Bayesian imputation of time-varying covariates in linear mixed models - Online Supplement

Nicole S. Erler[1,2], Dimitris Rizopoulos[1,2], Vincent W.V. Jaddoe[2,3,4], Oscar H. Franco[2] and Emmanuel M.E.H. Lesaffre[1,5]

Web Appendix A

Details on the implied exo- or endogeneity

In this section, we provide details on the implications that the approaches introduced in Section 4 of the manuscript have for the exo- or endogeneity of the time-varying covariate $\boldsymbol{s}$. For $\boldsymbol{s}$ to be exogenous, the conditions from Section 3.3 of the manuscript need to be fulfilled, i.e.,

$$\begin{cases} p(y_i(t), f(H_i^s(t), t, \nu) \mid H_i^y(t^-), H_i^s(t^-), \boldsymbol{\theta}) &= p(y_i(t) \mid f(H_i^s(t), t, \nu), H_i^y(t^-), H_i^s(t^-), \boldsymbol{\theta}_1) \times \\ & \quad\; p(s_i(t) \mid H_i^y(t^-), H_i^s(t^-), \boldsymbol{\theta}_2) \\ p(s_i(t) \mid H_i^s(t^-), H_i^y(t^-), \boldsymbol{x}_i, \boldsymbol{\theta}) &= p(s_i(t) \mid H_i^s(t^-), \boldsymbol{x}_i, \boldsymbol{\theta}) \end{cases}$$

with $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$ and $\boldsymbol{\theta}_1 \perp\!\!\!\perp \boldsymbol{\theta}_2$, where $H_i^y(t^-)$ and $H_i^s(t^-)$ denote the history of $\boldsymbol{y}$ and $\boldsymbol{s}$, respectively, up to, but excluding measurements at time t. To abbreviate the notation, we will drop the index i in the following sections.

A.1 Exogeneity of the sequential approach with independent random effects

As can easily be seen, the first condition is fulfilled in the sequential approach with independent random effects, since the joint distribution of $\boldsymbol{y}$ and $\boldsymbol{s}$ is specified as the product of the conditional and the marginal distribution and the parameters of these distributions are usually specified to be a priori independent.

[1]Department of Biostatistics, Erasmus MC, Wytemaweg 80, 3015CN Rotterdam, The Netherlands
[2]Department of Epidemiology, Erasmus MC, Wytemaweg 80, 3015CN Rotterdam, The Netherlands
[3]Department of Pediatrics, Erasmus MC, Wytemaweg 80, 3015CN Rotterdam, The Netherlands
[4]Generation R Study Group, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[5]L-Biostat, KU Leuven, Kapucijnenvoer 35, 3000 Leuven, Belgium

Corresponding author:
Nicole S. Erler, Erasmus MC, PO Box 2040, 3000 CA Rotterdam, The Netherlands
Email: n.erler@erasmusmc.nl

To show that the second condition is fulfilled as well, several steps are necessary. Under the assumption that $s(t)$ is independent from $H^s(t^-)$ and $H^y(t)$ given the random effects we can write

$$\begin{aligned} p(s(t) \mid H^s(t^-), H^y(t), \boldsymbol{x}, \boldsymbol{\theta}) &= \int p(s(t), \boldsymbol{b} \mid H^s(t^-), H^y(t), \boldsymbol{x}, \boldsymbol{\theta})\, d\boldsymbol{b} \\ &= \int p(s(t) \mid \boldsymbol{b}, H^s(t^-), H^y(t), \boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{b} \mid H^s(t^-), H^y(t), \boldsymbol{x}, \boldsymbol{\theta})\, d\boldsymbol{b} \\ &= \int\int p(s(t) \mid \boldsymbol{b}^s, \boldsymbol{x}, \boldsymbol{\theta}) \\ &\qquad p(\boldsymbol{b}^s, \boldsymbol{b}^y \mid H^s(t^-), H^y(t), \boldsymbol{x}, \boldsymbol{\theta})\, d\boldsymbol{b}^s\, d\boldsymbol{b}^y. \end{aligned} \tag{A-1}$$

Using Bayes theorem, the conditional distribution of the random effects can be rewritten as

$$\begin{aligned} p(\boldsymbol{b}^s, \boldsymbol{b}^y \mid H^s(t^-), H^y(t), \boldsymbol{x}, \boldsymbol{\theta}) &= \frac{p(\boldsymbol{b}^s, \boldsymbol{b}^y, H^s(t^-), H^y(t) \mid \boldsymbol{x}, \boldsymbol{\theta})}{p(H^s(t^-), H^y(t) \mid \boldsymbol{x}, \boldsymbol{\theta})} \\ &= \frac{p(H^s(t^-), H^y(t) \mid \boldsymbol{b}^s, \boldsymbol{b}^y, \boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{b}^s, \boldsymbol{b}^y \mid \boldsymbol{\theta})}{p(H^s(t^-), H^y(t) \mid \boldsymbol{x}, \boldsymbol{\theta})}. \end{aligned} \tag{A-2}$$

With prior independence of $\boldsymbol{b}^y$ and $\boldsymbol{b}^s$, i.e., $p(\boldsymbol{b}^s, \boldsymbol{b}^y \mid \boldsymbol{\theta}) = p(\boldsymbol{b}^s \mid \boldsymbol{\theta})\, p(\boldsymbol{b}^y \mid \boldsymbol{\theta})$, and assuming conditional independence of $H^s(t^-)$ and $H^y(t)$ given the random effects, the denominator of (A-2) can be split in two factors,

$$\begin{aligned} p(H^s(t^-), H^y(t) \mid \boldsymbol{x}, \boldsymbol{\theta}) &= \int\int p(H^s(t^-), H^y(t) \mid \boldsymbol{b}^y, \boldsymbol{b}^s, \boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{b}^y, \boldsymbol{b}^s \mid \boldsymbol{\theta})\, d\boldsymbol{b}^y\, d\boldsymbol{b}^s \\ &= \int p(H^s(t^-) \mid \boldsymbol{b}^s, \boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{b}^s \mid \boldsymbol{\theta})\, d\boldsymbol{b}^s \\ &\quad \int p(H^y(t) \mid \boldsymbol{b}^y, \boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{b}^y \mid \boldsymbol{\theta})\, d\boldsymbol{b}^y \\ &= p(H^s(t^-) \mid \boldsymbol{x}, \boldsymbol{\theta})\, p(H^y(t) \mid \boldsymbol{x}, \boldsymbol{\theta}). \end{aligned} \tag{A-3}$$

Substituting (A-2) and (A-3) into (A-1), and recognizing that

$$p(s(t) \mid \boldsymbol{b}^s, \boldsymbol{x}, \boldsymbol{\theta})\, p(H^s(t^-) \mid \boldsymbol{b}^s, \boldsymbol{x}, \boldsymbol{\theta}) = p(H^s(t) \mid \boldsymbol{b}^s, \boldsymbol{x}, \boldsymbol{\theta}),$$

allows us to factorize the integrand from (A-1) so that each factor only depends on either $\boldsymbol{b}^y$ or $\boldsymbol{b}^s$. The two integrals can then be solved separately and Equation (A-1) can be simplified as

$$\begin{aligned} p(s(t) \mid H^s(t^-), H^y(t), \boldsymbol{x}, \boldsymbol{\theta}) &= \frac{\int p(H^s(t) \mid \boldsymbol{b}^s, \boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{b}^s \mid \boldsymbol{\theta})\, d\boldsymbol{b}^s}{p(H^s(t^-) \mid \boldsymbol{x}, \boldsymbol{\theta})} \\ &\quad \frac{\int p(H^y(t) \mid \boldsymbol{b}^y, \boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{b}^y \mid \boldsymbol{\theta})\, d\boldsymbol{b}^y}{p(H^y(t) \mid \boldsymbol{x}, \boldsymbol{\theta})} \\ &= \frac{p(H^s(t) \mid \boldsymbol{x}, \boldsymbol{\theta})}{p(H^s(t^-) \mid \boldsymbol{x}, \boldsymbol{\theta})}\, \frac{p(H^y(t) \mid \boldsymbol{x}, \boldsymbol{\theta})}{p(H^y(t) \mid \boldsymbol{x}, \boldsymbol{\theta})} \\ &= p(s(t) \mid H^s(t^-), \boldsymbol{x}, \boldsymbol{\theta})\, \frac{p(H^s(t^-) \mid \boldsymbol{x}, \boldsymbol{\theta})}{p(H^s(t^-) \mid \boldsymbol{x}, \boldsymbol{\theta})} \\ &= p(s(t) \mid H^s(t^-), \boldsymbol{x}, \boldsymbol{\theta}), \end{aligned}$$

which shows that for the sequential approach with independent random effects also the second condition for exogeneity is fulfilled.

A.2 Endogeneity of the sequential approach with correlated random effects

Prior independence of the random effects of the models for $\boldsymbol{y}$ and $\boldsymbol{s}$ is a crucial assumption for exogeneity. When $p(\boldsymbol{b}^s, \boldsymbol{b}^y \mid \boldsymbol{\theta}) \neq p(\boldsymbol{b}^s \mid \boldsymbol{\theta})\, p(\boldsymbol{b}^y \mid \boldsymbol{\theta})$, neither the numerator nor the denominator in (A-2) can be factorized into independent factors and Equation (A-1) can not be simplified as in the case with independent random effects. Either of the two possible factorizations,

$$p(\boldsymbol{b}^s, \boldsymbol{b}^y \mid \boldsymbol{\theta}) = p(\boldsymbol{b}^s \mid \boldsymbol{b}^y, \boldsymbol{\theta})\, p(\boldsymbol{b}^y \mid \boldsymbol{\theta}) \qquad \text{or} \qquad p(\boldsymbol{b}^s, \boldsymbol{b}^y \mid \boldsymbol{\theta}) = p(\boldsymbol{b}^y \mid \boldsymbol{b}^s, \boldsymbol{\theta})\, p(\boldsymbol{b}^s \mid \boldsymbol{\theta}),$$

violate the requirements of exogeneity.

Using the former factorization, Equation (A-1) leads to

$$\begin{aligned} p(s(t) \mid H^s(t^-), H^y(t), \boldsymbol{x}, \boldsymbol{\theta}) &= \dots \\ &= \frac{1}{p(H^s(t^-), H^y(t) \mid \boldsymbol{x}, \boldsymbol{\theta})} \\ &\quad \int \int p(H^s(t) \mid \boldsymbol{b}^s, \boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{b}^s \mid \boldsymbol{b}^y, \boldsymbol{\theta})\, d\boldsymbol{b}^s \\ &\quad\quad p(H^y(t) \mid \boldsymbol{b}^y, \boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{b}^y \mid \boldsymbol{\theta})\, d\boldsymbol{b}^y, \end{aligned}$$

where $\boldsymbol{s}$ is conditioned on $\boldsymbol{b}^y$ which, since the random effects depend on all observations of $\boldsymbol{y}$, implies that $\boldsymbol{s}$ is not independent of the history of $\boldsymbol{y}$, thus violating condition 2).

Using the latter factorization leads to

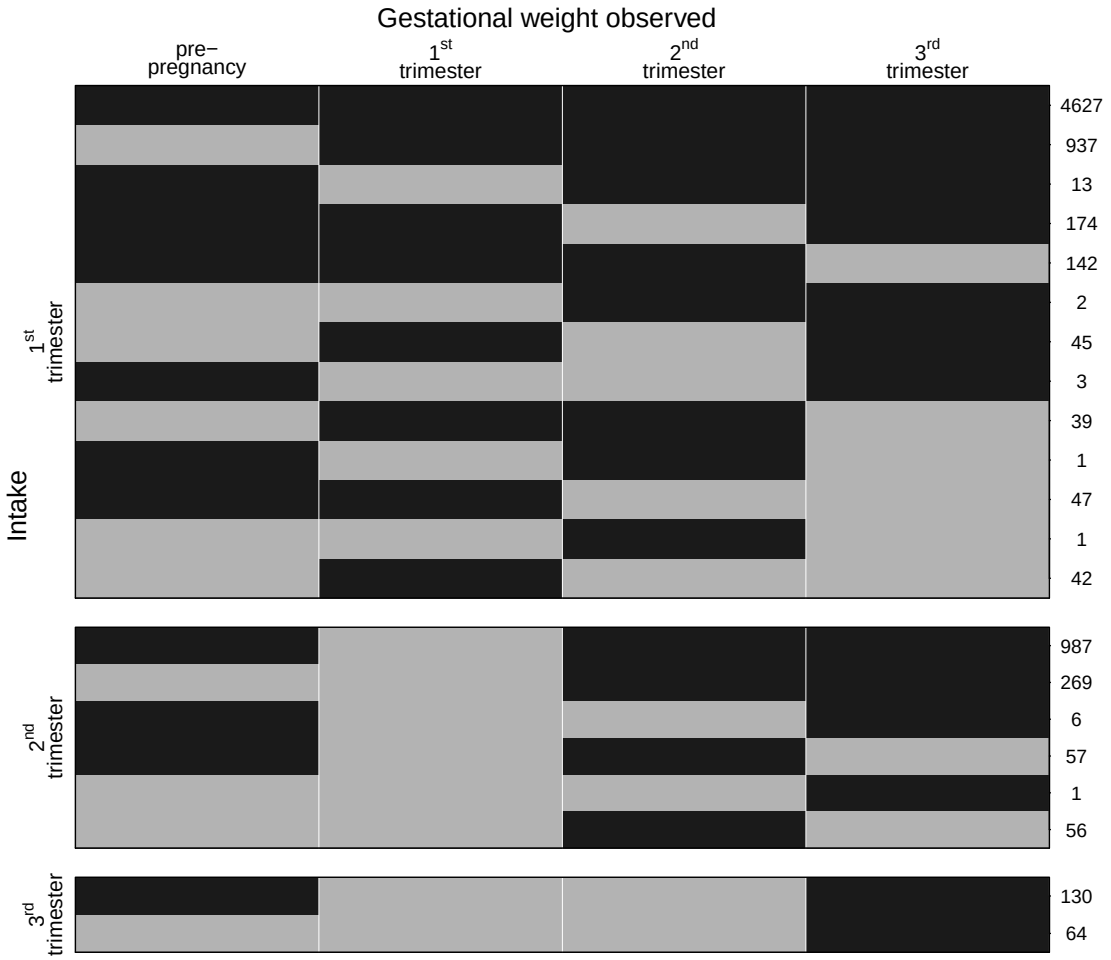
$$\begin{aligned} p(s(t) \mid H^s(t^-), H^y(t), \boldsymbol{x}, \boldsymbol{\theta}) &= \ldots \\ &= \frac{1}{p(H^s(t^-), H^y(t) \mid \boldsymbol{x}, \boldsymbol{\theta})} \\ &\quad \int \int p(H^y(t) \mid \boldsymbol{b}^y, \boldsymbol{x}, \boldsymbol{\theta})\; p(\boldsymbol{b}^y \mid \boldsymbol{b}^s, \boldsymbol{\theta})\; d\boldsymbol{b}^y \\ &\quad\quad p(H^s(t) \mid \boldsymbol{b}^s, \boldsymbol{x}, \boldsymbol{\theta})\; p(\boldsymbol{b}^s \mid \boldsymbol{\theta})\; d\boldsymbol{b}^s, \end{aligned}$$

where $\boldsymbol{y}$ depends on all observed values of $\boldsymbol{s}$ via $\boldsymbol{b}^s$, i.e. also on future values of $\boldsymbol{s}$, which also is at conflict with the exogeneity assumption. The sequential model with correlated random effects hence implies that $\boldsymbol{s}$ is endogenous.

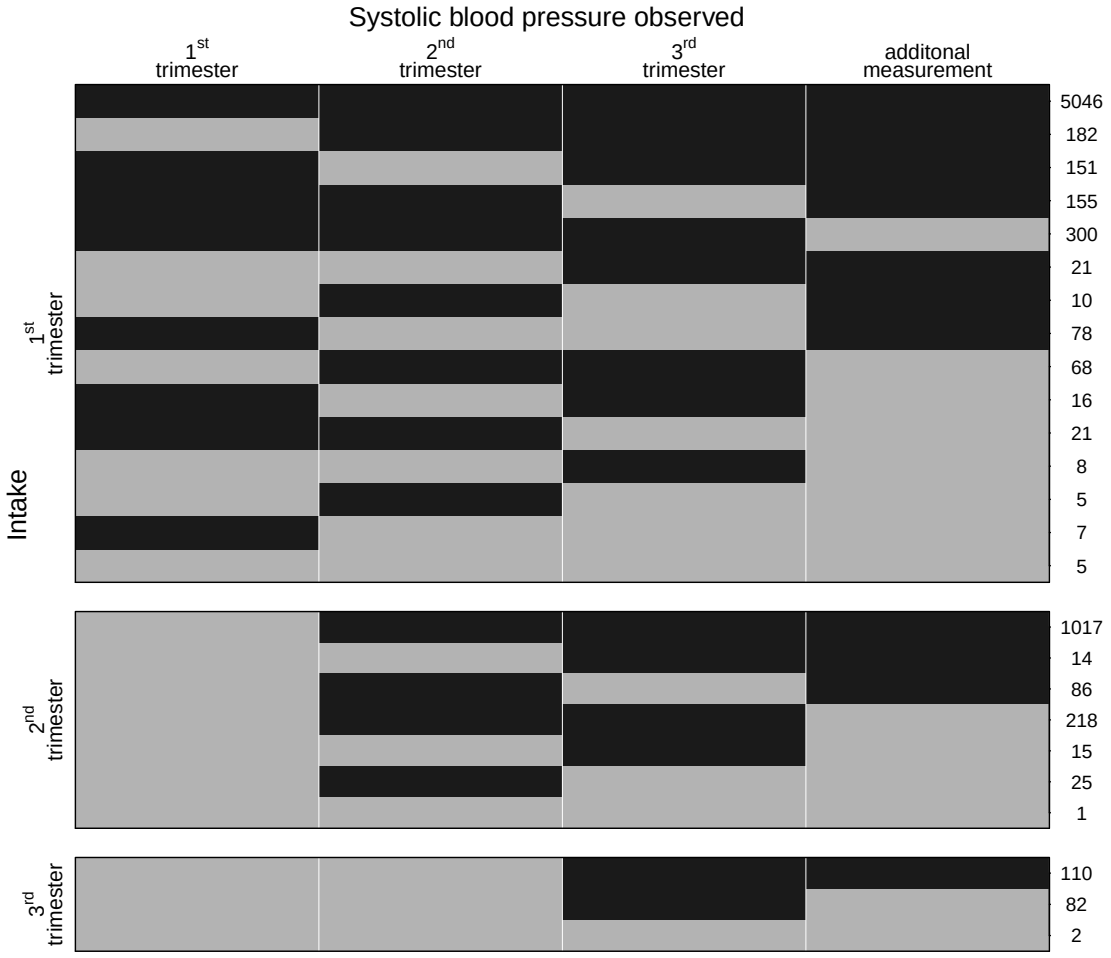
Web Appendix B

Analysis of the Generation R data

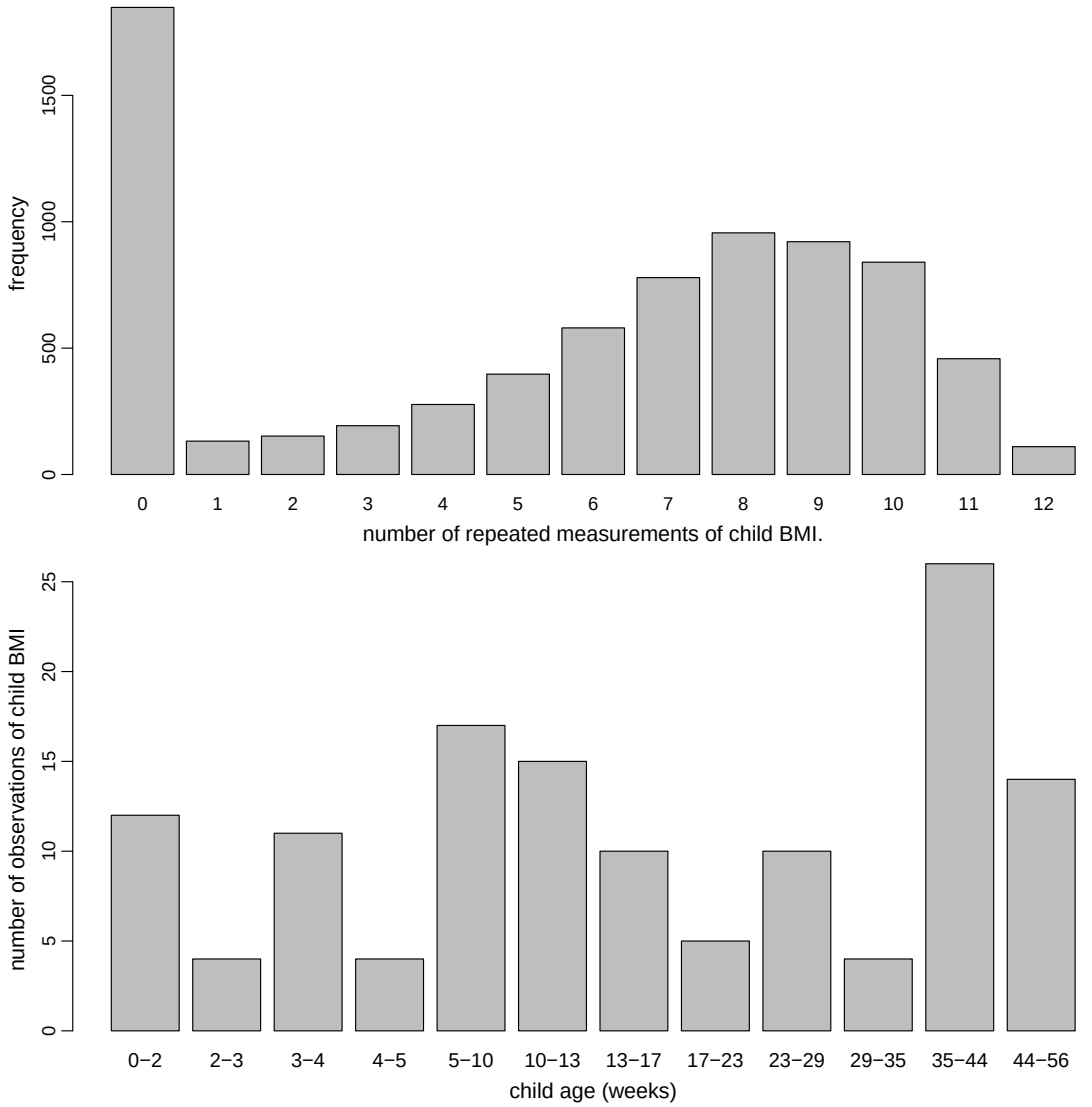
B.1 Missing data patterns



Web Figure 1. Missing data pattern for gestational weight. Dark color depicts observed values, light color missing values.



Web Figure 2. Missing data pattern for systolic blood pressure. Dark color depicts observed values, light color missing values.

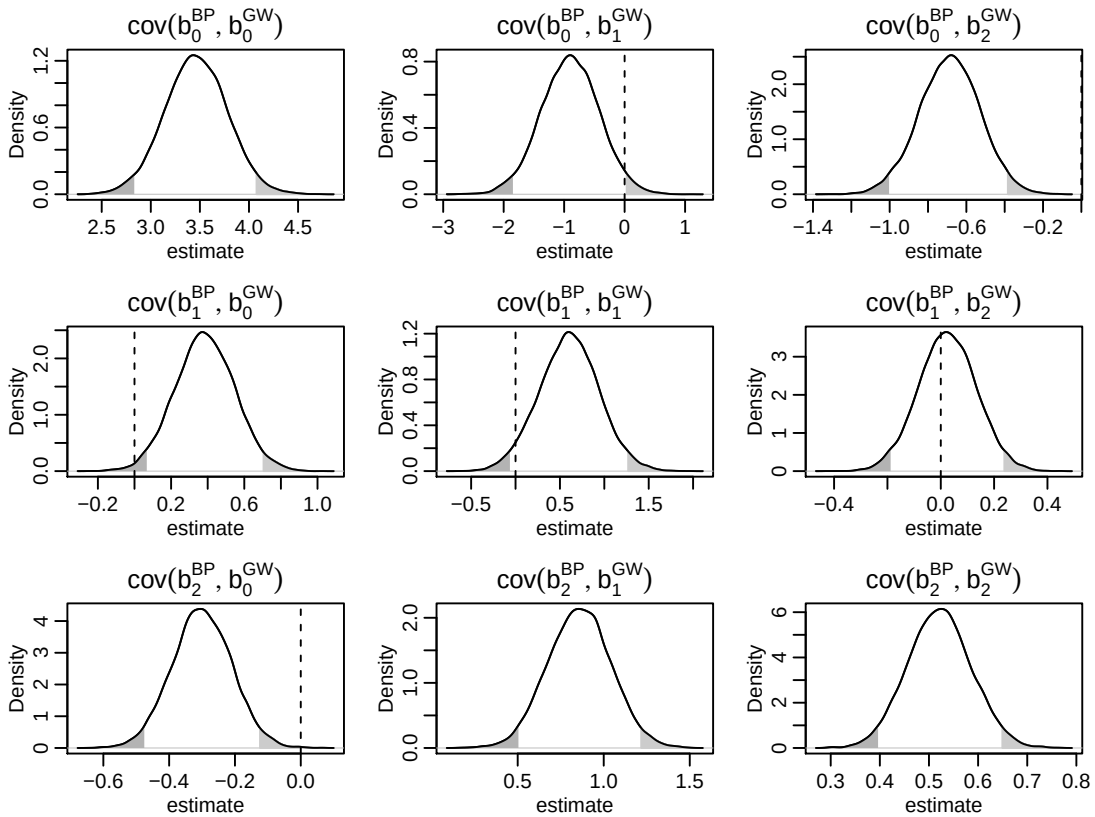


Web Figure 3. Distribution of the number of repeated BMI measurements per child (top) and number of observed values of child BMI per age category.

B.2 Prior distributions

$$
\begin{aligned}
\beta_k, \gamma &\overset{iid}{\sim} N(0, 1000), \qquad k = 0, \ldots, 10 \\
\sigma^2_{\mathrm{BP}}, \sigma^2_{\mathrm{GW}} &\overset{iid}{\sim} \text{inv-Gamma}(0.01, 0.01) \\
\alpha_k &\sim N(0, 1000), \qquad k = 0, \ldots, 10 \\
\mathbf{D}_{\mathrm{BP}} &\sim \text{inv-Wishart}(R_{\mathrm{BP}}, 3) \\
\mathbf{D}_{\mathrm{GW}} &\sim \text{inv-Wishart}(R_{\mathrm{GW}}, 3) \\
(\mathrm{diag}(R_{\mathrm{BP}})^T, \mathrm{diag}(R_{\mathrm{GW}})^T) &\overset{iid}{\sim} \text{Gamma}(1, 0.001)
\end{aligned}
$$

B.3 Additional graphics



Web Figure 4. Posterior distributions of the covariance between the random effects $\boldsymbol{b}^{\mathrm{BP}}$ and $\boldsymbol{b}^{\mathrm{GW}}$ from the endogenous setting in the first motivating question from the Generation R data presented in Section 5.1 in the manuscript. The dashed vertical line marks zero, i.e., the implied covariance in the exogenous setting, the shaded areas mark values outside the 95% credible interval.

Web Appendix C

Simulation Study

To evaluate the performance of the two imputation approaches described in Section 4 of the manuscript with regards to mis-specification of the endo- or exogeneity of a time-varying covariate and the bias introduced by mis-specification of the functional form in a more controlled setting, we performed a simulation study in which we compared results from correctly specified models with those that are mis-specified, for data generated in a range of different scenarios and different missing mechanisms.

The simulation study was set up to

1. confirm that both approaches provide unbiased estimates when the models are correctly specified during imputation and analysis,
2. investigate how mis-specification of the endo- or exogeneity influences the results, and
3. to explore bias due to mis-specification of the functional form, specifically

    - the bias introduced during imputation due to the implied linearity assumption of the multivariate normal approach when the true functional form is non-linear, and
    - the bias introduced when the imputation model as well as the analysis model are mis-specified as linear.

C.1 Design

We simulated 200 datasets in each of six scenarios that differed in the endo-/exogeneity of the covariate, the functional form and the model (sequential or multivariate normal) that was used. Common in all scenarios was that ten repeated measurements of a normally distributed time-varying covariate and a (conditionally) normal outcome variable, with measurements at the same, unbalanced time points, were created. Web Table 1 gives an overview of the different simulation scenarios. Under the sequential approach data was generated with a linear or a quadratic relation between covariate and outcome, where the covariate was either exogenous or endogenous. The multivariate normal model always generates data with a linear relation between the outcome and an endogenous covariate, however, there we considered two scenarios with regards to the correlation of the error terms, where in one scenario the error terms of outcome and covariate were independent and in the other correlated.

	linear relation	**quadratic relation**
exogenous	• sequential approach *seq. (exo., lin.)*	• sequential approach *seq. (exo., qdr.)*
endogenous	• sequential approach *seq. (endo., lin.)* • multivariate normal approach – independent error terms *mvn. (indep. err.)* – correlated error terms *mvn. (corr. err.)*	• sequential approach *seq. (endo., lin.)*

Web Table 1. Overview of data generating models, with their abbreviation written in italic font.

The general model used for simulation from the sequential approach was

$$\begin{aligned} y_{ij} &= (\beta_{y0} + b^y_{i0}) + (\beta_{y1} + b^y_{i1})t_{ij} + \boldsymbol{\gamma} f(s_{ij}) + \varepsilon^y_{ij} \\ s_{ij} &= (\beta_{s0} + b^s_{i0}) + (\beta_{s1} + b^s_{i1})t_{ij} + \varepsilon^s_{ij} \end{aligned} \tag{A-4}$$

with

$$\begin{aligned} \varepsilon^y_{ij} &\sim N(0, \sigma^2_y) \\ \varepsilon^s_{ij} &\sim N(0, \sigma^2_s) \end{aligned}$$

and

$$\begin{bmatrix} (b^y_{i0}, b^y_{i1})^T \\ (b^s_{i0}, b^s_{i1})^T \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{D}_y & \boldsymbol{D}_{y,s} \\ \boldsymbol{D}_{y,s} & \boldsymbol{D}_s \end{bmatrix}\right),$$

where t_{ij} is from a uniform distribution on the interval $[0, 5]$. The specific values of all parameters can be found in Web Table 3 in C.3. The general model to simulate from a multivariate normal distribution was

$$\begin{aligned} y_{ij} &= (\tilde{\beta}_{y0} + \tilde{b}^y_{i0}) + (\tilde{\beta}_{y1} + \tilde{b}^y_{i1})t_{ij} + \tilde{\varepsilon}^y_{ij} \\ s_{ij} &= (\tilde{\beta}_{s0} + \tilde{b}^s_{i0}) + (\tilde{\beta}_{s1} + \tilde{b}^s_{i1})t_{ij} + \tilde{\varepsilon}^s_{ij} \end{aligned} \tag{A-5}$$

with

$$\begin{bmatrix} \tilde{\varepsilon}^y_{ij} \\ \tilde{\varepsilon}^s_{ij} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \tilde{\sigma}^2_y & \tilde{\sigma}_{y,s} \\ \tilde{\sigma}_{y,s} & \tilde{\sigma}^2_s \end{bmatrix}\right)$$

and

$$\begin{bmatrix} (\tilde{b}^y_{i0}, \tilde{b}^y_{i1})^T \\ (\tilde{b}^s_{i0}, \tilde{b}^s_{i1})^T \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \tilde{\boldsymbol{D}}_y & \tilde{\boldsymbol{D}}_{y,s} \\ \tilde{\boldsymbol{D}}_{y,s} & \tilde{\boldsymbol{D}}_s \end{bmatrix}\right).$$

With these models, the six different data scenarios were specified as

- *seq (exo., lin.)*: simulation from model (A-4), with $\boldsymbol{\gamma} f(s_{ij}) = \gamma_1 s_{ij}$ and $\boldsymbol{D}_{y,s} = \mathbf{0}$,
- *seq. (exo., qdr.)*: simulation from model (A-4), with $\boldsymbol{\gamma} f(s_{ij}) = \gamma_1 s_{ij} + \gamma_2 s_{ij}^2$ and $\boldsymbol{D}_{y,s} = \mathbf{0}$,
- *seq (endo., lin.)*: simulation from model (A-4), with $\boldsymbol{\gamma} f(s_{ij}) = \gamma_1 s_{ij}$ and $\boldsymbol{D}_{y,s} \neq 0$,
- *seq (endo., qdr.)*: simulation from model (A-4), with $\boldsymbol{\gamma} f(s_{ij}) = \gamma_1 s_{ij} + \gamma_2 s_{ij}^2$ and $\boldsymbol{D}_{y,s} \neq 0$,
- *mvn (indep. err.)*: simulation from model (A-5), with $\tilde{\boldsymbol{D}}_{y,s} \neq \mathbf{0}$ and $\tilde{\sigma}^2_{y,s} = 0$,
- *mvn (corr. err.)*: simulation from model (A-5), with $\tilde{\boldsymbol{D}}_{y,s} \neq \mathbf{0}$ and $\tilde{\sigma}^2_{y,s} \neq 0$.

Missing values were created in $\boldsymbol{s}$ according to two MAR mechanisms. In missingness scenario *MAR.1*, the probability of s_{ij} being missing depended on y_{ij} only, while in missingness scenario *MAR.2* this probability depended on y_{ij} as well as s_{ij-1}, specifically

$$\begin{aligned} &\textit{MAR.1:} \qquad \Pr(s_{ij} = \text{NA}) = \text{expit}(y_{ij} + \zeta_1), \\ &\textit{MAR.2:} \qquad \Pr(s_{ij} = \text{NA}) = \text{expit}(y_{ij} + \zeta_1 + \zeta_2 I(s_{ij-1} < \zeta_3)), \end{aligned}$$

where $\text{expit}(x) = \exp(x)/(1 + \exp(x))$ and $I(\cdot)$ is the identity function which is one if the statement is true and zero otherwise. The values for the parameters ζ were chosen so that approximately 40% of $\boldsymbol{s}$ were missing and can be found in Web Table 4 in Web Appendix C.3.

C.2 Analysis models

Each of the datasets was analysed using both approaches with different assumptions regarding the endo- or exogeneity of the covariate and the functional form, before values were deleted, and for both missing mechanisms. An overview of the analysis methods used under the assumptions of either exo- or endogeneity is given in Web Table 2. The complete data was analysed using function `lmer()` from the R-package **lme4**[1,2] as well as with the sequential approach with independent random effects, $\boldsymbol{D}_{y,s} = \mathbf{0}$, when the covariate was assumed to be exogenous, and with the sequential approach with correlated random effects, i.e. $\boldsymbol{D}_{y,s} \neq \mathbf{0}$, when it was assumed to be endogenous. The functional form was specified to be either linear or quadratic, depending on the current assumption. Incomplete data from both missing mechanisms was imputed and analyzed with the sequential approach, again with either independent or correlated random effects. The imputation was repeated with the multivariate normal approach, using the model with independent error terms as well as the model with correlated error terms, and imputed datasets were created by drawing two values that were at least 50 iterations apart from each of the posterior chains of the incomplete covariate. The resulting ten imputed datasets were analysed analogous to the analysis of the complete data. When `lmer()` was used, the coefficients from the ten corresponding analyses were pooled using Rubin's Rules[3] and when the sequential approach was used, the ten sets of posterior Markov chains were combined to calculate posterior summary measures.

Assumption	**compl. data**	**MAR.1 & MAR.2**
exogenous	• sequential • `lmer()`	• sequential • multivariate normal (indep. err.) + `lmer()` • multivariate normal (corr. err.) + `lmer()`
endogenous	• sequential	• sequential • multivariate normal (indep. err.) + sequential • multivariate normal (corr. err.) + sequential

Web Table 2. Overview of imputation and analysis methods used under the assumption of exo- or endogeneity, for the complete as well as incomplete data.

All Bayesian analyses used five parallel Markov chains and were implemented in **JAGS**[4], using the R-package **rjags**[5]. Chains were assumed to have converged when the Gelman-Rubin criterion[6] was not more than 1.1 for the parameters of interest, $(\beta_{y0}, \beta_{y1}, \boldsymbol{\gamma}^T)$, and not more than 1.25 for other parameters. The number of iterations in the posterior sample necessary to obtain sufficient precision of the posterior estimate was determined based on the sampling error of the chains for $(\beta_{y0}, \beta_{y1}, \boldsymbol{\gamma}^T)$, which was required to be less than 5% of the corresponding standard error.

C.3 Parameter values

	sequential model	multivariate normal model
coefficients in the model for $\boldsymbol{y}$	$\beta_{y0} = -1.00$ $\beta_{y1} = 0.80$ $\gamma_1 = 1.20$ $\gamma_2 = -0.35$	$\tilde{\beta}_{y0} = -1.00$ $\tilde{\beta}_{y1} = 0.80$
coefficients in the model for $\boldsymbol{s}$	$\beta_{s0} = 2.00$ $\beta_{s1} = -0.30$	$\tilde{\beta}_{s0} = 2.00$ $\tilde{\beta}_{s1} = -0.30$
(co)variances of the error terms	$\sigma_y^2 = 0.40$ $\sigma_s^2 = 0.30$	$\tilde{\sigma}_y^2 = 0.40$ $\tilde{\sigma}_s^2 = 0.30$ $\tilde{\sigma}_{y,s} = 0.20$
(co)variances of random effects	$\boldsymbol{D}_y = \begin{bmatrix} 1.413 & 0.166 \\ 0.166 & 0.106 \end{bmatrix}$ $\boldsymbol{D}_s = \begin{bmatrix} 1.500 & -0.202 \\ -0.202 & 0.165 \end{bmatrix}$ $\boldsymbol{D}_{y,s} = \begin{bmatrix} 0.022 & -0.015 \\ -0.015 & -0.083 \end{bmatrix}$	$\tilde{\boldsymbol{D}}_y = \begin{bmatrix} 1.413 & 0.166 \\ 0.166 & 0.106 \end{bmatrix}$ $\tilde{\boldsymbol{D}}_s = \begin{bmatrix} 1.500 & -0.202 \\ -0.202 & 0.165 \end{bmatrix}$ $\tilde{\boldsymbol{D}}_{y,s} = \begin{bmatrix} 0.022 & -0.015 \\ -0.015 & -0.083 \end{bmatrix}$

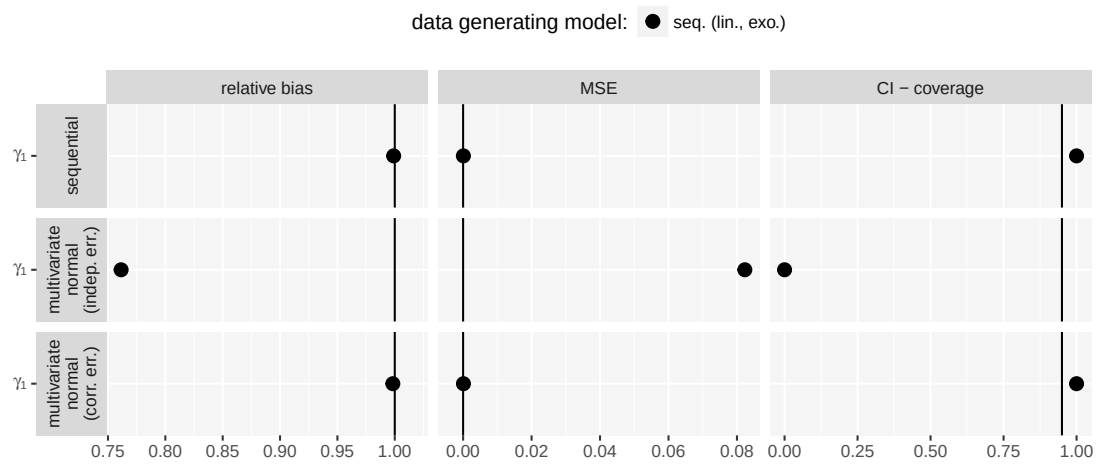
Web Table 3. Parameter values that were used in the data generating models.

	ζ_1	ζ_2	ζ_3
seq. (exo., lin.)	−3.20	−0.5	0.0
seq. (endo., lin.)	−3.20	−0.5	0.0
seq. (exo., qdr.)	−1.85	−0.7	−0.3
seq. (endo., qdr.)	−1.85	−0.7	−0.3
mvn. (indep. err.)	−1.60	−0.5	−0.3
mvn. (corr. err.)	−1.60	−0.5	−0.3

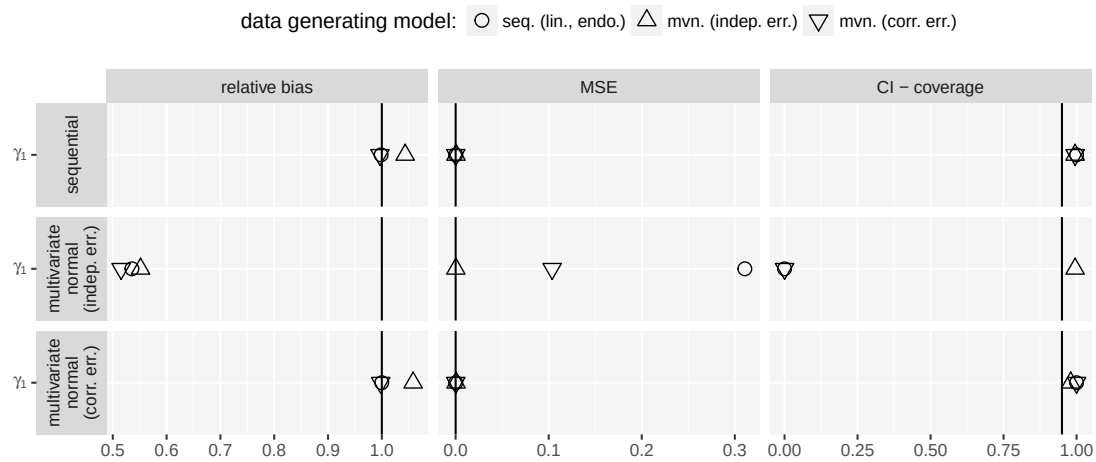
Web Table 4. Values for $\boldsymbol{\zeta}$ that were used to create missing values in each of the six data scenarios.

C.4 Detailed discussion of the results

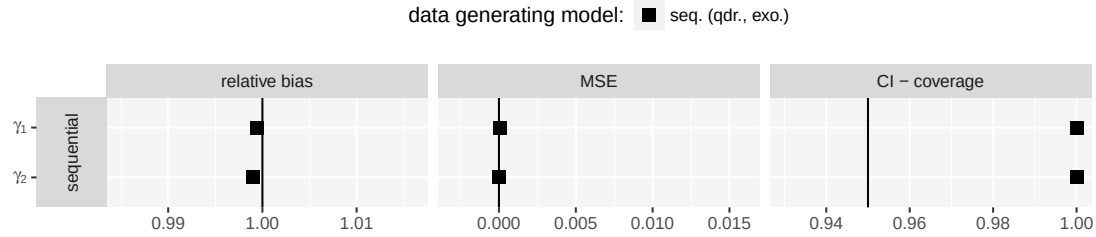
In this section, we present and discuss the results from the simulation study in detail. We adhere to the list of aims of the simulation study stated above and start with the comparison of the sequential and the multivariate normal approach when exo- or endogeneity and functional form are specified correctly. The relevant part of the results is shown in Web Figure 5 and grouped into four figures, according to whether the data generating model implied an exogenous or endogenous covariate and a linear or quadratic functional relation with the outcome. The six data generating models are indicated by different plotting symbols. The different approaches (sequential and multivariate normal, with independent or correlated error terms) are represented by rows, whereas the columns show three evaluation measures, summarized over all 200 simulated datasets. The relative bias was calculated as the median of the ratios of the estimate (REML estimate or posterior mean) from the analysis of the incomplete data and the estimate from the corresponding complete data analysis, the mean square error (MSE) as the average of the squared differences between the estimates from the missing and complete data analyses, and the CI-coverage as the proportion of CIs from the missing data analysis that covered the estimate from the corresponding complete data analysis. The desired value for each of these measures is indicated by the vertical line. Since results were similar under both missing mechanisms in most settings and to facilitate readability of the figures, we present results under *MAR.1* only. Furthermore, we only present results for the parameter related to the time-varying covariate, γ.



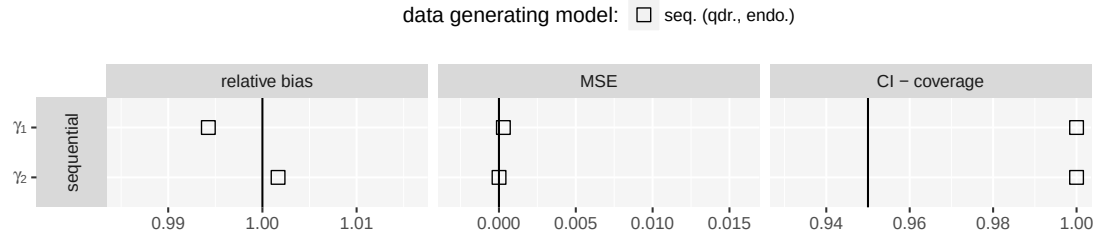
(a) Setting: linear relation, exogenous covariate



(b) Setting: linear relation, endogenous covariate



(c) Setting: quadratic relation, exogenous covariate



(d) Setting: quadratic relation, endogenous covariate

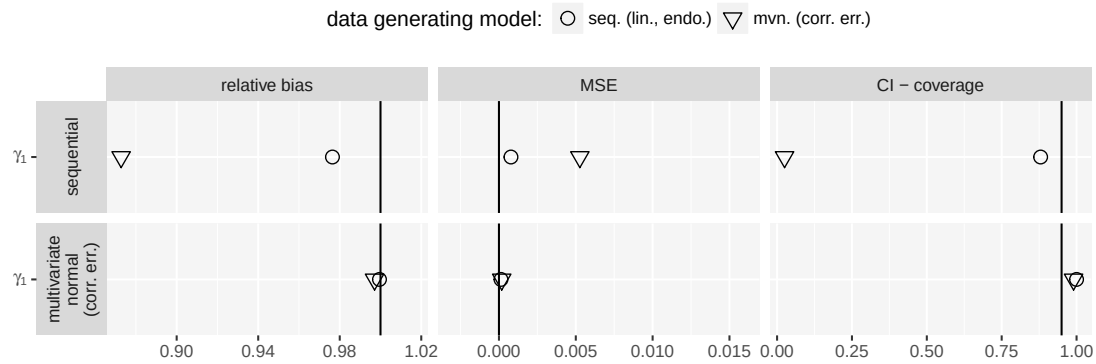
Web Figure 5. Relative bias, mean squared error (MSE), and proportion of CIs that covered the estimate from the analysis of the complete data, when imputation and analysis models were correctly specified with regards to exo- or endogeneity and functional form. The vertical lines mark the respective desired values.

Web Figure 5a shows that the sequential approach as well as the multivariate normal approach with correlated error terms were both unbiased and had CIs that covered the posterior mean from the complete data analysis in all 200 simulations, when data was generated with a linear

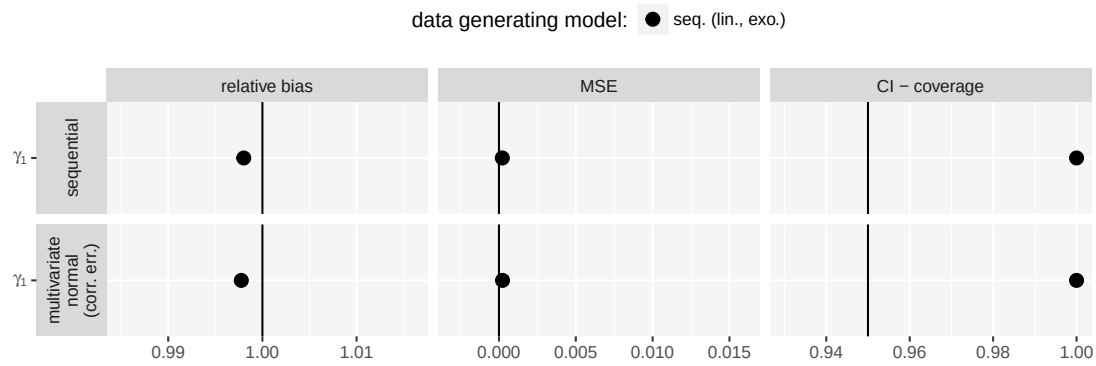
functional form and exogenous covariate. The multivariate normal approach with independent error terms, however, was clearly biased and had CIs that did not cover the estimate from the analysis of the complete data for any of the simulated datasets. It performed even worse in the endogenous setting (shown in Web Figure 5b), unless data was generated by this very model. The multivariate normal approach with correlated error terms and the sequential approach performed well in data that was simulated from the sequential model or the multivariate normal model with correlated error terms, but were biased for data that was generated by the multivariate normal model with independent error terms (relative bias approx. 1.05 under *MAR.1* and approx. 1.3 under *MAR.2*, for both approaches), however, coverage of the CIs was above 95%.

Corresponding results in the settings with a quadratic relation are shown in Web Figures 5c and 5d for the sequential approach, since a correct specification of a quadratic functional form during the imputation procedure is not possible in the multivariate normal approach. Also in these settings, the sequential approach performed well.

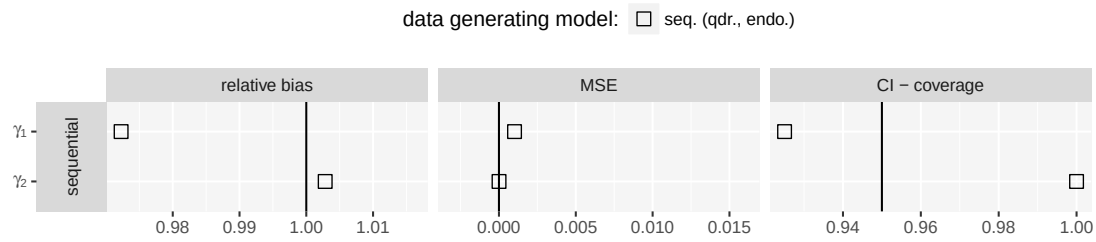
The results from the linear setting demonstrate that mis-specification of the correlation structure of the error terms may have great impact on the results and we will therefore exclude the multivariate normal model with independent error terms in the subsequent comparisons. Note that, since in our simulation ε_{ij}^s enters the model for y_{ij} through s_{ij}, the sequential approach also implies correlation between the error terms.



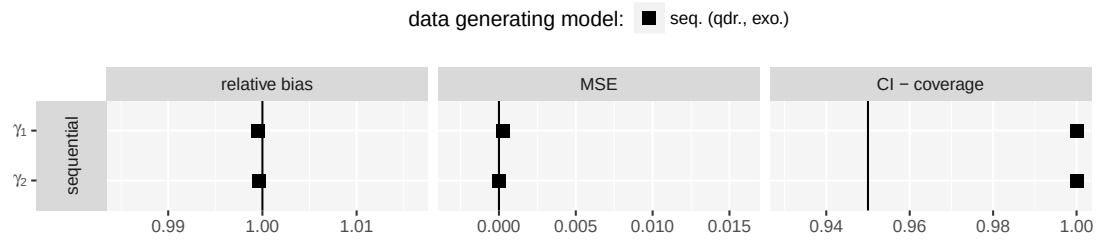
(a) Setting: linear relation, covariate mis-specified as exogenous



(b) Setting: linear relation, covariate mis-specified as endogenous



(c) Setting: quadratic relation, covariate mis-specified as exogenous



(d) Setting: quadratic relation, covariate mis-specified as endogenous

Web Figure 6. Relative bias, mean squared error (MSE), relative width of the 95% CI and proportion of CIs that covered the estimate from the analysis of the complete data, when imputation and analysis models were correctly specified with regards to the functional form but mis-specified with regards to the exo- or endogeneity of the time-varying covariate. The vertical lines mark the respective desired values.

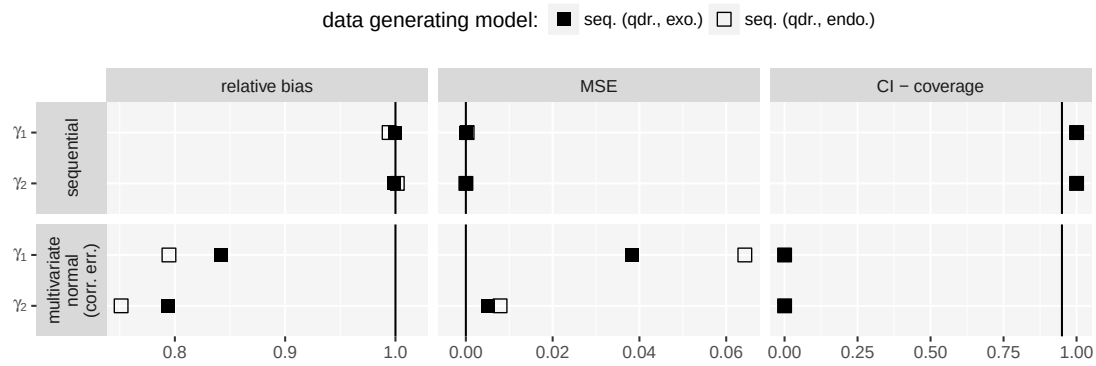
The second aim of this simulation study was to investigate the effect of mis-specification of the exo- or endogeneity during imputation and/or analysis. Web Figure 6 summarizes the

relevant results, again split into the different settings with regards to the assumption of exo- or endogeneity and the functional form. The setting in which a linear endogenous covariate is mis-specified as exogenous is shown in Figure 6a. There, imputation with the multivariate normal approach did not add any bias compared to the (in the same way mis-specified) analysis of the complete data, irrespective if the data were generated by the multivariate normal model or the (endogenous) sequential model. Imputation and analysis with the sequential approach with independent random effects introduced some additional bias: for data that was simulated from the sequential approach with correlated random effects, the relative bias was 0.98 and the coverage 88%. When data was simulated from the multivariate normal model, however, the relative bias worsened to 0.87 and the coverage to 2.5%. In the reversed case, where a linear, exogenous covariate was mis-specified as endogenous, estimates from both the sequential and the multivariate normal approach were unbiased (see Web Figure 6b). Simulations in the quadratic setting led to corresponding results (for the sequential approach), although the estimates of γ_2 were less biased than the estimates of γ_1, as can be seen in Web Figures 6c and 6d.

With regards to our third aim, the exploration of bias due to mis-specification of the functional form, we investigated two specific issues. First, we explored to what extend mis-specification of a quadratic functional form as linear during imputation with the multivariate normal approach, when the functional relation is correctly specified during the subsequent analysis, had an impact on the results. Our findings are summarized in Web Figure 7a and show that this type of mis-specification resulted in a relative bias of approximately 0.8 and very bad coverage of the CIs of both parameters, γ_1 and γ_2, which did not cover the true parameters in any of the simulations.

Second, we extended the mis-specification to the analysis model, i.e., assumed that the functional form was linear when the true structure was quadratic. In both approaches, this resulted in only little additional bias compared to the mis-specified analysis of the complete data, in the exogenous as well as in the endogenous setting, and had no influence on the coverage of the CIs (see Web Figure 7b).

Summarizing the results from our simulation study, we saw that mis-specification of the error terms in the multivariate normal approach as independent had the largest impact on the results, leading to estimates that were on average half the value of the estimate from the analysis of the complete data and CIs that had coverage of 0%. For mis-specification of an endogenous covariate as exogenous the most severe relative bias that was observed was 0.87 (and a corresponding coverage of the 95% CI of only 2.5%) in the setting where the sequential approach (with independent random effects) was used to impute data that was generated by the multivariate normal model. Imputation with the multivariate normal approach (with correlated error terms) in a setting where the functional form was correctly assumed to be quadratic had the second largest impact, with a relative bias of approximately 0.8, and resulted in CIs that only covered the parameter estimated in the analysis of the complete data in one single simulation for *MAR.2*. The bias that was added due to mis-specification of the functional form as linear during imputation and analysis as compared to the results from the analysis of the complete data under the same mis-specification, was small (between 0.96 and 1.05) and overall comparable between the multivariate normal and the sequential approach.



(a) Mis-specification of the functional form as linear during imputation with the multivariate normal approach, while correctly specifying it as quadratic in the analysis model in comparison with the correctly specified sequential model.

(b) Mis-specification of the functional form as linear in both, imputation and analysis model.

Web Figure 7. Relative bias, mean squared error (MSE), relative width of the 95% CI and proportion of CIs that covered the estimate from the analysis of the complete data, when imputation and analysis models were correctly specified with regards to exo- or endogeneity but mis-specified with regards to the functional form. The vertical lines mark the respective desired values.

C.5 JAGS syntax

We provide here an example of the **JAGS**[4] syntax that was used in the simulation study. The first part of syntax shows the sequential model assuming $\boldsymbol{s}$ is exogenous and has a linear relation with $\boldsymbol{y}$. We then show how to adapt that syntax for the endogenous setting. The second part contains example syntax that shows how the multivariate normal approach can be implemented in **JAGS** There, we present the syntax used in the setting with correlated error terms.

When implementing mixed models in **JAGS**, it is convenient to use hierarchical centering. This means that the fixed effects enter the linear predictor through the random effects, i.e. the

random effects are not centered around zero but around the fixed effects. The syntax differs from the formulas provided in the main manuscript to this regard. Note also that normal distributions have to be specified using the precision instead of the variance when using **JAGS**.

Data / Notation:

- `TN`: number of observations in the dataset
- `N`: number of individuals
- `priorR`: 4×4 diagonal matrix of NA values

```
model {
  for (j in 1:TN) {
    # linear mixed effects models for y and s
    y[j] ~ dnorm(mu.y[j], tau.y)
    s[j] ~ dnorm(mu.s[j], tau.s)

    # linear predictors
    # (with hierarchical centering specification for baseline effects)
    mu.y[j] <- inprod(b[subj[j], 1:2], Z[j, ]) + beta[3] * s[j]
    mu.s[j] <- inprod(b[subj[j], 3:4], Z[j, ])
  }

  # priors for the precision of y and s
  tau.y ~ dgamma(0.01, 0.01)
  tau.s ~ dgamma(0.01, 0.01)

  # specification of the random effects
  for (i in 1:N) {
    # random effects in the model for y
    b[i, 1:2] ~ dmnorm(mu.b[i, 1:2], inv.D.y[ , ])
    mu.b[i, 1] <- beta[1] # random intercept
    mu.b[i, 2] <- beta[2] # random slope

    # random effects in the model for s
    b[i, 3:4] ~ dmnorm(mu.b[i, 3:4], inv.D.s[ , ])
    mu.b[i, 3] <- alpha[1] # random intercept
    mu.b[i, 4] <- alpha[2] # random slope
  }

  # priors for the fixed effects
  for (k in 1:3) {
    beta[k] ~ dnorm(0, 0.001)
  }
  for (k in 1:2) {
    alpha[k] ~ dnorm(0, 0.001)
  }
```

```
  # priors for the precision of the random effects
  for (k in 1:4) {
    priorR[k, k] ~ dgamma(0.1, 0.01)
  }

  # precision and covariance matrix of the random effects in the model for y
  inv.D.y[1:2, 1:2] ~ dwish(priorR[1:2, 1:2], 2)
  D.y[1:2, 1:2] <- inverse(inv.D.y[ , ])

  # precision and covariance matrix of the random effects in the model for s
  inv.D.s[1:2, 1:2] ~ dwish(priorR[3:4, 3:4], 2)
  D.s[1:2, 1:2] <- inverse(inv.D.s[ , ])
}
```

C.5.1 Sequential approach In the endogenous setting, the specification of the random effects and related priors changes to:

```
  # specification of the random effects
  for (i in 1:N) {
    b[i, 1:4] ~ dmnorm(mu.b[i, 1:4], inv.D[ , ])

    # random effects in the model for y
    mu.b[i, 1] <- beta[1]
    mu.b[i, 2] <- beta[2]

    # random effects in the model for s
    mu.b[i, 3] <- alpha[1]
    mu.b[i, 4] <- alpha[2]
  }

  # priors for the precision of the random effects
  for(k in 1:4){
    priorR[k, k] ~ dgamma(0.1, 0.01)
  }

  # precision and covariance matrix of the random effects
  inv.D[1:4, 1:4] ~ dwish(priorR[1:4, 1:4], 4)
  D[1:4, 1:4] <- inverse(inv.D[ , ])
}
```

C.5.2 Multivariate normal approach A natural specification for the distribution of $\boldsymbol{y}$ and $\boldsymbol{s}$ in the setting with correlated random effects would be to specify $\{y(t), s(t)\}^T$ as multivariate normal. However, **JAGS** cannot sample from a multivariate normal distribution when only one component is missing. We therefore have to specify separate univariate normal distributions for $y(t)$ and $s(t)$ and additionally a multivariate normal distribution for the error terms $\{\varepsilon_i^y(t), \varepsilon_i^s(t)\}^T$. The precision in the univariate normal distributions for $y(t)$ and $s(t)$ is then set

to a large value to "force" the variation to be modeled in the multivariate normal distribution of the error terms rather than the two univariate normal distributions.

Data / Notation:

- `mu.eps` = c(0,0)
- `tau` = 10000

```
model {
  for (j in 1:TN) {
    # linear mixed effects models for y and s
    y[j] ~ dnorm(mu.y[j], tau)
    s[j] ~ dnorm(mu.s[j], tau)

    # linear predictors
    # (with hierarchical centering specification and "epsilon trick")
    mu.y[j] <- inprod(b[subj[j], 1:2], Z.y[j, ]) + eps[j, 1]
    mu.s[j] <- inprod(b[subj[j], 3:4], Z.s[j, ]) + eps[j, 2]

    eps[j, 1:2] ~ dmnorm(mu.eps[], inv.Sig[ , ])
  }

  # priors for the precision of y and s
  for (k in 1:2) {
    priorR.invSig[k, k] ~ dgamma(0.1, 0.01)
  }
  inv.Sig[1:2, 1:2] ~ dwish(priorR.invSig[1:2, 1:2], 2)
  Sig[1:2, 1:2] <- inverse(inv.Sig[ , ])


  # specification of the random effects
  for (i in 1:N) {
    # random effects in the model for y
    b[i, 1:4] ~ dmnorm(mu.b[i, 1:4], inv.D[ , ])
    mu.b[i, 1] <- beta[1, 1]
    mu.b[i, 2] <- beta[2, 1]

    # random effects in the model for s
    mu.b[i, 3] <- beta[1, 2]
    mu.b[i, 4] <- beta[2, 2]
  }

  # priors for the precision of the random effects
  for (k in 1:4) {
    priorR.invD[k, k] ~ dgamma(0.1, 0.01)
  }

  # precision and covariance matrix of the random effects
  inv.D[1:4, 1:4] ~ dwish(priorR.invD[1:4, 1:4], 4)
```

```
  D[1:4, 1:4] <- inverse(inv.D[ , ])

  # priors for the fixed effects
  for (k in 1:2) {
    beta[k, 1] ~ dnorm(0, 0.001)
    beta[k, 2] ~ dnorm(0, 0.001)
  }
}
```

References

1. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL `http://www.R-project.org/`.
2. Bates D, Mächler M, Bolker B et al. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 2015; 67(1): 1–48. DOI:10.18637/jss.v067.i01.
3. Rubin D. *Multiple Imputation for Nonresponse in Surveys*. Wiley Series in Probability and Statistics, Wiley, 1987.
4. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Hornik K, Leisch F and Zeileis A (eds.) *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. ISSN: 1609-395X.
5. Plummer M. *rjags: Bayesian graphical models using MCMC*, 2014. URL `http://CRAN.R-project.org/package=rjags`. R package version 3-13.
6. Gelman A, Meng XL and Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 1996; 6(4): 733–760.