4). The combination formulae yield point estimates and standard errors for each model parameter θ_p , $p = 1, \dots, P$. The combined point estimate is simply the average of the M estimates. Corresponding standard errors are obtained by combining a within-imputation and a between-imputation variance component. The latter component reflects the amount of missing information with respect to θ_p : Fewer information to predict the unobserved values leads to higher uncertainty in the imputations and a larger between variance. Note that if the missing values could be predicted without any uncertainty, the estimator $\hat{\theta}_{p,m}$ obtained from the $m = 1, \dots, M$ imputed datasets would not vary over the datasets – resulting in a between variance of zero.

More comprehensive introductions to MI and helpful resources for applied researchers are given in: Kleinke, Reinecke, Salfrán et al. (2020), who provide a general introduction and review recent developments and various applications of MI. Chapters 1, 4, and 5 in Little and Rubin (2020) summarize the problem of missing values, missing value patterns and mechanisms; additionally, a taxonomy of methods for imputing missing values and a discussion of the rationale of MI in comparison to other methods to account for uncertainty in parameter estimates due to missing values are given. Applications of MI in various contexts including linear and generalized linear models, categorical data analysis, survival analysis, structural equation modelling, longitudinal data analysis, and complex survey settings are discussed in Raghunathan et al. (2018).

Assumptions

Approaches for MI typically assume that the *mechanism generating the missing values* in a variable y is *ignorable* (Rubin, 1976, 1987; see also Kleinke, Reinecke, Salfrán et al., 2020, Chapter 2): Observing or not observing values in y can be seen as the result of a dichotomous process generating a binary response r , which obtains the value ‘1’ if the value of the corresponding variable is observed and ‘0’ otherwise. Transitioning from the realized sample to the observed subsample can be interpreted as an additional sampling step. The underlying process, however, is usually unknown during data collection and assumed to be a random process. When the mechanism generating the missing values is ignorable, datasets can be analysed without the need to explicitly model the mechanism. Stated differently, an ignorable mechanism can be assumed when the missing values are generated at random, given all observed information (i.e., the probability of not observing certain values of a variable depends on observed information only). To ensure the credibility of this assumption, applied researchers typically specify *rich* imputation models and include all variables related to *patterns of missing values* into the imputation model – even if these variables are omitted from the analysis model (cf. Collins et al., 2001).

How Imputations Are Generated

MI based on [Rubin's \(1987\)](#) theory can be considered a Bayesian approach: Imputations are generated by drawing from the posterior predictive distribution of the variables whose values are missing given all observed values. From a mathematical point of view, the posterior predictive distribution is usually difficult to approximate, because in most datasets different types of variables are present, for example, continuous, truncated, categorical or count, and missing value patterns are usually not monotone.

A popular approach to handle these situations is the mice (multiple imputation by chained equations) framework ([van Buuren, 2018](#); [van Buuren & Groothuis-Oudshoorn, 2011](#); [van Buuren et al., 2006](#)), where users can specify separate conditional imputation models for each variable with missing values. The models can be specifically tailored to the distribution and type of each variable (e.g., methods `norm` for a standard linear regression model, `logreg` for a logistic regression model or `polyreg` for a polytomous regression model). The software then cycles iteratively through all variables with missing values and imputes them step by step using predictions from a pre-specified regression model. Values of all other (observed and imputed) variables are used by the procedure. The approach is implemented in the R package `mice` ([van Buuren & Groothuis-Oudshoorn, 2011](#)). Extensions of the functionality of the package are available in other packages such as `Qtools` ([Geraci, 2016a, 2016b](#)) for QR-based MI and `ImputeRobust` ([Salfrán & Spiess, 2018](#)) for GAMLSS-based MI. The methodology underlying both approaches is described in the next two subsections.

QR and QR-Based MI of Missing Values

The QR model assumes a linear additive functional form for the τ -quantiles of the response y_i conditional on the P predictors \mathbf{x}_i , where $i = 1, \dots, n$ are the sample observations. Define the conditional τ quantile as

$$Q_{y_i | \mathbf{x}_i}(\tau) = \mathbf{x}_i^T \boldsymbol{\beta}_\tau, \quad (1)$$

where $\boldsymbol{\beta}_\tau$ denotes the parameter corresponding to the τ th quantile. Parameter estimates $\hat{\boldsymbol{\beta}}_\tau$ are obtained by minimizing the asymmetrically weighted loss function

$$\sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau), \quad \text{with} \quad \rho_\tau(\cdot) = \begin{cases} \tau, & \text{if } (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau) \geq 0 \\ \tau - 1, & \text{otherwise,} \end{cases} \quad (2)$$

where $\rho_\tau(\cdot)$ governs the weighting. Choosing $\tau = 0.5$ yields the conditional median (for further details, see [Koenker, 2005](#)).

The following data generating process illustrates the exact fit property of QR:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_\tau + \epsilon_{\tau,i}, \quad \text{with } \epsilon_{\tau,i} \sim F_{\tau,i} \quad \text{and} \quad F_{\tau,i}^{-1}(\tau) = 0, \quad (3)$$

where $F_{\tau,i}$ is some distribution function and $F_{\tau,i}^{-1}$ its inverse. Note that if q is the conditional τ quantile, such that $\tau = F_{\tau,i}(q)$, it holds that $q = F_{\tau,i}^{-1}(\tau)$ and the conditional τ quantile can be recovered by inverting the distribution function at τ . The weak assumptions about the distribution of the errors and the fact that quantiles of the conditional distribution of y_i are modelled implies advantages over restrictive parametric methods when distributional assumptions of the latter are violated (e.g., in case of outliers or heteroscedasticity). A disadvantage of QR is that a sufficiently large sample size is required for the model parameters to be identified because all parameters are estimated for each quantile.²

To illustrate QR-based MI of missing values, denote the response vector by \mathbf{y} , the matrix of P predictors by \mathbf{X} , and assume that all variables contain missing values (or incomplete cases). Define the dataset with missing values by $\mathbf{U} = (\mathbf{y}, \mathbf{X})$, where each of the $i = 1, \dots, n$ lines of the dataset is at least partially observed. The columns of the dataset \mathbf{U} are referred to by $q = 1, \dots, Q$, with $Q = P + 1$. Note that the total number of observations in the dataset is $N = n \cdot Q$. Denote the total number of missing values as N_{mis} and the number of observed values (or complete cases) as $N_{\text{obs}} = N - N_{\text{mis}}$. The number of missing values in each column of the dataset is referred to as $n_{q,\text{mis}}$, such that $N_{\text{mis}} = n_{1,\text{mis}} + \dots + n_{Q,\text{mis}}$.

We aim to impute the missing values in \mathbf{U} and use them in the analysis model. The matrix of imputed values is represented by \mathbf{V} . To simplify the exposition, assume that \mathbf{U} is sorted such that the lines with the observed values are given first and refer to the individual lines via \mathbf{u}_j , $j = 1, \dots, J$ ($J = \min(n_{1,\text{obs}}, \dots, n_{Q,\text{obs}}$). The remaining lines with the missing values are denoted by \mathbf{u}_k , $k = 1, \dots, K$ ($K = N - J$). The definition of \mathbf{v}_j and \mathbf{v}_k is analogous.

In the following, we describe function `mice.impute.rq` from package `Qtools` for QR-based MI in greater detail. Note that we do not consider approaches restricting regression quantiles to avoid quantile crossing. Based on the defined notation, function `mice.impute.rq` proceeds according to the [Algorithm 1 \(Geraci, 2016a\)](#).

[Algorithm 1](#) details the steps required for QR-based MI of missing values. First, choose the number of iterations C and the number of bootstrap samples M . For bootstrap iteration $m = 1, \dots, M$, proceed as follows: Define a temporary dataset $D_{m,c}$: in the first iteration $c = 1$, the lines of the dataset without missing values \mathbf{u}_j are used. At subsequent iterations $2, \dots, C$, the imputed values from the previous iteration $V_{m,c-1}$ are employed

² See also [Wilcox \(2016, Chapter 8\)](#) for a discussion of standard OLS regression, outliers in the response, and heteroscedasticity; [Wilcox \(2017\)](#) for an extensive overview of robust statistical methods and the precision of estimators when using robust methods with ordinal or (quasi-)quantitative data (Chapters 10 and 11); and [Koenker \(2005\)](#) for a comprehensive introduction to QR.

Algorithm 1

Algorithm of function *mice.impute.rq*

Data: $U = (y, X)$, with $X = (x_1, \dots, x_p)$, $Q = P + 1$, and N_{mis} missing values

Result: Datasets with imputed values V_1, \dots, V_M

Initialization: Set number of iterations C and number of bootstrap samples M ;

for $c = 1, \dots, C$ **do**

for $m = 1, \dots, M$ **do**

if $c=1$ **then**

 assign complete cases u_j to temporary dataset $D_{m,c}$

else

 update temporary dataset $D_{m,c}$ with imputed dataset from previous iteration

$V_{m,c-1}$

end

for $q = 1, \dots, Q$ **do**

 (1) Draw bootstrap sample $B_{m,c}$ of size $n_{q,\text{mis}}$ with replacement from lines of $D_{m,c}$;

 (2) Partition $B_{m,c}$ into its q th column $b_{m,c,q}$ and $B_{m,c,-q}$;

 (3) Draw K random values τ_1, \dots, τ_K from the uniform distribution $\mathcal{U}(\varepsilon, 1 - \varepsilon)$, with $0 < \varepsilon < 0.5$;

 (4) Estimate the QR $Q_{b_{i,q} | b_{i,-q}}(\tau) = b_{i,-q}^T \beta_\tau$, with $i \in B_{m,c}$;

 (5) Use $\tilde{\beta}_\tau$ to impute missing values via $\tilde{u}_{i,q} = u_{i,-q}^T \tilde{\beta}_\tau$, with $i \in U_K$;

end

 assign imputed values for all columns $q = 1, \dots, Q$ to V_m

end

end

return V_1, \dots, V_M

and the following five steps are used on all columns $q = 1, \dots, Q$: In (1), draw a bootstrap sample $B_{m,c}$ of size $n_{q,\text{mis}}$ from the lines of $D_{m,c}$ with replacement. In (2), partition $B_{m,c}$ into its q th column $b_{m,c,q}$ and all other columns $B_{m,c,-q}$. In (3), draw K random values τ_1, \dots, τ_K from the uniform distribution $\mathcal{U}(\varepsilon, 1 - \varepsilon)$, with $0 < \varepsilon < 0.5$ (note that the function default is $\varepsilon = 0.001$). In (4), fit a QR model based on the partitioned bootstrap sample $Q_{b_{i,q} | b_{i,-q}}(\tau) = b_{i,-q}^T \beta_\tau$ based on the bootstrap sample $B_{m,c}$. In (5), use the parameter estimates $\tilde{\beta}_\tau$ from (4) to impute the missing values for each column of the dataset with missing values U_K . Collect the imputed values for all $q = 1, \dots, Q$ columns and assign them to V_m . Then, iterate the five steps until the pre-specified number of iterations C is reached.³ The algorithm yields M datasets with imputed values V_1, \dots, V_M .

GAMLSS and GAMLSS-Based MI of Missing Values

De Jong (2012) and de Jong et al. (2016) have proposed a MI procedure based on a class of semi-parametric GAMLSS (Rigby & Stasinopoulos, 2005; Stasinopoulos & Rigby, 2007), which is implemented in R package `ImputeRobust` (Salfrán & Spiess, 2018) and extends the functionality of the `mice` package. Compared to standard normal model-based MI, which only considers modelling the location of the distribution, GAMLSS-based MI also allows to account for the scale, and the shape of the distribution and can be better tailored to a given empirical dataset.

In the following, we slightly extend the notation from the previous subsection to illustrate GAMLSS-based MI briefly.⁴ Assume that the conditional distribution of the q th column \mathbf{u}_q of dataset \mathbf{U} given the observed values $\mathbf{u}_{J,q}$ and the other columns \mathbf{U}_{-q} is

$$\mathbf{u}_q \mid \mathbf{u}_{J,q}, \mathbf{U}_{-q} \sim \mathcal{D}(\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \dots, \theta^{(S)}). \quad (4)$$

Parameters $\theta^{(1)}$ and $\theta^{(2)}$ are the location and scale parameters, and $\theta^{(3)}, \dots, \theta^{(S)}$ are shape parameters – depending on the specified distribution \mathcal{D} . Imputations for the missing values in column q of the dataset \mathbf{U}_K can be obtained by

$$g_q(\boldsymbol{\theta}_q) = \mathbf{U}_q \boldsymbol{\gamma}_q + \sum_{j=1}^J h_{j,q}(\mathbf{u}_{j,q}), \quad (5)$$

where $\boldsymbol{\theta}_q$ is a parameter vector and $g_q(\cdot)$ a link function relating the location, scale, and shape parameter of the missing values in column q to the dataset \mathbf{U} . Equation 5 is additively decomposed into a linear parametric part with design matrix \mathbf{U}_q and corresponding coefficient vector $\boldsymbol{\gamma}_q$ and a sum of smooth functions $h_{j,q}(\cdot)$ of the observed values $\mathbf{u}_{j,q}$.

Since package `gamlss` provides different approaches to model $h(\cdot)$, such as cubic smoothing splines, penalized splines, and local regression, the following algorithm provides a general description of GAMLSS-based MI (Salfrán & Spiess, 2018).

Algorithm 2 details GAMLSS-based MI of missing values. First, choose the number of bootstrap samples M , set $m = 1$, use the lines of the dataset without missing values \mathbf{U}_J as initial bootstrap sample \mathbf{B}_1 , and choose starting values $\theta_{0,k,q}^{(1)}, \theta_{0,k,q}^{(2)}, \theta_{0,k,q}^{(3)}, \dots, \theta_{0,k,q}^{(S)}$ for the parameters of the distribution \mathcal{D} . In (1), the penalized likelihood function ℓ_{pen} is specified based on the distribution \mathcal{D} . In (2), maximise ℓ_{pen} based on \mathbf{B}_1 to obtain parameter

3) Usually around 20 iterations are sufficient (for further details, see van Buuren, 2018; van Buuren & Groothuis-Oudshoorn, 2011).

4) More thorough descriptions of the method and its theoretical underpinnings are provided in Salfrán and Spiess (2018) and Kleinke, Reinecke, Salfrán, et al. (2020). Chapter 6 of the latter reference discusses settings where GAMLSS-based MI is suitable.

Algorithm 2*Algorithm of Function ImputeRobust*

Data: $U = (y, X)$, with $X = (x_1, \dots, x_p)$, $Q = P + 1$, and N_{mis} missing values;

assume $u_q | u_{J,-q} \sim \mathcal{D}(\hat{\theta}_{k,q}^{(1)}, \hat{\theta}_{k,q}^{(2)}, \hat{\theta}_{k,q}^{(3)}, \dots, \hat{\theta}_{k,q}^{(S)})$

Result: Datasets with imputed values V_1, \dots, V_M

Initialization: Choose number of bootstrap samples M and set $m = 1$;

assign $B_1 = U_J$ and set starting values $\hat{\theta}_{0,k,q}^{(1)}, \hat{\theta}_{0,k,q}^{(2)}, \hat{\theta}_{0,k,q}^{(3)}, \dots, \hat{\theta}_{0,k,q}^{(S)}$;

for $q = 1, \dots, Q$ **do**

(1) Specify penalized likelihood function ℓ_{pen} based on chosen \mathcal{D} ;

(2) Maximize ℓ_{pen} based on B_1 to obtain parameter estimates

$\hat{\theta}_{1,k,q}^{(1)}, \hat{\theta}_{1,k,q}^{(2)}, \hat{\theta}_{1,k,q}^{(3)}, \dots, \hat{\theta}_{1,k,q}^{(S)}$ and fitted values $\hat{u}_{J,q}$;

for $m = 2, \dots, M$ **do**

(3) Draw bootstrap sample B_m of size $n_{q,\text{mis}}$ with replacement from $(\hat{u}_{m-1,q}, U_{J,-q})$;

(4) Maximize ℓ_{pen} based on B_m and update parameter estimates

$\hat{\theta}_{m,k,q}^{(1)}, \hat{\theta}_{m,k,q}^{(2)}, \hat{\theta}_{m,k,q}^{(3)}, \dots, \hat{\theta}_{m,k,q}^{(S)}$ and fitted values $\hat{u}_{m,q}$

end

assign fitted values for all columns $q = 1, \dots, Q$ to V_m ;

end

return V_1, \dots, V_M

estimates $\hat{\theta}_{1,k,q}^{(1)}, \hat{\theta}_{1,k,q}^{(2)}, \hat{\theta}_{1,k,q}^{(3)}, \dots, \hat{\theta}_{1,k,q}^{(S)}$ and fitted values for column q of the dataset $\hat{u}_{J,q}$.

Next, set $m = 2$. In (3), the bootstrap sample B_m is updated by drawing $n_{q,\text{mis}}$ observations with replacement from the fitted values from the previous bootstrap iteration $\hat{u}_{m-1,q}$ and the corresponding observed columns of the dataset besides the q th column $U_{J,-q}$. In (4), maximise ℓ_{pen} based on B_m and update the parameter estimates $\hat{\theta}_{1,k,q}^{(1)}, \hat{\theta}_{1,k,q}^{(2)}, \hat{\theta}_{1,k,q}^{(3)}, \dots, \hat{\theta}_{1,k,q}^{(S)}$. Then, repeat (3) and (4) until $m = M$. The algorithm yields M datasets with imputed values V_1, \dots, V_M .

Imputation Based on PMM

Another approach that was recommended for a wide range of scenarios is PMM (cf. van Buuren, 2018, Chapter 3): The method samples from a pool of observed cases, which are ‘similar’ to the incomplete case. This so-called *donor* is then used to fill in the corresponding missing value of the *donee*. The similarity of observations is defined by similarity in predicted means for the observed and unobserved part of the variable

based on a linear regression model. Details about the implementation in package `mice` (function `mice.impute.pmm`) are given in [van Buuren \(2018, Chapter 3\)](#).

PMM is considered a semi-parametric imputation procedure: The parametric linear model is used only as a means for matching the incomplete case to a potential donor. Then, actually observed values are imputed. This preserves many properties of the observed part of the data, which makes PMM robust against mild to moderate violations of distributional assumptions of the normal linear regression model. This was systematically evaluated by [Kleinke \(2017\)](#) for different degrees of skewness, different sample sizes, and varying percentages of missing values. Their results indicated that: 1) especially when sample size was small, small donor pools yielded better results in terms of unbiased parameter estimates and standard errors compared to large donor pools; 2) PMM generally worked well, unless the data were highly skewed and/or more than roughly 20% to 30% of the data were imputed; 3) PMM generally performed better when sample size was large.

Disadvantages of PMM are that the approach can neither extrapolate beyond the range of the observed data, nor interpolate when regions of the data are sparsely populated. Additionally, PMM fails when there are substantial violations of the assumptions of the normal linear regression model.

Monte Carlo Evaluation of QR-Based MI

Overview

In this section, we evaluate QR-based MI by Monte Carlo Simulation based on a typical setting in the social sciences: Skewed and heteroskedastic data. The underlying data were taken from [Haupt et al. \(2014\)](#), who investigated risk factors for corporal punishment by fathers with a QR model. We proceed as follows: First, draw 1000 bootstrap samples with replacement from the complete cases and fit a QR model to each of the bootstrap samples. Second, use the average parameter estimates as a proxy for the unknown ‘true’ population parameters. Third, introduce that values are missing at random (MAR) to the bootstrap samples by deleting a certain fraction of values. Finally, evaluate QR-based MI regarding bias in point estimates and coverage rates (CR) of the 95% confidence intervals of the parameters and contrast the results with the complete data estimates and results obtained from other established MI methods (standard normal model-based MI, PMM, and GAMLSS-based MI).

Hypotheses

The advantages and disadvantages of QR in comparison to GAMLSS are discussed in [Kneib \(2013\)](#) and [Rigby et al. \(2013\)](#). QR, for example, is distribution-free, while GAMLSS introduce model structure and are more parsimonious regarding the data requirements

– while also bearing a higher risk of mis-specification. To our best knowledge, QR- and GAMLSS-based MI have not been compared systematically in Monte Carlo simulations, yet. We expect the preferred method to depend on the specific setting. In our simulation, we expect QR- and GAMLSS-based MI to yield comparable results. We assume that both approaches will outperform PMM and standard normal model-based MI – especially when quantiles farther away from the center of the distribution of the response are considered: PMM and normal model-based MI create imputations based on predictions for the conditional mean. Thus, both normal model-based MI and PMM should perform well in the vicinity of the center of the distribution.

Method

Design

Our simulation has a $2 \times 2 \times 4$ (percentage of missing values \times missing value pattern \times imputation method) design: From the bootstrap samples, we deleted 25% or 40% of values either from the response only, or from the response and one of the predictors following a MAR pattern (for details, see Section “Procedure”). We imputed missing values by QR- and GAMLSS-based MI, PMM, and standard normal model-based MI.

Evaluation Criteria

We evaluated the approaches based on the goal to yield unbiased statistical inference: Bias in the parameter estimate was defined as the difference between the true parameter θ and the average estimate $\hat{\theta}$ across the bootstrap samples. Bias was reported relative to the true population parameter as

$$\%Bias = \frac{Bias}{\theta} \cdot 100\% . \quad (6)$$

Coverage Rate (CR) was defined as the percentage of 95% confidence intervals that included the true population parameter. Schafer and Graham (2002) deem CR below 90% as seriously low. Undercoverage may result from biased point estimates and/or an underestimated variance. The confidence interval could either be too narrow to include the true parameter, and/or it can be shifted too much to the left or to the right to include the true parameter. Together, bias and CR allow to assess the accuracy of statistical inferences.

Sample and Measures

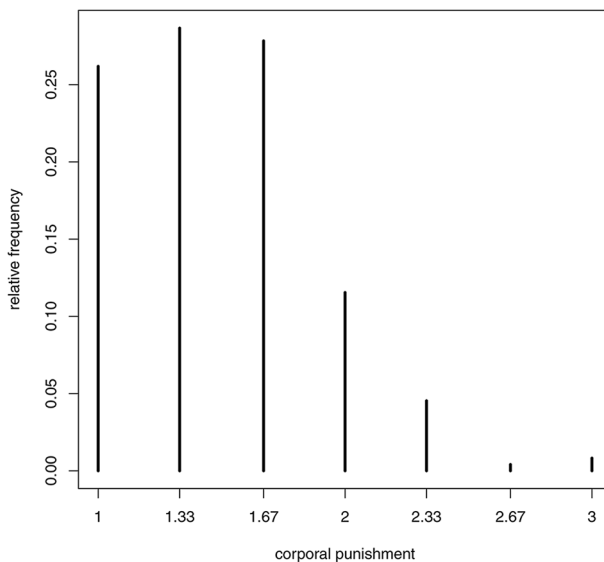
Our simulation is based on a sample taken from the longitudinal Erlangen-Nuremberg Development and Prevention Study. The data include 609 families (675 preschool chil-

dren) from the North-Bavarian cities of Erlangen and Nuremberg (for more details, see Lösel et al., 2009).

The response in our model *corporal punishment* by fathers (CP), as well as two risk factors *inconsistent discipline* (ID) and *other discipline practices* (OD) were assessed using a German version of the Alabama Parenting Questionnaire (APQ; Lösel et al., 2003). Items were rated on five-point scales ranging from 1 (“never”) to 5 (“always”). Scale scores were obtained by calculating the means of the respective items. The corporal punishment scale was based on three items such as “You spank your child with your hand when he/she has done something wrong”. The empirical distribution of the corporal punishment scale is summarized in Figure 1.⁵ Inconsistent discipline (six items) summarizes parenting behaviors like “The punishment you give your child depends on your mood”. Other discipline practices (seven items) refer to actions such as “You take away privileges or money from your child as a punishment”. A further risk factor of interest was *socioeconomic status* (SES). The computation of the SES index was based on Geißler (1994). Finally, our model included the child’s gender by the *gender indicator* (FEMALE).

Figure 1

Needle Plot of the Empirical Distribution of the Response Corporal Punishment



5) Note that Erlangen-Nuremberg is a relatively affluent area. More extreme CP scores might be expected in more deprived urban areas.

Procedure

We simulated 1000 bootstrap samples from the original $N = 485$ complete cases. Each bootstrap sample had a size corresponding to the original sample size $N = 675$. We then introduced MAR patterns (Rubin, 1976). Depending on the respective condition, we deleted either 25% or 40% of the values of the response CP, and/or the predictor OD. Values were deleted using the following rule: For each observation i we computed the probability of a missing value Pr_i^{mis} with the logit model

$$Pr_i^{mis} = \text{invlogit}(5 - 2 \cdot \text{SES}_i) \quad (7)$$

Coefficients in Equation 7 were chosen to ensure probabilities of missing values ranging from about 0.2 to 0.8 – depending on the individual SES scores. In our simulation, participants from lower social classes were assigned larger probabilities of missing values.

Depending on the respective condition, we sampled the necessary number of cases with selection probabilities Pr_i^{mis} and deleted the respective observations from the datasets. An overview of the resulting missing value patterns is displayed in Table 1.

Table 1

Overview on Simulated Missing Value Patterns and Average Observed Percentages of Each Pattern

Percentage of each pattern	SES	FEMALE	ID	OD	CP
25% missing values in CP					
74.96%	1	1	1	1	1
25.04%	1	1	1	1	0
25% missing values in CP and OD					
56.54%	1	1	1	1	1
18.42%	1	1	1	1	0
18.42%	1	1	1	0	1
6.61%	1	1	1	0	0
40% missing values in CP					
60.00%	1	1	1	1	1
40.00%	1	1	1	1	0
40% missing values in CP and OD					
36.71%	1	1	1	1	1
23.29%	1	1	1	1	0
23.29%	1	1	1	0	1
16.71%	1	1	1	0	0

Note. Observed values denoted by ‘1’, ‘0’ indicates missing value for variables: SES = socio-economic status; FEMALE = gender indicator; ID = inconsistent discipline; OD = other discipline practices; CP = corporal punishment. Missing values were MAR and based on individuals’ SES scores.

Missing values were imputed with QR- and GAMLSS-based MI, PMM, and normal model-based MI by using functions implemented in packages `mice`, `Qtools`, and `ImputeRobust`. Unless stated otherwise, we used default settings of the functions. All other columns of the dataset were used to predict missing values in a particular column. As common practice (e.g., Kleinke, 2018), the number of imputations was set to the percentage of missing values.

We then compared parameter estimates based on the imputed datasets with the means of the 1000 complete data estimates. We obtained coefficient estimates from a QR model, in which we regressed the response CP on SES, FEMALE, ID, and OD regarding quantiles $\tau \in \{.2, .3, .4, .5, .6, .7, .8\}$. Heteroscedasticity-robust standard errors were computed as in Haupt et al. (2014).

Results

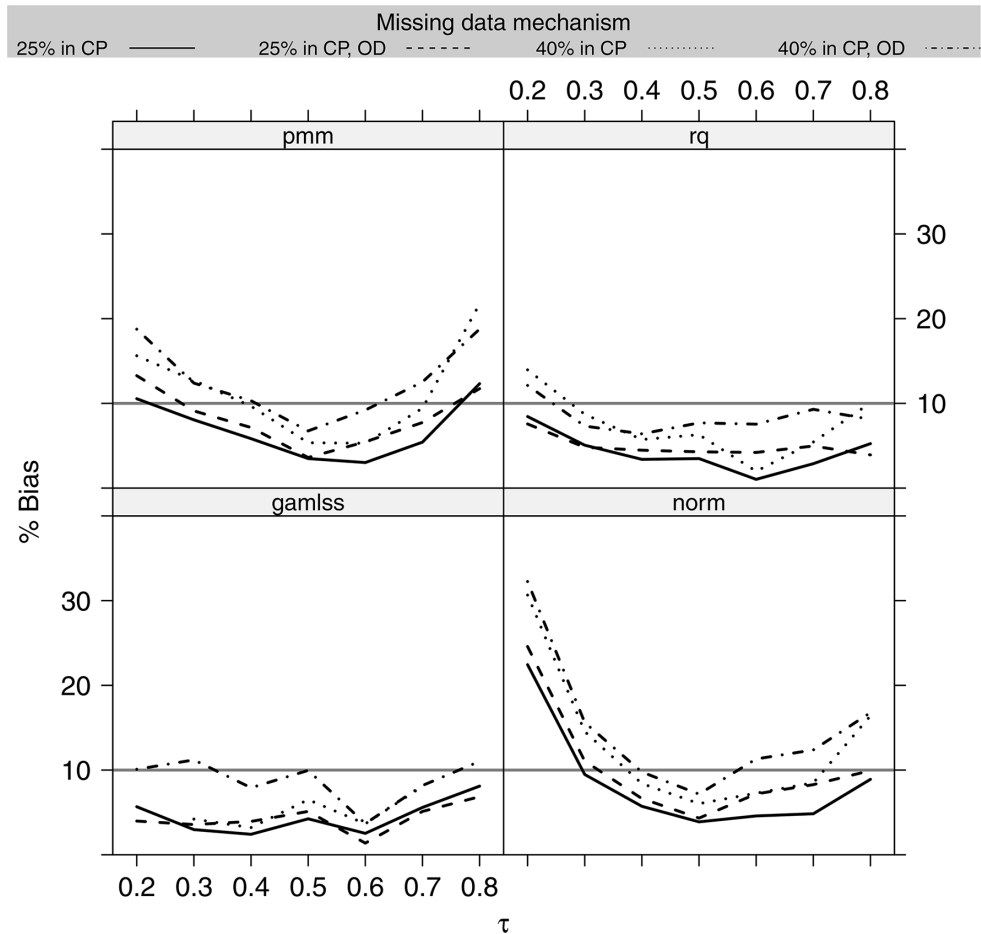
Figure 2 summarizes the results on bias in the point estimates and displays the average amount of relative absolute bias across all model parameters. Following Forero and Maydeu Olivares (2009), bias was considered substantial if it exceeded 10% of the true population parameter. Overall, bias generally increased: (i) when the response and additionally predictor values were missing; (ii) with the number of missing values.

When considering the different approaches for MI, both PMM and normal model-based MI had only a small bias in the center part of the distribution of the response (i.e., $\tau \in \{.4, .5, .6\}$), on average. As expected, bias increased when moving towards the tails of the distribution for both approaches. QR- and GAMLSS-based MI produced widely unbiased results across all quantiles and Monte Carlo designs.

Figure 3 displays the corresponding CR. All approaches produced large enough confidence intervals (some even slightly above the nominal $(1 - \alpha)$ percent), such that a bias in point estimates was alleviated by a wider interval. CR was sufficiently large across all parameter estimates and quantiles, even when the percentage of missing values was 40%.

Figure 2

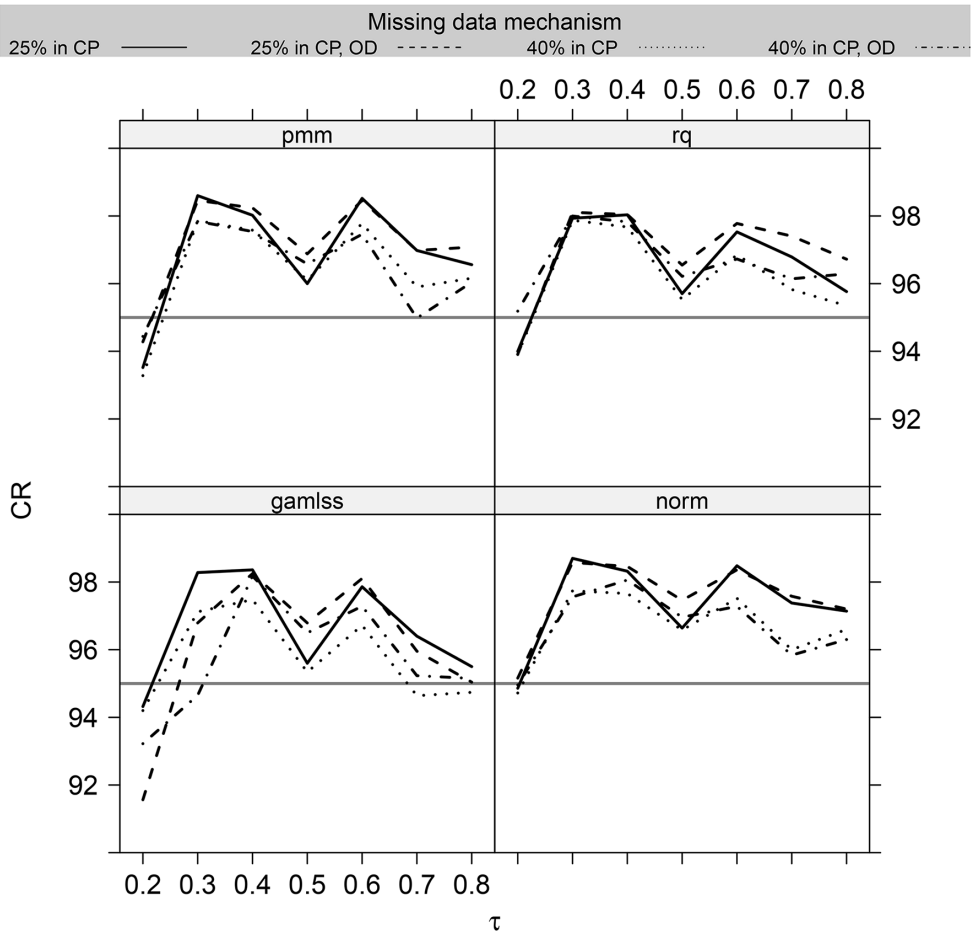
Percent Parameter Bias Produced by the Four Missing Value Methods in the Four Missing Value Scenarios



Note. Results were averaged across all parameter estimates; pmm is predictive mean matching; rq is quantile regression-based MI; gamlss is MI based on generalized additive models for location, scale, and shape; norm is Bayesian linear regression. % Bias above 10% of the true population parameter was considered substantial.

Figure 3

Coverage Rates (CR) of the 95% Confidence Intervals Produced by the Four Missing Value Methods in the Four Missing Value Scenarios



Note. Results were averaged across all parameter estimates; pmm is predictive mean matching; rq is quantile regression-based MI; gamlss is MI based on generalized additive models for location, scale, and shape; norm is Bayesian linear regression; horizontal line indicates nominal CR of 95%. CR below 90% were deemed substantial.

Discussion

Normal model-based MI produced unbiased inferences in the center region of the response. While point estimates in the off-center regions were biased, CR were nevertheless above 90%. The observed behaviour of normal model-based MI is due to the so-called *self-correcting property of MI*: A mild to moderate bias in the parameter estimate resulting

from the mis-specification of the imputation model by assuming normality and homoscedasticity of the error distribution is alleviated by an increase in the corresponding variance estimate. This also lead to accurate CR produced by PMM. This result should be of considerable interest to applied researchers facing datasets with incomplete cases: In practice, valid inference in off-center regions of the response are often of paramount interest – for example, when investigating risk factors for corporal punishment. Typically, applied researchers are not only interested in a suitable CR, but also in obtaining accurate point estimates regarding the tails of the distribution. To achieve this, both, the imputation and analysis model need to be able to capture the underlying properties of the data generating process. Our simulations show that this is widely the case for QR- and GAMLSS-based MI: Valid inference can be obtained from both approaches under different missing value patterns and percentages. Following previous recommendations (cf. Kleinke, Reinecke, Salfrán et al., 2020) that the imputation method and the subsequent analysis model should coincide, we recommend to use QR-based MI when the analysis model is a QR model.

Application

In this section, we compare the findings of the two empirical applications by Haupt et al. (2014) and Fritsch et al. (2019), who used the complete cases only with results obtained by QR-based MI.

Substantive Models

Haupt et al. (2014) considered further predictors in their QR model that we did not include in our simulations above. These were 1) fathers' *aggressiveness* (AG), measured by the aggressiveness scale of the FPI-R personality questionnaire 2) *parental distress* (PD), measuring the perceived stress of being a parent, 3) *dysfunctional interaction* (DI), assessing the parents' perception that the child does not meet their expectations or that they are manipulated by the child, and 4) the age of the father (AGE). For further details on the predictors, see Haupt et al. (2014). CP was regressed on the aforementioned predictors and heteroscedasticity-robust standard errors were computed (Powell, 1991). Regression coefficients were estimated for quantiles $\tau \in \{.2, .3, .4, .5, .6, .7, .8\}$.

In addition to Haupt et al. (2014), Fritsch et al. (2019) explored the relevance of further predictors (for an overview and a detailed description, see Table 1 and Section 3 in Fritsch et al., 2019). These variables were *age difference of the parents* (AGE.DIFF), as well as categorical indicators if the mother or father were older than the other parent respectively, *parental involvement*, *positive parenting*, *monitoring and supervision by the parents*, a measure of the '*difficulty*' of the child, *fathers' life satisfaction*, *social orientation*, *performance orientation*, *excitability*, *inhibition*, *levels of stress*, *health concerns* and *physi-*

cal complaints, candor, extraversion, and emotionality. Fritsch et al. (2019) focused on a prediction-centered approach and modelled the conditional mean of CP. We considered changes in the predictive performance when using the dataset obtained from QR-based MI.

Missing Values

Complete case analysis can be expected to yield unbiased statistical inferences in certain setting only, for example when only values in the response are missing or when the missing values are missing completely at random (MCAR) in the sense of Rubin (1976). MCAR means the probability of a missing value in a variable neither depends on observed information in the dataset (e.g., on SES), nor on unobserved information (e.g., on the unobserved CP values). The MCAR assumption is violated for example, when variables in the dataset have non-zero correlations with the missing value indicators of incompletely observed variables. We computed correlations of each variable with the missing value indicators of other variables in the dataset and found only small correlations between $-.16$ and $.15$. Effects of MAR should therefore be rather minor and we expect only marginal differences from the results reported in Haupt et al. (2014).

For the present analysis we included data from 575 children from families where both parents were present at the beginning of the study and for whom we had complete information regarding AGE, AGE.DIFF, and SES. Missing value patterns (e.g., van Buuren & Groothuis-Oudshoorn, 2011) are displayed in Table 2. About 82% of all cases are complete cases. No answers about CP were available from 79 fathers (about 14%). As many of these fathers also did not answer the other questionnaires (e.g., about parental distress or their aggressiveness), we could use mainly sociodemographic information (e.g., SES and FEMALE) to impute the missing values about CP. For some cases, we could also make use of observed information from the Freiburg Personality Inventory.

We imputed missing values using package `mice`. Missing values in the response were imputed by function `mice.impute.rq` from R package `Qtools` (Geraci, 2016a, 2016b). All other variables were imputed using `mice`'s default settings. Imputation models were automatically generated using `mice`'s default settings (for details, see van Buuren, 2018; van Buuren & Groothuis Oudshoorn, 2011). We then ran 20 iterations of `mice`'s Gibbs sampler and created a total of $m = 100$ imputations. We fitted the same QR model as in Haupt et al. (2014) to the imputed data and used standard combination rules (Barnard & Rubin, 1999; Rubin, 1987) to obtain the multiple imputation inferences.

Table 2*Missing Value Patterns*

Frequency	474	48	31	9
FEMALE	1	1	1	1
SES	1	1	1	1
AGE	1	1	1	1
AGE.DIFF	1	1	1	1
OLDER.M	1	1	1	1
OLDER.F	1	1	1	1
FPI	1	1	0	0
APQ	1	0	0	1
PSI	1	0	0	1

Note. Observed values denoted by '1', '0' indicates missing value: FEMALE = indicator for child's gender; SES = socioeconomic status; AGE = fathers' age; AGE.DIFF = age difference of the parents; OLDER.M and OLDER.F = indicators if mother or father is older than the other parent respectively; FPI = Freiburg Personality Inventory; APQ = Alabama Parenting Questionnaire; PSI = Parenting Stress Index. Column header gives observed frequency of each pattern. Only patterns with observed frequency larger than 5 are shown.

MI Results

The top part of [Table 3](#) displays point estimates and corresponding standard errors, the bottom part deviations from the complete case results reported in [Haupt et al. \(2014\)](#). Coefficient estimates and standard errors are similar and the qualitative results across the two analysis are identical: Predictors SES, FEMALE, and PD have negative coefficient estimates, which are close to zero for PD for small values of τ . The strength of the absolute effects relative to the standard errors is highest for SES and smaller for FEMALE. For the latter predictor, the relative strength of the absolute effect decreases with increasing τ . For PD, the relative strength of the effect increases with increasing τ until $\tau = 0.5$ and slightly decreases for larger τ . Positive coefficient estimates are obtained for all other predictors. The largest coefficient estimates relative to the standard errors are obtained for DI, ID, and OD. For AG and AGE, coefficient estimates are generally smaller or only slightly larger than the standard errors. This confirms the risk factors for CP identified by [Haupt et al. \(2014\)](#) for the dataset imputed by QR-based MI: Low SES, DI, ID, and OD – while the limitations of the cross-sectional analysis and the bidirectional design of the analysis noted by [Haupt et al. \(2014\)](#) remain unchanged.

Additionally, we reproduced the results of the prediction-centered modelling approach in [Fritsch et al. \(2019\)](#). The paper models the conditional expectation of the response CP with multiple linear regression, regression trees and random forests.⁶ An

6) For details on the methods, see [Fritsch et al. \(2019\)](#) and the references given there.

Table 3

Parameter Estimates and Heteroscedasticity-Robust Standard Errors of Linear τ -Quantile Regression (QR) Model Based on QR-Based MI, and Comparisons With Complete Case Results of Haupt et al. (2014)

Parameter	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$
MI results							
Intercept	-0.804 (0.056)	-0.532 (0.055)	-0.290 (0.055)	-0.076 (0.055)	0.157 (0.056)	0.417 (0.06)	0.711 (0.062)
AGE	0.014 (0.06)	0.032 (0.062)	0.041 (0.061)	0.025 (0.059)	0.065 (0.066)	0.076 (0.068)	0.085 (0.068)
SES	-0.137 (0.065)	-0.207 (0.062)	-0.239 (0.061)	-0.228 (0.058)	-0.220 (0.061)	-0.233 (0.065)	-0.249 (0.07)
FEMALE	-0.120 (0.057)	-0.143 (0.058)	-0.125 (0.058)	-0.129 (0.058)	-0.111 (0.058)	-0.115 (0.062)	-0.071 (0.064)
AG	0.061 (0.055)	0.059 (0.057)	0.051 (0.056)	0.044 (0.06)	0.037 (0.059)	0.051 (0.063)	0.066 (0.062)
PD	-0.017 (0.055)	-0.031 (0.057)	-0.103 (0.056)	-0.117 (0.06)	-0.153 (0.059)	-0.148 (0.063)	-0.127 (0.062)
DI	0.169 (0.068)	0.202 (0.07)	0.273 (0.07)	0.297 (0.067)	0.326 (0.066)	0.353 (0.078)	0.351 (0.073)
ID	0.155 (0.076)	0.164 (0.076)	0.156 (0.072)	0.154 (0.075)	0.173 (0.077)	0.172 (0.087)	0.191 (0.089)
OD	0.055 (0.055)	0.095 (0.057)	0.135 (0.057)	0.168 (0.058)	0.198 (0.059)	0.199 (0.065)	0.187 (0.066)
Deviations from complete case results							
Intercept	-0.009	-0.009	-0.013	-0.011	0.007	0.029	-0.008
AGE	0.002	-0.008	-0.008	-0.018	0.054	0.065	0.022
SES	0.017	0.005	0.001	0.021	0.008	-0.005	-0.022
FEMALE	0.009	0.006	0.017	-0.005	0.013	-0.002	0.003
AG	0.001	-0.009	-0.014	-0.015	-0.016	-0.016	0.001
PD	0.018	0.034	0.008	-0.006	0.005	-0.006	-0.014
DI	-0.026	-0.032	-0.009	0.012	-0.020	-0.001	0.010
ID	0.017	0.021	0.004	0.012	0.007	-0.027	-0.008
OD	-0.007	-0.010	-0.006	-0.008	-0.007	-0.014	-0.012

Note. The top part of the table shows MI point estimates and heteroscedasticity-robust standard errors (in parentheses). All model variables were standardized. The bottom part of the table displays deviations from complete case results reported in Haupt et al. (2014) reproduced based on standardized variables.

out-of-sample evaluation of the three approaches is carried out based on root mean square error, mean square error, and mean absolute error. The paper reports superior predictive performance of random forest-based modelling. We reproduced the analysis

with the 100 datasets obtained from QR-based MI as follows: First we split the datasets into training and test data; for the test data, we set aside the exact same observations as in the original analysis across all imputed datasets. Second, we fitted a multiple linear regression model, a regression tree, and a random forest to each of the 100 imputed datasets. Note that for the regression trees and the random forests, we tuned the meta-parameters of the algorithm for each of the 100 datasets. Third, we predicted the response and evaluated the predictive performance of all methods by the three prediction error criteria for each of the 100 datasets. Overall, 1) there were only minor differences across the 100 datasets in predictive performance; 2) the predictive performance of modelling based on random forests and multiple linear regression improved slightly; 3) the predictive performance of modelling based on regression trees decreased slightly. We do not report more detailed results here, as there were only minor differences compared to the results reported in [Fritsch et al. \(2019\)](#).

Discussion

Note that in this application, the true distribution of observed and missing values, as well as the mechanism generating the missing values is unknown. This is typical for empirical datasets, where applied researchers are generally unaware of the ‘true’ impact of missing values on the obtained inferences. We compared estimates based on the complete cases – under the assumption of MCAR with estimates obtained by QR-based MI under the assumption of MAR. Although we cannot rule out MNAR mechanisms, our analyses provide empirical evidence that MAR processes were negligible and did not substantially affect the modelling results. Given that the MAR assumption holds – at least approximately – and that both, the substantive model and the imputation model are correctly specified (an untestable assumption in the analysis of empirical datasets), the reported MI inferences can be considered as unbiased.

General discussion

We have evaluated QR-based MI in a scenario that is typical for the social sciences and imputed incomplete scale scores that were obtained from ordinal Likert scale items. Typically, bias in statistical inferences depends on multiple factors: The amount of missing values in the dataset, the mechanism generating the missing values and the resulting missing value patterns, the inter-relationships in the given dataset, and the selected imputation method and model ([Raghunathan, 2016](#); [Schafer, 1997](#)). In our simulation, we held the percentage of missing values and the missing value pattern constant within a certain design so that differences in bias and CR can be attributed to the employed imputation methods. Overall, QR- and GAMLSS-based MI yielded the most accurate inferences in comparison to the established MI approaches normal model-based MI and

PMM. While the two latter methods yielded unbiased parameter estimates only in the center parts of the distributions, QR- and GAMLSS-based MI also performed well in the off-center regions. Only in the more extreme scenarios (percentage of missing values of 40%) some small biases regarding the lower quantiles were found. Based on our findings, we conclude that if researchers are interested in modelling effects beyond the conditional mean, imputation methods beyond standard normal model-based MI are promising. For example, when considering conditional quantiles, QR-based MI is particularly useful. Overall our findings support the recommendation that the imputation model and the analysis model should coincide.

We see the following directions for future research: First, as a sufficiently large sample is typically required by QR to estimate all model parameters in all quantiles of interest sufficiently well, poor performance of QR-based MI could be a byproduct of the relatively small employed sample size of $N = 675$. It might be instructive to replicate our simulation using a larger sample size to estimate all quantiles of interest of the conditional distribution of the response. Practical guidelines on the required sample size to obtain unbiased statistical inferences should be established by future research – ideally by taking the model complexity into account. Second, QR generally assumes a continuous response. The corporal punishment variable employed in this paper, for example was obtained by computing the mean of discrete 5-Point Likert scale items. Obviously, this does not yield a continuous variable, but creates point masses around certain values (see also [Figure 1](#)). Treating scale scores based on discrete items as (quasi-)continuous is a common practice in the social sciences. To our best knowledge, systematic evaluations of the behaviour of QR do not exist, when the response is not (perfectly) continuous. Our results suggest, that QR-based MI works sufficiently well for missing values that are based on 5-Point Likert scale items, given an adequately sized sample. In future research, the present study could be replicated and the results could be compared to settings where the response is continuous. Additionally, it may be helpful for practitioners to establish guidelines on the number of levels required for a discrete response (e.g., derived from Likert scale items), to consider the response in a QR model (quasi-)continuous. Third, QR-based MI should be evaluated more broadly and more complex settings, such as models including non-linear terms and interaction effects, should be considered. Finally, it will be of interest to applied researchers to compare QR-based MI with other imputation methods implemented in `ImputeRobust`. Further systematic evaluations and comparisons will help to single out settings in which to use QR-based MI, and when to use other semi-, or even fully parametric approaches.

Funding: The authors have no funding to report.

Acknowledgments: We thank two anonymous referees and the editor for helpful comments and suggestions. All remaining errors are ours.

Competing Interests: Jost Reinecke is one of the Editors-in-Chief of *Methodology*, but played no editorial role in this particular article or intervened in any form in the peer review process.

References

- Barnard, J., & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, 86(4), 948-955. <https://doi.org/10.1093/biomet/86.4.948>
- Bottai, M., & Zhen, H. (2013). Multiple imputation based on conditional quantile estimation. *Epidemiology, Biostatistics and Public Health*, 10(1), 1-18.
- Chodkiewicz, J., Talarowska, M., Miniszewska, J., Nawrocka, N., & Bilinski, P. (2020). Alcohol consumption reported during the COVID-19 pandemic: the initial stage. *International Journal of Environmental Research and Public Health*, 17(13), 4677-4688. <https://doi.org/10.3390/ijerph17134677>
- Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive missing-data strategies in modern missing-data procedures. *Psychological Methods*, 6(4), 330-351. <https://doi.org/10.1037/1082-989X.6.4.330>
- de Jong, R. (2012). *Robust multiple imputation* [Doctoral dissertation, University of Hamburg]. <http://ediss.sub.uni-hamburg.de/volltexte/2012/5971/>
- de Jong, R., van Buuren, S., & Spiess, M. (2016). Multiple imputation of predictor variables using generalized additive models. *Communication in Statistics – Simulation and Computation*, 45(3), 968-985. <https://doi.org/10.1080/03610918.2014.911894>
- Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*, 14(3), 275-299. <https://doi.org/10.1037/a0015825>
- Fritsch, M., Haupt, H., Lösel, F., & Stemmler, M. (2019). Regression trees and random forests as alternatives to classical regression modeling: Investigating the risk factors for corporal punishment. *Psychological Test and Assessment Modeling*, 61(4), 389-417. https://www.psychologie-aktuell.com/fileadmin/Redaktion/Journale/ptam-2019-4/02_Fritsch.pdf
- Geißler, R. (Ed.).(1994). *Soziale Schichtung und Lebenschance in Deutschland* [Social stratification and life chances in Germany]. Enke.
- Geraci, M. (2016a). Estimation of regression quantiles in complex surveys with data missing at random: An application to birthweight determinants. *Statistical Methods in Medical Research*, 25(4), 1393-1421. <https://doi.org/10.1177/0962280213484401>

- Geraci, M. (2016b). Qtools: A collection of models and tools for quantile inference. *The R Journal*, 8(2), 117-138. <https://doi.org/10.32614/RJ-2016-037>
- Geraci, M. (2017). Qtools: Utilities for quantiles [Computer software manual]. <https://cran.r-project.org/package=Qtools> (R package version 1.2)
- Geraci, M., & McLain, A. (2018). Multiple imputation for bounded variables. *Psychometrika*, 83(4), 919-940. <https://doi.org/10.1007/s11336-018-9616-y>
- Haupt, H., Lösel, F., & Stemmler, M. (2014). Quantile regression analysis and other alternatives to ordinary least squares regression: A Methodological Comparison on Corporal Punishment. *Methodology*, 10(3), 81-91. <https://doi.org/10.1027/1614-2241/a000077>
- Kleinke, K. (2017). Multiple imputation under violated distributional assumptions – A systematic evaluation of the assumed robustness of predictive mean matching. *Journal of Educational and Behavioral Statistics*, 42(4), 371-404. <https://doi.org/10.3102/1076998616687084>
- Kleinke, K. (2018). Multiple imputation by predictive mean matching when sample size is small. *Methodology*, 14(1), 3-15. <https://doi.org/10.1027/1614-2241/a000141>
- Kleinke, K., Reinecke, J., Salfrán, D., & Spiess, M. (2020). *Applied multiple imputation. Advantages, pitfalls, new developments and applications in R*. Springer Nature.
- Kleinke, K., Reinecke, J., & Weins, C. (2020). The development of delinquency during adolescence: a comparison of missing data techniques revisited. *Quality & Quantity*, 55, 877-895. <https://doi.org/10.1007/s11135-020-01030-5>
- Kneib, T. (2013). Beyond mean regression. *Statistical Modelling*, 13(4), 275-303. <https://doi.org/10.1177/1471082X13494159>
- Koenker, R. (2005). *Quantile regression*. Cambridge University Press.
- Koenker, R., & Hallock, K. F. (2001, December.). Quantile Regression. *Journal of Economic Perspectives*, 15(4), 143-156. <https://doi.org/https://doi.org/10.1257/jep.15.4.143>
- Little, R. J. A., & Rubin, D. B. (2020). *Statistical analysis with missing data*. John Wiley & Sons.
- Lösel, F., Beelmann, A., Jaurusch, S., Scherer, S., Stemmler, M., & Wallner, S. (2003). *Skalen zur Messung elterlichen Erziehungsverhaltens bei Vorschul- und Grundschulkindern. Die deutsche Version des Alabama Parenting Questionnaires (APQ)* [unpublished manuscript]. University of Erlangen-Nürnberg, Department of Psychology.
- Lösel, F., Stemmler, M., Jaurusch, S., & Beelmann, A. (2009). Universal prevention of antisocial development: Short- and long-term effects of a child- and parent-oriented program. *Monatsschrift für Kriminologie und Strafrechtsreform*, 92(2-3), 289-307. <https://doi.org/10.1515/mks-2009-922-314>
- Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In W. Barnett, J.L. Powell, & G. Tauchen (Eds.), *Nonparametric and semiparametric methods in econometrics and statistics* (pp. 357-384). Cambridge University Press.
- Raghunathan, T. (2016). *Missing data analysis in practice*. CRC Press.
- Raghunathan, T., Berglund, P.A., & Solenberger, P. W. (2018). *Multiple imputation in practice: With examples using IVEware*. CRC Press.

- Ranger, J., & Wolgast, A. (2019). Using response times as collateral information about latent traits in psychological tests. *Methodology*, 95(4), 185-196. <https://doi.org/10.1027/1614-2241/a000181>
- Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507-554. <https://doi.org/10.1111/j.1467-9876.2005.00510.x>
- Rigby, R. A., Stasinopoulos, D. M., & Voudouris, V. (2013). Discussion: A comparison of GAMLSS with quantile regression. *Statistical Modelling*, 13(4), 335-348. <https://doi.org/10.1177/1471082X13494316>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592. <https://doi.org/10.1093/biomet/63.3.581>
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434), 473-489. <https://doi.org/10.1080/01621459.1996.10476908>
- Salfrán, D., & Spiess, M. (2018). Generalized Additive Model Multiple Imputation by Chained Equations with Package ImputeRobust. *The R Journal*, 10(1), 61-72. <https://doi.org/10.32614/RJ-2018-014>
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman and Hall.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177. <https://doi.org/10.1037/1082-989X.7.2.147>
- Stasinopoulos, D. M., & Rigby, R. A. (2007). Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, 23(7), 1-46. <https://doi.org/10.18637/jss.v023.i07>
- Sugiyama, Y., Matsushima, M., & Yoshimoto, H. (2020). Association between alcohol consumption/ alcohol use disorders and patient complexity: a cross-sectional study. *BMJ Open*, 10(8), 1-10. <https://doi.org/10.1136/bmjopen-2019-034665>
- van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). CRC Press.
- van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12), 1049-1064. <https://doi.org/10.1080/10629360600810434>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67. <https://doi.org/10.18637/jss.v045.i03>
- Wayman, J. C. (2002). The utility of educational resilience for studying degree attainment in school dropouts. *The Journal of Educational Research*, 95(3), 167-178. <https://doi.org/10.1080/00220670209596587>
- Wilcox, R. R. (2016). *Understanding and applying basic statistical methods using R*. John Wiley & Sons.
- Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing* (4th ed.). Academic press.

Yu, L. M., Burton, A., & Rivero-Arias, O. (2007). Evaluation of software for multiple imputation of semi-continuous data. *Statistical Methods in Medical Research*, 16, 243-258.
<https://doi.org/10.1177/0962280206074464>



Methodology is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.