A Model-Based Imputation Procedure for Multilevel Regression Models with Random Coefficients, Interaction Effects, and Non-linear Terms

Craig K. Enders
Han Du
Brian T. Keller
University of California – Los Angeles

Author Note

Correspondence concerning this article should be addressed to Craig K. Enders, UCLA, Department of Psychology, Franz Hall, 502 Portola Plaza, Los Angeles, CA 90095, cenders@psych.ucla.edu.

This work was supported by Institute of Educational Sciences award R305D150056

Online First Publication, July 1, 2019, *Psychological Methods*

Abstract

Despite the broad appeal of missing data handling approaches that assume a missing at random (MAR) mechanism (e.g., multiple imputation and maximum likelihood estimation), some very common analysis models in the behavioral science literature are known to cause bias-inducing problems for these approaches. Regression models with incomplete interactive or polynomial effects are a particularly important example because they are among the most common analyses in behavioral science research applications. In the context of single-level regression, fully Bayesian (model-based) imputation approaches have shown great promise with these popular analysis models. The purpose of this paper is to extend model-based imputation to multilevel models with up to three levels, including functionality for mixtures of categorical and continuous variables. Computer simulation results suggest that this new approach can be quite effective when applied to multilevel models with random coefficients and interaction effects. In most scenarios that we examined, imputation-based parameter estimates were quite accurate and tracked closely with those of the complete data. The new procedure is available in the Blimp software application for macOS, Windows, and Linux, and the paper includes a data analysis example illustrating its use.

A good deal of methodological literature supports missing data handling methods that assume a missing at random (MAR) mechanism whereby the probability of missingness is unrelated to an incomplete variable's scores after conditioning on the observed data (Little & Rubin, 2002; Rubin, 1976). Full information maximum likelihood estimation and multiple imputation are MAR-based methods that enjoy widespread use in behavioral science applications. When missing values are restricted to the outcome variable, maximum likelihood solutions abound in popular software packages (e.g., mixed modeling packages in SPSS, Stata, R, etc.) and are probably preferable because valid estimates are obtained by simply fitting the analysis model to the observed data (Little, 1992; von Hippel, 2007). For additive analysis models with incomplete explanatory variables (e.g., multiple regression, multilevel models with random intercepts), classic multiple imputation routines are similarly plentiful and effective (Schafer, 1997; Schafer & Yucel, 2002; van Buuren, 2012; Van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006), and maximum likelihood estimation can be implemented by specifying the predictors as random normal variables in structural equation modeling software such as M*plus* (Muthén & Muthén, 1998–2017).

Despite their widespread appeal and availability, classic missing data handling procedures are known to induce bias when applied to a broad class of single-level and multilevel regression models featuring interactive effects, polynomial terms, or random coefficients. For example, conventional imputation approaches typically invoke reverse regressions where the outcome variable predicts incomplete covariates. This specification is appropriate for additive models with normally distributed variables, but reverse regressions are often statistically incompatible with analyses that include interactive or non-linear terms because they may define an implausible distribution of missing values (Bartlett, Seaman, White, & Carpenter, 2015; Kim, Sugar, & Belin, 2015; Liu, Gelman, Hill, Su, & Kropko, 2014). This incompatibility is the source of biases reported in the literature (Bartlett et al., 2015; Enders, Baraldi, & Cham, 2014; Enders, Hayes, & Du, 2019; Grund, Ludtke, & Robitzsch, 2018; Seaman, Bartlett, & White, 2012; Zhang & Wang, 2017). Although our focus is multiple imputation, it is important to note that problems associated with non-linearities are also germane to maximum likelihood estimation (Enders et al., 2014; Yuan & Savalei, 2014).

A growing body of recent missing data research has on Bayesian approaches that tailor multiple imputation to a particular analysis model (Bartlett et al., 2015; Erler, Rizopoulos, Jaddoe, Franco, & Lesaffre, 2017; Erler et al., 2016; Goldstein, Carpenter, & Browne, 2014; Kim, Belin, & Sugar, 2018; Kim et al., 2015; Zhang & Wang, 2017). Roughly speaking, these methods specify a distribution for the explanatory variables (e.g., a normal distribution), then selectively choose imputations from this distribution that fit well (i.e., produce a high likelihood) when evaluated in an analysis model with interactive or non-linear terms. Simulation results from the aforementioned studies suggest that Bayesian approaches to imputation offer substantial improvement over older missing data handling methods for interactive and non-linear effects (e.g., just-another-variable imputation, passive imputation). Because the Bayesian estimation explicitly incorporates the substantive analysis model, we refer to this approach simply as model-based imputation.

Building on recent developments, the purpose of our paper is to outline and evaluate a model-based imputation procedure that correctly handles incomplete predictor variables in a wide range of single-level and multilevel regression models with non-linear effects (e.g., interactions, polynomial terms, random coefficients). Our approach is related to other Bayesian imputation methods (e.g., substantive model-compatible imputation the sequential Bayesian approach) described in the literature (Bartlett et al., 2015; Grund et al., 2018; Ibrahim, Chen, & Lipsitz, 2002; Kim et al., 2018; Kim et al., 2015; Zhang & Wang, 2017)[1]. In particular, we extend recent work by Erler and colleagues (Erler et al., 2017; Erler et al., 2016) and Goldstein et al. (2014)[2] to accommodate general missing data patterns in data sets with up to three levels, including functionality for incomplete categorical variables.

The structure of the paper is as follows. First, we discuss the issue of compatibility, as this helps clarify why conventional imputation approaches fail when applied to analysis models with interactive or non-linear effects. Second, we provide an overview of model-based imputation in the context of a single-level moderated regression analysis. Third, we outline an extension of model-based imputation that accommodates data structures with up to three levels, and we then show how to accommodate categorical variables in the procedure. Fourth, we report

[1] Model-based imputation for single-level regression models is available in the R packages 'smcfcs' (Bartlett & Keogh, 2018) and 'mdmb' (Robitzsch & Lüdke, 2018) and in dedicated Bayesian analysis packages such as OpenBUGS (Zhang & Wang, 2017) and JAGS , among others.
[2] The two-level imputation procedure from Goldstein et al. (2014) is available in the REALCOM (Carpenter, Goldstein, & Kenward, 2011) software and the R packages 'jomo' (Quartagno & Carpenter, 2018) and 'mdmb' (Robitzsch & Lüdke, 2018).

the results from three simulation studies that evaluate the proposed procedure. Finally, we use the Blimp application to apply model-based imputation to a real data analysis.

Compatibility of the Analysis and Imputation Models

An important concern with multiple imputation is whether the distributions implied by an imputation procedure match those induced by the analysis model. This issue, known as compatibility, has roots in mathematical statistics (Arnold, Castillo, & Sarabia, 1999, 2001; Arnold & Press, 1989) and is a topic of recent interest in the missing data literature (Bartlett et al., 2015; Carpenter & Kenward, 2013; Hughes et al., 2014; Liu et al., 2014).

To illustrate the concept of compatibility, consider a linear regression analysis model with an incomplete predictor *X*.

$$\begin{gathered} y_i = \beta_0 + \beta_1(x_i) + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \end{gathered} \tag{1}$$

A typical application of multiple imputation uses a reverse linear regression that to define a distribution of missing values, given the outcome variable.

$$\begin{gathered} x_i = \gamma_0 + \gamma_1(y_i) + e_i \\ x_{i(mis)} \sim N(\gamma_0 + \gamma_1(y_i), \sigma_e^2) \end{gathered} \tag{2}$$

Together, analysis and imputation form a set of conditional models, and these conditional models imply a set of conditional distributions – the analysis model assumes that *Y* is normal given *X*, and the imputation model assumes that *X* is normal given *Y*. Compatibility is concerned with whether a set of conditional models and their corresponding distributions relate to one another in a coherent way.

The formal definition of compatibility given by Arnold and colleagues (Arnold et al., 1999, 2001; Arnold & Press, 1989) and more recently by Liu et al. (2014)[3] and Bartlett et al.

---

[3] Definition 1 of Liu et al., 2014, p. 160) states the following: "A set of conditional models $\{g_j(x_j|x_{-j}, \theta_j): \theta_j \in \Theta_j, j = 1, \dots, p\}$ is said to be compatible if there exists a joint model $f\{(x|\theta): \theta \in \Theta\}$ and a collection of surjective maps $\{t_j: \Theta \to \Theta_j: j = 1, \dots, p\}$ such that for each j, $\theta_j \in \Theta_j$ and $\theta \in t_j^{-1}(\theta_j) = \{\theta: t_j(\theta) = \theta_j\}$, we have $g_j(x_j|x_{-j}, \theta_j) = f_j(x_j|x_{-j}, \theta)$. Otherwise, $\{g_j: j = 1, \dots, p\}$ is said to be incompatible."

(2015) is complex, so we sketch the basic ideas here and refer interested readers to these sources for additional details. Distilled at a basic level, the definition of compatibility says that a joint distribution exists, and the conditional distributions induced by this joint distribution are exactly the same as those we specify in our analyses. Returning to the previous example, the bivariate normal distribution has the property that its corresponding conditional distributions are also normal with constant variance. Thus, analysis and imputation models from Equations 1 and 2 are compatible because they are identical to those induced by a bivariate normal joint distribution. The practical implication of compatibility is that the imputation model should generate appropriate imputations for the analysis because the two models are functionally linked to a common joint distribution.

Next, consider a moderated regression model (Aiken & West, 1991; Cohen, Cohen, West, & Aiken, 2002), examples of which abound in the applied literature.

$$y_i = \beta_0 + \beta_1(x_{1i}) + \beta_2(x_{2i}) + \beta_3(x_{1i})(x_{2i}) + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2) \tag{3}$$

Further, assume that X_1 and the product are incomplete. The moderated regression analysis assumes the outcome variable is conditionally normal, given the covariates and their interaction. However, the interactive effect in the analysis model precludes the possibility that X_1 is normal when conditioning on Y (Arnold et al., 1999, 2001; Arnold & Press, 1989; Bartlett et al., 2015; Liu et al., 2014; Sarabia, Castillo, & Arnold, 2001; Seaman et al., 2012). As such, an imputation model based on reverse linear regression is incompatible with the moderated regression because it specifies a distribution of missing values that is implausible given the interaction term in the analysis model. This incompatibility is the source of the biases noted in the literature (Bartlett et al., 2015; Enders et al., 2014; Kim et al., 2015; Seaman et al., 2012; Zhang & Wang, 2017). Model-based imputation attempts to remedy this problem by sampling imputations from a set of compatible models.

Model-Based Imputation for Single-Level Regression

So that readers can better understand our multilevel imputation scheme, this section summarizes model-based imputation for a single-level moderated regression analysis such as that in Equation 3. The methodology we describe here is closely related to substantive model-compatible imputation and the sequential Bayesian approach from the literature (Bartlett et al., 2015; Goldstein et al., 2014; Grund et al., 2018; Ibrahim et al., 2002; Kim et al., 2018; Kim et al., 2015; Lüdke, Robitzsch, & West, 2019; Zhang & Wang, 2017)[4]. Note that imputation relies on a set of regression models, the parameters of which are obtained via an iterative Bayesian estimation algorithm (the Gibbs sampler). We discuss the estimation steps later in the manuscript, but for now assume that the necessary quantities have been estimated.

A variety of older imputation methods can be applied to the moderated regression analysis (Grund et al., 2018; Kim et al., 2018; Kim et al., 2015; Seaman et al., 2012; van Buuren, 2012; van Buuren et al., 2018; Vink & van Buuren, 2013; von Hippel, 2009), including passive imputation (e.g., impute X_1 conditional on Y and X_2, then compute the product term deterministically) and just-another-variable imputation (e.g., treat the product of X_1 and X_2 as variable to be imputed). The limitations of these approaches are well documented, so we refer interested readers to the literature for additional information (e.g., recent work by Kim and colleagues provides a comprehensive evaluation of several imputation strategies; Kim et al., 2018; Kim et al., 2015).

The idea behind model-based imputation is to parameterize the imputation problem as a set of compatible univariate distributions, one of which aligns with the analysis model. To frame the procedure that we adopt throughout the manuscript, consider the analysis model from Equation 3. We motivate our procedure by applying the conditional probability rule to factor the joint distribution of the analysis variables as

$$p(Y, X_1, X_2) = p(Y|X_1, X_2) \times p(X_1, X_2) \tag{4}$$

where $p(Y, X_1, X_2)$ is the joint distribution, $p(Y|X_1, X_2)$ is the distribution of Y induced by the analysis model (i.e., a normal distribution, conditional on the covariates and their interaction), and $p(X_1, X_2)$ is the joint distribution of the covariates. The above expression readily generalizes to a scenario with R covariates, in which case the factorization is $p(Y, X_1, \dots, X_R) = p(Y|X_1, \dots, X_R) \times p(X_1, \dots, X_R)$.

[4] The key differences among these procedures are how they partition the covariate distribution and whether they treat complete covariates as random variables. Our approach specifies all predictor variables as normally distributed variables, regardless of whether they are incomplete.

Although it is not the only way to do so, we assure that imputation models arising from Equation 4 are mutually compatible by specifying a multivariate normal distribution for the explanatory variables. That is, we do not allow non-linear relations among covariates. Although conceptually similar to what we are doing, the so-called "sequential" parameterization of the joint distribution (Erler et al., 2017; Erler et al., 2016; Ibrahim et al., 2002; Lüdke et al., 2019) is somewhat more flexible in that it can accommodate non-linear relations among covariates. In our view, specifying linear relations among the covariates is not a substantial practical limitation because most substantive researchers would not have a theoretical basis for specifying non-linear relations among variables that would otherwise have been treated as fixed in a complete-data analysis.

To impute X_1, we must derive its conditional distribution given the other analysis variables. Applying Bayes theorem gives the expression from Bartlett et al. (2015) and Kim et al. (2015), among others.

$$p(X_1|Y, X_2) \propto p(Y|X_1, X_2) \times p(X_1|X_2) \tag{5}$$

A benefit of assuming multivariate normality for the covariates is that the joint distribution of dependent variable and covariates must exist (a critical component of compatibility) when the analysis model is specified as a linear regression with normal errors and constant variance, as follows (Arnold et al., 1999, 2001; Arnold & Press, 1989; Liu et al., 2014).

$$\begin{gathered} x_{1i} = \gamma_0 + \gamma_1(x_{2i}) + e_i \\ e_i \sim N(0, \sigma_e^2) \end{gathered} \tag{6}$$

Because X_1 appears in both terms on the right side of Equation 5 – once as a covariate and once as an outcome – the posterior distribution of missing values has a complex form that depends on the product of two normal distributions.

$$p(X_1|Y, X_2) \propto N(\beta_0 + \beta_1(x_{1i}) + \beta_2(x_{2i}) + \beta_3(x_{1i})(x_{2i}), \sigma_\varepsilon^2) \times N(\gamma_0 + \gamma_1(x_{2i}), \sigma_e^2) \tag{7}$$

We can derive the exact distribution of missing values by comparing the product of the two normal kernels to the form of a univariate normal distribution and matching powers of X_1, which gives an expression algebraically equivalent to the one in Kim et al. (2015, p. 1878).

$$p(X_1|Y,X_2) = N\left(\frac{\sigma_{e_1}^2(\beta_1+\beta_3 x_{2i})(y_i-\beta_0-\beta_2 x_{2i})+\sigma_\varepsilon^2(\gamma_0+\gamma_1 x_{2i})}{\sigma_e^2(\beta_1+\beta_3 x_{2i})^2+\sigma_\varepsilon^2}, \frac{\sigma_\varepsilon^2\sigma_e^2}{\sigma_e^2(\beta_1+\beta_3 x_{2i})^2+\sigma_\varepsilon^2}\right). \quad (8)$$

Equation 8 shows that the mean and variance of the incomplete variable are non-linear functions of the other variables and two sets of model parameters. Comparing the correct conditional distribution to the misspecified one implied by a linear model with constant error variance (e.g., one similar to Equation 2) underscores the problem with standard imputation schemes.

To perform imputation, a Bayesian estimation sequence (discussed later) generates the coefficients and residual variances for the regression models (i.e., $\boldsymbol{\beta}$, σ_ε^2, $\boldsymbol{\gamma}$, and σ_e^2), after which the algorithm samples X_1 imputations from the distribution in Equation 8. The product is then computed by multiplying the resulting imputations by the corresponding observed values of X_2. Kim et al. (2015) show that computing the product in this manner is equivalent to drawing X_1 and its interaction term as a pair. In the general situation with more than one missing covariate, a comparable distribution can be derived for $p(X_2|Y,X_1)$ – all that is needed are the model parameters from the regression of X_2 on X_1. Finally, missing outcome scores do not require a special procedure, as an imputation model with the same form as the analysis model generates replacement values (i.e., replacement values are drawn from a normal distribution with mean and variance equal to $\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\beta_3 x_{1i}x_{2i}$ and σ_ε^2, respectively).

Model-Based Imputation for Multilevel Regression Analyses

Having illustrated model-based imputation in the context of a familiar single-level regression analysis, we now extend the procedure to multilevel regression models with missing values at any level. To keep the ensuing discussion as simple as possible, we describe the procedure for a two-level analysis, but the extension to three levels is straightforward. To provide an analytic context, consider a two-level random coefficient analysis with a pair of level-1 covariates and a single level-2 explanatory variable

$$
\begin{gathered}
y_{ij} = \beta_0 + \beta_1(x_{1ij}) + \beta_2(x_{2ij}) + \beta_3(x_{3j}) + b_{0j} + b_{1j}(x_{1ij}) + \varepsilon_{ij} \\
\begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} \sim MN(0, \mathbf{\Sigma}_b) \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)
\end{gathered} \tag{9}
$$

where the β's are fixed effects, b_{0j} and b_{1j} are random intercept and random slope residuals, respectively, for level-2 cluster *j*, and ε_{ij} is a within-cluster residual for observation *i* in cluster *j*. For simplicity, we do not center predictor variables in the analysis model, but the resulting imputations can be grand or group mean centered in the subsequent analysis (Enders & Tofighi, 2007; Kreft, de Leeuw, & Aiken, 1995).

This model, which features an interaction between a manifest variable and a random effect (i.e., latent slope variable), has been the focus of recent missing data research (Enders et al., 2019; Enders, Keller, & Levy, 2018; Grund, Ludtke, & Robitzsch, 2016; Grund et al., 2018; Kunkle & Kaizer, 2017; Lüdke, Robitzsch, & Grund, 2017). A standard method for imputing X_1 is to specify a "reverse random coefficient model" that features *Y* as a random slope predictor of X_1.

$$
\begin{gathered}
x_{1ij} = \gamma_0 + \gamma_1(y_{ij}) + \gamma_2(x_{2ij}) + \gamma_3(x_{3j}) + u_{0j} + u_{1j}(y_{ij}) + e_{ij} \\
\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim MN(0, \mathbf{\Sigma}_u) \quad e_{ij} \sim N(0, \sigma_e^2) \\
x_{1ij(mis)} \sim N(\gamma_0 + \gamma_1(y_{ij}) + \gamma_2(x_{2ij}) + \gamma_3(x_{3j}) + u_{0j} + u_{1j}(y_{ij}), \sigma_e^2)
\end{gathered} \tag{10}
$$

Consistent with the moderated regression analysis, this reverse regression model is incompatible with the analysis model because it incorrectly assumes that the conditional distribution of X_1 given *Y* is normal. As such, it gives imputations that are implausible given the random coefficient in the analysis model. Later in this section we show that the correct (compatible) conditional distribution is a complex non-linear function similar to that in Equation 8.

To begin, we specify a multivariate normal distribution for the explanatory variables because this ensures that we can derive a set of mutually compatible imputation models using pairs of regression models, e.g., $p(Y|X_1, \dots, X_R) \times p(X_r|X_1, \dots, X_{r-1}, X_{r+1}, \dots, X_R)$. Further, we divide the *R* covariates into a set of *P* predictors at level-1 and a second set of *Q* covariates at level-2. Note that auxiliary variables enter both conditional models in the same way as the substantive covariates, so we do not differentiate substantively-motivated predictor variables from auxiliary variables. The joint distribution for a two-level analysis model is

$$\boldsymbol{x}_i^{(1)} \sim MN(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_1) \quad \boldsymbol{x}_j^{(2)} \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2) \tag{11}$$

where $\boldsymbol{x}_i^{(1)}$ is a *P*-element vector of level-1 scores for observation *i*, $\boldsymbol{\mu}_j$ is the corresponding a *P*-element vector of latent cluster means (i.e., random intercepts) for group *j*, $\boldsymbol{x}_j^{(2)}$ is an *R*-element between-cluster score vector that includes the *P* latent group means in $\boldsymbol{\mu}_j$ and the *Q* level-2 covariate scores, $\boldsymbol{\mu}$ contains the *R* grand means, $\boldsymbol{\Sigma}_1$ is a *P* by *P* within-cluster covariance matrix, and $\boldsymbol{\Sigma}_2$ is an *R* by *R* between-cluster covariance matrix. For simplicity, we focus on two-level analyses, but the extension to three levels is straightforward. In this case, the covariate distribution becomes

$$\boldsymbol{x}_i^{(1)} \sim MN(\boldsymbol{\mu}_{jk}, \boldsymbol{\Sigma}_1) \quad \boldsymbol{x}_j^{(2)} \sim MN(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_2) \quad \boldsymbol{x}_k^{(3)} \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma}_3) \tag{12}$$

where *k* indexes the third level. The online supplemental document gives a description of the three-level procedure. Note that Equations 11 and 12 are slightly different from recent related work that treats complete covariates as fixed (Erler et al., 2017; Erler et al., 2016).

Returning to the random coefficient analysis from Equation 9, the predictors follow a multivariate normal distribution

$$\begin{pmatrix} x_{1ij} \\ x_{2ij} \end{pmatrix} \sim MN(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_1) \quad \begin{pmatrix} \mu_{1j} \\ \mu_{2j} \\ x_{3j} \end{pmatrix} \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2) \tag{13}$$

where $\boldsymbol{x}_i^{(1)} = (x_{1ij}, x_{2ij})$, $\boldsymbol{\mu}_j = (\mu_{1j}, \mu_{2j})$, $\boldsymbol{x}_j^{(2)} = (\mu_{1j}, \mu_{2j}, x_{3j})$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$, $\boldsymbol{\Sigma}_1$ is a 2 by 2 within-cluster covariance matrix, and $\boldsymbol{\Sigma}_2$ is a 3 by 3 between-cluster covariance matrix. Importantly, the cluster means in $\boldsymbol{\mu}_j$ can be represented as arithmetic averages of the level-1 scores, or they can be modeled as normally distributed latent variables (Lüdke, Marsh, Robitzsch, & Trautwein, 2011; Lüdke et al., 2008; Shin & Raudenbush, 2010). In this context, latent group means are just random intercepts from a multilevel model and should not be confused with latent means from a factor analysis model. Following the recommendation of Grund, Lüdke, and Robitzsch (2017), we use latent cluster means because these quantities readily accommodate unequal group sizes, whereas taking arithmetic average of level-1 scores assumes balanced data (Carpenter & Kenward, 2013; Grund et al., 2017; Resche-Rigon & White, 2018)[5]. The online supplement gives the full conditional distribution of the latent cluster means for our two-level example, and we point interested readers to Keller, Du, and Enders (2019) for a detailed treatment of fully conditional specification imputation with latent means.

Next, we must derive the conditional distribution of X_r given all other predictors, $X_{-r} = (X_1, \dots, X_{r-1}, X_{r+1}, \dots, X_R)$. Applying Bayes' theorem gives $p(X_r|Y, X_{-r}) \propto p(Y|X_r, X_{-r}) \times p(X_r|X_{-r})$, which is a more general expression for Equation 5. The multivariate normality assumption for the covariates again ensures that specifying each $p(X_r|X_{-r})$ as a linear regression with normal errors and constant variance yields mutually compatible imputation models (Arnold et al., 1999, 2001; Arnold & Press, 1989; Liu et al., 2014). Returning to the random coefficient analysis example, the within-cluster associations in $\mathbf{\Sigma}_1$ can equivalently be expressed as a pair of regression models

$$\begin{gathered} x_{1ij} = \mu_{1j} + \gamma_{11}(x_{2ij} - \mu_{2j}) + e_{1ij} \\ e_{1ij} \sim N(0, \sigma^2_{e_1}) \end{gathered} \tag{14}$$

$$\begin{gathered} x_{2ij} = \mu_{2j} + \gamma_{21}(x_{1ij} - \mu_{1j}) + e_{2ij} \\ e_{2ij} \sim N(0, \sigma^2_{e_2}) \end{gathered} \tag{15}$$

where the leading subscript on the slope coefficient indexes the outcome variable. Three points are worth noting. First, because predictor variables are centered at their latent group means, γ_{11} and γ_{21} are "pure" within-cluster regression slopes (Kreft et al., 1995; Raudenbush & Bryk, 2002). Second, the group means, μ_{1j} and μ_{2j}, are normally distributed latent variables that are

[5] The Blimp application described later can implement latent or manifest group means, with latent cluster means set by default. In our extensive test simulations, we have rarely observed meaningful differences between the two methods.

identical to random intercepts (i.e., random effects). Finally, although neither equation appears to condition on X_3, this variable does in fact influence X_1 and X_2 indirectly via their cluster means, μ_{1j} and μ_{2j}. The equations below illustrate this point.

Next, the between-cluster variances and covariances in $\mathbf{\Sigma}_2$ can be modeled with the following set of level-2 regression models.

$$\begin{gathered}\mu_{1j} = \mu_1 + \eta_{11}(\mu_{2j} - \mu_2) + \eta_{12}(x_{3j} - \mu_3) + \zeta_{1j} \\ \zeta_{1j} \sim N(0, \sigma^2_{\zeta_1})\end{gathered} \tag{16}$$

$$\begin{gathered}\mu_{2j} = \mu_2 + \eta_{21}(\mu_{1j} - \mu_1) + \eta_{22}(x_{3j} - \mu_3) + \zeta_{2j} \\ \zeta_{2j} \sim N(0, \sigma^2_{\zeta_2})\end{gathered} \tag{17}$$

$$\begin{gathered}x_{3j} = \mu_3 + \eta_{31}(\mu_{1j} - \mu_1) + \eta_{32}(\mu_{2j} - \mu_2) + \zeta_{3j} \\ \zeta_{3j} \sim N(0, \sigma^2_{\zeta_3})\end{gathered} \tag{18}$$

Two points are worth noting. First, we include regressions for the latent group means (i.e., random intercepts) because these quantities are integral to level-1 and level-2 imputation and must be estimated at every iteration of estimation. Second, the explanatory variables in each equation are again centered, such that the grand means function as fixed intercepts. Here again, our procedure is different from previous work (Erler et al., 2017; Erler et al., 2016; Goldstein et al., 2014) because we explicitly model the level-1 and level-2 parts of all predictors, treating the latter terms as normally distributed latent variables.

Having defined the necessary regressions, we can now derive the posterior distribution of the missing values in the same manner as we did for a single-level analysis. To illustrate, consider the random slope predictor X_1. As before, X_1's distribution has a complex form that depends on the product of two normal distributions (and two sets of model parameters)[6].

$$p(X_1|Y, X_2, X_3) \propto p(Y|X_1, X_2, X_3) \times p(X_1|X_2, X_3) \tag{19}$$

[6] The generic probability notation may not convey the fact that X_1 conditions on X_3 via its latent group mean in the between-cluster part of the model (see Equation 16). We could instead write the conditional distribution as $p(X_1|Y, X_2, X_3, X_{1(\mathrm{B})}, X_{2(\mathrm{B})}) \propto p(Y\ |X_1, X_2, X_3, X_{1(\mathrm{B})}, X_{2(\mathrm{B})}) \times p(X_1|X_2, X_3, X_{1(\mathrm{B})}, X_{2(\mathrm{B})})$, where $X_{1(\mathrm{B})}$ and $X_{2(\mathrm{B})}$ are the between-cluster parts of the covariate.

$$\times N\left(\mu_{1j} + \gamma_{11}\left(x_{2ij} - \mu_{2j}\right), \sigma_{e_1}^2\right)$$

Because x_{1ij} is conditionally normal given the other predictors, we can derive the exact distribution of the missing values by comparing the product of the two normal kernels to the form of a univariate normal distribution and matching powers of X_1, as follows.

$$x_{1ij(mis)} = N\left(\begin{array}{c}\frac{\sigma_{e_1}^2\left(\beta_1 + b_1 + \beta_4 x_{3j}\right)\left(y_i - \beta_0 - b_0 - \beta_2 x_{2ij} - \beta_3 x_{3j}\right) + \sigma_\varepsilon^2\left(\mu_{1j} + \left(x_{2,ij} + \mu_{2,j}\right)\right)}{\sigma_{e_1}^2\left(\beta_1 + b_1 + \beta_4 x_{3j}\right)^2 + \sigma_\varepsilon^2}, \\ \frac{\sigma_\varepsilon^2 \sigma_{e_1}^2}{\sigma_{e_1}^2\left(\beta_1 + b_1 + \beta_4 x_{3j}\right)^2 + \sigma_\varepsilon^2}\end{array}\right) \quad (20)$$

Examining the correct conditional distribution highlights the fact that the reverse random coefficient imputation model from Equation 10 is incompatible with the analysis model (e.g., because it assumes constant variance, whereas the correct variance is a non-linear function). Also, the previous expression highlights that missing values condition on all variables in the analysis, not just the covariates. For this reason, we would expect the procedure to accommodate a range of MAR processes that depend on the analysis variables (e.g., missingness induced by the outcome).

Analogous distributions can be derived for the other explanatory variables, although the form of each distribution will generally depend on the level at which a covariate is measured as well as its role in the analysis. For example, consider the level-2 predictor X_3. Because this variable is constant for all observations in level-2 cluster j, its conditional distribution features a product over all observations in that group.

$$\begin{aligned} p(X_3|Y, X_1, X_2) &\propto p(Y\,|X_1, X_2, X_3) \times p(X_3|X_1, X_2) \\ &\propto \prod_{i=1}^{n_j} N\left(\left(\beta_0 + b_{0j}\right) + \left(\beta_1 + b_{1j}\right)x_{1ij} + \beta_2\left(x_{2ij}\right) + \beta_3\left(x_{3j}\right), \sigma_\varepsilon^2\right) \quad (21) \\ &\times N\left(\mu_3 + \eta_{31}\left(\mu_{1j} - \mu_1\right) + \eta_{32}\left(\mu_{2j} - \mu_2\right), \sigma_{\zeta_3}^2\right) \end{aligned}$$

In practice, it is more straightforward to use a Metropolis sampling step (Hastings, 1970) to draw imputations from $p(Y|X_r, X_{-r}) \times p(X_r|X_{-r})$ because this approach can approximate a target distribution such as Equation 20 by simply evaluating candidate imputations in both normal likelihood functions. This eliminates the need to derive exact distributions for every unique application. As described in the next section, the Metropolis algorithm is one computational step in the Gibbs sampler that generates parameter values, random effects, and latent group means. Interested readers can find the technical details for the Metropolis sampler in the online supplemental material.

A brief sidebar about centering is warranted given its important role in multilevel analyses. We use latent group mean centering in the covariate models to partition explanatory variables into within- and between-cluster components, and we do not impose any structure on their covariance matrices (e.g., the association between $x_{1ij} - \mu_{1j}$ and $x_{2ij} - \mu_{2j}$ need not be the same as that between μ_{1j} and μ_{2j}). For simplicity, we did not center predictors in the analysis model. If the substantive goal seeks to disentangle within- and between-cluster influences of a level-1 covariate, latent group means can also be added to the Y part of the imputation model (e.g., μ_{1j} and μ_{2j} could be specified as level-2 predictors of Y) and the resulting imputations can then be centered at their grand or group means (Enders & Tofighi, 2007; Kreft et al., 1995).

Gibbs Sampler Algorithm for Bayesian Estimation

Model-based imputation requires parameter values and random effect estimates for the substantive model and a set of parameter values and latent group means (i.e., random intercepts) for each explanatory variable. We use an iterative Bayesian estimation algorithm known as the Gibbs sampler to generate these quantities at each computational cycle, and a final Metropolis within Gibbs step draws incomplete covariate scores from their target distributions. The Bayesian paradigm views the model parameters, random effects, latent group means, and missing values as random variables that have a joint distribution, and the Gibbs sampler estimates one quantity at a time, drawing values from a probability distribution that conditions on a prior distribution and the current values of all other variables. In the interest of space, we sketch the major algorithmic steps here and refer interested readers to the online supplemental material for a technical exposition of full conditional distributions (e.g., the form and parameters of each distribution, default and user-specifiable priors).

The full cadre of estimation steps for a two-level model with continuous variables is given below. The first five steps generate estimates for the substantive analysis model, and the remaining steps target explanatory variables. To simplify the presentation, we index the entire set of covariates as $r = 1, \ldots, R$, such that $r \leq P$ corresponds to either a level-1 observation or its corresponding level-2 group mean, and $r > P$ refers to a manifest level-2 variable. The computational steps for a single iteration t are as follows.

1. Draw regression coefficients in $\boldsymbol{\beta}^{(t)}$ from $p(\boldsymbol{\beta}|\cdot)$
2. Draw the residual variance $\sigma_{\varepsilon}^{2(t)}$ from $p(\sigma_{\varepsilon}^{2}|\cdot)$
3. Draw random effects $\boldsymbol{b}_{j}^{(t)}$ from $p(\boldsymbol{b}_{j}|\cdot)$
4. Draw the random effect covariance matrix $\boldsymbol{\Sigma}_{b}^{(t)}$ from $p(\boldsymbol{\Sigma}_{b}|\cdot)$
5. Draw missing outcome scores $y_{(mis)}^{(t)}$ from $p(Y\,|X_r, X_{-r})$
6. Draw latent cluster means $\boldsymbol{\mu}_{rj}^{(t)}$ from $p(\mu_{rj}|\cdot)$ for $r = 1$ to P
7. Draw the grand mean $\mu_{r}^{(t)}$ from $p(\mu_{r}|\cdot)$ for $r = 1$ to R
8. Draw within-cluster regression coefficients in $\boldsymbol{\gamma}_{r}^{(t)}$ from $p(\boldsymbol{\gamma}_{r}|\cdot)$ for $r = 1$ to P
9. Draw the within-cluster residual variance $\sigma_{e_r}^{2(t)}$ from $p(\sigma_{e_r}^{2}|\cdot)$ for $r = 1$ to P
10. Draw between-cluster regression coefficients $\boldsymbol{\eta}_{r}^{(t)}$ from $p(\boldsymbol{\eta}_{r}|\cdot)$ for $r = 1$ to R
11. Draw the between-cluster residual variance $\sigma_{\zeta_r}^{2(t)}$ from $p(\sigma_{\zeta_r}^{2}|\cdot)$ for $r = 1$ to R
12. Using a Metropolis sampler, draw missing covariates $X_{r(mis)}^{(t)}$ from $p(Y|X_r, X_{-r}) \times p(X_r|X_{-r})$ for $r = 1$ to R

The dot after the vertical pipe conveys the idea that the entire set of variables being conditioned on are fixed at their current values (e.g., the first step draws the substantive model's regression coefficients from a distribution that conditions on random effects, variance components, and imputed data from the previous iteration). Finally, note that a given model or variable may only require a subset of these steps. For example, if the outcome variable is complete, steps 1 through 4 are needed for the posterior distributions of the covariates, but imputation step 5 is omitted.

Categorical Variables

Thus far we have focused on continuous explanatory variables, but model-based imputation readily accommodates incomplete ordinal and nominal variables. We use a cumulative probit model for ordinal variables and a multinomial probit model for nominal responses (Agresti, 2012; Albert & Chib, 1993; Carpenter & Kenward, 2013; Jiao & van Dyk, 2015; Johnson & Albert, 1999; Mcculloch & Rossi, 1994). In the interest of space, we describe the procedure for binary and ordered responses here and take up the multinomial probit model in a separate work. Additional details on probit imputation schemes (also known as latent variable imputation) are widely available in the literature (Asparouhov & Muthén, 2010; Carpenter & Kenward, 2013; Enders et al., 2018; Goldstein, Bonnet, & Rocher, 2007; Goldstein, Carpenter, Kenward, & Levin, 2009).

Probit regression imagines discrete responses arising from an underlying normal latent variable distribution (Agresti, 2012; Albert & Chib, 1993; Johnson & Albert, 1999). We denote the discrete and latent versions of covariate r as X_r and X_r^*, respectively. The probit model defines the underlying latent variable as a *z*-score, with the linear predictor from a regression model defining the center of the distribution and the variance fixed to establish a scale. For a binary covariate, the model additionally incorporates a threshold parameter, κ, that divides the latent distribution into two segments, such that X_r^* is below the threshold when X_r equals zero and above the threshold when X_r equals one. This threshold parameter, which is typically fixed at zero to avoid redundancy with the fixed regression intercept, can be viewed as the *z*-score cutoff, above which the discrete score changes from zero to one. The probit model for ordinal variables has an identical formulation but incorporates additional threshold parameters. For example, an ordered categorical variable with $c = 1, \ldots, C$ response options requires $C - 1$ threshold parameters, such that $X_r = c$ if $\kappa_{c-1} < X_r^* \leq \kappa_c$. In this situation, the first threshold is still fixed at zero, but the remaining thresholds are updated at each iteration of the Gibbs sampler. We use the Metropolis-Hastings step described by Cowles (1996) for this purpose because it converges more quickly than other algorithms (Albert & Chib, 1993).

To illustrate imputation for categorical explanatory variables, reconsider the random coefficient analysis from Equation 9, this time treating the level-1 covariate X_2 as binary. We previously motivated model-based imputation by assuming a multivariate normal distribution for the explanatory variables because this ensures that we can derive mutually compatible imputation models. For categorical covariates, the normality assumption applies to the underlying latent scores. This necessitates a change to the level-1 normal distribution, where diagonal elements corresponding to the categorical variables are fixed at unity, as follows.

$$\begin{pmatrix} x_{1ij} \\ x^*_{2ij} \end{pmatrix} \sim MN(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_1) \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma^2_{X_1} & \sigma_{X_1 X_2^*} \\ \sigma_{X_2^* X_1} & 1 \end{pmatrix} \tag{22}$$

Introducing latent variables changes the within-cluster regressions, which now model associations on the latent metric.

$$\begin{gathered} x_{1ij} = \mu_{1j} + \gamma_{11}\left(x^*_{2ij} - \mu_{2j}\right) + e_{1ij} \\ e_{1ij} \sim N(0, \sigma^2_{e_1}) \end{gathered} \tag{23}$$

$$\begin{gathered} x^*_{2ij} = \mu_{2j} + \gamma_{21}\left(x_{1ij} - \mu_{1j}\right) + e_{2ij} \\ e_{2ij} \sim N(0, 1 - \gamma^2_{21}\sigma^2_{X_1}) \end{gathered} \tag{24}$$

Importantly, the residual variance in X_2^*'s equation is no longer a free parameter but is a deterministic subtraction of explained variance from the total within-cluster variance[7], which is fixed at unity in the joint distribution of the covariates (Equation 22). An alternate parameterization fixes the residual variance in Equation 24 to unity, which then defines total variance as $1 + \gamma^2_{21}\sigma^2_{X_1}$. No changes are needed for the between-cluster part of the model, so Equations 16 to 18 also apply to this example.

Given a full sample of latent variable scores, standard Bayesian estimation steps generate the parameter values, random effects, and latent group means required for imputation. The Gibbs sampler algorithm outlined in the previous section is augmented with an additional step that draws latent variable scores for each case, after which it performs the estimation steps treating the synthetic scores as real data. For cases with observed data, latent variable scores are drawn from a truncated normal distribution, such that observed scores of zero and one have latent scores below and above the threshold, respectively (e.g., $x^*_{2ij} < \kappa$ if $x_{2ij} = 0$, and $x^*_{2ij} \geq \kappa$ when $x_{2ij} = 1$). Robert (1995) describes an efficient approach for sampling from tails of a truncated normal distribution, but synthetic values can also be generated by repeatedly drawing values from a normal distribution until obtaining a score in the desired range.

[7] In Equation 28, $\sigma^2_{X_1}$ is the within-cluster variance of X_1 because latent group mean centering removes all between-cluster variance from the level-1 predictors. In the case of a level-2 or level-3 categorical variable, $\sigma^2_{X_1}$ would reflect the total between-cluster variation at that level. With two or more predictors in a covariate model, $\gamma^2_{21}\sigma^2_{X_1}$ is replaced by the analogous matrix expression, $\boldsymbol{\gamma}'\boldsymbol{\Sigma}_X\boldsymbol{\gamma}$, where $\boldsymbol{\Sigma}_X$ is the relevant within- or between-cluster covariance matrix.

For cases with missing data, the final Metropolis sampling step generates latent variable imputations from an unrestricted normal distribution, and it subsequently creates discrete imputes for the analysis model by comparing the continuous values to the threshold(s). As such, the posterior distribution of the missing values has a complex form that now depends on the product of two normal distributions and a function that reflects this categorization process. To illustrate, the posterior distribution of X_2 is

$$\begin{aligned} p(X_2|Y, X_1, X_3) &\propto p(Y|X_1, X_2, X_3) \times p(X_2^*|X_1, X_3) \times p(X_2|X_2^*) \\ &\propto N\left(\left(\beta_0 + b_{0j}\right) + \left(\beta_1 + b_{1j}\right)x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3j}, \sigma_\varepsilon^2\right) \\ &\times N\left(\mu_{2j} + \gamma_{21}\left(x_{1ij} - \mu_{1j}\right), 1 - \gamma_{21}^2 \sigma_{X_1}^2\right) \\ &\times \left(I\left(x_{2ij}^* \geq 0\right)I\left(x_{2ij} = 1\right) + I\left(x_{2ij}^* < 0\right)I\left(x_{2ij} = 0\right)\right) \end{aligned} \quad (25)$$

where $p(X_2^*|X_1, X_3)$ is the distribution induced by the probit model from Equation 24, and the indicator functions that comprise the final term reflect the link between the latent and discrete imputes (i.e., X_r = 1 if $X_r^* \geq \kappa$ and X_r = 0 if $X_r^* < \kappa$). Consistent with the procedure for continuous variables, we use a Metropolis sampler to draw candidate pairs of latent and discrete imputations, retaining those that are likely to originate from the distribution in Equation 25. The technical details for the Metropolis step are given in the online supplemental materials.

Simulation Study 1: Two-Level Random Coefficient Analysis

This section describes the first of three Monte Carlo simulation studies used to evaluate model-based imputation. For this simulation, a random coefficient model with a single covariate at each level served as the population model.

$$y_{ij} = \beta_0 + \beta_1\left(x_{1ij}\right) + \beta_2\left(x_{2j}\right) + b_{0j} + b_{1j}\left(x_{1ij}\right) + \varepsilon_{ij} \quad (26)$$

We generated 1000 artificial data sets within each cell of a design that varied five between-subjects factors: the intraclass correlation of the level-1 variables (ICC = .10 and .50), number of level-2 clusters (J = 30 and 100), within-cluster sample size (n_j = 10 and 30), missing data rate (15% or 25% missing data on each predictor), and distribution shape of the level-2 predictor (normally distributed versus a single-degree of freedom chi-square). These conditions routinely arise in behavioral science research (e.g., the low and high ICCs are typical of cross-sectional and longitudinal data, respectively; the sample size combinations are distributed around the 30/30 rule-of-thumb from the literature), and the missing data rates are high enough to reveal biases for a reverse random coefficient imputation strategy (Enders et al., 2019; Enders et al., 2018; Grund et al., 2016).

We derived population parameters following variance decompositions given in Snijders and Bosker (2012, p. 116-117) and Rights and Sterba (2018). In particular, Rights and Sterba (2018) define variance-explained effect size measures that we used to derive meaningful parameter values for the simulation. In line with our treatment of explanatory variables (e.g., the normal distributions from Equation 11), these authors treat covariates as random variables that have within-cluster and between-cluster (i.e., level-1 and level-2) covariance matrices. This is convenient for modifying the intraclass correlations and defining variance explained effect sizes for the fixed and random parts of the model at each level. To establish a metric for the covariates, we constrained the within-cluster variance of X_1 at one and solved for the between-cluster variance that gives the desired ICC. The total variance of the level-2 predictor X_2 was also set to one, and its correlation with the between-cluster part of X_1 was r = .30. Finally, we set the total variance of Y to 100 and solved for the within- and between-cluster variances that gave the same ICC as X_1. These arbitrary constraints on the variances allowed us to specify effect sizes and solve for the corresponding model parameters.

Applying expressions from Rights and Sterba (2018), we chose a value for the slope variance that explained 10% of the within-cluster variance in the outcome, and we derived the fixed level-1 regression slope that accounted for an additional 10% of this variance. Given these parameters, the residual within-cluster variance is fully determined. Moving to the level-2 parameters, the slope variance and between-cluster variance of X_1 determine a portion of the between-cluster variance (this is a consequence of grand mean centering). Similarly, the β_1 coefficient and the between-cluster variance of X_1 determine part the variance attributable to the fixed effects. Next, we solved for the β_2 coefficient that explained an additional 10% of the level-2 variance, and we computed the residual (intercept) variance by subtracting out explained variance due to the fixed and random effects. Finally, we set the correlation between the random intercepts and slopes at r = .30. Our goal for the simulation was to implement effect sizes that are meaningful to substantive researchers and large enough to expose potential problems with the imputation procedures. Equations 27 and 28 give the resulting population parameters for the ICC = .10 and .50 conditions, respectively.

$$y_{ij} = 50 + 3.162\left(x_{1ij}\right) + .744\left(x_{2ij}\right) + b_{0j} + b_{1j}\left(x_{1ij}\right) + \varepsilon_{ij}$$
$$\begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} \sim MN(0, \begin{pmatrix} 7.000 & 2.510 \\ 2.510 & 10.000 \end{pmatrix}) \quad \varepsilon_{ij} \sim N(0{,}72) \tag{27}$$

$$y_{ij} = 50 + 3.162\left(x_{1ij}\right) + 1.664\left(x_{2ij}\right) + b_{0j} + b_{1j}\left(x_{1ij}\right) + \varepsilon_{ij}$$
$$\begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} \sim MN(0, \begin{pmatrix} 35.000 & 5.612 \\ 5.612 & 10.000 \end{pmatrix}) \quad \varepsilon_{ij} \sim N(0{,}40) \tag{28}$$

All variables or terms on the right side of the population regression equation were first generated by sampling deviation scores from univariate or multivariate normal distributions with the desired variances and covariances. To simulate nonnormal data, we created a chi-square variate by squaring X_2 and rescaling it to have a zero mean and unit variance (on average) prior to inducing its between-cluster correlation with X_1. This resulted in a variable with skewness and excess kurtosis values of approximately 2.0 and 9.0, respectively. After generating the predictor variables and residual terms, the outcome variable was computed as the weighted sum in Equation 27 or 28.

We imposed a 15% or 25% missing data rate on both explanatory variables, such that missing values on X_1 and X_2 were generated as a function of Y and the Y group means, respectively[8]. We used a logistic regression equation to link missingness probabilities to the outcome variable. Using a latent variable formulation for logistic regression (Agresti, 2012; Johnson & Albert, 1999), we derived intercept and slope coefficients that produced the desired missing data rate and a pseudo R^2 (McKelvey & Zavoina, 1975) value equal to .50 between the cause of missingness and the latent propensities. Finally, we sampled a missing data indicator for each observation (0 = observed, 1 = missing) from a binomial distribution with success rate

[8] In a second set of simulations not reported here, we created level-1 and level-2 auxiliary variables, A_1 and A_2, that were responsible for missingness on X_1 and X_2, respectively. We set the correlation between each covariate-auxiliary pair at approximately .50. The model-based imputation results were similar to those presented here, although listwise deletion performed much better since the MAR selection mechanism was much weaker.

equal to that observation's missingness probability from the logistic regression model, and we deleted scores with indicator values of one.

We used the Blimp 2.0 application (Enders et al., 2018; Keller & Enders, 2019) to apply reverse random coefficient imputation (i.e., conventional fully conditional specification) and model-based imputation[9]. The reverse random coefficient approach uses a model similar to that in Equation 10 to impute X_1, and it uses the Y and X_1 cluster means (computed as arithmetic averages) as predictors of the missing X_2 scores. The algorithmic steps for this version of fully conditional specification (Blimp offers others) are identical to invoking the mice.impute.2l.pan and mice.impute.2lonly.norm functions in the R package MICE (van Buuren, 2011; van Buuren et al., 2018). The model-based approach is identical to the procedure described earlier except that the simulation model includes a single level-1 covariate rather than two. Similar model-based procedures for this particular random coefficient model can be implemented in specialized Bayesian analysis programs such as JAGS (Erler et al., 2017; Erler et al., 2016; Grund et al., 2018; Plummer, 2016) and in the R packages jomo (Quartagno & Carpenter, 2018) and mdmb (Robitzsch & Lüdke, 2018).

After examining potential scale reduction factors (Gelman et al., 2014; Gelman & Rubin, 1992) from several artificial data sets, we generated 10 imputations from a Gibbs sampler algorithm with 1000 burn-in and thinning iterations (i.e., imputed data sets were saved at 1000-iteration increments). We used the complete-data maximum likelihood estimator in M*plus* 8 (Muthén & Muthén, 1998–2017) to fit the random slope model to the multiply imputed data sets, and we wrote a custom R program to pool estimates and standard errors. To provide additional comparisons, we also report the complete-data (pre-deletion) and listwise deletion results. The online supplemental material also presents limited simulation results for the full information maximum likelihood estimator for missing data in M*plus*, which can accommodate incomplete covariates with numerical integration. Computational tasks were executed on UCLA's Hoffman2 supercomputer, and we used a Linux shell script to coordinate simulation tasks. All simulation code is available upon request.

Simulation Study 1 Results

[9] We implemented the PRIOR2 options for the substantive analysis and the XPRIOR3 option for the covariate models. The technical appendix from the online supplement describes these priors in detail.

Although not part of our main simulation design, we began by examining the large-sample behavior of the imputation methods in data sets with 1000 level-2 clusters and 50 observations per cluster. Figure 1 gives trellis plots displaying relative bias values (the difference between an average estimate and its true value expressed as a percent of the true value) for each combination of intraclass correlation, distribution shape, and missing data rate. As a rough heuristic, published simulations often define bias values less than 10% in absolute value as acceptable (Finch, West, & MacKinnon, 1997; Kaplan, 1988), so the figures display these thresholds as dashed lines. As seen in the figure, model-based imputation estimates were effectively indistinguishable from those of the complete data. Perhaps somewhat surprisingly, violating normality by including a skewed level-2 predictor had no material impact on parameter recovery. In contrast, there was no situation where fully conditional specification (reverse random coefficient imputation) produced uniformly accurate estimates, as the random slope variance was consistently underestimated by 10% to 20% (the covariance was also biased). Previous studies have also noted this bias (Enders et al., 2018; Grund et al., 2016), which is a consequence of incompatibility. Finally, listwise deletion estimates were uniformly and substantially biased. Because there is no reason to expect deletion to improve with smaller samples, we omit this approach from further discussion.

Turning to the full simulation design, Figures 2 and 3 give trellis plots displaying relative bias values with normally distributed predictors and 15% and 25% missing data, respectively. The online supplemental material gives a full set of graphical and tabular displays of the model-based estimates and their bias values. Considered as a whole, model-based imputation estimates tracked closely with those of the complete data, and the procedure was clearly preferable to reverse random coefficient imputation (fully conditional specification). The combination of ICC = .10, small sample size (30 clusters with 10 observations per group), and 25% missing data produced the largest bias values, but performance was still quite good for most parameters. For comparison, Figure 4 displays the relative bias values from the simulation conditions with a skewed level-2 predictor and 25% missing data rate (see the online supplemental material for a few set of graphical displays). Consistent with the large-sample simulation, violating normality at level-2 had little to no impact on parameter recovery. Because the results did not materially differ from those in Figures 2 and 3, no further discussion is warranted. Although imputation is our main focus, we also applied the full information maximum likelihood estimator in M*plus* to the normally distributed simulation data because this option would be widely available to substantive researchers. The maximum likelihood estimates were prone to large biases, particularly in the ICC = .50 conditions. These results, which appear in the online supplement, underscore that problems related to non-linear terms are not restricted to multiple imputation.

The trellis plot in Figure 5 displays frequentist confidence interval coverage values for the 25% missing data rate condition. Because the other conditions were quite similar, we give the full set of plots in the online supplemental material. Coverage is the proportion of estimates where the 95% symmetric confidence interval included the true parameter, and the dashed lines at .925 and .975 correspond to Bradley's (1978) so-called liberal criterion. Coverage values lower than the nominal 95% rate reflect Type I error inflation (e.g., a coverage value of 90% suggests a twofold increase in Type I errors), whereas values greater than 95% reflect conservative inference. We restrict our attention to the fixed effects because the literature argues that symmetric confidence intervals are inappropriate for variance estimates (Maas & Hox, 2005; Snijders & Bosker, 2012). As seen in the figure, coverage values for the level-2 slope coefficient were often too low (about 90%) in conditions with only 30 clusters, but the complete-data coverage rates exhibited the same pattern.

Simulation Study 2:
Two-Level Random Coefficient Analysis with a Categorical Predictor

The second simulation study evaluated model-based imputation with a categorical explanatory variable. The random coefficient analysis from Equation 26 again served as the population model, but the level-2 covariate X_2 was a dichotomous variable with equal category proportions, on average. Creating the binary variable required two small changes to the data-generating process, but the simulation design and procedures were otherwise identical. First, to derive the true population parameters, we set the variance of the underlying continuous X_2^* scores at .25, which is the same value we would expect from the binary variable. Second, after deriving the true parameter values, we generated continuous data and subsequently dichotomized X_2 by splitting the underlying continuous distribution at zero (the population mean). We again used Blimp 2 to implement conventional fully conditional specification and model-based imputation[10], respectively, and we used listwise deletion as an additional comparison.

[10] We implemented the PRIOR2 options for the substantive analysis and the XPRIOR3 option for the covariate models. The technical appendix from the online supplement describes these priors in detail.

Simulation Study 2 Results

As a precursor to the full simulation study, we again examined the large-sample behavior of the imputation methods in data sets with 1000 level-2 clusters and 50 observations per cluster. The trellis plots in Figure 6 show that model-based imputation estimates were virtually unbiased, whereas reverse random coefficient imputation (fully conditional specification) consistently underestimated the slope variance by 10% to 20% of its true value. Because the full simulation produced results that were virtually identical to those from the first simulation study, we give the graphical summaries in the online supplement. The results can be summarized as follows: (a) model-based estimates largely tracked with those of the complete data, (b) the combination of small within-cluster sample size and low intraclass correlation produced the largest bias values (similar to those from Figures 2 and 3), and (c) parameter recovery for fully conditional specification was meaningfully worse. Finally, coverage values were comparable to those from Figure 5 (e.g., complete-data and imputation-based estimates of the level-2 slope coefficient were often too low).

Simulation Study 3:
Three-Level Analysis with a Cross-Level Interaction

Thus far we have considered random coefficient models because they have been the focus of recent multilevel imputation literature. However, model-based imputation can accommodate a much broader range of interactive and non-linear effects. To illustrate its performance in a different context, the final simulation examined a three-level random coefficient analysis with a covariate at each level and a cross-level interaction involving a level-1 and level-3 predictor.

$$\begin{aligned} y_{ijk} = \beta_0 + \beta_1(x_{1ijk}) + \beta_2(x_{2jk}) + \beta_3(x_{3k}) + \beta_4(x_{1ijk})(x_{3k}) \\ + b_{0k} + b_{1k}(x_{1ijk}) + b_{0jk} + b_{1jk}(x_{1ijk}) + \varepsilon_{ijk} \end{aligned} \quad (29)$$

For this analysis, the imputation procedure applies the multivariate normal distribution from Equation 12 to the covariates, which again induces a set of linear regression models at each level. The analysis model functions in the same way as it did before (i.e., defining the distribution of Y given the predictors and interaction). In the interest of space, the online supplement gives the posterior distributions of the missing data for this problem.

The simulation generated 1000 artificial data sets within each cell of a design that varied four between-subjects factors: the distribution of variance across levels (20% and 50% of the variation distributed across level-2 and level-3), number of level-3 clusters (K = 30 and 100), level-1 within-cluster sample size (n_j = 10 and 50), and missing data rate for the covariates (15% or 25%). The two variability configurations featured 80% or 50% of Y and X_1's variance at level-1, with between-cluster variability distributed evenly across the two higher levels (we refer to these as the ICC = .20 and .50 conditions, respectively). In these same conditions, 80% or 50% of X_2's variance was assigned to level-2 with the rest at level-3. The number of level-2 clusters within each level-3 cluster was held constant at 5, resulting in a range of sample sizes between 1500 (30 level-3 clusters and 10 level-1 observations per level-2 group) and 25,000 (100 level-3 clusters and 50 level-1 observations per level-2 group). As before, missingness on the covariates was imposed as a function of the within- and between-cluster parts of the outcome variable. Because this simulation is meant as a proof of concept under ideal circumstances, we do not investigate the impact of nonnormality. It was somewhat surprising that imputing a skewed predictor did not impact parameter recovery, but we would be hesitant to assume that the same holds true when the nonnormal variable is part of an interaction. Keller (2019) provides a thorough investigation of model-based imputation for multilevel interactive effects, and his work examines this issue.

The effect size measures from Rights and Sterba (2018) readily extend to three-level models, so our data-generating process largely mimicked the previous simulations. As before, we set the covariate correlations to r = .30, and we specified fixed effects hierarchically, such that each coefficient incremented the explained variance at a particular level by 10%. Because interaction effects tend to be smaller in magnitude (Chaplin, 1991), we set the cross-level interaction coefficient to explain an additional 5% of the within-cluster variance. Finally, we identified random slope variances that explained 10% of the outcome's variance at level-1, and we determined the residual variance at each level by subtracting out the explained portions of variance due to the fixed and random effects. Equations 30 and 31 give the resulting population parameters for the ICC = .20 and .50 conditions, respectively.

$$y_{ijk} = 49.836 + 3.098(x_{1ijk}) + .724(x_{2jk}) + .654(x_{3k}) + 1.549(x_{1ijk})(x_{3k}) \qquad (30)$$

$$+ b_{0k} + b_{1k}(x_{1ijk}) + \; b_{0jk} + b_{1jk}(x_{1ijk}) + \varepsilon_{ijk}$$
$$\begin{pmatrix} b_{0k} \\ b_{1k} \end{pmatrix} \sim MN(0, \begin{pmatrix} 5.104 & 1.917 \\ 1.917 & 8.000 \end{pmatrix}) \begin{pmatrix} b_{0jk} \\ b_{1jk} \end{pmatrix} \sim MN(0, \begin{pmatrix} 5.500 & 1.990 \\ 1.990 & 8.000 \end{pmatrix})$$
$$\varepsilon_{ijk} \sim N(0{,}52.000)$$

$$y_{ijk} = 49.740 + 2.449(x_{1ijk}) + 1.445(x_{2jk}) + .938(x_{3k}) + 1.225(x_{1ijk})(x_{3k})$$
$$+ b_{0k} + b_{1k}(x_{1ijk}) + \; b_{0jk} + b_{1jk}(x_{1ijk}) + \varepsilon_{ijk}$$
$$\begin{pmatrix} b_{0k} \\ b_{1k} \end{pmatrix} \sim MN(0, \begin{pmatrix} 11.183 & 2.243 \\ 2.243 & 5 \end{pmatrix}) \begin{pmatrix} b_{0jk} \\ b_{1jk} \end{pmatrix} \sim MN(0, \begin{pmatrix} 13.750 & 2.487 \\ 2.487 & 5.000 \end{pmatrix}) \qquad (31)$$
$$\varepsilon_{ijk} \sim N(0{,}32.500)$$

We again used Blimp 2 to apply just-another-variable imputation (i.e., conventional fully conditional specification that treats the cross-level product as a variable to be imputed) and model-based imputation[11]. We are unaware of other software packages that apply these approaches to three-level data. After examining potential scale reduction factors (Gelman et al., 2014; Gelman & Rubin, 1992) from several artificial data sets, we generated 10 imputations from a Gibbs sampler algorithm with 1000 burn-in and thinning iterations (i.e., imputed data sets were saved at 1000-iteration increments). As before, we used M*plus*'s complete-data maximum likelihood estimator to fit the analysis model to the multiply imputed data sets, and we wrote a custom R program to pool estimates and standard errors. Computational tasks were executed on UCLA's Shared Hoffman2 Cluster, and we used a Linux shell script to coordinate simulation tasks. All simulation code is available upon request.

Simulation Study 3 Results

Figures 7 and 8 display average relative bias values for the fixed effects and variance components, respectively. In the interest of space, we focus on the 25% missing data rate and point readers to the online supplement for a full set of graphical displays. For clarity we omit listwise deletion from Figures 7 and 8, but these results are in the supplement. As seen in the figures, imputing the incomplete product term with fully conditional specification (i.e., just-

[11] We implemented the PRIOR2 options for the substantive analysis and the XPRIOR3 option for the covariate models. The technical appendix from the online supplement describes these priors in detail.

another-variable imputation) introduced substantial biases that did not dissipate as sample size increased (e.g., the interaction and level-3 slope coefficients were 20% to 40% lower than their true values, and a number of variance estimates were also distorted). In contrast, model-based imputation estimates were generally quite accurate, with most bias values well below 10%. Estimating this model with only 30 level-3 units gave negatively biased estimates of the level-3 slope and the interaction coefficient, but increasing the number of groups to 100 effectively eliminated this issue. Variance estimates were generally quite accurate, and increasing either the within-cluster sample size at level-1 (e.g., from 10 to 50) or the number of level-3 groups improved parameter recovery. Not surprisingly, listwise deletion estimates were uniformly and severely biased (see the online supplement).

Finally, Figure 9 displays frequentist confidence interval coverage values for the 25% missing data rate condition, and the online supplemental material gives a full set of plots. Model-based coverage values were almost always within Bradley's liberal bounds. The complete-data coverage values for the level-3 slope coefficient and the interaction coefficient were lower (worse) in some cases (e.g., with 30 level-3 groups), which presumably occurred because missing data uncertainty widened confidence intervals and counteracted the complete-data estimator's natural tendency for under-coverage.

Real Data Example

This section illustrates model-based imputation in the context of a cluster-randomized trial of a novel math problem-solving intervention (Montague, Krawec, Enders, & Dietz, 2014). The data set, which is available on the Blimp website, features three levels: 6874 repeated measurements at level-1 nested in 982 students at level-2, and students nested in 29 schools at level-3. Schools were randomly assigned to an intervention (novel curriculum) or control (standard curriculum) condition, such that all students within a given school received the same treatment. To keep the example relatively simple, we ignore nesting at the school level and focus on a two-level model (in fact, there was a relatively small proportion of variation at the school level). The analysis is a two-level growth curve model with a cross-level interaction involving condition and months (i.e., the group-by-time interaction). In addition to an intervention dummy code and its interaction with the temporal predictor, the substantive analysis model features a time-varying math self-efficacy rating scale and a lunch assistance dummy code as covariates.

$$probsolv_{ti} = \beta_0 + \beta_1(matheff_{ti}) + \beta_2(month0_{ti}) + \beta_3(frlunch_i) + \beta_4(condition_i) + \beta_5(month0_{ti})(condition_i) + b_{0i} + b_{1i}(matheff_{ti}) + b_{2i}(month0_{ti}) + e_{ti} \quad (32)$$

The percentage of missing observations for each incomplete variable were as follows: problem-solving (11.5%), self-efficacy (11.45%), lunch assistance status (4.7%). A substantial proportion of problem-solving and self-efficacy scores resulted from planned missingness where the control group was assessed bimonthly instead of monthly. Regardless of mechanism, time-varying variables were imputed by fixing the corresponding temporal predictor at the planned assessment dates.

Appendix A gives the Blimp 2 script for model-based imputation. To add the third level of nesting, one would simply list the school-level identifier on the CLUSTERID line (to our knowledge, Blimp is the only application that can apply this procedure to three-level data structures). The Blimp 2 user guide (Keller & Enders, 2019) provides a detailed description of the scripting language, including a number of new conventions and commands that differ from its predecessor. A byproduct of our procedure is that the software also gives Bayesian estimates (e.g., posterior means and standard deviations) as optional output, so we provide these results as a comparison to illustrate a simple sensitivity analysis. We also used Blimp to implement a second model-based imputation procedure that incorporates gender (complete), standardized math scores (4.7% missing), and the school-level percentage of non-native English speakers (complete) as auxiliary variables (these additional variables are added to the MODEL line, and gender must be declared as ORDINAL). Finally, although theoretical and computer simulation results generally argue against it, we also included fully conditional specification. The procedure should achieve near-optimal performance in this example because the interacting variables are complete. The software is available as a free download for macOS, Windows, and Linux at www.appliedmissingdata.com/multilevel-imputation, and the full set of analysis scripts and the data are available from the same URL.

As a starting point, it is important to recall that Blimp specifies a multivariate normal joint distribution for the explanatory variables (or their latent scores, in the case of discrete predictors), and the software does not allow users to specify non-linear relations (e.g., quadratic terms or random coefficients) among pairs of covariates. The so-called sequential version of model-based imputation outlined by Ibrahim et al. (2002) and more recently by Erler and colleagues (Erler et al., 2017; Erler et al., 2016) offers this flexibility, but this approach is not yet available for three-level data structures. Assuming normality for explanatory variables simplifies imputation because the user need only specify the substantive analysis on the MODEL line, and the software automatically constructs the appropriate covariate models. The cross-level product term in the analysis model is specified by joining the interacting variables with an asterisk (e.g., `month0*condition`), and the random coefficients are specified by listing the random predictors to the right of the vertical pipe (e.g., `MODEL: … | matheff month0;`). In a three-level model, Blimp would automatically include random coefficients at level-2 and level-3.

Executing the script in Appendix A generates imputations for the lower-order variables that are consistent with the specified interaction effect, but the product term is not added to the filled-in data sets. Rather, Blimp saves uncentered lower-order variables so that users can apply group mean or grand mean centering (or leave variables uncentered) prior to computing interactions. This is in contrast to fully conditional specification, which treats the interaction as a variable to be imputed. In this framework, it may be necessary to work with uncentered variables then rescale the imputed product term to approximate a centered solution (Enders et al., 2014). In addition to its theoretical and empirical benefits, model-based imputation is highly convenient because it allows researchers to apply familiar procedures for probing interaction effects (Aiken & West, 1991; Bauer & Curran, 2005).

Blimp can print a table of potential scale reduction factor diagnostics (Gelman & Rubin, 1992), and it optionally saves parameter values that can readily be converted to trace plots in other software (e.g., an R plotting script is provided with the other files for this example). The potential scale reduction factors suggest that the Gibbs sampler converges in approximately 2000 iterations (i.e., across all models and all parameters, the highest potential scale reduction factor is approximately 1.05). Based on this information, we requested 20 imputations from a Gibbs sampler with 4000 burn-in and 2000 thinning iterations (i.e., we saved the first data set after 4000 computational cycles and saved additional data sets every 2000 iterations thereafter). The job takes approximately one minute on a 2018 10-core iMac Pro and about two minutes on a 2017 two-core Macbook Pro. The resulting imputations are compatible with all major analysis packages (the SAVE command outputs imputations in stacked format or as separate files), and the set of files for this example includes analysis scripts for M*plus*, R, SPSS, and Stata.

Table 1 gives the parameter estimates from the analysis, with random effect covariances omitted from the table for brevity. The key finding is that the intervention-by-time interaction coefficient is positive and significant, meaning that students in the intervention schools exhibited more rapid problem-solving gains than students in control schools. The Bayesian slope variance estimates are slightly larger than those of the imputation procedures (these quantities can be viewed as estimates taken across more than 30,000 imputations), but differences were generally slight. In particular, fully conditional specification estimates were not dramatically different from those of model-based imputation. As noted previously, fully conditional specification should achieve its optimal performance in this example because the random predictor (months since baseline) and the interacting variables are complete. Theoretical and simulation results suggest that this would not be true in general.

Discussion

Despite the broad appeal of multiple imputation and other MAR-based approaches, a broad class of regression models featuring interactive effects, polynomial terms, or random coefficients are known to cause bias-inducing problems for popular missing data handling procedures (Bartlett et al., 2015; Enders et al., 2014; Seaman et al., 2012; Zhang & Wang, 2017). A growing body of recent missing data research has focused on fully Bayesian multiple imputation methods that are appropriate for interactive and non-linear effects (Bartlett et al., 2015; Erler et al., 2017; Erler et al., 2016; Goldstein et al., 2014; Kim et al., 2018; Kim et al., 2015; Zhang & Wang, 2017). Building on these recent developments, this paper outlined a model-based multiple imputation methodology designed to handle a wide range of interactive and non-linear effects in single-level and multilevel regression models with up to three levels. This procedure offers a number of compelling advantages: it (a) has a strong theoretical foundation in the Bayesian framework, (b) readily extends to three-level data structures, (c) uses latent variables (i.e., random effects) to model between-cluster variation and covariation, (d) readily accommodates categorical variables, and (e) produces Bayesian analysis results (e.g., posterior means and standard deviations) as a byproduct of estimation. The primary downside of the procedure is that covariates can exert non-linear or random influences on the outcome but not each other. The so-called sequential approach to fully Bayesian imputation (Erler et al., 2017; Erler et al., 2016; Ibrahim et al., 2002) can accommodate certain patterns of non-linearities, but this approach has not been extended to the range of applications that we consider here (e.g., three-level models, categorical variables).

Computer simulation results suggest that model-based imputation is quite effective when applied to multilevel models with random coefficients and interaction effects. In most scenarios that we examined, estimates tracked closely with those of a complete data analysis and either reduced or eliminated biases associated with conventional approaches such as fully conditional specification imputation (and maximum likelihood estimation, although our investigation of this procedure was quite limited). These improvements were particularly salient for a model with a cross-level interaction term.

Importantly, model-based imputation in Blimp assumes that covariates are multivariate normal. It was somewhat surprising that imputing a skewed predictor did not impact parameter recovery, but we are hesitant to assume that the same holds true when the nonnormal variable is part of an interaction. Further, an obvious avenue for future research is to examine the impact of non-normal level-1 covariates. Keller (2019) provides a thorough investigation of model-based imputation for multilevel interactive effects, and his work examines these cases. Second, the normality assumption implies that covariates are linearly related, and we did not investigate scenarios where covariates are non-linearly related. At least for two-level models, it is possible to implement a comparable sequential decomposition of the covariate distribution (Ibrahim et al., 2002) in dedicated Bayesian analysis software such as JAGS (Erler et al., 2017; Erler et al., 2016; Grund et al., 2018). Because that approach allows some covariates to be non-linear functions of others, it could be more robust to normality violations than our method (Lüdke et al., 2019). In practice, we suspect that researchers would rarely have the information needed to correctly specify such non-linearities, but this alternative is important to consider.

This paper was an initial foray and is necessarily limited in scope and generalizability. First, we investigated a small subset of non-linear effects that are possible. The framework can readily accommodate three-way and higher interactions, polynomial effects, and combinations of interactive and polynomial terms. Virtually nothing is known about the application of model-based imputation to these models, and a great deal of research is needed to clarify the procedure's limitations. Second, our paper offered a very limited glimpse into categorical variable imputation. Simulations conducted while developing Blimp suggest that latent variable imputation can work well in a wide range of situations, but the procedure is almost certainly sensitive to the number and type of categorical variables, the distribution of response options, and sample size, to name a few. Finally, the simulation conditions we examined were necessarily limited in scope, and multilevel models offer myriad possibilities for manipulating sample sizes, intraclass correlations, effect sizes, and number of random effects.

In sum, our paper outlined a new imputation approach for multilevel models with interactive or non-linear effects. Limited computer simulations suggest that model-based imputation can offer substantial improvement over conventional imputation methods and maximum likelihood estimation. The Blimp application offers a user-friendly environment for implementing model-based imputation, and the software's website has a number of resource materials, including analysis scripts for all major software platforms.

References

Agresti, A. (2012). *Categorical data analysis* (3rd ed.). Hoboken, NJ: Wiley.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Albert, J. H., & Chib, S. (1993). Bayesian-Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association, 88*(422), 669-679. doi:10.2307/2290350

Arnold, B. C., Castillo, E., & Sarabia, J. M. (1999). *Conditional Specification of Statistical Models*. New York: Springer.

Arnold, B. C., Castillo, E., & Sarabia, J. M. (2001). Conditionally specified distributions: An introduction. *Statistical Science, 16*, 249-274. doi:www.jstor.org/stable/2676688

Arnold, B. C., & Press, S. J. (1989). Compatible Conditional Distributions. *Journal of the American Statistical Association, 84*(405), 152-156. doi:10.2307/2289858

Asparouhov, T., & Muthén, B. (2010). Multiple imputation with Mplus. Retrieved from www.statmodel.com/download/Imputations7.pdf

Bartlett, J., & Keogh, R. (2018). Package 'smcfcs'. Retrieved from cran.r-project.org/web/packages/smcfcs/smcfcs.pdf

Bartlett, J. W., Seaman, S. R., White, I. R., & Carpenter, J. R. (2015). Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical Methods in Medical Research, 24*(4), 462-487. doi:10.1177/0962280214521348

Bauer, D. J., & Curran, P. J. (2005). Probing Interactions in Fixed and Multilevel Regression: Inferential and Graphical Techniques. *Multivariate Behav Res, 40*(3), 373-400. doi:10.1207/s15327906mbr4003_5

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31*(2), 144-152. doi:10.1111/j.2044-8317.1978.tb00581.x

Browne, W. J. (1998). *Applying MCMC methods to multi-level models.* (PhD thesis), University of Bath, United Kingdom.

Browne, W. J., & Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics, 15*(3), 391-420. doi:10.1007/s001800000041

Carpenter, J. R., Goldstein, H., & Kenward, M. G. (2011). REALCOM-IMPUTE Software for Multilevel Multiple Imputation with Mixed Response Types. *Journal of Statistical Software, 45*(5), 1-14. doi:10.18637/jss.v045.i05

Carpenter, J. R., & Kenward, M. G. (2013). *Multiple imputation and its application*. West Sussex, UK: Wiley.

Chaplin, W. F. (1991). The next generation in moderation research in personality psychology. *Journal of Personality, 59*, 143-178. doi:10.1111/j.1467-6494.1991.tb00772.x

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences* (Third Edition ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cowles, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing, 6*(2), 101-111. doi:10.1007/Bf00162520

Enders, C. K., Baraldi, A. N., & Cham, H. (2014). Estimating interaction effects with incomplete predictor variables. *Psychological Methods, 19*(1), 39-55. doi:10.1037/a0035314

Enders, C. K., Hayes, T., & Du, H. (2019). A comparison of multilevel imputation schemes for random coefficient models: Fully conditional specification and joint model imputation with random covariance matrices. *Multivariate Behav Res, Advanced online publication*. doi:doi.org/10.1080/00273171.2018.1477040

Enders, C. K., Keller, B. T., & Levy, R. (2018). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods, 23*(2), 298-317. doi:10.1037/met0000148

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121-138. doi:10.1037/1082-989X.12.2.121

Erler, N. S., Rizopoulos, D., Jaddoe, V. W., Franco, O. H., & Lesaffre, E. M. (2017). Bayesian imputation of time-varying covariates in linear mixed models. *Statistical Methods in Medical Research*, 962280217730851. doi:10.1177/0962280217730851

Erler, N. S., Rizopoulos, D., Rosmalen, J., Jaddoe, V. W., Franco, O. H., & Lesaffre, E. M. (2016). Dealing with missing covariates in epidemiologic studies: a comparison between multiple imputation and a full Bayesian approach. *Statistics in Medicine, 35*(17), 2955-2974. doi:10.1002/sim.6944

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of Sample Size and Nonnormality on the Estimation of Mediated Effects in Latent Variable Models. *Structural Equation Modeling-a Multidisciplinary Journal, 4*(2), 87-107. doi:10.1080/10705519709540063

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC Press.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457-472. doi:10.1214/ss/1177011136

Goldstein, H., Bonnet, G., & Rocher, T. (2007). Multilevel structural equation models for the analysis of comparative data on educational performance. *Journal of Educational and Behavioral Statistics, 32*(3), 252-286. doi:10.3102/1076998606298042

Goldstein, H., Carpenter, J., Kenward, M. G., & Levin, K. A. (2009). Multilevel models with multivariate mixed response types. *Statistical Modelling, 9*(3), 173-197. doi:10.1177/1471082x0800900301

Goldstein, H., Carpenter, J. R., & Browne, W. J. (2014). Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *Journal of the Royal Statistical Society Series a-Statistics in Society, 177*(2), 553-564. doi:10.1111/rssa.12022

Grund, S., Lüdke, O., & Robitzsch, A. (2017). Multiple imputation of missing data at level 2: A comparison of fully conditional and joint modeling in multilevel designs. *Journal of Educational and Behavioral Statistics, 43*, 316-353. doi:doi.org/10.3102/1076998617738087

Grund, S., Ludtke, O., & Robitzsch, A. (2016). Multiple imputation of missing covariate values in multilevel models with random slopes: a cautionary note. *Behavior Research Methods, 48*(2), 640-649. doi:10.3758/s13428-015-0590-3

Grund, S., Ludtke, O., & Robitzsch, A. (2018). Multiple Imputation of Missing Data for Multilevel Models: Simulations and Recommendations. *Organizational Research Methods, 21*(1), 111-149. doi:10.1177/1094428117703686

Hastings, W. K. (1970). Monte-Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika, 57*(1), 97-109. doi:10.2307/2334940

Hughes, R. A., White, I. R., Seaman, S. R., Carpenter, J. R., Tilling, K., & Sterne, J. A. C. (2014). Joint modelling rationale for chained equations. *BMC Medical Research Methodology, 14*(1-10). doi:10.1186/1471-2288-14-28

Ibrahim, J. G., Chen, M. H., & Lipsitz, S. R. (2002). Bayesian methods for generalized linear models with covariates missing at random. *Canadian Journal of Statistics-Revue Canadienne De Statistique, 30*(1), 55-78. doi:10.2307/3315865

Jiao, X., & van Dyk, D. A. (2015). *A corrected and more efficient suite of MCMC samplers for the multinomal probit model*. (eprint arXiv:1504.07823).

Johnson, V. E., & Albert, J. H. (1999). *Ordinal data modeling*. New York: Springer.

Kaplan, D. (1988). The Impact of Specification Error on the Estimation, Testing, and Improvement of Structural Equation Models. *Multivariate Behav Res, 23*(1), 69-86. doi:10.1207/s15327906mbr2301_4

Keller, B. T. (2019). *Substantive model-compatible imputation for multilevel interaction effects.* (PhD thesis), UCLA, Los Angeles, CA.

Keller, B. T., Du, H., & Enders, C. K. (2019). Multilevel multiple imputation with latent group means: A fully conditional specification approach. *Manuscript in preparation*.

Keller, B. T., & Enders, C. K. (2019). Blimp User's Guide (Version 2).

Kim, S., Belin, T. R., & Sugar, C. A. (2018). Multiple imputation with non-additively related variables: Joint-modeling and approximations. *Statistical Methods in Medical Research, 27*(6), 1683-1694. doi:10.1177/0962280216667763

Kim, S., Sugar, C. A., & Belin, T. R. (2015). Evaluating model-based imputation methods for missing covariates in regression models with interactions. *Statistics in Medicine, 34*(11), 1876-1888. doi:10.1002/sim.6435

Kreft, I. G., de Leeuw, J., & Aiken, L. S. (1995). The Effect of Different Forms of Centering in Hierarchical Linear Models. *Multivariate Behavioral Research, 30*(1), 1-21. doi:10.1207/s15327906mbr3001_1

Kunkle, D., & Kaizer, E. E. (2017). A comparison of existing methods for multiple imputation in individual participant data meta-analysis. *Statistics in Medicine, 36*, 3507-3532. doi:doi.org/10.1002/sim.7388

Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association, 87*, 1227-1237. doi:10.2307/2290664

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.

Liu, J. C., Gelman, A., Hill, J., Su, Y. S., & Kropko, J. (2014). On the stationary distribution of iterative imputations. *Biometrika, 101*(1), 155-173. doi:10.1093/biomet/ast044

Lüdke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2 x 2 taxonomy of multilevel latent contextual models: accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods, 16*(4), 444-467. doi:10.1037/a0024376

Lüdke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new approach to group-level effects in contextual studies. *Psychological Methods, 13*(3), 201-229. doi:10.1037/a0012869

Lüdke, O., Robitzsch, A., & Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods, 22*(1), 141-165. doi:10.1037/met0000096

Lüdke, O., Robitzsch, A., & West, S. G. (2019). Regression models involving nonlinear effects with missing data: A sequential modeling approach using Bayesian estimation. *Manuscript submitted for publication*.

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. Berlin: Springer.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*, 86-92. doi:10.1027/1614-1881.1.3.86

Mcculloch, R., & Rossi, P. E. (1994). An Exact Likelihood Analysis of the Multinomial Probit Model. *Journal of Econometrics, 64*(1-2), 207-240. doi:10.1016/0304-4076(94)90064-7

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology, 4*(1), 103-120. doi:10.1080/0022250x.1975.9989847

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide. Eighth edition.* Los Angeles, CA: Muthén & Muthén.

Plummer, M. (2016). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling (Version 4.2.0). Retrieved from Retrieved from http://sourceforge.net/projects/mcmc-jags/

Quartagno, M., & Carpenter, J. (2018). Package 'jomo'. Retrieved from cran.r-project.org/web/packages/jomo/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Resche-Rigon, M., & White, I. R. (2018). Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Statistical Methods in Medical Research, 27*(6), 1634-1649. doi:10.1177/0962280216666564

Rights, J. D., & Sterba, S. K. J. P. m. (2018). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures.

Robert, C. P. (1995). Simulation of Truncated Normal Variables. *Statistics and Computing, 5*(2), 121-125. doi:10.1007/Bf00143942

Robitzsch, A., & Lüdke, O. (2018). Package 'mdmb'. Retrieved from cran.r-project.org/web/packages/mdmb/mdmb.pdf

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592. doi:10.1093/biomet/63.3.581

Sarabia, J. M. a., Castillo, E., & Arnold, B. C. (2001). comments and a rejoinder by the authors). *Statistical Science, 16*(3), 249-274. doi:10.1214/ss/1009213728

Savalei, V., & Bentler, P. M. (2009). A Two-Stage Approach to Missing Data: Theory and Application to Auxiliary Variables. *Structural Equation Modeling-a Multidisciplinary Journal, 16*(3), 477-497. doi:10.1080/10705510903008238

Savalei, V., & Falk, C. F. (2014). Robust Two-Stage Approach Outperforms Robust Full Information Maximum Likelihood With Incomplete Nonnormal Data. *Structural Equation Modeling-a Multidisciplinary Journal, 21*(2), 280-302. doi:10.1080/10705511.2014.882692

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall.

Schafer, J. L., & Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics, 11*(2), 437-457. doi:10.1198/106186002760180608

Seaman, S. R., Bartlett, J. W., & White, I. R. (2012). Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC Medical Research Methodology, 12*, 46. doi:10.1186/1471-2288-12-46

Shin, Y. Y., & Raudenbush, S. W. (2010). A Latent Cluster-Mean Approach to the Contextual Effects Model With Missing Data. *Journal of Educational and Behavioral Statistics, 35*(1), 26-53. doi:10.3102/1076998609345252

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage.

van Buuren, S. (2011). Multiple imputation of multilevel data. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis* (pp. 173-196). New York: Routledge.

van Buuren, S. (2012). *Flexible imputation of missing data*. New York: Chapman & Hall.

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. doi:10.1080/10629360600810434

van Buuren, S., Groothuis-Oudshoorn, K., Robitzsch, A., Vink, G., Doove, L., Jolani, S., . . . Gray, B. (2018). Package 'mice'. Retrieved from cran.r-project.org/web/packages/mice/mice.pdf

Vink, G., & van Buuren, S. (2013). Multiple imputation of squared terms. *Sociological Methods & Research, 42*(4), 598-607. doi:10.1177/0049124113502943

von Hippel, P. T. (2007). Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology, 37*, 83-117. doi:doi.org/10.1111/j.1467-9531.2007.00180.x

von Hippel, P. T. (2009). How to impute interactions, squares, and other transformed variables. *Sociological Methodology, 39*, 265-291. doi:doi.org/10.1111/j.1467-9531.2009.01215.x

Yuan, K. H., & Savalei, V. (2014). Consistency, bias and efficiency of the normal-distribution-based MLE: The role of auxiliary variables. *Journal of Multivariate Analysis, 124*, 353-370. doi:10.1016/j.jmva.2013.11.006

Yucel, R. M. (2008). Multiple imputation inference for multivariate multilevel continuous data with ignorable non-response. *Philos Trans A Math Phys Eng Sci, 366*(1874), 2389-2403. doi:10.1098/rsta.2008.0038

Zhang, Q., & Wang, L. (2017). Moderation analysis with missing data in the predictors. *Psychological Methods, 22*(4), 649-666. doi:10.1037/met0000104

Appendix A

Blimp Syntax for Applying Bayesian Estimation and

Model-Based Imputation with the Real Data Example

```
DATA: ~/desktop/example.dat;
VARIABLES: school student wave condition eslpct
   ethnic male frlunch achgroup stanmath month0
   month7 probsolv matheff condbymonth;
ORDINAL: frlunch condition;
CLUSTERID: student;
MISSING: 999;
MODEL: probsolv ~ matheff month0 frlunch condition
   month0*condition | matheff month0;
SEED: 90291;
NIMPS: 20;
BURN: 4000;
THIN: 2000;
CHAINS: 10 processors 10;
OPTIONS: estimates latent psr;
SAVE: stacked = ~/desktop/imps.csv;
```

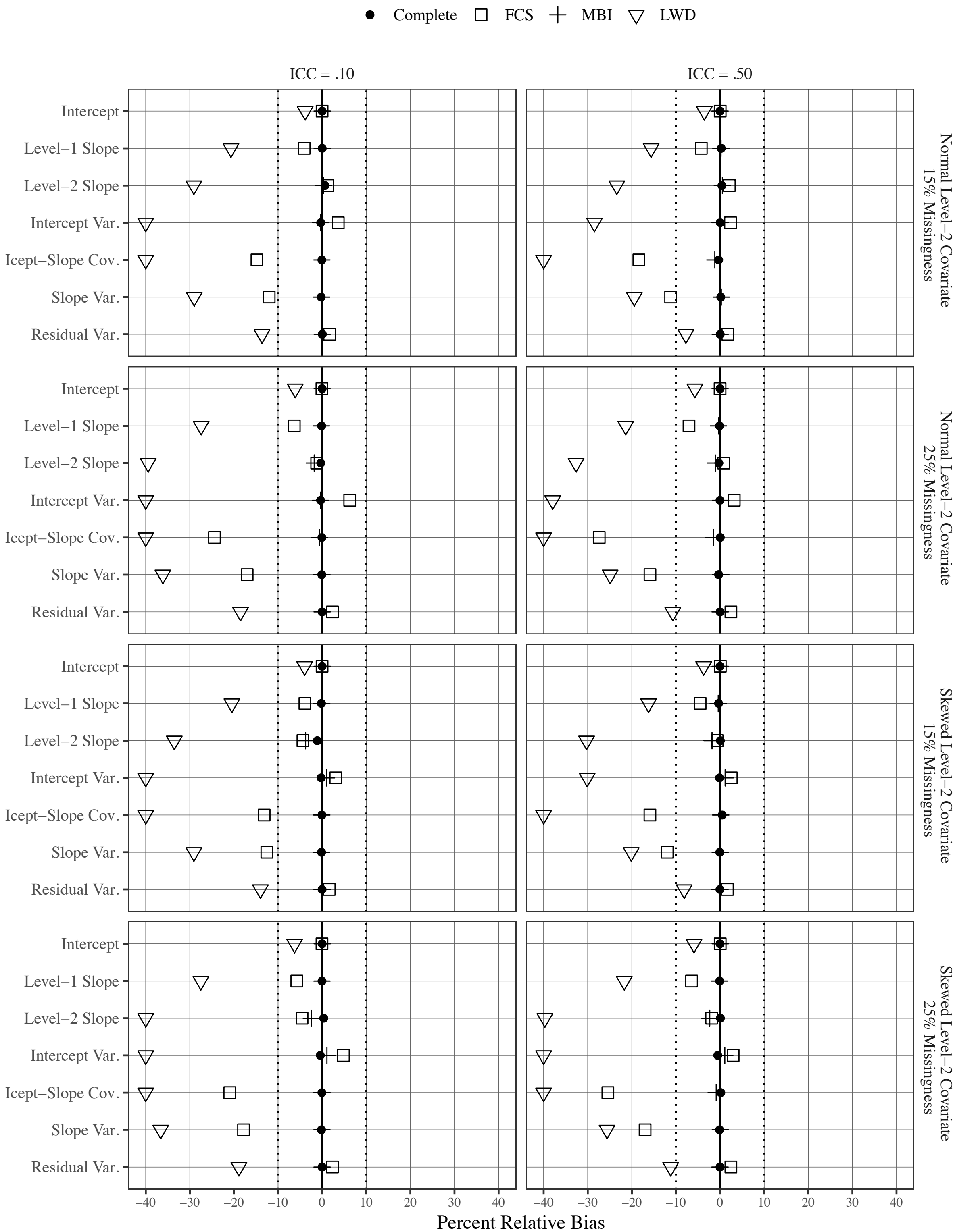
Table 1

Parameter Estimates from the Real Data Analysis Example

	Bayes		MBI		MBI + AV		FCS	
Parameter	Mean	*SD*	Est.	*SE*	Est.	*SE*	Est.	*SE*
Fixed Effect Coefficients								
Intercept	46.468	0.594	46.516	0.577	46.541	0.587	46.613	0.571
Self-Efficacy Slope	0.511	0.055	0.508	0.054	0.499	0.056	0.498	0.053
Monthly Change Slope	0.406	0.043	0.410	0.039	0.405	0.039	0.406	0.042
Lunch Assistance Slope	-0.916	0.306	-0.933	0.285	-0.899	0.290	-0.925	0.290
Condition Slope	-0.467	0.270	-0.474	0.267	-0.441	0.266	-0.487	0.272
Months by Condition Slope	0.330	0.053	0.326	0.051	0.331	0.051	0.331	0.054
Level-2 (Student-Level) Variance Components								
Intercept Variance	33.877	8.268	30.225	7.853	31.993	8.120	30.862	8.034
Self-Efficacy Slope Variance	0.447	0.105	0.400	0.098	0.408	0.100	0.394	0.100
Months Slope Variance	0.137	0.032	0.129	0.033	0.128	0.032	0.128	0.032
Level-1 Residual Variance	12.225	0.274	12.206	0.347	12.209	0.346	12.238	0.348

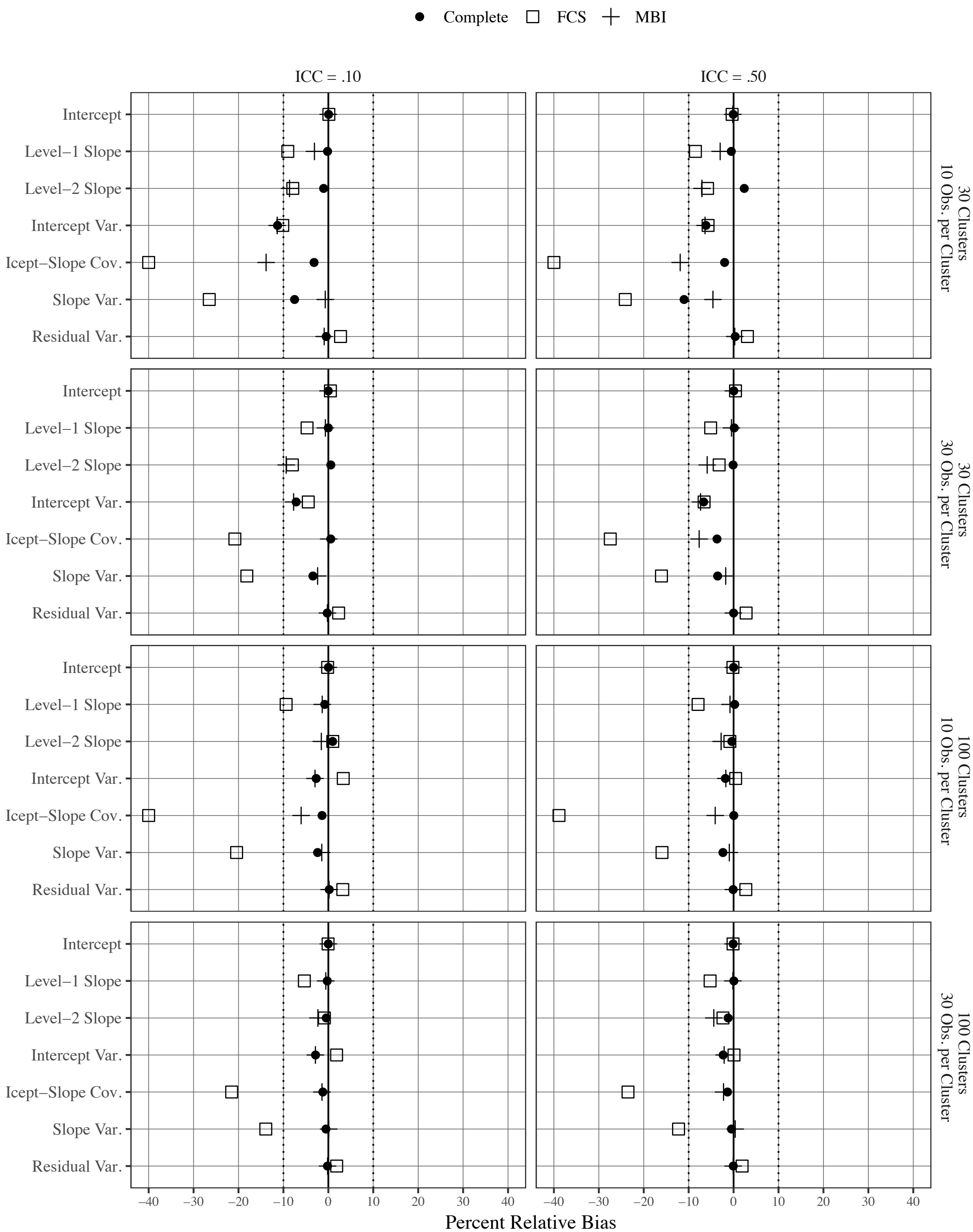
Bayes = Bayesian posterior summary, MBI = model-based imputation, MBI + AV = model-based imputation with auxiliary variables, FCS = fully conditional specification.

Figure 1. Average relative bias values from the large-sample simulation featuring a random coefficient model with either a normal or skewed level-2 predictor. The dashes represent bias values of ± 0.10. FCS = fully conditional specification ("reverse random coefficient" imputation), MBI = model-based imputation, LWD = listwise deletion.



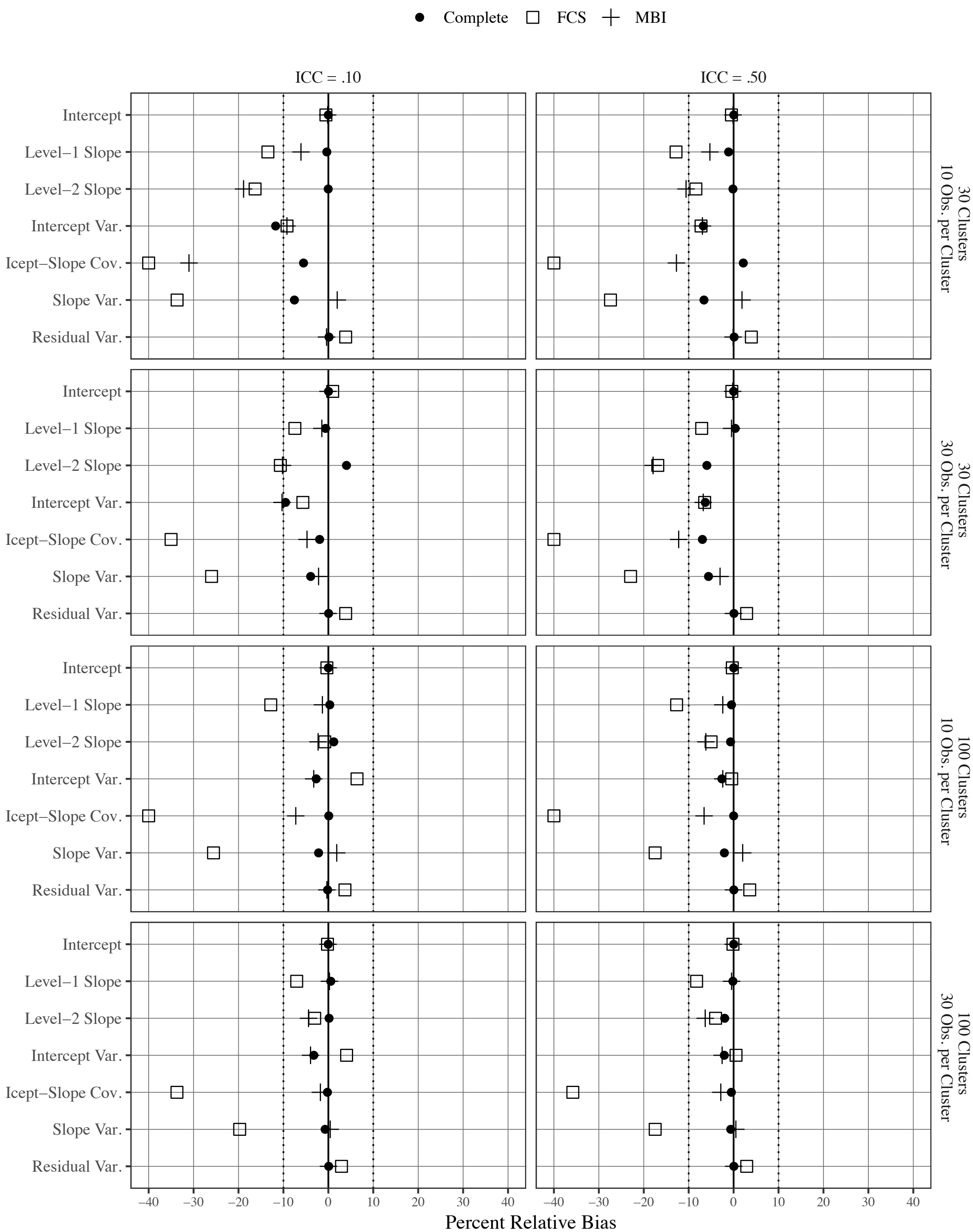
Complete
FCS
MBI
LWD
ICC = .10
ICC = .50
Intercept
Level−1 Slope
Level−2 Slope
Intercept Var.
Icept−Slope Cov.
Slope Var.
Residual Var.
Normal Level−2 Covariate 15% Missingness
Normal Level−2 Covariate 25% Missingness
Skewed Level−2 Covariate 15% Missingness
Skewed Level−2 Covariate 25% Missingness
−40 −30 −20 −10 0 10 20 30 40
Percent Relative Bias

Figure 2. Average relative bias values from the simulation featuring a random coefficient model with normally distributed predictors and 15% missing data. The dashes represent bias values of ± 0.10. FCS = fully conditional specification ("reverse random coefficient" imputation), MBI = model-based imputation.



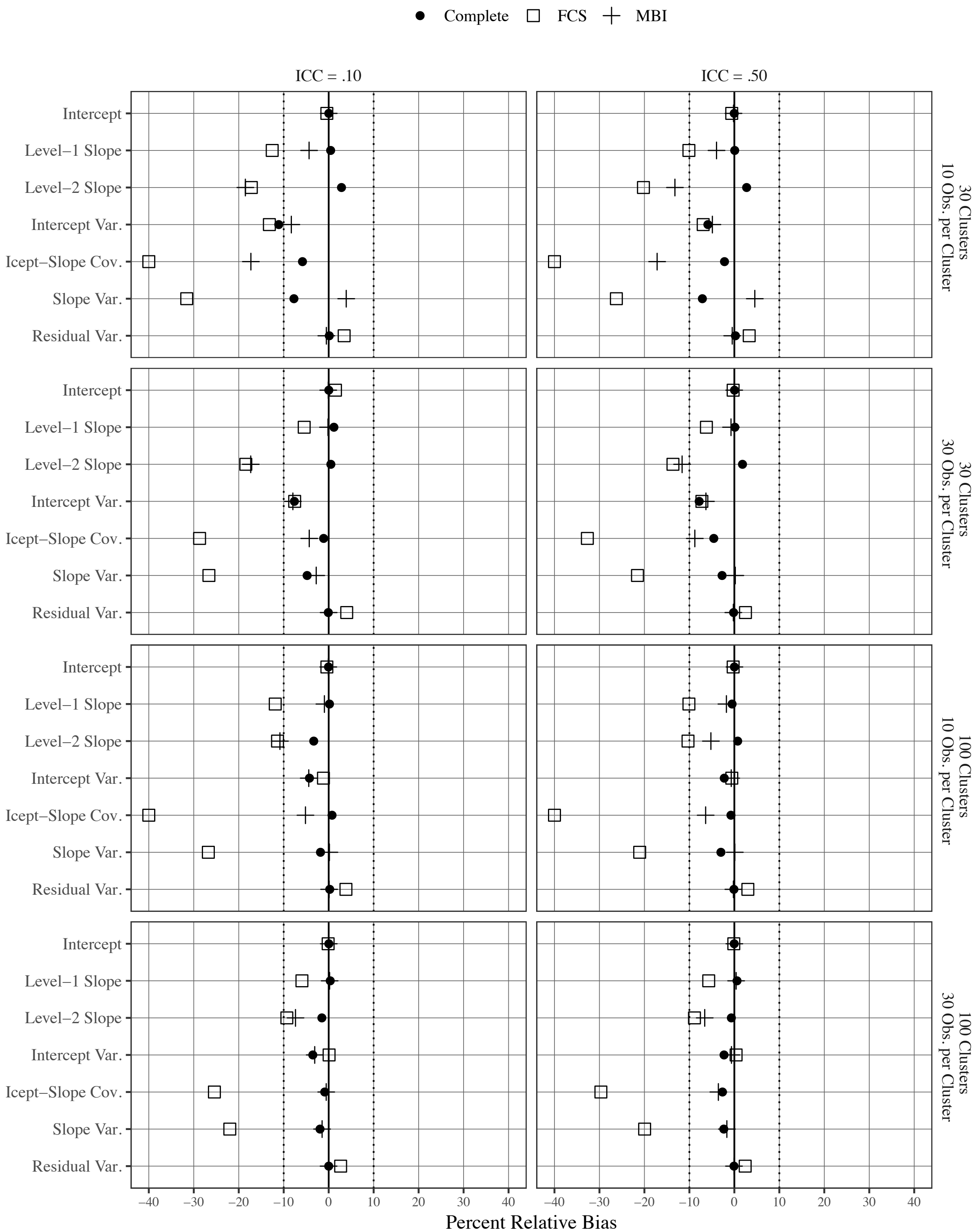
Complete
FCS
MBI
ICC = .10
ICC = .50
Intercept
Level−1 Slope
Level−2 Slope
Intercept Var.
Icept−Slope Cov.
Slope Var.
Residual Var.
30 Clusters
10 Obs. per Cluster
30 Clusters
30 Obs. per Cluster
100 Clusters
10 Obs. per Cluster
100 Clusters
30 Obs. per Cluster
−40
−30
−20
−10
0
10
20
30
40
Percent Relative Bias

Figure 3. Average relative bias values from the simulation featuring a random coefficient model with normally distributed predictors and 25% missing data. The dashes represent bias values of ± 0.10. FCS = fully conditional specification ("reverse random coefficient" imputation), MBI = model-based imputation.



Complete
FCS
MBI
ICC = .10
ICC = .50
Intercept
Level–1 Slope
Level–2 Slope
Intercept Var.
Icept–Slope Cov.
Slope Var.
Residual Var.
30 Clusters
10 Obs. per Cluster
30 Clusters
30 Obs. per Cluster
100 Clusters
10 Obs. per Cluster
100 Clusters
30 Obs. per Cluster
–40
–30
–20
–10
0
10
20
30
40
Percent Relative Bias

Figure 4. Average relative bias values from the simulation featuring a random coefficient model with a skewed level-2 predictor and 25% missing data. The dashes represent bias values of ± 0.10. FCS = fully conditional specification ("reverse random coefficient" imputation), MBI = model-based imputation.



Complete
FCS
MBI
ICC = .10
ICC = .50
Intercept
Level−1 Slope
Level−2 Slope
Intercept Var.
Icept−Slope Cov.
Slope Var.
Residual Var.
30 Clusters
10 Obs. per Cluster
30 Clusters
30 Obs. per Cluster
100 Clusters
10 Obs. per Cluster
100 Clusters
30 Obs. per Cluster
−40 −30 −20 −10 0 10 20 30 40
Percent Relative Bias

Figure 5. Confidence interval coverage for the fixed effects from a random coefficient model with 25% missing data. The dashes at .925 and .975 represent Bradley's (1978) so-called liberal criterion. FCS = fully conditional specification ("reverse random coefficient" imputation), MBI = model-based imputation.

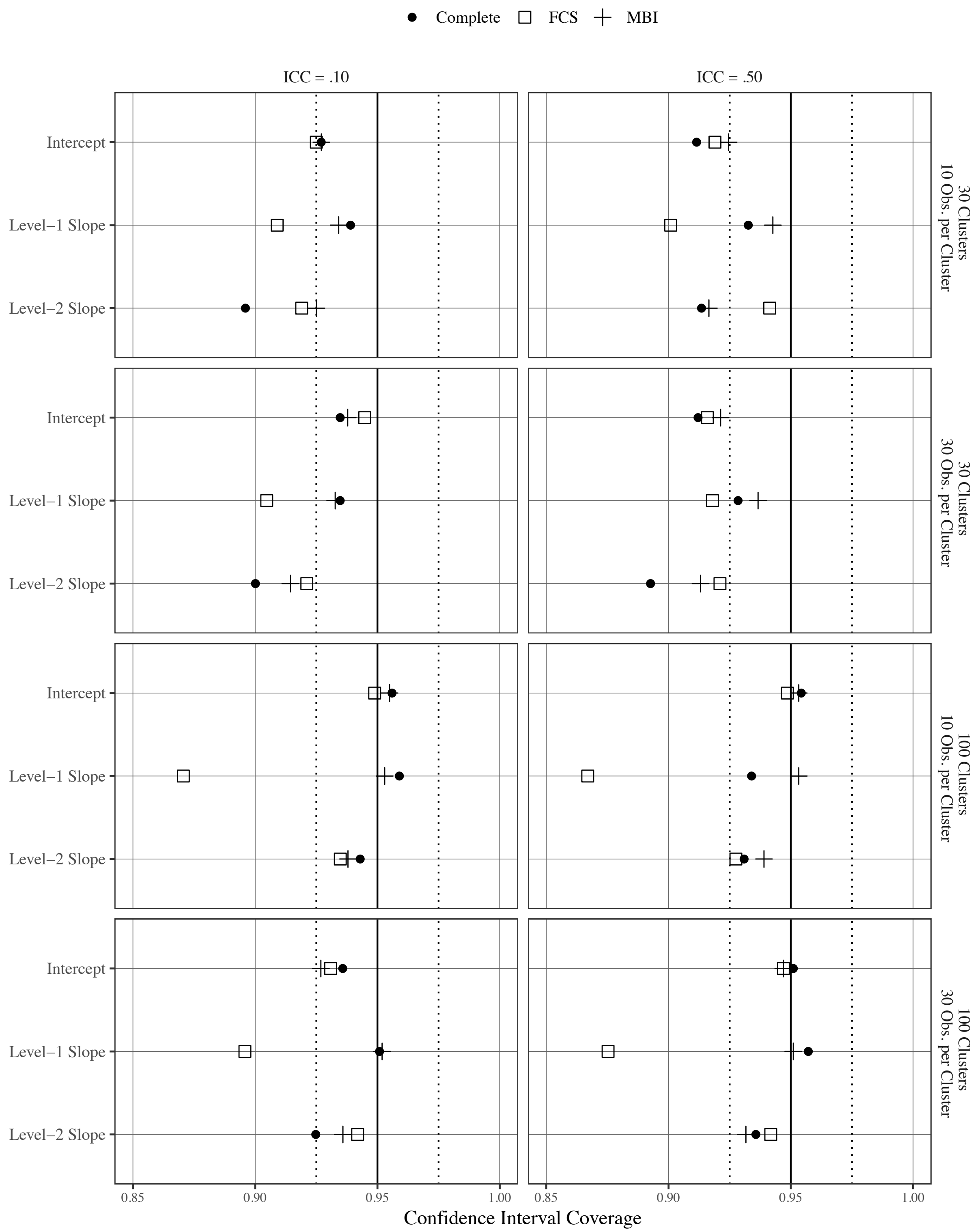
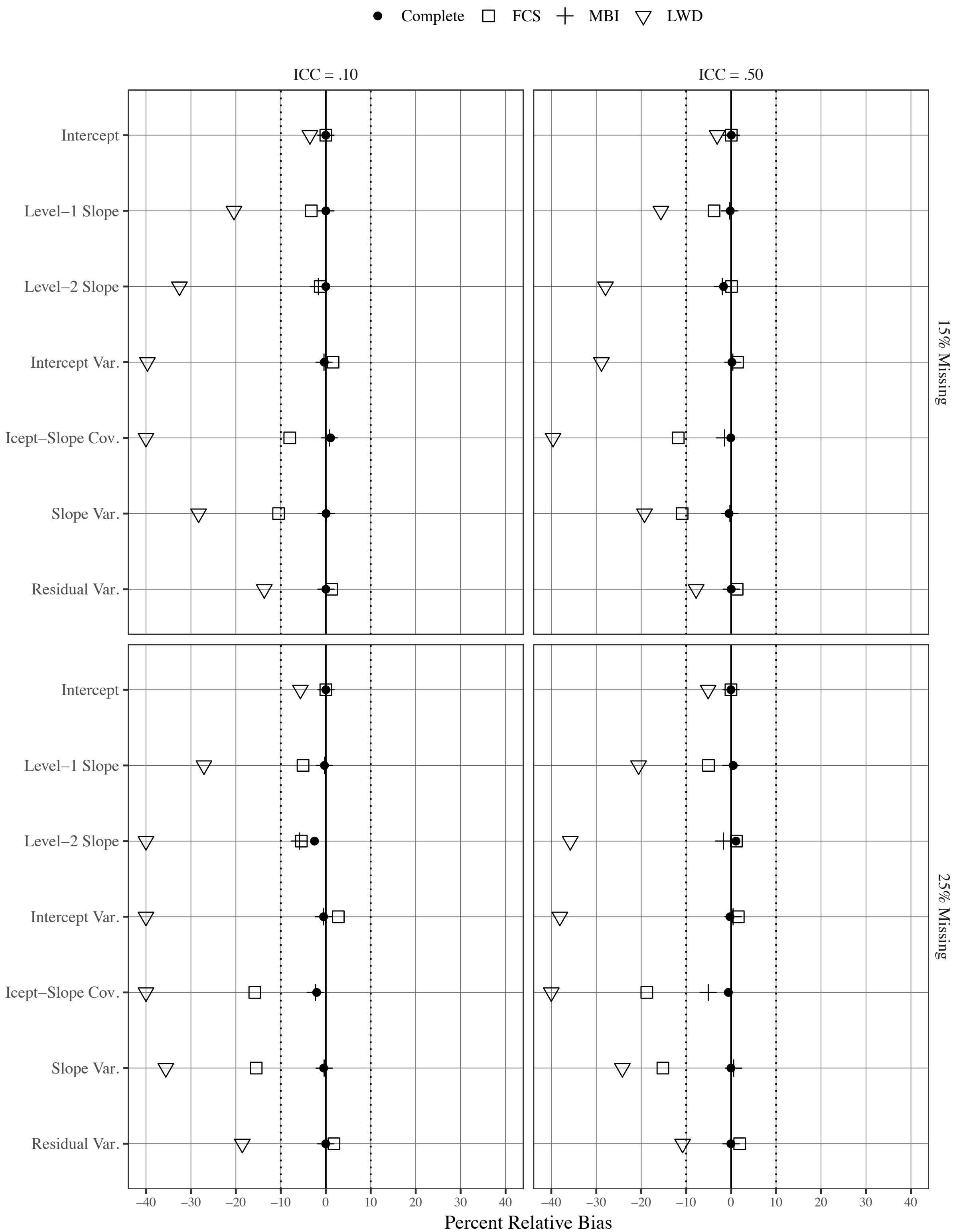

Complete
FCS
MBI
ICC = .10
ICC = .50
Intercept
Level−1 Slope
Level−2 Slope
30 Clusters
10 Obs. per Cluster
30 Clusters
30 Obs. per Cluster
100 Clusters
10 Obs. per Cluster
100 Clusters
30 Obs. per Cluster
0.85
0.90
0.95
1.00
Confidence Interval Coverage

Figure 6. Average relative bias values from the large-sample simulation featuring a random coefficient model with an incomplete binary level-2 predictor. The dashes represent bias values of ± 0.10. FCS = fully conditional specification ("reverse random coefficient" imputation), MBI = model-based imputation, LWD = listwise deletion.



Complete
FCS
MBI
LWD
ICC = .10
ICC = .50
Intercept
Level–1 Slope
Level–2 Slope
Intercept Var.
Icept–Slope Cov.
Slope Var.
Residual Var.
15% Missing
25% Missing
–40
–30
–20
–10
0
10
20
30
40
Percent Relative Bias

Figure 7. Average relative bias values for the fixed effects from the three-level simulation featuring random coefficients, a cross-level interaction, and 25% missing data. The dashes represent bias values of ± 0.10. FCS = fully conditional specification ("just another variable" imputation), MBI = model-based imputation.

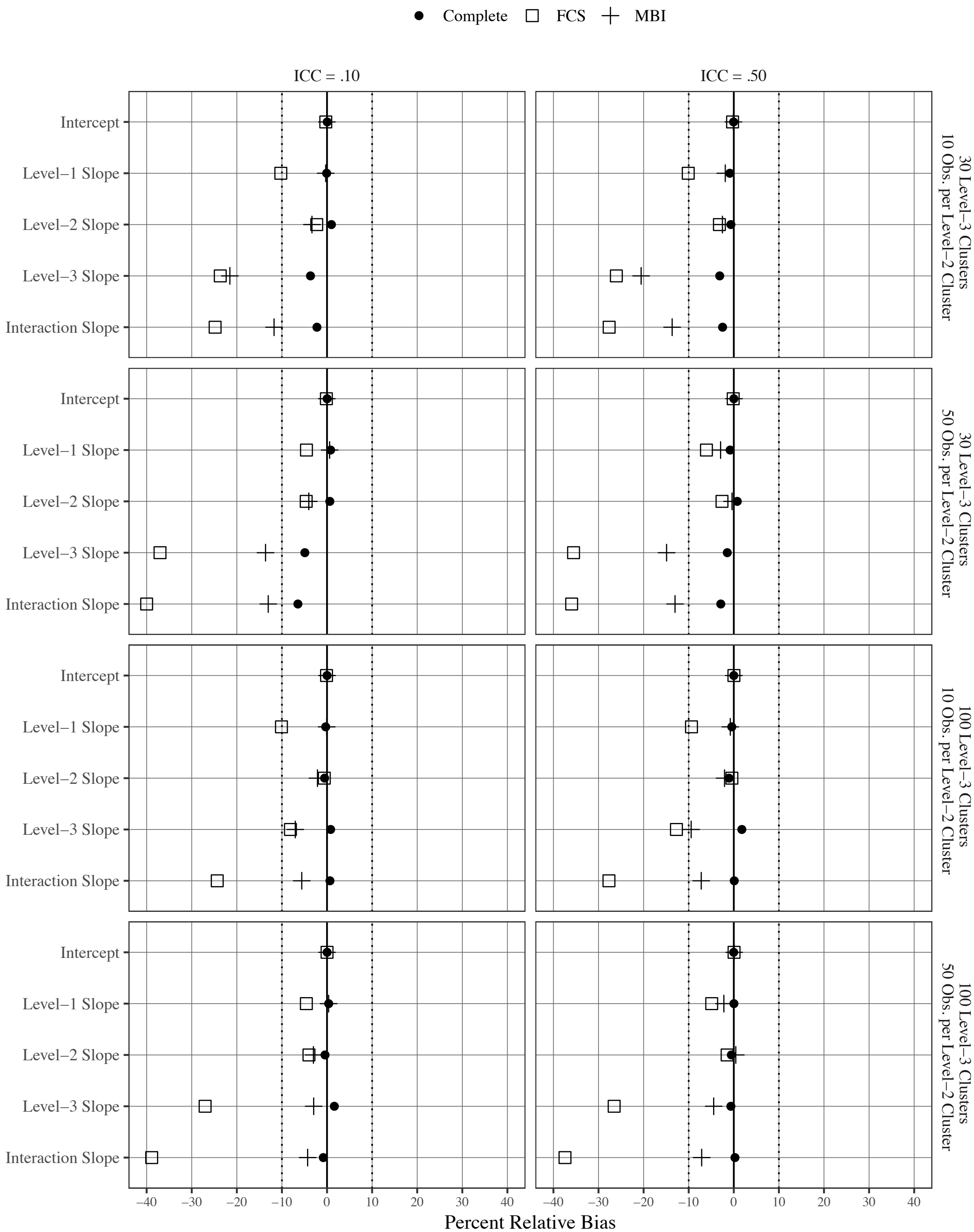
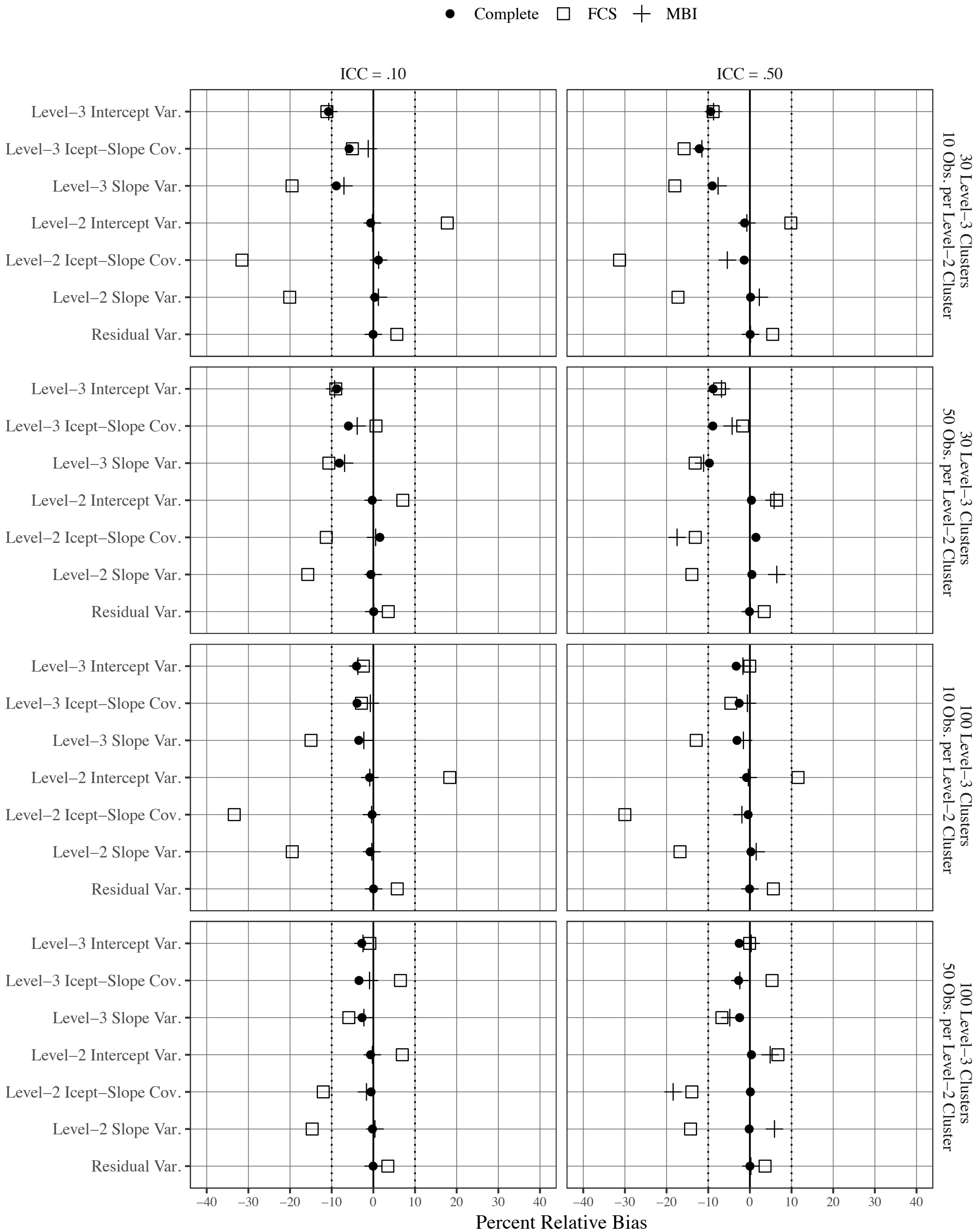
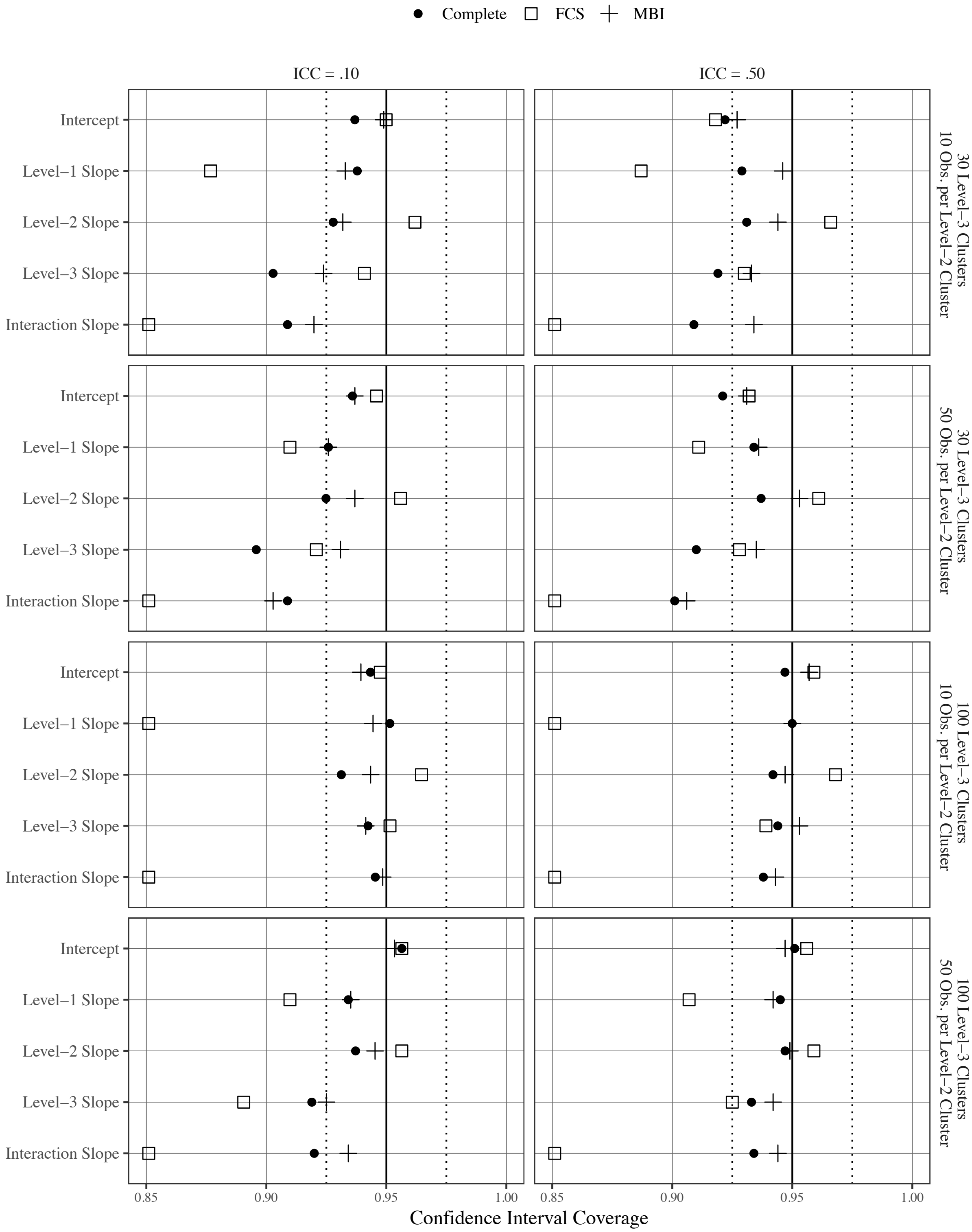

Complete
FCS
MBI
ICC = .10
ICC = .50
Intercept
Level−1 Slope
Level−2 Slope
Level−3 Slope
Interaction Slope
30 Level−3 Clusters
10 Obs. per Level−2 Cluster
30 Level−3 Clusters
50 Obs. per Level−2 Cluster
100 Level−3 Clusters
10 Obs. per Level−2 Cluster
100 Level−3 Clusters
50 Obs. per Level−2 Cluster
−40
−30
−20
−10
0
10
20
30
40
Percent Relative Bias

Figure 8. Average relative bias values for the variance components from the three-level simulation featuring random coefficients, a cross-level interaction, and 25% missing data. The dashes represent bias values of ± 0.10. FCS = fully conditional specification ("just another variable" imputation), MBI = model-based imputation.



Complete
FCS
MBI
ICC = .10
ICC = .50
Level−3 Intercept Var.
Level−3 Icept−Slope Cov.
Level−3 Slope Var.
Level−2 Intercept Var.
Level−2 Icept−Slope Cov.
Level−2 Slope Var.
Residual Var.
30 Level−3 Clusters 10 Obs. per Level−2 Cluster
30 Level−3 Clusters 50 Obs. per Level−2 Cluster
100 Level−3 Clusters 10 Obs. per Level−2 Cluster
100 Level−3 Clusters 50 Obs. per Level−2 Cluster
−40 −30 −20 −10 0 10 20 30 40
Percent Relative Bias

Figure 9. Confidence interval coverage for the fixed effects from the three-level simulation featuring random coefficients, a cross-level interaction, and 25% missing data. The dashes at .925 and .975 represent Bradley's (1978) so-called liberal criterion. FCS = fully conditional specification ("just another variable" imputation), MBI = model-based imputation.


Complete
FCS
MBI
ICC = .10
ICC = .50
Intercept
Level−1 Slope
Level−2 Slope
Level−3 Slope
Interaction Slope
30 Level−3 Clusters
10 Obs. per Level−2 Cluster
30 Level−3 Clusters
50 Obs. per Level−2 Cluster
100 Level−3 Clusters
10 Obs. per Level−2 Cluster
100 Level−3 Clusters
50 Obs. per Level−2 Cluster
0.85
0.90
0.95
1.00
Confidence Interval Coverage

Online Supplemental Material

Craig K. Enders, Han Du, and Brian T. Keller

- Section A: Bayesian estimation steps and full conditional distributions for model-based imputation
- Section B: Description of model-based imputation for 3-level computer simulation model
- Section C: Trellis plots displaying relative bias and confidence interval coverage from all conditions of the computer simulation.

A. MCMC Sampling Steps and Distributions for Two-Level Imputation

This document gives technical details of the two-level Gibbs sampler, specifically the full conditional distributions used to draw model parameters, random effects, latent means, and missing values for model-based imputation in Blimp.

Gibbs Sampler Steps for the Analysis Model

In this section we abandon the scalar notation from the manuscript in favor of a more succinct matrix representation of the multilevel model

$$\boldsymbol{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{b}_j + \boldsymbol{\varepsilon}_j \tag{SA1}$$

where $\boldsymbol{y}_j$ is the vector of outcome scores for cluster j, $\mathbf{X}_j$ is the corresponding matrix of predictor variables (level-1 or level-2), including a unit vector for the intercept, $\mathbf{Z}_j$ is a subset of the level-1 variables in $\mathbf{X}_j$ that have a random influence on the outcome (e.g., a unit vector and any random coefficient predictors), $\boldsymbol{b}_j$ is the column vector of level-2 residuals for cluster j, and $\boldsymbol{\varepsilon}_j$ is a vector of within-cluster residuals. A variety of sources give the full conditional distributions for this model (Browne, 1998; Browne & Draper, 2000; Enders et al., 2018; Lynch, 2007; Schafer & Yucel, 2002; Yucel, 2008), which we summarize here for completeness.

To illustrate the following steps more concretely, we will refer to the following substantive model, which includes within-cluster and cross-level interaction effects. Between-cluster interactions involving pairs of level-2 variables are also possible, but it should become evident that the composition of the analysis model has no bearing on the covariate models.

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1(x_{1ij}) + \beta_2(x_{2ij}) + \beta_3(x_{3ij}) + \beta_4(x_{1ij})(x_{2ij}) + \beta_5(x_{4j}) \\ &\quad + \beta_6(x_{5j}) + \beta_7(x_{3ij})(x_{5j}) + b_{0j} + b_{1j}(x_{1ij}) + \varepsilon_{ij} \\ &\begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} \sim MN(0, \mathbf{\Sigma}_b) \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \end{aligned} \tag{SA2}$$

Step 1: Draw regression coefficients from $p(\boldsymbol{\beta}|\cdot) \propto p(Y|\boldsymbol{\beta}, \boldsymbol{b}_j, \mathbf{\Sigma}_b, \sigma_\varepsilon^2, X)p(\boldsymbol{\beta})$. Assuming a uniform prior, $p(\boldsymbol{\beta}) \propto 1$, the full conditional distribution is a multivariate normal distribution.

$$\begin{aligned} \boldsymbol{\beta} &\sim MN\left(\hat{\boldsymbol{\beta}}, \mathbf{\Sigma}_{\hat{\boldsymbol{\beta}}}\right) \\ \hat{\boldsymbol{\beta}} &= \left(\sum_{j=1}^{J} \mathbf{X}_j' \mathbf{X}_j\right)^{-1} \sum_{j=1}^{J} \mathbf{X}_j'(\boldsymbol{y}_j - \mathbf{Z}_j \boldsymbol{b}_j) \\ \mathbf{\Sigma}_{\hat{\boldsymbol{\beta}}} &= \sigma_\varepsilon^2 \left(\sum_{j=1}^{J} \mathbf{X}_j' \mathbf{X}_j\right)^{-1} \end{aligned} \tag{SA3}$$

Applied to the model from Equation SA2, the **X** matrix includes a unit vector for the intercept and a column for each explanatory variable and interaction term, the **Z** matrix is comprised of a unit vector and X_1.

Step 2: Draw random effects for cluster j from a multivariate normal distribution.

$$\begin{aligned} \boldsymbol{b}_j &\sim MN\left(\hat{\boldsymbol{b}}_j, \mathbf{V}_{\boldsymbol{b}_j}\right) \\ \mathbf{V}_{\boldsymbol{b}_j} &= (\sigma_\varepsilon^{-2}\ \mathbf{Z}_j' \mathbf{Z}_j + \mathbf{\Sigma}_b^{-1})^{-1} \\ \hat{\boldsymbol{b}}_j &= \sigma_\varepsilon^{-2}\ \mathbf{V}_{\boldsymbol{b}_j} \mathbf{Z}_j' (\boldsymbol{y}_j - \mathbf{X}_j \boldsymbol{\beta}) \end{aligned} \tag{SA4}$$

Step 3: Draw the residual variance from $p(\sigma_\varepsilon^2|\cdot) \propto p(Y|\boldsymbol{\beta}, \boldsymbol{b}_j, \boldsymbol{\Sigma}_b, \sigma_\varepsilon^2, X)p(\sigma_\varepsilon^2)$. We define $1/\sigma_\varepsilon^2$ as a gamma random variable and draw the reciprocal of the residual variance (i.e., the precision) from a gamma distribution

$$
\begin{aligned}
1/\sigma_\varepsilon^2 &\sim \mathrm{G}\left(\frac{N + df_p}{2}, \frac{S + S_p}{2}\right) \\
S &= \sum_{j=1}^{J} \boldsymbol{\varepsilon}_j' \boldsymbol{\varepsilon}_j \\
\boldsymbol{\varepsilon}_j &= \boldsymbol{y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mathbf{Z}_j\boldsymbol{b}_j
\end{aligned}
\tag{SA5}
$$

with hyperparameters df_p and S_p for the prior distribution. The default setting in Blimp specifies $S_p = 1$ and $df_p = 2$, which corresponds to a gamma(1,.5) prior. Two other options are to set $S_p = 0$ and $df_p = -2$ (the PRIOR2 keyword of the OPTIONS command) and $S_p = 0$ and $df_p = 0$ (the PRIOR3 keyword), a Jeffreys prior.

Step 4: Draw the between-cluster covariance matrix variance from $p(\boldsymbol{\Sigma}_b|\cdot) \propto p(Y|\boldsymbol{\beta}, \boldsymbol{b}_j, \boldsymbol{\Sigma}_b, \sigma_\varepsilon^2, X)p(\boldsymbol{\Sigma}_b)$. We define the inverse of the covariance matrix (i.e., the precision matrix, $\boldsymbol{\Sigma}_b^{-1}$) as a Wishart random variable. The level-2 precision matrix is sampled from a Wishart distribution, conditional on the current parameter estimates, level-2 residuals, and imputations.

$$
\begin{aligned}
\boldsymbol{\Sigma}_b^{-1} &\sim \mathrm{W}((\mathbf{S} + \mathbf{S}_p^{-1})^{-1}, J + df_p) \\
\mathbf{S} &= \sum_{j=1}^{J} \boldsymbol{b}_j' \boldsymbol{b}_j
\end{aligned}
\tag{SA6}
$$

$\mathbf{S}_p$ can be viewed as the *inverse* of the prior sums of squares matrix based on df_p degrees of freedom (i.e., prior observations). As such, $\mathbf{S} + \mathbf{S}_p^{-1}$ is a sums of squares and cross products matrix based on $J + df_p$ observations. The default prior sets $\mathbf{S}_p^{-1}$ $= \mathbf{I}$ and $df_p = p + 1$, where p is the dimension of $\mathbf{\Sigma}_b$. This prior corresponds to marginal uniform priors between -1 and 1 for all correlations and a marginal inverse gamma prior IG(1,.5) for variance elements. Specifying the PRIOR2 keyword of the OPTIONS command sets $\mathbf{S}_p^{-1} = 0$ and $df_p = -p - 1$, which is equivalent to a uniform prior on the elements in $\mathbf{\Sigma}_b$. Finally, the PRIOR3 keyword sets $\mathbf{S}_p^{-1} = 0$ and $df_p = 0$. For random intercept models with a single level-2 variance component, we draw the reciprocal of the variance, $1/\sigma_b^2$, from a gamma random variable with analogous univariate priors based on $p = 1$.

Step 5: Draw the imputation for observation i in cluster j from a univariate normal posterior distribution.

$$y_{ij(mis)} = N(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_j, \sigma_\varepsilon^2) \tag{SA7}$$

Gibbs Sampler Steps for the Covariate Model r

To convey the estimation steps in the most general way possible, we introduce new notation that differs from that in the manuscript. To begin, index the P level-1 predictors as $p = 1, \ldots, P$, and index the Q level-2 predictors as $q = 1, \ldots, Q$. As explained in the paper, each level-1 variable has a regression model for the observations and a regression model for its latent cluster means. Thus, level-1 estimation involves P computational cycles, and level-2 estimation requires $R = P + Q$ computational cycles. To simplify the notation, we index the entire set of variables as $r = 1, \ldots, R$, such that $r \leq P$ corresponds to either a level-1 observation or its corresponding level-2 group mean, and $r > P$ refers to a manifest level-2 variable.

Using generic notation, the level-1 and level-2 regression models are given below.

$$\begin{gathered} x_{r,ij} = \mu_{r,j} + \tilde{\boldsymbol{x}}_{-r,ij}\boldsymbol{\gamma}_r + e_{r,ij} \\ e_{r,ij} \sim N(0, \sigma^2_{e_r}) \end{gathered} \tag{SA8}$$

$$\begin{gathered} x_{r,j} = \mu_r + \tilde{\boldsymbol{x}}_{-r,j}\boldsymbol{\eta}_r + \zeta_{r,j} \\ \zeta_{r,j} \sim N(0, \sigma^2_{\zeta_r}) \end{gathered} \tag{SA9}$$

where $x_{r,ij}$ is the level-1 score for covariate r, $\tilde{\boldsymbol{x}}_{-r,ij}$ denotes the $P - 1$ row vector of all other level-1 predictor variables except r, centered at their latent group means. Turning to the between-cluster regression, the outcome, $x_{r,j}$, is either a latent group mean (e.g., $x_{r,j} = \mu_{r,j}$) when $r \leq P$ or a manifest level-2 variable when $r > P$, and $\tilde{\boldsymbol{x}}_{-r,j}$ is a $R - 1$ row vector of grand-mean centered level-2 variables other than r. For some of the Gibbs sampling steps, it is convenient to concatenate observation-level quantities into matrices. For example, the N-row vector of level-1 outcome scores is $\boldsymbol{x}_{r,ij}$ and the corresponding N by $P - 1$ matrix of centered level-1 predictors is $\widetilde{\boldsymbol{X}}_{-r,ij}$. Similarly, the J-row vector of level-2 outcome scores is $\boldsymbol{x}_{r,j}$ and the J by R matrix of mean-centered predictors is $\widetilde{\boldsymbol{X}}_{-r,j}$.

To make the notation more concrete, consider the analysis model from Equation SA2. The multivariate normality assumption induces the following level-1 regression models.

$$\begin{aligned} x_{1ij} &= \mu_{1j} + \gamma_{11}(x_{2ij} - \mu_{2j}) + \gamma_{12}(x_{3ij} - \mu_{3j}) + e_{1ij} \\ x_{2ij} &= \mu_{2j} + \gamma_{21}(x_{1ij} - \mu_{1j}) + \gamma_{22}(x_{3ij} - \mu_{3j}) + e_{2ij} \\ x_{3ij} &= \mu_{3j} + \gamma_{31}(x_{1ij} - \mu_{1j}) + \gamma_{32}(x_{2ij} - \mu_{2j}) + e_{3ij} \end{aligned} \tag{SA10}$$

For any given model, $\boldsymbol{x}_{r,ij}$ is N-row vector of outcome scores, $\boldsymbol{\gamma}_r$ is a 2-element vector of level-1 regression coefficients, and $\widetilde{\boldsymbol{X}}_{-r,ij}$ is the N by 2 matrix of latent group mean centered predictors on the right side of each equation. The multivariate normality assumption also induces the following level-2 regression models.

$$\begin{aligned}
\mu_{1j} &= \mu_1 + \eta_{11}(\mu_{2j} - \mu_2) + \eta_{12}(\mu_{3j} - \mu_3) + \eta_{13}(x_{4j} - \mu_4) + \eta_{14}(x_{5j} - \mu_5) + \zeta_{1j} \\
\mu_{2j} &= \mu_2 + \eta_{21}(\mu_{1j} - \mu_1) + \eta_{22}(\mu_{3j} - \mu_3) + \eta_{23}(x_{4j} - \mu_4) + \eta_{24}(x_{5j} - \mu_5) + \zeta_{2j} \\
\mu_{3j} &= \mu_3 + \eta_{31}(\mu_{1j} - \mu_1) + \eta_{32}(\mu_{2j} - \mu_2) + \eta_{33}(x_{4j} - \mu_4) + \eta_{34}(x_{5j} - \mu_5) + \zeta_{3j} \\
x_{4j} &= \mu_4 + \eta_{41}(\mu_{1j} - \mu_1) + \eta_{42}(\mu_{2j} - \mu_2) + \eta_{43}(x_{3j} - \mu_3) + \eta_{44}(x_{5j} - \mu_5) + \zeta_{4j} \\
x_{5j} &= \mu_5 + \eta_{51}(\mu_{1j} - \mu_1) + \eta_{52}(\mu_{2j} - \mu_2) + \eta_{53}(x_{3j} - \mu_3) + \eta_{54}(x_{4j} - \mu_4) + \zeta_{5j}
\end{aligned} \tag{SA11}$$

For any given model, $\boldsymbol{x}_{r,j}$ is the J-row vector of outcome scores (latent means or manifest variables), $\boldsymbol{\eta}_r$ is a 4-element vector of level-2 regression coefficients, and $\widetilde{\boldsymbol{X}}_{-r,j}$ is the J by 4 matrix of grand mean centered predictors on the right side of each equation.

Step 6. If variable r is measured at level-1 (i.e., $r \leq P$), draw its latent cluster means from $p(\mu_{r,j}|\cdot) \propto p(X_r|\mu_r, \mu_{r,j}, \boldsymbol{\gamma}_r, \sigma^2_{e_r}, X_{-r})p(\mu_{r,j}|, \boldsymbol{x}_{-r,j}, \boldsymbol{\eta}_r, \sigma^2_{\zeta_r})$, which is the univariate normal distribution below.

$$p(\mu_{r,j}|\cdot) = N\left(\frac{\sigma^2_{\zeta_r}\sum_{i=1}^{n_j}(x_{r,ij} - \tilde{\boldsymbol{x}}_{-r,ij}\boldsymbol{\gamma}_r) + \sigma^2_{e_r}(\mu_r + \tilde{\boldsymbol{x}}_{-r,j}\boldsymbol{\eta}_r)}{\sigma^2_{e_r} + n_j\sigma^2_{\zeta_r}}, \frac{\sigma^2_{e_r}\sigma^2_{\zeta_r}}{\sigma^2_{e_r} + n_j\sigma^2_{\zeta_r}}\right) \tag{SA12}$$

Note that $\sigma^2_{e_r}$ and $\sigma^2_{\zeta_r}$ residual variance for covariate r's observations and latent means, respectively.

Step 7: Draw the grand mean of variable r from $p(\mu_r | \cdot) \propto p\left(X_r | \mu_r, \boldsymbol{\eta}_r, \sigma^2_{\zeta_r}, \mu_{r,j}, \boldsymbol{\gamma}_r, \sigma^2_{e_r}, X_{-r}\right) p(\mu_r)$. Assuming a uniform prior, $p(\mu_r) \propto 1$, the full conditional distribution is a univariate normal distribution.

$$p(\mu_r | \cdot) = N\left(\frac{\sum_{j=1}^{J}\left(x_{r,j} - \tilde{\boldsymbol{x}}_{-r,j}\boldsymbol{\eta}_r\right)}{J}, \frac{\sigma^2_{\zeta_r}}{J}\right) \tag{SA13}$$

where J the number of level-2 units. As noted previously, $x_{r,j}$ is a latent cluster mean when variable r is measured at level-1, and it is a score when r is at level-2.

Step 8: If variable r is measured at level-1 (i.e., $r \leq P$), draw its within-cluster regression slopes from $p(\boldsymbol{\gamma}_r | \cdot) \propto p\left(X_r | \mu_r, \boldsymbol{\eta}_r, \sigma^2_{\zeta_r}, \mu_{r,j}, \boldsymbol{\gamma}_r, \sigma^2_{e_r}, X_{-r}\right) p(\boldsymbol{\gamma}_r)$. Because latent group means replace the regression intercept, the level-1 regression requires that the outcome (i.e., a level-1 score) is centered at its current group mean from step 9. In line with our previous notation, we denote the N-row vector of centered outcome scores as $\tilde{\boldsymbol{x}}_{r,ij}$. Assuming independent uniform priors, $p(\boldsymbol{\gamma}_r) \propto 1$, the full conditional distribution is a multivariate normal distribution.

$$p(\boldsymbol{\gamma}_r | \cdot) = MN\left(\hat{\boldsymbol{\gamma}}_r, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\gamma}}_r}\right) \tag{SA14}$$

$$\hat{\boldsymbol{\gamma}}_r = \left(\widetilde{\boldsymbol{X}}'_{-r,ij}\widetilde{\boldsymbol{X}}_{-r,ij}\right)^{-1}\widetilde{\boldsymbol{X}}'_{-r,ij}\tilde{\boldsymbol{x}}_{r,ij} \tag{SA15}$$

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\gamma}}_r} = \sigma^2_{e_r}\left(\widetilde{\boldsymbol{X}}'_{-r,ij}\widetilde{\boldsymbol{X}}_{-r,ij}\right)^{-1} \tag{SA16}$$

Note that it is not necessary to account for clustering here because all scores are centered at their latent group means.

Step 9: If variable r is measured at level-1 (i.e., $r \leq P$), draw the within-cluster residual variance from $p(\sigma^2_{e_r}|\cdot) \propto p(X_r|\mu_r, \boldsymbol{\eta}_r, \sigma^2_{\zeta_r}, \mu_{r,j}, \boldsymbol{\gamma}_r, \sigma^2_{e_r}, X_{-r})p(\sigma^2_{e_r})$. First, define a level-1 residual as follows

$$\hat{e}_{r,ij} = x_{r,ij} - \mu_{r,j} - \tilde{\boldsymbol{x}}_{-r,ij}\boldsymbol{\gamma}_r \tag{SA17}$$

and stack these residuals into an N-row vector define a N-row vector $\hat{\boldsymbol{e}}_{r,ij}$. We define $1/\sigma^2_e$ as a gamma random variable and draw the reciprocal of the residual variance (i.e., precision) from a gamma distribution

$$1/\sigma^2_e \sim \mathrm{G}\left(\frac{N + df_p}{2}, \frac{\hat{\boldsymbol{e}}'_{r,j}\hat{\boldsymbol{e}}_{r,j} + S_p}{2}\right) \tag{SA18}$$

with hyperparameters df_p and S_p for the prior distribution. The default setting in Blimp (XPRIOR1) specifies $S_p = 1$ and $df_p = 2$, which corresponds to a gamma(1,.5) prior. Two other options are to set $S_p = 0$ and $df_p = -2$ (the XPRIOR2 keyword of the OPTIONS command) and $S_p = 0$ and $df_p = 0$ (the XPRIOR3 keyword), a Jeffreys prior. Our simulation used a Jeffreys prior, but we have found that the default setting prevents between-cluster variances from collapsing when covariates have very low intraclass correlations.

Step 10: Draw between-cluster regression slopes for covariate r from $p(\boldsymbol{\eta}_r|\cdot) \propto p(X_r|\mu_r, \boldsymbol{\eta}_r, \sigma^2_{\zeta_r}, \mu_{r,j}, \boldsymbol{\gamma}_r, \sigma^2_{e_r}, X_{-r})p(\boldsymbol{\eta}_r)$. Because the grand mean replaces the fixed regression intercept, the level-2 regression requires the outcome (i.e., a latent group mean or level-2 score) to be centered at its grand mean from step 6. In line with our previous notation, we denote the J-row vector of centered outcome scores as $\tilde{\boldsymbol{x}}_{r,j}$.

Assuming independent uniform priors, $p(\boldsymbol{\eta}_r) \propto 1$, the full conditional distribution is a multivariate normal distribution

$$p(\boldsymbol{\eta}_r|\cdot) = MN(\hat{\boldsymbol{\eta}}_r, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\eta}}_r}) \tag{SA19}$$

$$\hat{\boldsymbol{\eta}}_r = \left(\widetilde{\boldsymbol{X}}'_{-r,j}\widetilde{\boldsymbol{X}}_{-r,j}\right)^{-1}\widetilde{\boldsymbol{X}}'_{-r,j}\tilde{\boldsymbol{x}}_{r,j} \tag{SA20}$$

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\eta}}_r} = \sigma^2_{\zeta_r}\left(\widetilde{\boldsymbol{X}}'_{-r,j}\widetilde{\boldsymbol{X}}_{-r,j}\right)^{-1} \tag{SA21}$$

Step 11: Draw the between-cluster residual variance for covariate r from $p(\sigma^2_{\zeta_r}|\cdot) \propto p(X_r|\mu_r, \boldsymbol{\eta}_r, \sigma^2_{\zeta_r}, \mu_{r,j}, \boldsymbol{\gamma}_r, \sigma^2_{e_r}, X_{-r})p(\sigma^2_{\zeta_r})$. First, define a J-row vector of level-2 residuals as

$$\hat{\boldsymbol{\zeta}}_{r,j} = \boldsymbol{x}_{r,j} - \mathbf{1}\mu_r - \widetilde{\boldsymbol{X}}_{-r,j}\boldsymbol{\eta}_r \tag{SA22}$$

where $\mathbf{1}$ is a J-row unit vector. To reiterate, $\boldsymbol{x}_{r,j}$ contains latent group means ($r \leq P$) or manifest level-2 variables ($r > P$). We define $1/\sigma^2_{\zeta_r}$ as a gamma random variable and draw the reciprocal of the residual variance (i.e., the precision) from a gamma distribution

$$1/\sigma^2_{\zeta_r} \sim \mathrm{G}\left(\frac{J + df_p}{2}, \frac{\hat{\boldsymbol{\zeta}}'_{r,j}\hat{\boldsymbol{\zeta}}_{r,j} + S_p}{2}\right) \tag{SA23}$$

with hyperparameters df_p and S_p for the prior distribution. The default setting in Blimp (XPRIOR1) specifies $S_p = 1$ and $df_p = 2$, which corresponds to a gamma(1,.5) prior. Two other options are to set $S_p = 0$ and $df_p = -2$ (the XPRIOR2 keyword of the OPTIONS command) and $S_p = 0$ and $df_p = 0$ (the XPRIOR3 keyword), a

Jeffreys prior. Note keywords induce the same priors at all levels of the data hierarchy.

Step 12: Draw missing values from $p(X_r) \propto p(Y|X_r, X_{-r}) \times p(X_r|X_{-r})$. For each missing score, the Metropolis algorithm draws a candidate imputation $X_{r(candidate)}$ from a normal proposal distribution

$$X^{(t)}_{r(candidate)} = N\left(X^{(t)}_{r(current)}, \sigma^2_{(proposal)}\right) \tag{SA24}$$

where the mean $X^{(t)}_{r,i(current)}$ is the current imputation i at iteration t, and the variance $\sigma^2_{(proposal)}$ is chosen to optimize the acceptance rate of the candidate imputations. We have found that setting $\sigma^2_{(proposal)} = 9\left(\sigma^2_{e_r}\right)$ for level-1 variables and $\sigma^2_{(proposal)} = 2.25\left(\sigma^2_{\zeta_r}\right)$ for level-2 variables tends to give optimal acceptance rates, although the Blimp application adaptively tunes the spread of the proposal distribution by increasing or decreasing the constant multiplier at regular intervals during the burn-in phase in an attempt to achieve an acceptance rate for the imputations between 0.25 and 0.45 (Gelman et al., 2014; Johnson & Albert, 1999; Lynch, 2007). For each incomplete variable, Blimp checks the acceptance rate every 50 iterations by computing the proportion of accepted imputations for a particular variable across all incomplete observations during the 50-iteration interval (e.g., for 10 incomplete cases, the acceptance rate for a particular variable is the proportion of the 500 draws that are accepted). If the acceptance rate does not fall between 0.25 and 0.45, the program increases (if the acceptance rate is too high) or decreases (if the acceptance rate is too low) the variance multiplier. Once the burn-in iterations are complete, tuning checks are turned off. Normal proposal distributions are routinely used in Bayesian analysis texts (Gelman et al., 2014; Lynch, 2007), but a strong rationale for adopting this distribution in our context is that the correct conditional distribution is, in fact, normal (see manuscript Equations 8 and 20). Thus, the Metropolis algorithm provides a way to model the true distribution of missing values without deriving the complex non-linear functions that define its mean and variance.

After drawing a candidate imputation from the proposal distribution, the Metropolis algorithm calculates the natural logarithm of an importance ratio (IR) that quantifies the height of the target density evaluated at the candidate imputation proportional to its height when evaluated at the current imputation.

$$\begin{aligned} ln(\text{IR}) = & \Big[ln[p(Y \mid X_1, \ldots, X_{r(candiate)}, \ldots, X_R)] + ln[p(X_{r(candiate)} \mid X_{-r})]\Big] \\ & - \Big[ln[p(Y \mid X_1, \ldots, X_{r(current)}, \ldots, X_R)] + ln[p(X_{r(current)} \mid X_{-r})]\Big] \end{aligned} \quad \text{(SA25)}$$

Note that $p(Y \mid X_1, \ldots, X_{r(candiate)}, \ldots, X_R)$ and $(Y \mid X_1, \ldots, X_{r(current)}, \ldots, X_R)$ involve the product of likelihoods (or the sum of log likelihoods) when X_r is at level-2; $p(X_{r(candiate)} \mid X_{-r})$ and $p(X_{r(current)} \mid X_{-r})$ always evaluate a single observation. The importance ratio defines the probability of a Bernoulli random variable that determines whether the candidate value is retained as the current imputation for the next iteration. If the importance ratio exceeds unity, the candidate imputation is automatically accepted. If the ratio is large but less than one, the candidate imputation is likely to be accepted because it has a high probability of originating from $p(Y|X_R) \times p(X_r|X_{-r})$. As the ratio decreases, so too does the chance of retaining the candidate value because it is unlikely to originate from the target density. To account for the natural logarithm, the Metropolis sampler draws a random number u_i from a uniform distribution $U(0,1)$ and accepts $X^{(t)}_{r(candidate)}$ as the new current imputation for the next iteration $t+1$ if $ln(\text{IR}) > ln(u_i)$.

For incomplete categorical variables, the Metropolis sampler computes the importance ratio as follows

$$\text{IR} = \frac{p(Y|X_1, \dots, X_{r(candiate)}, \dots, X_R)p\left(X^*_{r(candiate)}|X_{-r}\right)}{p(Y|X_1, \dots, X_{r(current)}, \dots, X_R)p\left(X^*_{r(current)}|X_{-r}\right)}. \quad \text{(SA26)}$$

where $X^*_{r(candiate)}$ is a candidate imputation on the underlying latent variable metric. Consistent with the procedure for continuous variables, the algorithm draws a candidate latent variable score from a normal proposal distribution, and the current and candidate synthetic scores, $X^*_{r(current)}$ and $X^*_{r(candiate)}$, respectively, are then evaluated in the $p(X^*_r|X_{-r})$ components of the ratio. The corresponding discrete candidate $X_{r(candiate)}$ for the first component of the numerator product is generated by comparing the latent candidate to the threshold parameter, such that $X_{r(candiate)}$ $= 0$ if $X^*_{r(candiate)} < \kappa$ and $X_{r(candiate)} = 1$ if $X^*_{r(candiate)} \geq \kappa$.

B.

Model-Based Imputation for a Three-Level Model

The three-level analysis model for the third computer simulation is as follows.

$$\begin{gathered} y_{ijk} = \beta_0 + \beta_1(x_{1ijk}) + \beta_2(x_{2jk}) + \beta_3(x_{3k}) + \beta_4(x_{1ijk})(x_{3k}) \\ + b_{0k} + b_{1k}(x_{1ijk}) + \; b_{0jk} + b_{1jk}(x_{1ijk}) + \varepsilon_{ijk} \\ \begin{pmatrix} b_{0k} \\ b_{1k} \end{pmatrix} \sim MN(0, \boldsymbol{\Sigma}_{b_k}) \begin{pmatrix} b_{0jk} \\ b_{1jk} \end{pmatrix} \sim MN(0, \boldsymbol{\Sigma}_{b_{jk}}) \;\; \varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon) \end{gathered} \tag{SB1}$$

Auxiliary variables enter the model as additional explanatory variables.

The covariate distribution is multivariate normal

$$\boldsymbol{x}^{(1)} \sim MN(\boldsymbol{\mu}_{jk}, \boldsymbol{\Sigma}_1) \;\; \boldsymbol{x}^{(2)} \sim MN(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_2) \quad \boldsymbol{x}^{(3)} \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma}_3) \tag{SB2}$$

where $\boldsymbol{x}^{(1)} = (x_{1ijk})$, $\boldsymbol{\mu}_{jk} = (\mu_{1jk})$, $\boldsymbol{x}^{(2)} = (\mu_{1jk}, x_{2jk})$, $\boldsymbol{\mu}_k = (\mu_{1k}, \mu_{2k})$, $\boldsymbol{x}^{(3)} = (\mu_{1k}, \mu_{2k}, x_{3k})$ $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$, $\boldsymbol{\Sigma}_1$ is a scalar within-cluster variance, $\boldsymbol{\Sigma}_2$ is a 2 by 2 between-cluster covariance matrix at level-2, and $\boldsymbol{\Sigma}_3$ is a 3 by 3 between-cluster covariance matrix at level-3.

The covariate regression models are as follows, and analogous regressions are constructed for the latent variable cluster means.

$$\begin{gathered} x_{1ijk} = \mu_{1jk} + e_{1ijk} \\ e_{1ijk} \sim N(0, \sigma^2_{e_1}) \\ x_{2jk} = \mu_{2k} + (\mu_{1jk} - \mu_{1k})\eta_{21} + \zeta_{2jk} \\ \zeta_{2jk} \sim N\left(0, \sigma^2_{\zeta_2}\right) \\ x_{3k} = \mu_3 + (\mu_{1k} - \mu_1)\alpha_{31} + (\mu_{2k} - \mu_2)\alpha_{32} + r_{3k} \\ r_{3k} \sim N\left(0, \sigma^2_{r_3}\right) \end{gathered} \tag{SB3}$$

To make the following expressions more succinct, define $\hat{y}_{ijk}$ as the linear predictor (i.e., predictor value) of Y for observation i in level-2 and level-3 clusters j and k, respectively.

$$\begin{aligned}\hat{y}_{ijk} = \beta_0 + \beta_1(x_{1ijk}) + \beta_2(x_{2jk}) + \beta_3(x_{3k}) + \beta_4(x_{1ijk})(x_{3k}) \\ + b_{0k} + b_{1k}(x_{1ijk}) + \; b_{0jk} + b_{1jk}(x_{1ijk})\end{aligned} \quad \text{(SB4)}$$

The Metropolis-Hasting algorithm samples imputations from the following functions

$$\begin{aligned}p(X_1|Y, X_2, X_3) &\propto p(Y\,|X_1, X_2, X_3) \times p(X_1|X_2, X_3) \\ &\propto N(\hat{y}_{ijk}, \sigma^2_\varepsilon) \times N(\mu_{1jk}, \sigma^2_{e_1})\end{aligned} \quad \text{(SB5)}$$

$$\begin{aligned}p(X_2|Y, X_1, X_3) &\propto p(Y\,|X_1, X_2, X_3) \times p(X_2|X_1, X_3) \\ &\propto \prod_{i=1}^{n_{i|j}} N(\hat{y}_{ijk}, \sigma^2_\varepsilon) \times N\big(\mu_{2k} + (\mu_{1jk} - \mu_{1k})\eta_{21}, \sigma^2_{\zeta_2}\big)\end{aligned} \quad \text{(SB6)}$$

$$\begin{aligned}p(X_3|Y, X_1, X_2) &\propto p(Y\,|X_1, X_2, X_3) \times p(X_3|X_1, X_2) \\ &\propto \prod_{i=1}^{n_{ij|k}} N(\hat{y}_{ijk}, \sigma^2_\varepsilon) \\ &\times N\big(\mu_3 + (\mu_{1k} - \mu_1)\alpha_{31} + (\mu_{2k} - \mu_2)\alpha_{32}, \sigma^2_{r_3}\big)\end{aligned} \quad \text{(SB7)}$$

where $n_{i|j}$ denotes all level-1 observations in a given cluster j (i.e., all observations with a particular X_2 score in common), and $n_{ij|k}$ represents the number of observations in a given level-3 cluster k (i.e., all observations with a particular X_3 score in common).

C. Full Results from Computer Simulation Studies

C. Full Results from Computer Simulation Studies

Large Sample Simulation 1: Average Estimates and Bias Values for Model-Based Imputation

ICC	Missing %	Distribution	Parameter	True Value	Avg. Est.	Bias
0.1	15%	Normal	Intercept	50.000	50.002	0.003
0.1	15%	Normal	Level-1 Slope	3.162	3.163	0.020
0.1	15%	Normal	Level-2 Slope	0.744	0.746	0.271
0.1	15%	Normal	Intercept Var.	7.000	6.978	-0.308
0.1	15%	Normal	Icept-Slope Cov.	2.510	2.509	-0.055
0.1	15%	Normal	Slope Var.	10.000	9.988	-0.116
0.1	15%	Normal	Residual Var.	72.000	72.019	0.027
0.1	15%	Skewed	Intercept	50.000	50.006	0.011
0.1	15%	Skewed	Level-1 Slope	3.162	3.158	-0.120
0.1	15%	Skewed	Level-2 Slope	0.744	0.716	-3.767
0.1	15%	Skewed	Intercept Var.	7.000	7.068	0.977
0.1	15%	Skewed	Icept-Slope Cov.	2.510	2.508	-0.090
0.1	15%	Skewed	Slope Var.	10.000	9.983	-0.168
0.1	15%	Skewed	Residual Var.	72.000	71.984	-0.023
0.5	15%	Normal	Intercept	50.000	49.994	-0.012
0.5	15%	Normal	Level-1 Slope	3.162	3.169	0.205
0.5	15%	Normal	Level-2 Slope	1.664	1.673	0.533
0.5	15%	Normal	Intercept Var.	35.000	35.004	0.012
0.5	15%	Normal	Icept-Slope Cov.	5.612	5.546	-1.182
0.5	15%	Normal	Slope Var.	10.000	10.027	0.271
0.5	15%	Normal	Residual Var.	40.000	40.015	0.039
0.5	15%	Skewed	Intercept	50.000	50.020	0.040
0.5	15%	Skewed	Level-1 Slope	3.162	3.149	-0.406
0.5	15%	Skewed	Level-2 Slope	1.664	1.633	-1.822
0.5	15%	Skewed	Intercept Var.	35.000	35.405	1.158
0.5	15%	Skewed	Icept-Slope Cov.	5.612	5.621	0.158
0.5	15%	Skewed	Slope Var.	10.000	10.003	0.034
0.5	15%	Skewed	Residual Var.	40.000	39.989	-0.027
0.1	25%	Normal	Intercept	50.000	49.999	-0.001
0.1	25%	Normal	Level-1 Slope	3.162	3.157	-0.175
0.1	25%	Normal	Level-2 Slope	0.744	0.731	-1.797
0.1	25%	Normal	Intercept Var.	7.000	6.975	-0.353
0.1	25%	Normal	Icept-Slope Cov.	2.510	2.494	-0.653
0.1	25%	Normal	Slope Var.	10.000	9.993	-0.070
0.1	25%	Normal	Residual Var.	72.000	71.996	-0.005
0.1	25%	Skewed	Intercept	50.000	50.011	0.022
0.1	25%	Skewed	Level-1 Slope	3.162	3.161	-0.039
0.1	25%	Skewed	Level-2 Slope	0.744	0.726	-2.470
0.1	25%	Skewed	Intercept Var.	7.000	7.076	1.084
0.1	25%	Skewed	Icept-Slope Cov.	2.510	2.509	-0.038
0.1	25%	Skewed	Slope Var.	10.000	9.995	-0.049

0.1	25%	Skewed	Residual Var.	72.000	71.987	-0.018
0.5	25%	Normal	Intercept	50.000	50.000	0.001
0.5	25%	Normal	Level-1 Slope	3.162	3.151	-0.359
0.5	25%	Normal	Level-2 Slope	1.664	1.646	-1.085
0.5	25%	Normal	Intercept Var.	35.000	35.040	0.114
0.5	25%	Normal	Icept-Slope Cov.	5.612	5.529	-1.480
0.5	25%	Normal	Slope Var.	10.000	10.016	0.164
0.5	25%	Normal	Residual Var.	40.000	40.009	0.022
0.5	25%	Skewed	Intercept	50.000	50.024	0.049
0.5	25%	Skewed	Level-1 Slope	3.162	3.157	-0.183
0.5	25%	Skewed	Level-2 Slope	1.664	1.625	-2.343
0.5	25%	Skewed	Intercept Var.	35.000	35.375	1.071
0.5	25%	Skewed	Icept-Slope Cov.	5.612	5.564	-0.866
0.5	25%	Skewed	Slope Var.	10.000	10.006	0.063
0.5	25%	Skewed	Residual Var.	40.000	39.993	-0.016

Simulation 1: Average Estimates and Bias Values for Model-Based Imputation

N	ICC	Missing %	Distribution	Parameter	True Value	Avg. Est.	Bias
J = 30, n = 10	0.1	15%	Normal	Intercept	50.000	49.987	-0.025
J = 30, n = 30	0.1	15%	Normal	Intercept	50.000	49.971	-0.059
J = 100, n = 10	0.1	15%	Normal	Intercept	50.000	49.991	-0.018
J = 100, n = 30	0.1	15%	Normal	Intercept	50.000	49.993	-0.014
J = 30, n = 10	0.1	15%	Normal	Level-1 Slope	3.162	3.064	-3.096
J = 30, n = 30	0.1	15%	Normal	Level-1 Slope	3.162	3.142	-0.657
J = 100, n = 10	0.1	15%	Normal	Level-1 Slope	3.162	3.119	-1.358
J = 100, n = 30	0.1	15%	Normal	Level-1 Slope	3.162	3.144	-0.569
J = 30, n = 10	0.1	15%	Normal	Level-2 Slope	0.744	0.680	-8.639
J = 30, n = 30	0.1	15%	Normal	Level-2 Slope	0.744	0.674	-9.382
J = 100, n = 10	0.1	15%	Normal	Level-2 Slope	0.744	0.732	-1.565
J = 100, n = 30	0.1	15%	Normal	Level-2 Slope	0.744	0.727	-2.317
J = 30, n = 10	0.1	15%	Normal	Intercept Var.	7.000	6.203	-11.388
J = 30, n = 30	0.1	15%	Normal	Intercept Var.	7.000	6.459	-7.730
J = 100, n = 10	0.1	15%	Normal	Intercept Var.	7.000	6.792	-2.971
J = 100, n = 30	0.1	15%	Normal	Intercept Var.	7.000	6.798	-2.883
J = 30, n = 10	0.1	15%	Normal	Icept-Slope Cov.	2.510	2.162	-13.859
J = 30, n = 30	0.1	15%	Normal	Icept-Slope Cov.	2.510	2.512	0.081
J = 100, n = 10	0.1	15%	Normal	Icept-Slope Cov.	2.510	2.358	-6.057
J = 100, n = 30	0.1	15%	Normal	Icept-Slope Cov.	2.510	2.474	-1.414
J = 30, n = 10	0.1	15%	Normal	Slope Var.	10.000	9.931	-0.686
J = 30, n = 30	0.1	15%	Normal	Slope Var.	10.000	9.761	-2.389
J = 100, n = 10	0.1	15%	Normal	Slope Var.	10.000	9.853	-1.465
J = 100, n = 30	0.1	15%	Normal	Slope Var.	10.000	10.008	0.075
J = 30, n = 10	0.1	15%	Normal	Residual Var.	72.000	71.334	-0.925
J = 30, n = 30	0.1	15%	Normal	Residual Var.	72.000	71.853	-0.204
J = 100, n = 10	0.1	15%	Normal	Residual Var.	72.000	72.124	0.172
J = 100, n = 30	0.1	15%	Normal	Residual Var.	72.000	71.880	-0.167
J = 30, n = 10	0.1	25%	Normal	Intercept	50.000	49.918	-0.164
J = 30, n = 30	0.1	25%	Normal	Intercept	50.000	49.960	-0.079
J = 100, n = 10	0.1	25%	Normal	Intercept	50.000	49.980	-0.041
J = 100, n = 30	0.1	25%	Normal	Intercept	50.000	49.963	-0.073
J = 30, n = 10	0.1	25%	Normal	Level-1 Slope	3.162	2.969	-6.098
J = 30, n = 30	0.1	25%	Normal	Level-1 Slope	3.162	3.116	-1.450
J = 100, n = 10	0.1	25%	Normal	Level-1 Slope	3.162	3.120	-1.322
J = 100, n = 30	0.1	25%	Normal	Level-1 Slope	3.162	3.170	0.245
J = 30, n = 10	0.1	25%	Normal	Level-2 Slope	0.744	0.604	-18.871
J = 30, n = 30	0.1	25%	Normal	Level-2 Slope	0.744	0.668	-10.216
J = 100, n = 10	0.1	25%	Normal	Level-2 Slope	0.744	0.727	-2.286
J = 100, n = 30	0.1	25%	Normal	Level-2 Slope	0.744	0.711	-4.410
J = 30, n = 10	0.1	25%	Normal	Intercept Var.	7.000	6.355	-9.218

J = 30, n = 30	0.1	25%	Normal	Intercept Var.	7.000	6.278	-10.311
J = 100, n = 10	0.1	25%	Normal	Intercept Var.	7.000	6.772	-3.255
J = 100, n = 30	0.1	25%	Normal	Intercept Var.	7.000	6.723	-3.959
J = 30, n = 10	0.1	25%	Normal	Icept-Slope Cov.	2.510	1.731	-31.027
J = 30, n = 30	0.1	25%	Normal	Icept-Slope Cov.	2.510	2.391	-4.745
J = 100, n = 10	0.1	25%	Normal	Icept-Slope Cov.	2.510	2.328	-7.255
J = 100, n = 30	0.1	25%	Normal	Icept-Slope Cov.	2.510	2.466	-1.772
J = 30, n = 10	0.1	25%	Normal	Slope Var.	10.000	10.197	1.975
J = 30, n = 30	0.1	25%	Normal	Slope Var.	10.000	9.782	-2.177
J = 100, n = 10	0.1	25%	Normal	Slope Var.	10.000	10.187	1.869
J = 100, n = 30	0.1	25%	Normal	Slope Var.	10.000	10.042	0.423
J = 30, n = 10	0.1	25%	Normal	Residual Var.	72.000	71.709	-0.404
J = 30, n = 30	0.1	25%	Normal	Residual Var.	72.000	71.959	-0.057
J = 100, n = 10	0.1	25%	Normal	Residual Var.	72.000	71.767	-0.323
J = 100, n = 30	0.1	25%	Normal	Residual Var.	72.000	72.022	0.030
J = 30, n = 10	0.5	15%	Normal	Intercept	50.000	49.908	-0.184
J = 30, n = 30	0.5	15%	Normal	Intercept	50.000	49.967	-0.066
J = 100, n = 10	0.5	15%	Normal	Intercept	50.000	49.975	-0.050
J = 100, n = 30	0.5	15%	Normal	Intercept	50.000	49.946	-0.107
J = 30, n = 10	0.5	15%	Normal	Level-1 Slope	3.162	3.068	-2.987
J = 30, n = 30	0.5	15%	Normal	Level-1 Slope	3.162	3.147	-0.472
J = 100, n = 10	0.5	15%	Normal	Level-1 Slope	3.162	3.137	-0.803
J = 100, n = 30	0.5	15%	Normal	Level-1 Slope	3.162	3.157	-0.164
J = 30, n = 10	0.5	15%	Normal	Level-2 Slope	1.664	1.547	-7.041
J = 30, n = 30	0.5	15%	Normal	Level-2 Slope	1.664	1.566	-5.881
J = 100, n = 10	0.5	15%	Normal	Level-2 Slope	1.664	1.617	-2.779
J = 100, n = 30	0.5	15%	Normal	Level-2 Slope	1.664	1.590	-4.403
J = 30, n = 10	0.5	15%	Normal	Intercept Var.	35.000	32.774	-6.359
J = 30, n = 30	0.5	15%	Normal	Intercept Var.	35.000	32.426	-7.354
J = 100, n = 10	0.5	15%	Normal	Intercept Var.	35.000	34.396	-1.726
J = 100, n = 30	0.5	15%	Normal	Intercept Var.	35.000	34.264	-2.102
J = 30, n = 10	0.5	15%	Normal	Icept-Slope Cov.	5.612	4.946	-11.883
J = 30, n = 30	0.5	15%	Normal	Icept-Slope Cov.	5.612	5.183	-7.657
J = 100, n = 10	0.5	15%	Normal	Icept-Slope Cov.	5.612	5.383	-4.090
J = 100, n = 30	0.5	15%	Normal	Icept-Slope Cov.	5.612	5.487	-2.242
J = 30, n = 10	0.5	15%	Normal	Slope Var.	10.000	9.540	-4.597
J = 30, n = 30	0.5	15%	Normal	Slope Var.	10.000	9.825	-1.755
J = 100, n = 10	0.5	15%	Normal	Slope Var.	10.000	9.905	-0.950
J = 100, n = 30	0.5	15%	Normal	Slope Var.	10.000	10.036	0.356
J = 30, n = 10	0.5	15%	Normal	Residual Var.	40.000	40.105	0.262
J = 30, n = 30	0.5	15%	Normal	Residual Var.	40.000	39.977	-0.058
J = 100, n = 10	0.5	15%	Normal	Residual Var.	40.000	39.975	-0.061
J = 100, n = 30	0.5	15%	Normal	Residual Var.	40.000	39.962	-0.094

J = 30, n = 10	0.5	25%	Normal	Intercept	50.000	49.931	-0.138
J = 30, n = 30	0.5	25%	Normal	Intercept	50.000	49.900	-0.200
J = 100, n = 10	0.5	25%	Normal	Intercept	50.000	49.974	-0.052
J = 100, n = 30	0.5	25%	Normal	Intercept	50.000	49.971	-0.057
J = 30, n = 10	0.5	25%	Normal	Level-1 Slope	3.162	2.995	-5.278
J = 30, n = 30	0.5	25%	Normal	Level-1 Slope	3.162	3.147	-0.468
J = 100, n = 10	0.5	25%	Normal	Level-1 Slope	3.162	3.086	-2.413
J = 100, n = 30	0.5	25%	Normal	Level-1 Slope	3.162	3.148	-0.457
J = 30, n = 10	0.5	25%	Normal	Level-2 Slope	1.664	1.488	-10.585
J = 30, n = 30	0.5	25%	Normal	Level-2 Slope	1.664	1.365	-17.924
J = 100, n = 10	0.5	25%	Normal	Level-2 Slope	1.664	1.561	-6.181
J = 100, n = 30	0.5	25%	Normal	Level-2 Slope	1.664	1.559	-6.322
J = 30, n = 10	0.5	25%	Normal	Intercept Var.	35.000	32.569	-6.945
J = 30, n = 30	0.5	25%	Normal	Intercept Var.	35.000	32.632	-6.765
J = 100, n = 10	0.5	25%	Normal	Intercept Var.	35.000	34.156	-2.413
J = 100, n = 30	0.5	25%	Normal	Intercept Var.	35.000	34.105	-2.557
J = 30, n = 10	0.5	25%	Normal	Icept-Slope Cov.	5.612	4.898	-12.724
J = 30, n = 30	0.5	25%	Normal	Icept-Slope Cov.	5.612	4.926	-12.228
J = 100, n = 10	0.5	25%	Normal	Icept-Slope Cov.	5.612	5.244	-6.572
J = 100, n = 30	0.5	25%	Normal	Icept-Slope Cov.	5.612	5.452	-2.851
J = 30, n = 10	0.5	25%	Normal	Slope Var.	10.000	10.187	1.871
J = 30, n = 30	0.5	25%	Normal	Slope Var.	10.000	9.699	-3.014
J = 100, n = 10	0.5	25%	Normal	Slope Var.	10.000	10.202	2.022
J = 100, n = 30	0.5	25%	Normal	Slope Var.	10.000	10.050	0.501
J = 30, n = 10	0.5	25%	Normal	Residual Var.	40.000	39.954	-0.116
J = 30, n = 30	0.5	25%	Normal	Residual Var.	40.000	39.990	-0.025
J = 100, n = 10	0.5	25%	Normal	Residual Var.	40.000	39.986	-0.035
J = 100, n = 30	0.5	25%	Normal	Residual Var.	40.000	40.005	0.013
J = 30, n = 10	0.1	15%	Skewed	Intercept	50.000	49.954	-0.092
J = 30, n = 30	0.1	15%	Skewed	Intercept	50.000	49.972	-0.056
J = 100, n = 10	0.1	15%	Skewed	Intercept	50.000	49.988	-0.024
J = 100, n = 30	0.1	15%	Skewed	Intercept	50.000	50.002	0.005
J = 30, n = 10	0.1	15%	Skewed	Level-1 Slope	3.162	3.095	-2.142
J = 30, n = 30	0.1	15%	Skewed	Level-1 Slope	3.162	3.145	-0.547
J = 100, n = 10	0.1	15%	Skewed	Level-1 Slope	3.162	3.150	-0.396
J = 100, n = 30	0.1	15%	Skewed	Level-1 Slope	3.162	3.165	0.090
J = 30, n = 10	0.1	15%	Skewed	Level-2 Slope	0.744	0.714	-3.995
J = 30, n = 30	0.1	15%	Skewed	Level-2 Slope	0.744	0.681	-8.427
J = 100, n = 10	0.1	15%	Skewed	Level-2 Slope	0.744	0.719	-3.371
J = 100, n = 30	0.1	15%	Skewed	Level-2 Slope	0.744	0.722	-3.001
J = 30, n = 10	0.1	15%	Skewed	Intercept Var.	7.000	6.481	-7.420
J = 30, n = 30	0.1	15%	Skewed	Intercept Var.	7.000	6.506	-7.051
J = 100, n = 10	0.1	15%	Skewed	Intercept Var.	7.000	6.695	-4.351

J = 100, n = 30	0.1	15% Skewed	Intercept Var.	7.000	6.860	-2.005
J = 30, n = 10	0.1	15% Skewed	Icept-Slope Cov.	2.510	2.220	-11.551
J = 30, n = 30	0.1	15% Skewed	Icept-Slope Cov.	2.510	2.400	-4.376
J = 100, n = 10	0.1	15% Skewed	Icept-Slope Cov.	2.510	2.512	0.094
J = 100, n = 30	0.1	15% Skewed	Icept-Slope Cov.	2.510	2.449	-2.443
J = 30, n = 10	0.1	15% Skewed	Slope Var.	10.000	9.936	-0.645
J = 30, n = 30	0.1	15% Skewed	Slope Var.	10.000	9.727	-2.733
J = 100, n = 10	0.1	15% Skewed	Slope Var.	10.000	10.136	1.357
J = 100, n = 30	0.1	15% Skewed	Slope Var.	10.000	9.826	-1.744
J = 30, n = 10	0.1	15% Skewed	Residual Var.	72.000	71.119	-1.224
J = 30, n = 30	0.1	15% Skewed	Residual Var.	72.000	72.040	0.055
J = 100, n = 10	0.1	15% Skewed	Residual Var.	72.000	71.994	-0.009
J = 100, n = 30	0.1	15% Skewed	Residual Var.	72.000	72.156	0.217
J = 30, n = 10	0.1	25% Skewed	Intercept	50.000	50.003	0.007
J = 30, n = 30	0.1	25% Skewed	Intercept	50.000	49.950	-0.100
J = 100, n = 10	0.1	25% Skewed	Intercept	50.000	49.954	-0.091
J = 100, n = 30	0.1	25% Skewed	Intercept	50.000	50.008	0.015
J = 30, n = 10	0.1	25% Skewed	Level-1 Slope	3.162	3.024	-4.358
J = 30, n = 30	0.1	25% Skewed	Level-1 Slope	3.162	3.157	-0.157
J = 100, n = 10	0.1	25% Skewed	Level-1 Slope	3.162	3.132	-0.949
J = 100, n = 30	0.1	25% Skewed	Level-1 Slope	3.162	3.169	0.207
J = 30, n = 10	0.1	25% Skewed	Level-2 Slope	0.744	0.606	-18.528
J = 30, n = 30	0.1	25% Skewed	Level-2 Slope	0.744	0.615	-17.334
J = 100, n = 10	0.1	25% Skewed	Level-2 Slope	0.744	0.663	-10.841
J = 100, n = 30	0.1	25% Skewed	Level-2 Slope	0.744	0.689	-7.381
J = 30, n = 10	0.1	25% Skewed	Intercept Var.	7.000	6.418	-8.309
J = 30, n = 30	0.1	25% Skewed	Intercept Var.	7.000	6.443	-7.963
J = 100, n = 10	0.1	25% Skewed	Intercept Var.	7.000	6.690	-4.434
J = 100, n = 30	0.1	25% Skewed	Intercept Var.	7.000	6.783	-3.107
J = 30, n = 10	0.1	25% Skewed	Icept-Slope Cov.	2.510	2.076	-17.299
J = 30, n = 30	0.1	25% Skewed	Icept-Slope Cov.	2.510	2.402	-4.321
J = 100, n = 10	0.1	25% Skewed	Icept-Slope Cov.	2.510	2.380	-5.166
J = 100, n = 30	0.1	25% Skewed	Icept-Slope Cov.	2.510	2.496	-0.549
J = 30, n = 10	0.1	25% Skewed	Slope Var.	10.000	10.390	3.900
J = 30, n = 30	0.1	25% Skewed	Slope Var.	10.000	9.723	-2.767
J = 100, n = 10	0.1	25% Skewed	Slope Var.	10.000	10.016	0.161
J = 100, n = 30	0.1	25% Skewed	Slope Var.	10.000	9.853	-1.465
J = 30, n = 10	0.1	25% Skewed	Residual Var.	72.000	71.625	-0.521
J = 30, n = 30	0.1	25% Skewed	Residual Var.	72.000	71.989	-0.016
J = 100, n = 10	0.1	25% Skewed	Residual Var.	72.000	72.067	0.094
J = 100, n = 30	0.1	25% Skewed	Residual Var.	72.000	71.985	-0.020
J = 30, n = 10	0.5	15% Skewed	Intercept	50.000	49.946	-0.108
J = 30, n = 30	0.5	15% Skewed	Intercept	50.000	49.972	-0.057

J = 100, n = 10	0.5	15% Skewed	Intercept	50.000	49.976	-0.048
J = 100, n = 30	0.5	15% Skewed	Intercept	50.000	50.028	0.057
J = 30, n = 10	0.5	15% Skewed	Level-1 Slope	3.162	3.092	-2.217
J = 30, n = 30	0.5	15% Skewed	Level-1 Slope	3.162	3.136	-0.818
J = 100, n = 10	0.5	15% Skewed	Level-1 Slope	3.162	3.123	-1.229
J = 100, n = 30	0.5	15% Skewed	Level-1 Slope	3.162	3.161	-0.055
J = 30, n = 10	0.5	15% Skewed	Level-2 Slope	1.664	1.547	-7.000
J = 30, n = 30	0.5	15% Skewed	Level-2 Slope	1.664	1.539	-7.523
J = 100, n = 10	0.5	15% Skewed	Level-2 Slope	1.664	1.539	-7.473
J = 100, n = 30	0.5	15% Skewed	Level-2 Slope	1.664	1.629	-2.071
J = 30, n = 10	0.5	15% Skewed	Intercept Var.	35.000	32.958	-5.835
J = 30, n = 30	0.5	15% Skewed	Intercept Var.	35.000	33.120	-5.372
J = 100, n = 10	0.5	15% Skewed	Intercept Var.	35.000	34.516	-1.383
J = 100, n = 30	0.5	15% Skewed	Intercept Var.	35.000	34.600	-1.142
J = 30, n = 10	0.5	15% Skewed	Icept-Slope Cov.	5.612	4.952	-11.771
J = 30, n = 30	0.5	15% Skewed	Icept-Slope Cov.	5.612	5.426	-3.317
J = 100, n = 10	0.5	15% Skewed	Icept-Slope Cov.	5.612	5.356	-4.572
J = 100, n = 30	0.5	15% Skewed	Icept-Slope Cov.	5.612	5.508	-1.863
J = 30, n = 10	0.5	15% Skewed	Slope Var.	10.000	9.946	-0.545
J = 30, n = 30	0.5	15% Skewed	Slope Var.	10.000	9.868	-1.319
J = 100, n = 10	0.5	15% Skewed	Slope Var.	10.000	9.700	-2.999
J = 100, n = 30	0.5	15% Skewed	Slope Var.	10.000	9.899	-1.013
J = 30, n = 10	0.5	15% Skewed	Residual Var.	40.000	39.917	-0.208
J = 30, n = 30	0.5	15% Skewed	Residual Var.	40.000	39.902	-0.246
J = 100, n = 10	0.5	15% Skewed	Residual Var.	40.000	39.964	-0.090
J = 100, n = 30	0.5	15% Skewed	Residual Var.	40.000	40.029	0.072
J = 30, n = 10	0.5	25% Skewed	Intercept	50.000	49.905	-0.189
J = 30, n = 30	0.5	25% Skewed	Intercept	50.000	49.981	-0.038
J = 100, n = 10	0.5	25% Skewed	Intercept	50.000	50.010	0.020
J = 100, n = 30	0.5	25% Skewed	Intercept	50.000	49.997	-0.005
J = 30, n = 10	0.5	25% Skewed	Level-1 Slope	3.162	3.037	-3.964
J = 30, n = 30	0.5	25% Skewed	Level-1 Slope	3.162	3.139	-0.742
J = 100, n = 10	0.5	25% Skewed	Level-1 Slope	3.162	3.106	-1.764
J = 100, n = 30	0.5	25% Skewed	Level-1 Slope	3.162	3.175	0.390
J = 30, n = 10	0.5	25% Skewed	Level-2 Slope	1.664	1.444	-13.215
J = 30, n = 30	0.5	25% Skewed	Level-2 Slope	1.664	1.470	-11.620
J = 100, n = 10	0.5	25% Skewed	Level-2 Slope	1.664	1.577	-5.215
J = 100, n = 30	0.5	25% Skewed	Level-2 Slope	1.664	1.554	-6.585
J = 30, n = 10	0.5	25% Skewed	Intercept Var.	35.000	33.290	-4.887
J = 30, n = 30	0.5	25% Skewed	Intercept Var.	35.000	32.789	-6.318
J = 100, n = 10	0.5	25% Skewed	Intercept Var.	35.000	34.758	-0.692
J = 100, n = 30	0.5	25% Skewed	Intercept Var.	35.000	34.776	-0.639
J = 30, n = 10	0.5	25% Skewed	Icept-Slope Cov.	5.612	4.649	-17.173

J = 30, n = 30	0.5	25%	Skewed	Icept-Slope Cov.	5.612	5.119	-8.789
J = 100, n = 10	0.5	25%	Skewed	Icept-Slope Cov.	5.612	5.254	-6.384
J = 100, n = 30	0.5	25%	Skewed	Icept-Slope Cov.	5.612	5.413	-3.555
J = 30, n = 10	0.5	25%	Skewed	Slope Var.	10.000	10.455	4.549
J = 30, n = 30	0.5	25%	Skewed	Slope Var.	10.000	10.021	0.212
J = 100, n = 10	0.5	25%	Skewed	Slope Var.	10.000	10.010	0.102
J = 100, n = 30	0.5	25%	Skewed	Slope Var.	10.000	9.833	-1.665
J = 30, n = 10	0.5	25%	Skewed	Residual Var.	40.000	39.811	-0.472
J = 30, n = 30	0.5	25%	Skewed	Residual Var.	40.000	39.910	-0.224
J = 100, n = 10	0.5	25%	Skewed	Residual Var.	40.000	39.915	-0.213
J = 100, n = 30	0.5	25%	Skewed	Residual Var.	40.000	39.967	-0.083

Simulation 2: Average Estimates and Bias Values for Model-Based Imputation

N	ICC	Missing %	Parameter	rue Value	Avg. Est.	Bias
J = 30, n = 10	0.1	15%	Intercept	50.000	50.073	0.147
J = 30, n = 30	0.1	15%	Intercept	50.000	50.044	0.088
J = 100, n = 10	0.1	15%	Intercept	50.000	50.034	0.068
J = 100, n = 30	0.1	15%	Intercept	50.000	50.006	0.012
J = 30, n = 10	0.1	15%	Level-1 Slope	3.162	3.052	-3.481
J = 30, n = 30	0.1	15%	Level-1 Slope	3.162	3.111	-1.611
J = 100, n = 10	0.1	15%	Level-1 Slope	3.162	3.108	-1.710
J = 100, n = 30	0.1	15%	Level-1 Slope	3.162	3.153	-0.296
J = 30, n = 10	0.1	15%	Level-2 Slope	0.861	0.711	-17.403
J = 30, n = 30	0.1	15%	Level-2 Slope	0.861	0.739	-14.163
J = 100, n = 10	0.1	15%	Level-2 Slope	0.861	0.797	-7.487
J = 100, n = 30	0.1	15%	Level-2 Slope	0.861	0.816	-5.265
J = 30, n = 10	0.1	15%	Intercept Var.	7.556	6.839	-9.495
J = 30, n = 30	0.1	15%	Intercept Var.	7.556	6.942	-8.130
J = 100, n = 10	0.1	15%	Intercept Var.	7.556	7.238	-4.207
J = 100, n = 30	0.1	15%	Intercept Var.	7.556	7.421	-1.783
J = 30, n = 10	0.1	15%	Icept-Slope Cov.	2.608	2.071	-20.583
J = 30, n = 30	0.1	15%	Icept-Slope Cov.	2.608	2.476	-5.073
J = 100, n = 10	0.1	15%	Icept-Slope Cov.	2.608	2.498	-4.223
J = 100, n = 30	0.1	15%	Icept-Slope Cov.	2.608	2.599	-0.333
J = 30, n = 10	0.1	15%	Slope Var.	10.000	9.824	-1.761
J = 30, n = 30	0.1	15%	Slope Var.	10.000	9.734	-2.660
J = 100, n = 10	0.1	15%	Slope Var.	10.000	9.921	-0.785
J = 100, n = 30	0.1	15%	Slope Var.	10.000	9.778	-2.222
J = 30, n = 10	0.1	15%	Residual Var.	72.000	71.452	-0.761
J = 30, n = 30	0.1	15%	Residual Var.	72.000	71.900	-0.139
J = 100, n = 10	0.1	15%	Residual Var.	72.000	72.016	0.022
J = 100, n = 30	0.1	15%	Residual Var.	72.000	71.873	-0.176
J = 30, n = 10	0.1	25%	Intercept	50.000	50.045	0.090
J = 30, n = 30	0.1	25%	Intercept	50.000	50.008	0.016
J = 100, n = 10	0.1	25%	Intercept	50.000	50.001	0.002
J = 100, n = 30	0.1	25%	Intercept	50.000	50.015	0.030
J = 30, n = 10	0.1	25%	Level-1 Slope	3.162	3.084	-2.475
J = 30, n = 30	0.1	25%	Level-1 Slope	3.162	3.137	-0.788
J = 100, n = 10	0.1	25%	Level-1 Slope	3.162	3.099	-1.989
J = 100, n = 30	0.1	25%	Level-1 Slope	3.162	3.162	-0.002
J = 30, n = 10	0.1	25%	Level-2 Slope	0.861	0.704	-18.248
J = 30, n = 30	0.1	25%	Level-2 Slope	0.861	0.759	-11.823
J = 100, n = 10	0.1	25%	Level-2 Slope	0.861	0.840	-2.514
J = 100, n = 30	0.1	25%	Level-2 Slope	0.861	0.815	-5.358
J = 30, n = 10	0.1	25%	Intercept Var.	7.556	6.756	-10.593

J = 30, n = 30	0.1	25%	Intercept Var.	7.556	6.822	-9.718
J = 100, n = 10	0.1	25%	Intercept Var.	7.556	7.160	-5.245
J = 100, n = 30	0.1	25%	Intercept Var.	7.556	7.338	-2.886
J = 30, n = 10	0.1	25%	Icept-Slope Cov.	2.608	2.260	-13.340
J = 30, n = 30	0.1	25%	Icept-Slope Cov.	2.608	2.523	-3.256
J = 100, n = 10	0.1	25%	Icept-Slope Cov.	2.608	2.438	-6.501
J = 100, n = 30	0.1	25%	Icept-Slope Cov.	2.608	2.587	-0.809
J = 30, n = 10	0.1	25%	Slope Var.	10.000	10.530	5.300
J = 30, n = 30	0.1	25%	Slope Var.	10.000	9.860	-1.404
J = 100, n = 10	0.1	25%	Slope Var.	10.000	9.960	-0.396
J = 100, n = 30	0.1	25%	Slope Var.	10.000	10.077	0.774
J = 30, n = 10	0.1	25%	Residual Var.	72.000	71.558	-0.613
J = 30, n = 30	0.1	25%	Residual Var.	72.000	71.715	-0.396
J = 100, n = 10	0.1	25%	Residual Var.	72.000	71.942	-0.081
J = 100, n = 30	0.1	25%	Residual Var.	72.000	71.920	-0.112
J = 30, n = 10	0.5	15%	Intercept	50.000	49.956	-0.088
J = 30, n = 30	0.5	15%	Intercept	50.000	50.009	0.017
J = 100, n = 10	0.5	15%	Intercept	50.000	50.043	0.085
J = 100, n = 30	0.5	15%	Intercept	50.000	50.041	0.083
J = 30, n = 10	0.5	15%	Level-1 Slope	3.162	3.071	-2.880
J = 30, n = 30	0.5	15%	Level-1 Slope	3.162	3.174	0.362
J = 100, n = 10	0.5	15%	Level-1 Slope	3.162	3.148	-0.448
J = 100, n = 30	0.5	15%	Level-1 Slope	3.162	3.169	0.224
J = 30, n = 10	0.5	15%	Level-2 Slope	1.926	1.865	-3.162
J = 30, n = 30	0.5	15%	Level-2 Slope	1.926	1.845	-4.207
J = 100, n = 10	0.5	15%	Level-2 Slope	1.926	1.816	-5.686
J = 100, n = 30	0.5	15%	Level-2 Slope	1.926	1.799	-6.596
J = 30, n = 10	0.5	15%	Intercept Var.	37.781	34.949	-7.495
J = 30, n = 30	0.5	15%	Intercept Var.	37.781	35.002	-7.355
J = 100, n = 10	0.5	15%	Intercept Var.	37.781	36.671	-2.939
J = 100, n = 30	0.5	15%	Intercept Var.	37.781	36.730	-2.784
J = 30, n = 10	0.5	15%	Icept-Slope Cov.	5.831	4.883	-16.255
J = 30, n = 30	0.5	15%	Icept-Slope Cov.	5.831	5.555	-4.736
J = 100, n = 10	0.5	15%	Icept-Slope Cov.	5.831	5.619	-3.635
J = 100, n = 30	0.5	15%	Icept-Slope Cov.	5.831	5.713	-2.020
J = 30, n = 10	0.5	15%	Slope Var.	10.000	9.847	-1.535
J = 30, n = 30	0.5	15%	Slope Var.	10.000	9.691	-3.091
J = 100, n = 10	0.5	15%	Slope Var.	10.000	9.998	-0.020
J = 100, n = 30	0.5	15%	Slope Var.	10.000	9.979	-0.210
J = 30, n = 10	0.5	15%	Residual Var.	40.000	40.050	0.125
J = 30, n = 30	0.5	15%	Residual Var.	40.000	39.824	-0.441
J = 100, n = 10	0.5	15%	Residual Var.	40.000	39.987	-0.032
J = 100, n = 30	0.5	15%	Residual Var.	40.000	40.015	0.038

J = 30, n = 10	0.5	25%	Intercept	50.000	50.068	0.135
J = 30, n = 30	0.5	25%	Intercept	50.000	50.045	0.090
J = 100, n = 10	0.5	25%	Intercept	50.000	50.052	0.105
J = 100, n = 30	0.5	25%	Intercept	50.000	49.984	-0.031
J = 30, n = 10	0.5	25%	Level-1 Slope	3.162	3.061	-3.216
J = 30, n = 30	0.5	25%	Level-1 Slope	3.162	3.119	-1.358
J = 100, n = 10	0.5	25%	Level-1 Slope	3.162	3.123	-1.241
J = 100, n = 30	0.5	25%	Level-1 Slope	3.162	3.107	-1.741
J = 30, n = 10	0.5	25%	Level-2 Slope	1.926	1.568	-18.571
J = 30, n = 30	0.5	25%	Level-2 Slope	1.926	1.581	-17.917
J = 100, n = 10	0.5	25%	Level-2 Slope	1.926	1.756	-8.829
J = 100, n = 30	0.5	25%	Level-2 Slope	1.926	1.832	-4.865
J = 30, n = 10	0.5	25%	Intercept Var.	37.781	34.805	-7.877
J = 30, n = 30	0.5	25%	Intercept Var.	37.781	34.800	-7.890
J = 100, n = 10	0.5	25%	Intercept Var.	37.781	36.606	-3.110
J = 100, n = 30	0.5	25%	Intercept Var.	37.781	37.031	-1.984
J = 30, n = 10	0.5	25%	Icept-Slope Cov.	5.831	4.736	-18.778
J = 30, n = 30	0.5	25%	Icept-Slope Cov.	5.831	5.327	-8.647
J = 100, n = 10	0.5	25%	Icept-Slope Cov.	5.831	5.582	-4.269
J = 100, n = 30	0.5	25%	Icept-Slope Cov.	5.831	5.624	-3.561
J = 30, n = 10	0.5	25%	Slope Var.	10.000	10.240	2.400
J = 30, n = 30	0.5	25%	Slope Var.	10.000	9.891	-1.088
J = 100, n = 10	0.5	25%	Slope Var.	10.000	10.004	0.037
J = 100, n = 30	0.5	25%	Slope Var.	10.000	9.936	-0.639
J = 30, n = 10	0.5	25%	Residual Var.	40.000	40.025	0.062
J = 30, n = 30	0.5	25%	Residual Var.	40.000	39.928	-0.181
J = 100, n = 10	0.5	25%	Residual Var.	40.000	40.076	0.190
J = 100, n = 30	0.5	25%	Residual Var.	40.000	40.026	0.064

Simulation 3: Average Estimates and Bias Values for Model-Based Imputation

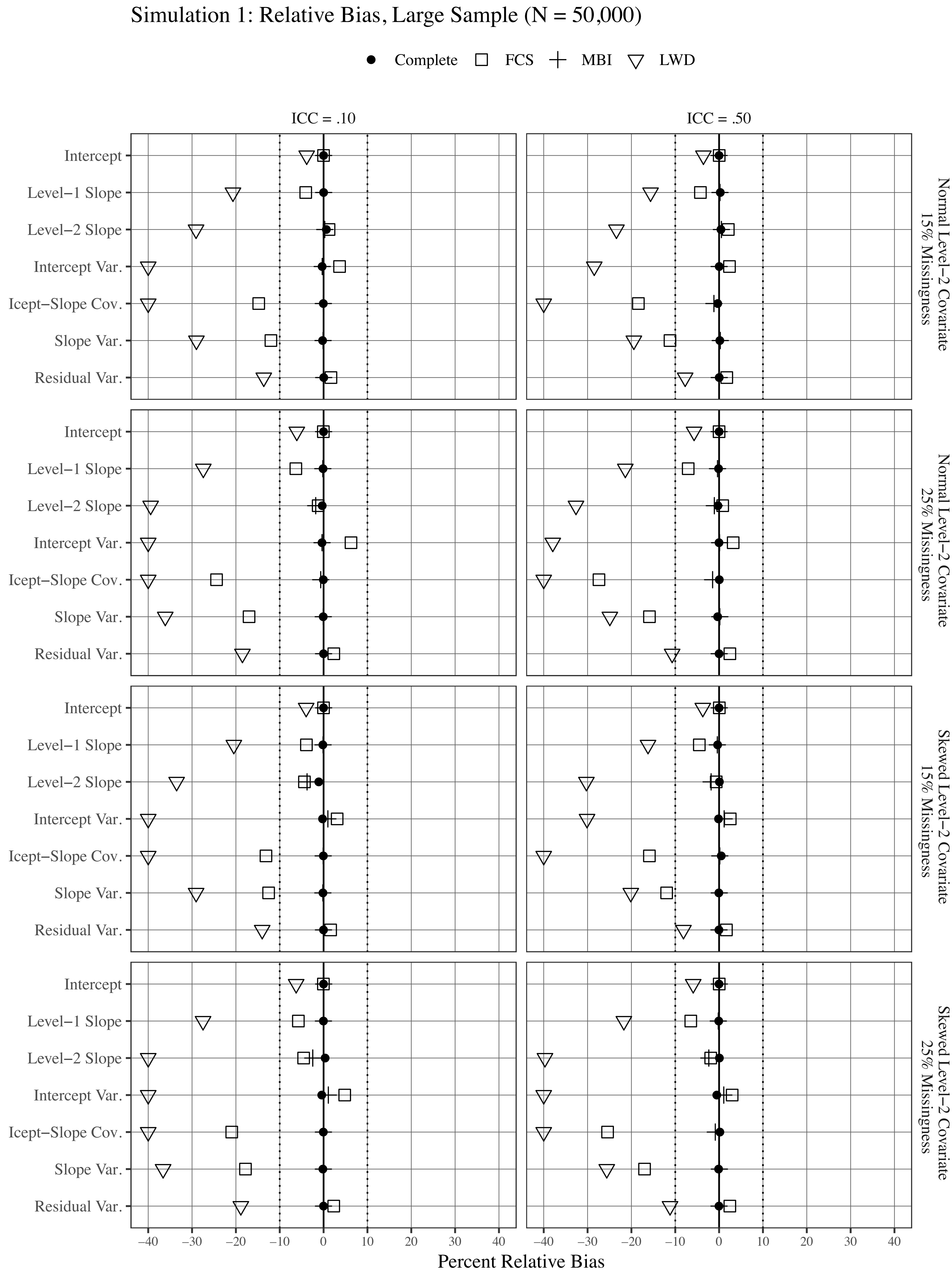
N	ICC	Missing %	Parameter	True Value	Avg. Est.	Bias
J = 30, n = 10	0.1	15%	Intercept	49.836	49.819	-0.033
J = 30, n = 10	0.1	15%	Level-1 Slope	3.098	3.099	0.023
J = 30, n = 10	0.1	15%	Level-2 Slope	0.724	0.706	-2.495
J = 30, n = 10	0.1	15%	Level-3 Slope	0.654	0.630	-3.644
J = 30, n = 10	0.1	15%	Interaction Slope	1.549	1.436	-7.276
J = 30, n = 10	0.1	15%	Level-3 Intercept Var.	5.104	4.557	-10.711
J = 30, n = 10	0.1	15%	Level-3 Icept-Slope Cov.	1.917	1.893	-1.231
J = 30, n = 10	0.1	15%	Level-3 Slope Var.	8.000	7.436	-7.055
J = 30, n = 10	0.1	15%	Level-2 Intercept Var.	5.500	5.487	-0.236
J = 30, n = 10	0.1	15%	Level-2 Icept-Slope Cov.	1.990	2.015	1.271
J = 30, n = 10	0.1	15%	Level-2 Slope Var.	8.000	8.096	1.205
J = 30, n = 10	0.1	15%	Residual Var.	52.000	51.989	-0.020
J = 30, n = 10	0.5	15%	Intercept	49.740	49.723	-0.034
J = 30, n = 10	0.5	15%	Level-1 Slope	2.449	2.437	-0.505
J = 30, n = 10	0.5	15%	Level-2 Slope	1.145	1.132	-1.098
J = 30, n = 10	0.5	15%	Level-3 Slope	0.938	0.857	-8.628
J = 30, n = 10	0.5	15%	Interaction Slope	1.225	1.138	-7.066
J = 30, n = 10	0.5	15%	Level-3 Intercept Var.	11.183	10.208	-8.718
J = 30, n = 10	0.5	15%	Level-3 Icept-Slope Cov.	2.243	1.985	-11.523
J = 30, n = 10	0.5	15%	Level-3 Slope Var.	5.000	4.618	-7.641
J = 30, n = 10	0.5	15%	Level-2 Intercept Var.	13.750	13.649	-0.732
J = 30, n = 10	0.5	15%	Level-2 Icept-Slope Cov.	2.487	2.353	-5.418
J = 30, n = 10	0.5	15%	Level-2 Slope Var.	5.000	5.113	2.261
J = 30, n = 10	0.5	15%	Residual Var.	32.500	32.548	0.148
J = 30, n = 10	0.1	25%	Intercept	49.836	49.818	-0.036
J = 30, n = 10	0.1	25%	Level-1 Slope	3.098	3.089	-0.297
J = 30, n = 10	0.1	25%	Level-2 Slope	0.724	0.700	-3.341
J = 30, n = 10	0.1	25%	Level-3 Slope	0.654	0.513	-21.541
J = 30, n = 10	0.1	25%	Interaction Slope	1.549	1.367	-11.749
J = 30, n = 10	0.1	25%	Level-3 Intercept Var.	5.104	4.647	-8.962
J = 30, n = 10	0.1	25%	Level-3 Icept-Slope Cov.	1.917	1.933	0.828
J = 30, n = 10	0.1	25%	Level-3 Slope Var.	8.000	7.560	-5.503
J = 30, n = 10	0.1	25%	Level-2 Intercept Var.	5.500	5.450	-0.908
J = 30, n = 10	0.1	25%	Level-2 Icept-Slope Cov.	1.990	1.983	-0.358
J = 30, n = 10	0.1	25%	Level-2 Slope Var.	8.000	8.219	2.738
J = 30, n = 10	0.1	25%	Residual Var.	52.000	51.964	-0.070
J = 30, n = 10	0.5	25%	Intercept	49.740	49.694	-0.092
J = 30, n = 10	0.5	25%	Level-1 Slope	2.449	2.403	-1.886
J = 30, n = 10	0.5	25%	Level-2 Slope	1.145	1.116	-2.517
J = 30, n = 10	0.5	25%	Level-3 Slope	0.938	0.745	-20.550
J = 30, n = 10	0.5	25%	Interaction Slope	1.225	1.057	-13.668

J = 30, n = 10	0.5	25%	Level-3 Intercept Var.	11.183	10.546	-5.694
J = 30, n = 10	0.5	25%	Level-3 Icept-Slope Cov.	2.243	2.157	-3.832
J = 30, n = 10	0.5	25%	Level-3 Slope Var.	5.000	4.748	-5.038
J = 30, n = 10	0.5	25%	Level-2 Intercept Var.	13.750	13.852	0.739
J = 30, n = 10	0.5	25%	Level-2 Icept-Slope Cov.	2.487	2.273	-8.625
J = 30, n = 10	0.5	25%	Level-2 Slope Var.	5.000	5.124	2.473
J = 30, n = 10	0.5	25%	Residual Var.	32.500	32.570	0.217
J = 100, n = 10	0.1	15%	Intercept	49.836	49.835	-0.002
J = 100, n = 10	0.1	15%	Level-1 Slope	3.098	3.099	0.017
J = 100, n = 10	0.1	15%	Level-2 Slope	0.724	0.726	0.322
J = 100, n = 10	0.1	15%	Level-3 Slope	0.654	0.620	-5.249
J = 100, n = 10	0.1	15%	Interaction Slope	1.549	1.490	-3.808
J = 100, n = 10	0.1	15%	Level-3 Intercept Var.	5.104	4.914	-3.714
J = 100, n = 10	0.1	15%	Level-3 Icept-Slope Cov.	1.917	1.903	-0.730
J = 100, n = 10	0.1	15%	Level-3 Slope Var.	8.000	7.817	-2.286
J = 100, n = 10	0.1	15%	Level-2 Intercept Var.	5.500	5.453	-0.863
J = 100, n = 10	0.1	15%	Level-2 Icept-Slope Cov.	1.990	1.981	-0.432
J = 100, n = 10	0.1	15%	Level-2 Slope Var.	8.000	7.970	-0.370
J = 100, n = 10	0.1	15%	Residual Var.	52.000	52.039	0.075
J = 100, n = 10	0.5	15%	Intercept	49.740	49.754	0.027
J = 100, n = 10	0.5	15%	Level-1 Slope	2.449	2.463	0.550
J = 100, n = 10	0.5	15%	Level-2 Slope	1.145	1.143	-0.154
J = 100, n = 10	0.5	15%	Level-3 Slope	0.938	0.876	-6.664
J = 100, n = 10	0.5	15%	Interaction Slope	1.225	1.179	-3.723
J = 100, n = 10	0.5	15%	Level-3 Intercept Var.	11.183	10.995	-1.674
J = 100, n = 10	0.5	15%	Level-3 Icept-Slope Cov.	2.243	2.230	-0.605
J = 100, n = 10	0.5	15%	Level-3 Slope Var.	5.000	4.922	-1.552
J = 100, n = 10	0.5	15%	Level-2 Intercept Var.	13.750	13.697	-0.383
J = 100, n = 10	0.5	15%	Level-2 Icept-Slope Cov.	2.487	2.440	-1.919
J = 100, n = 10	0.5	15%	Level-2 Slope Var.	5.000	5.076	1.516
J = 100, n = 10	0.5	15%	Residual Var.	32.500	32.497	-0.008
J = 100, n = 10	0.1	25%	Intercept	49.836	49.836	0.001
J = 100, n = 10	0.1	25%	Level-1 Slope	3.098	3.096	-0.080
J = 100, n = 10	0.1	25%	Level-2 Slope	0.724	0.709	-2.116
J = 100, n = 10	0.1	25%	Level-3 Slope	0.654	0.608	-7.039
J = 100, n = 10	0.1	25%	Interaction Slope	1.549	1.462	-5.605
J = 100, n = 10	0.1	25%	Level-3 Intercept Var.	5.104	4.985	-2.342
J = 100, n = 10	0.1	25%	Level-3 Icept-Slope Cov.	1.917	2.002	4.454
J = 100, n = 10	0.1	25%	Level-3 Slope Var.	8.000	8.017	0.208
J = 100, n = 10	0.1	25%	Level-2 Intercept Var.	5.500	5.500	0.002
J = 100, n = 10	0.1	25%	Level-2 Icept-Slope Cov.	1.990	1.941	-2.459
J = 100, n = 10	0.1	25%	Level-2 Slope Var.	8.000	8.065	0.818
J = 100, n = 10	0.1	25%	Residual Var.	52.000	52.030	0.058

J = 100, n = 10	0.5	25%	Intercept	49.740	49.749	0.018
J = 100, n = 10	0.5	25%	Level-1 Slope	2.449	2.430	-0.777
J = 100, n = 10	0.5	25%	Level-2 Slope	1.145	1.121	-2.058
J = 100, n = 10	0.5	25%	Level-3 Slope	0.938	0.850	-9.458
J = 100, n = 10	0.5	25%	Interaction Slope	1.225	1.136	-7.220
J = 100, n = 10	0.5	25%	Level-3 Intercept Var.	11.183	11.137	-0.408
J = 100, n = 10	0.5	25%	Level-3 Icept-Slope Cov.	2.243	2.198	-2.010
J = 100, n = 10	0.5	25%	Level-3 Slope Var.	5.000	5.038	0.769
J = 100, n = 10	0.5	25%	Level-2 Intercept Var.	13.750	13.850	0.729
J = 100, n = 10	0.5	25%	Level-2 Icept-Slope Cov.	2.487	2.413	-2.999
J = 100, n = 10	0.5	25%	Level-2 Slope Var.	5.000	5.119	2.381
J = 100, n = 10	0.5	25%	Residual Var.	32.500	32.508	0.026
J = 30, n = 30	0.1	15%	Intercept	49.836	49.837	0.002
J = 30, n = 30	0.1	15%	Level-1 Slope	3.098	3.102	0.123
J = 30, n = 30	0.1	15%	Level-2 Slope	0.724	0.713	-1.575
J = 30, n = 30	0.1	15%	Level-3 Slope	0.654	0.591	-9.673
J = 30, n = 30	0.1	15%	Interaction Slope	1.549	1.449	-6.463
J = 30, n = 30	0.1	15%	Level-3 Intercept Var.	5.104	4.629	-9.300
J = 30, n = 30	0.1	15%	Level-3 Icept-Slope Cov.	1.917	1.843	-3.881
J = 30, n = 30	0.1	15%	Level-3 Slope Var.	8.000	7.447	-6.908
J = 30, n = 30	0.1	15%	Level-2 Intercept Var.	5.500	5.498	-0.035
J = 30, n = 30	0.1	15%	Level-2 Icept-Slope Cov.	1.990	2.001	0.550
J = 30, n = 30	0.1	15%	Level-2 Slope Var.	8.000	7.995	-0.057
J = 30, n = 30	0.1	15%	Residual Var.	52.000	52.053	0.101
J = 30, n = 30	0.5	15%	Intercept	49.740	49.798	0.117
J = 30, n = 30	0.5	15%	Level-1 Slope	2.449	2.439	-0.432
J = 30, n = 30	0.5	15%	Level-2 Slope	1.145	1.139	-0.499
J = 30, n = 30	0.5	15%	Level-3 Slope	0.938	0.880	-6.174
J = 30, n = 30	0.5	0.15	Interaction Slope	1.225	1.083	-11.542
J = 30, n = 30	0.5	0.15	Level-3 Intercept Var.	11.183	10.421	-6.811
J = 30, n = 30	0.5	0.15	Level-3 Icept-Slope Cov.	2.243	2.148	-4.266
J = 30, n = 30	0.5	0.15	Level-3 Slope Var.	5.000	4.444	-11.113
J = 30, n = 30	0.5	0.15	Level-2 Intercept Var.	13.750	14.550	5.815
J = 30, n = 30	0.5	0.15	Level-2 Icept-Slope Cov.	2.487	2.053	-17.481
J = 30, n = 30	0.5	0.15	Level-2 Slope Var.	5.000	5.323	6.451
J = 30, n = 30	0.5	0.15	Residual Var.	32.500	32.533	0.100
J = 30, n = 30	0.1	0.25	Intercept	49.836	49.813	-0.046
J = 30, n = 30	0.1	0.25	Level-1 Slope	3.098	3.116	0.579
J = 30, n = 30	0.1	0.25	Level-2 Slope	0.724	0.695	-4.040
J = 30, n = 30	0.1	0.25	Level-3 Slope	0.654	0.565	-13.637
J = 30, n = 30	0.1	0.25	Interaction Slope	1.549	1.347	-13.034
J = 30, n = 30	0.1	0.25	Level-3 Intercept Var.	5.104	4.630	-9.280
J = 30, n = 30	0.1	0.25	Level-3 Icept-Slope Cov.	1.917	1.870	-2.436

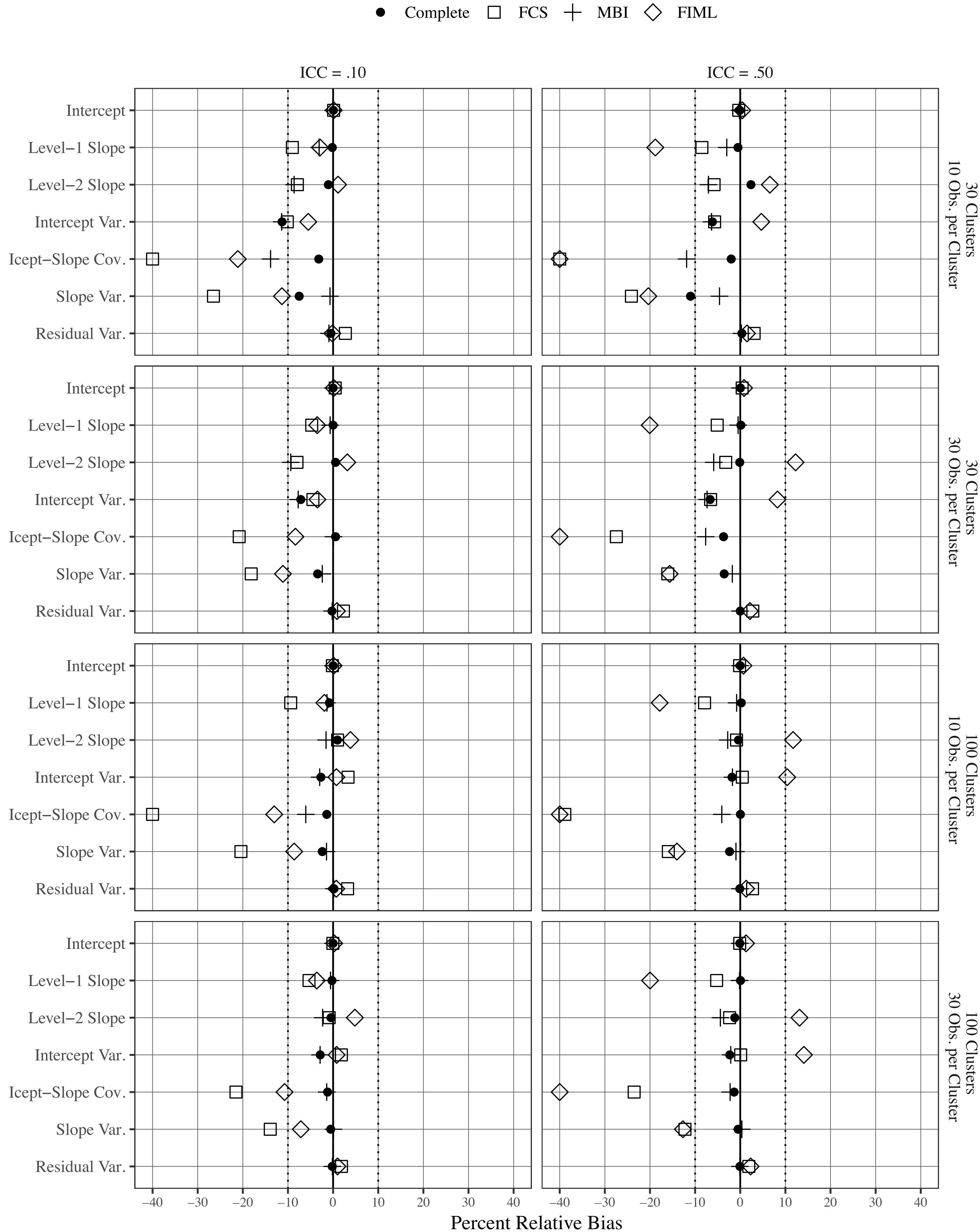
J = 30, n = 30	0.1	0.25	Level-3 Slope Var.	8.000	7.383	-7.716
J = 30, n = 30	0.1	0.25	Level-2 Intercept Var.	5.500	5.482	-0.327
J = 30, n = 30	0.1	0.25	Level-2 Icept-Slope Cov.	1.990	1.974	-0.793
J = 30, n = 30	0.1	0.25	Level-2 Slope Var.	8.000	8.055	0.692
J = 30, n = 30	0.1	0.25	Residual Var.	52.000	52.030	0.057
J = 30, n = 30	0.5	0.25	Intercept	49.740	49.759	0.038
J = 30, n = 30	0.5	0.25	Level-1 Slope	2.449	2.378	-2.926
J = 30, n = 30	0.5	0.25	Level-2 Slope	1.145	1.140	-0.395
J = 30, n = 30	0.5	0.25	Level-3 Slope	0.938	0.798	-14.907
J = 30, n = 30	0.5	0.25	Interaction Slope	1.225	1.065	-13.029
J = 30, n = 30	0.5	0.25	Level-3 Intercept Var.	11.183	10.593	-5.269
J = 30, n = 30	0.5	0.25	Level-3 Icept-Slope Cov.	2.243	2.175	-3.044
J = 30, n = 30	0.5	0.25	Level-3 Slope Var.	5.000	4.524	-9.525
J = 30, n = 30	0.5	0.25	Level-2 Intercept Var.	13.750	14.683	6.786
J = 30, n = 30	0.5	0.25	Level-2 Icept-Slope Cov.	2.487	1.777	-28.552
J = 30, n = 30	0.5	0.25	Level-2 Slope Var.	5.000	5.519	10.390
J = 30, n = 30	0.5	0.25	Residual Var.	32.500	32.659	0.490
J = 100, n = 30	0.1	0.15	Intercept	49.836	49.841	0.010
J = 100, n = 30	0.1	0.15	Level-1 Slope	3.098	3.092	-0.203
J = 100, n = 30	0.1	0.15	Level-2 Slope	0.724	0.717	-0.999
J = 100, n = 30	0.1	0.15	Level-3 Slope	0.654	0.615	-6.022
J = 100, n = 30	0.1	0.15	Interaction Slope	1.549	1.527	-1.459
J = 100, n = 30	0.1	0.15	Level-3 Intercept Var.	5.104	4.977	-2.487
J = 100, n = 30	0.1	0.15	Level-3 Icept-Slope Cov.	1.917	1.899	-0.914
J = 100, n = 30	0.1	0.15	Level-3 Slope Var.	8.000	7.819	-2.265
J = 100, n = 30	0.1	0.15	Level-2 Intercept Var.	5.500	5.486	-0.246
J = 100, n = 30	0.1	0.15	Level-2 Icept-Slope Cov.	1.990	1.957	-1.652
J = 100, n = 30	0.1	0.15	Level-2 Slope Var.	8.000	8.031	0.384
J = 100, n = 30	0.1	0.15	Residual Var.	52.000	51.981	-0.037
J = 100, n = 30	0.5	0.15	Intercept	49.740	49.778	0.076
J = 100, n = 30	0.5	0.15	Level-1 Slope	2.449	2.426	-0.939
J = 100, n = 30	0.5	0.15	Level-2 Slope	1.145	1.178	2.866
J = 100, n = 30	0.5	0.15	Level-3 Slope	0.938	0.907	-3.370
J = 100, n = 30	0.5	0.15	Interaction Slope	1.225	1.168	-4.645
J = 100, n = 30	0.5	0.15	Level-3 Intercept Var.	11.183	11.219	0.324
J = 100, n = 30	0.5	0.15	Level-3 Icept-Slope Cov.	2.243	2.189	-2.408
J = 100, n = 30	0.5	0.15	Level-3 Slope Var.	5.000	4.759	-4.829
J = 100, n = 30	0.5	0.15	Level-2 Intercept Var.	13.750	14.419	4.869
J = 100, n = 30	0.5	0.15	Level-2 Icept-Slope Cov.	2.487	2.029	-18.428
J = 100, n = 30	0.5	0.15	Level-2 Slope Var.	5.000	5.295	5.904
J = 100, n = 30	0.5	0.15	Residual Var.	32.500	32.578	0.239
J = 100, n = 30	0.1	0.25	Intercept	49.836	49.838	0.004
J = 100, n = 30	0.1	0.25	Level-1 Slope	3.098	3.109	0.357

J = 100, n = 30	0.1	0.25	Level-2 Slope	0.724	0.702	-3.018
J = 100, n = 30	0.1	0.25	Level-3 Slope	0.654	0.635	-2.960
J = 100, n = 30	0.1	0.25	Interaction Slope	1.549	1.483	-4.295
J = 100, n = 30	0.1	0.25	Level-3 Intercept Var.	5.104	4.967	-2.691
J = 100, n = 30	0.1	0.25	Level-3 Icept-Slope Cov.	1.917	1.929	0.604
J = 100, n = 30	0.1	0.25	Level-3 Slope Var.	8.000	7.861	-1.739
J = 100, n = 30	0.1	0.25	Level-2 Intercept Var.	5.500	5.517	0.305
J = 100, n = 30	0.1	0.25	Level-2 Icept-Slope Cov.	1.990	1.937	-2.675
J = 100, n = 30	0.1	0.25	Level-2 Slope Var.	8.000	8.098	1.227
J = 100, n = 30	0.1	0.25	Residual Var.	52.000	52.046	0.089
J = 100, n = 30	0.5	0.25	Intercept	49.740	49.784	0.088
J = 100, n = 30	0.5	0.25	Level-1 Slope	2.449	2.396	-2.200
J = 100, n = 30	0.5	0.25	Level-2 Slope	1.145	1.150	0.444
J = 100, n = 30	0.5	0.25	Level-3 Slope	0.938	0.896	-4.458
J = 100, n = 30	0.5	0.25	Interaction Slope	1.225	1.138	-7.122
J = 100, n = 30	0.5	0.25	Level-3 Intercept Var.	11.183	11.405	1.994
J = 100, n = 30	0.5	0.25	Level-3 Icept-Slope Cov.	2.243	2.156	-3.879
J = 100, n = 30	0.5	0.25	Level-3 Slope Var.	5.000	4.737	-5.261
J = 100, n = 30	0.5	0.25	Level-2 Intercept Var.	13.750	14.586	6.077
J = 100, n = 30	0.5	0.25	Level-2 Icept-Slope Cov.	2.487	1.833	-26.306
J = 100, n = 30	0.5	0.25	Level-2 Slope Var.	5.000	5.555	11.091
J = 100, n = 30	0.5	0.25	Residual Var.	32.500	32.598	0.303

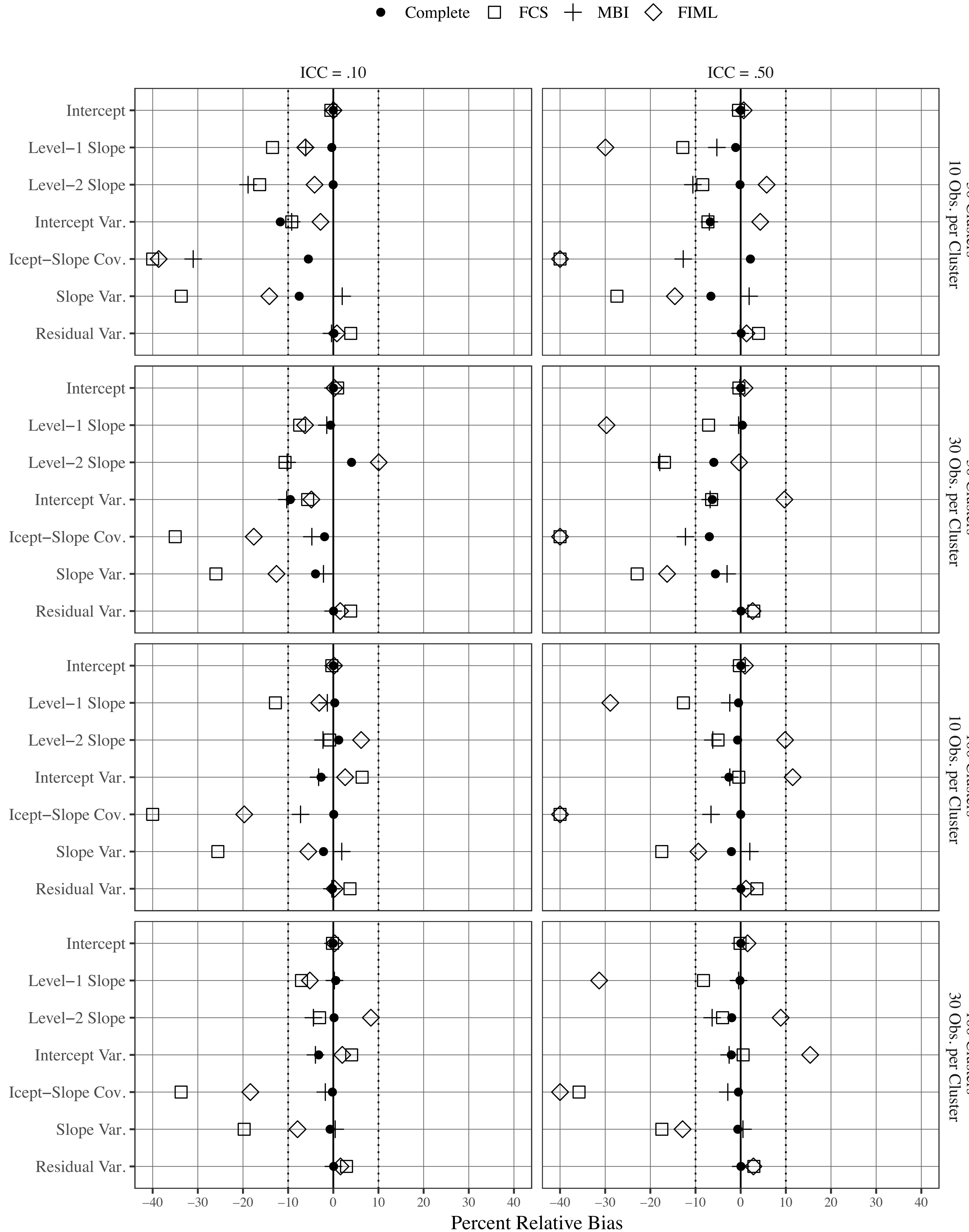


Simulation 1: Relative Bias, Large Sample (N = 50,000)
Complete
FCS
MBI
LWD
ICC = .10
ICC = .50
Intercept
Level−1 Slope
Level−2 Slope
Intercept Var.
Icept−Slope Cov.
Slope Var.
Residual Var.
Normal Level−2 Covariate 15% Missingness
Normal Level−2 Covariate 25% Missingness
Skewed Level−2 Covariate 15% Missingness
Skewed Level−2 Covariate 25% Missingness
−40
−30
−20
−10
0
10
20
30
40
Percent Relative Bias

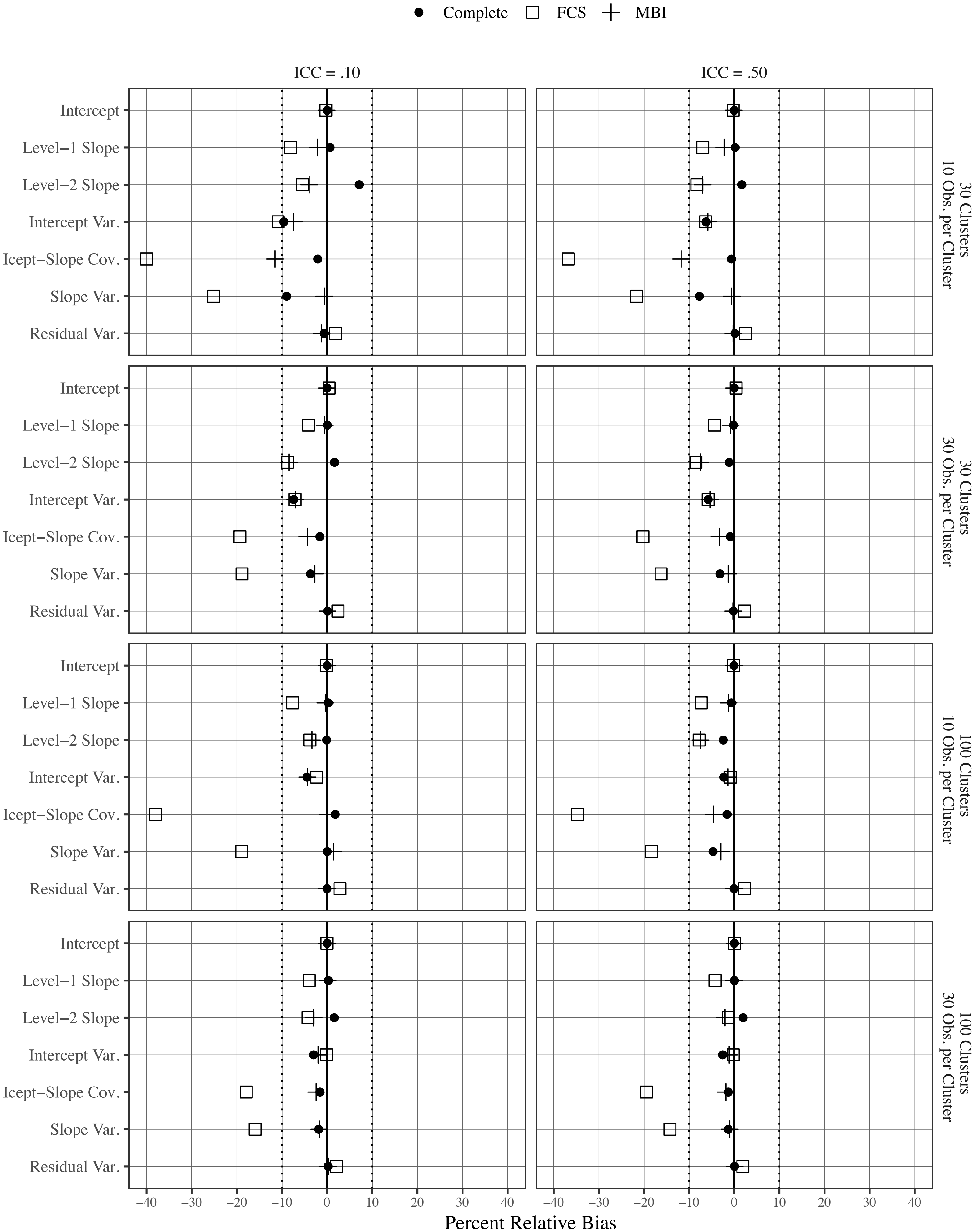
Simulation 1
Relative Bias: Normal Distribution, 15% Missingness



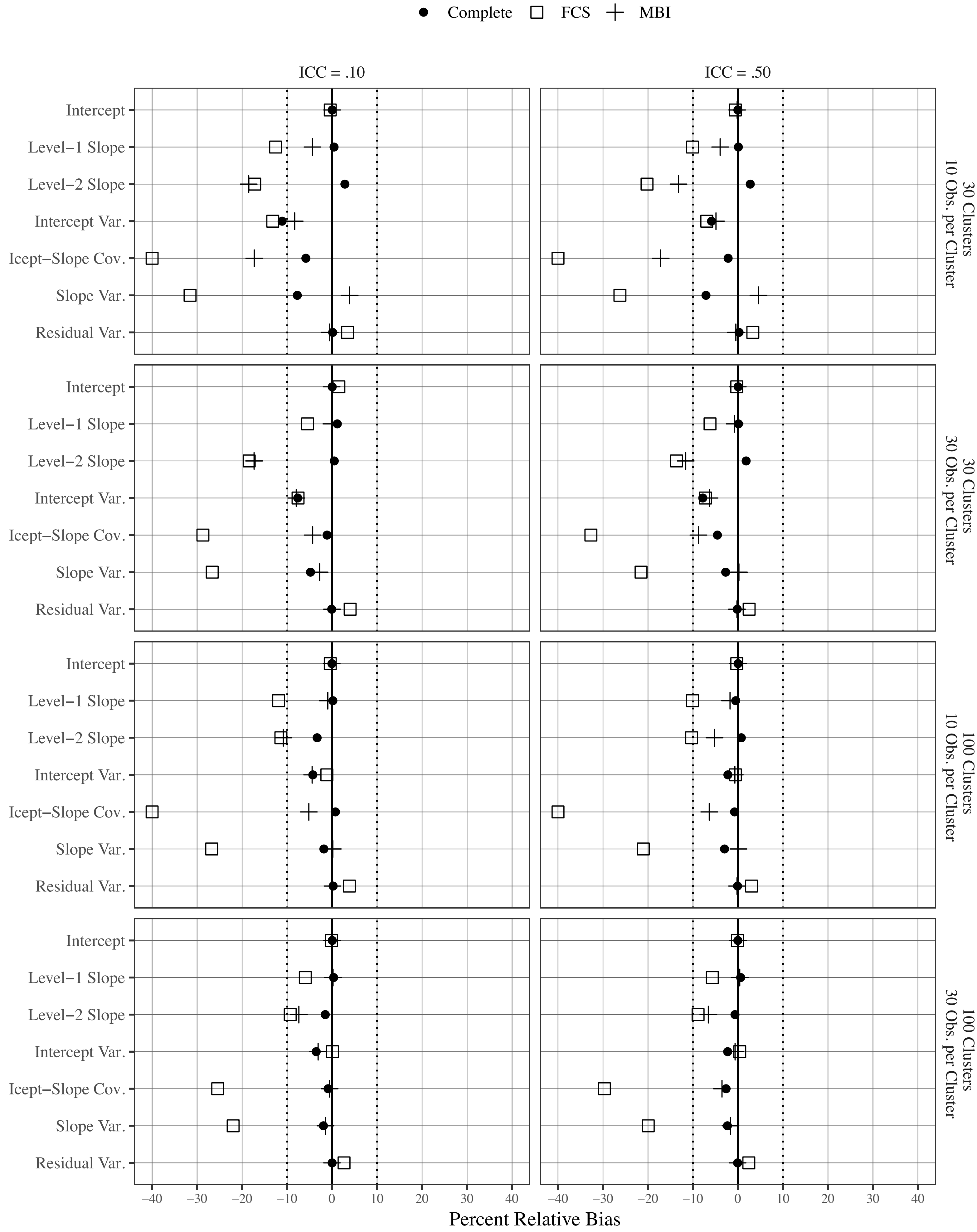
Simulation 1
Relative Bias: Normal Distribution, 25% Missingness



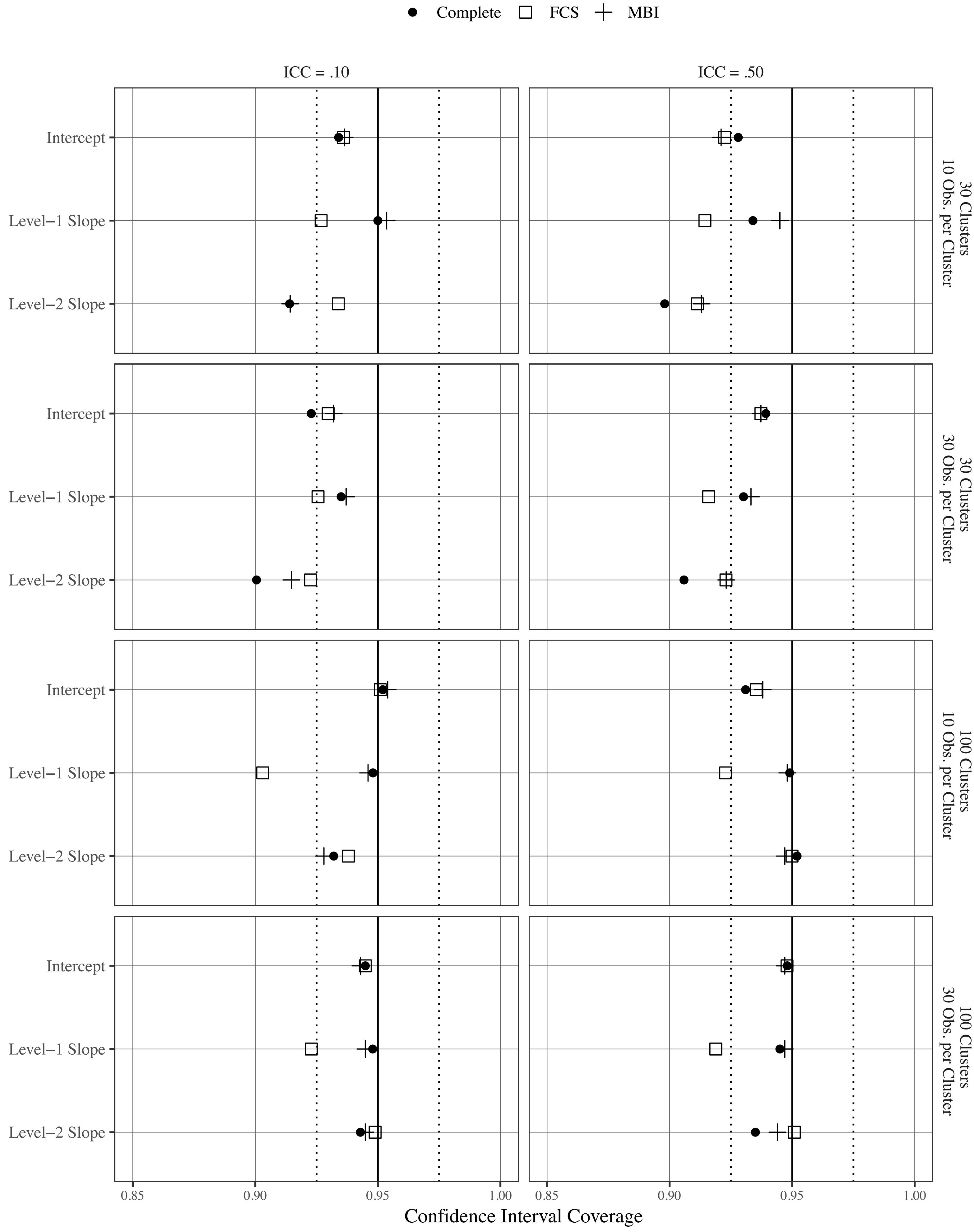
Simulation 1
Relative Bias: Skewed Distribution, 15% Missingness



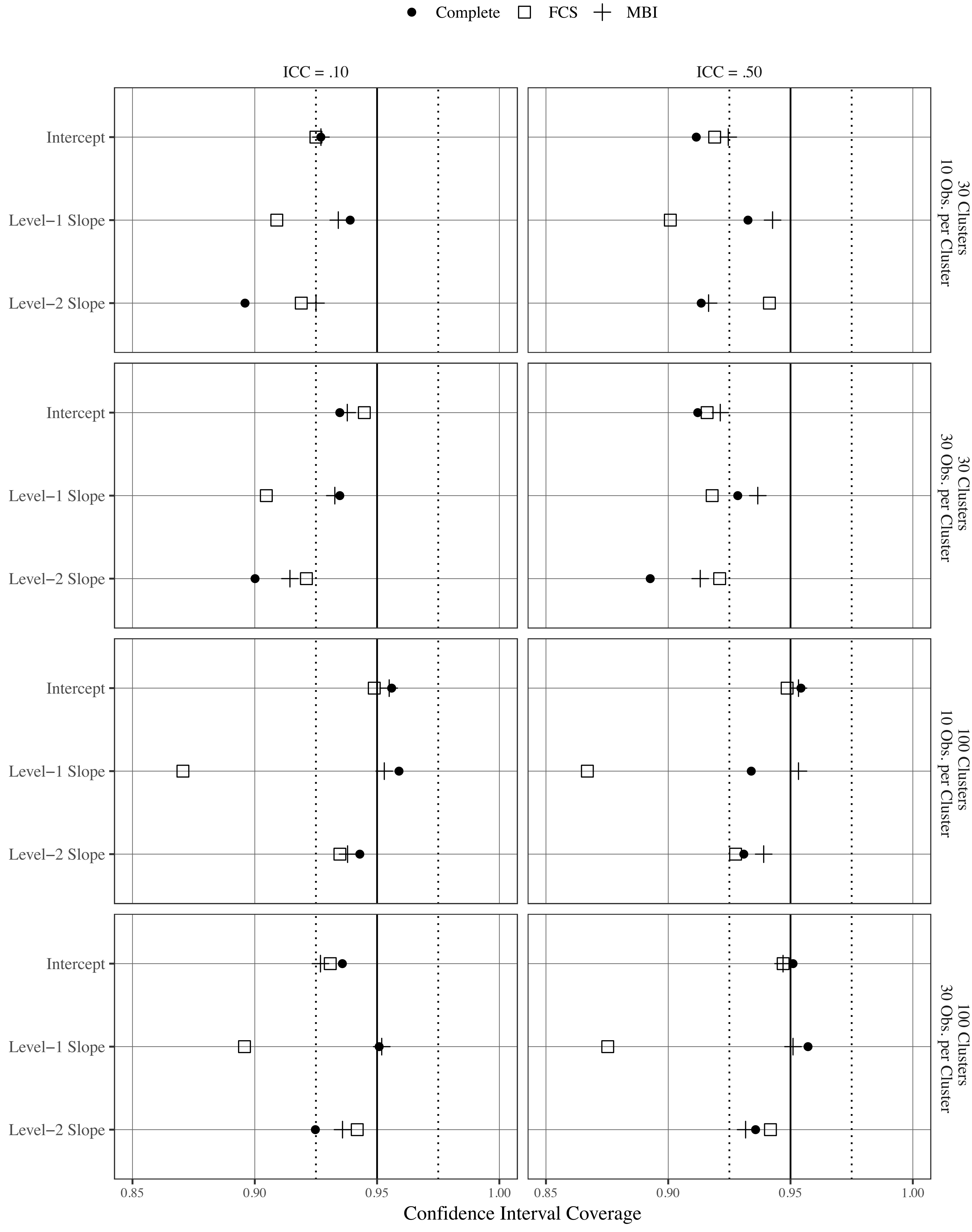
Simulation 1
Relative Bias: Skewed Distribution, 25% Missingness



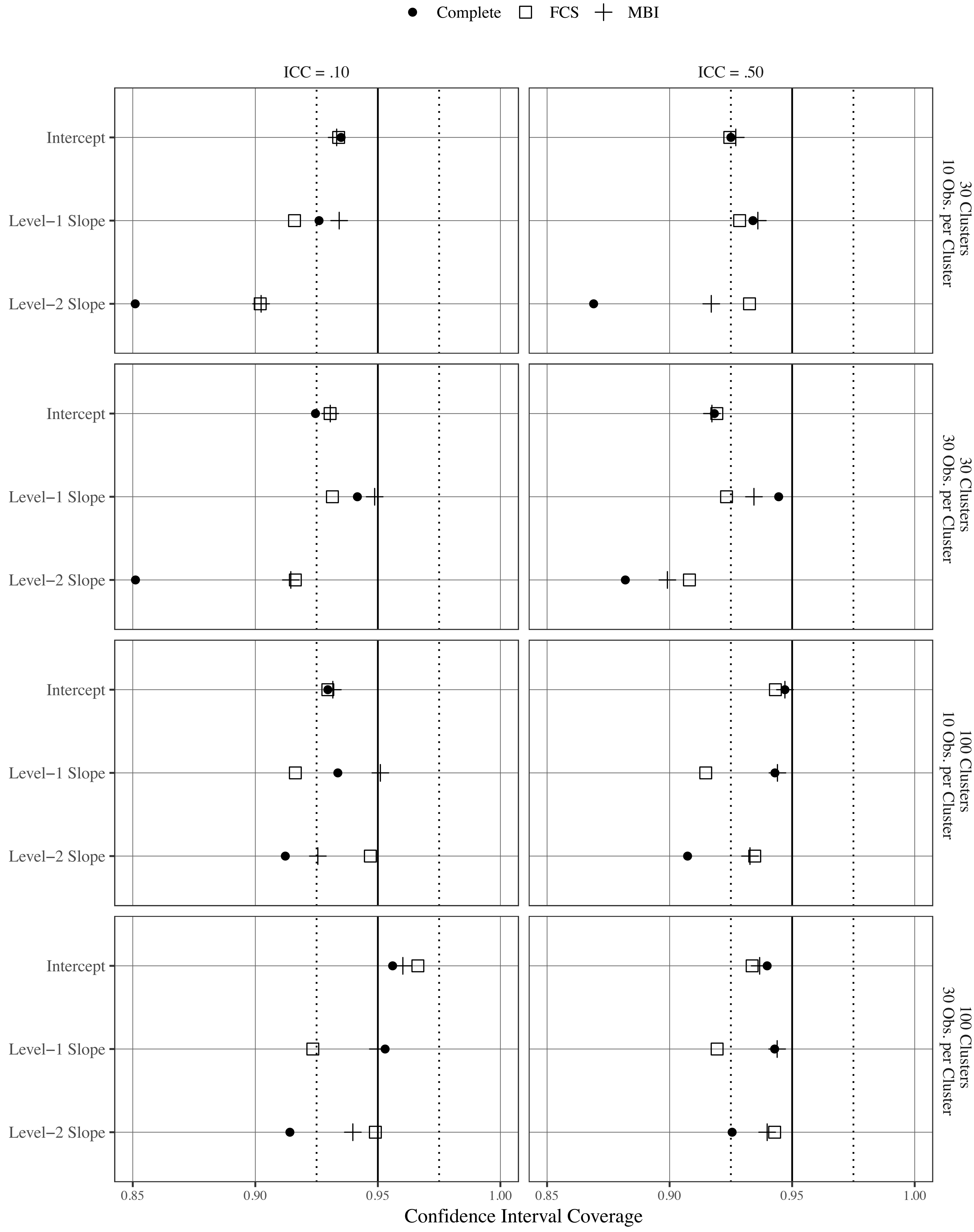
Simulation 1
Interval Coverage: Normal Distribution, 15% Missingness



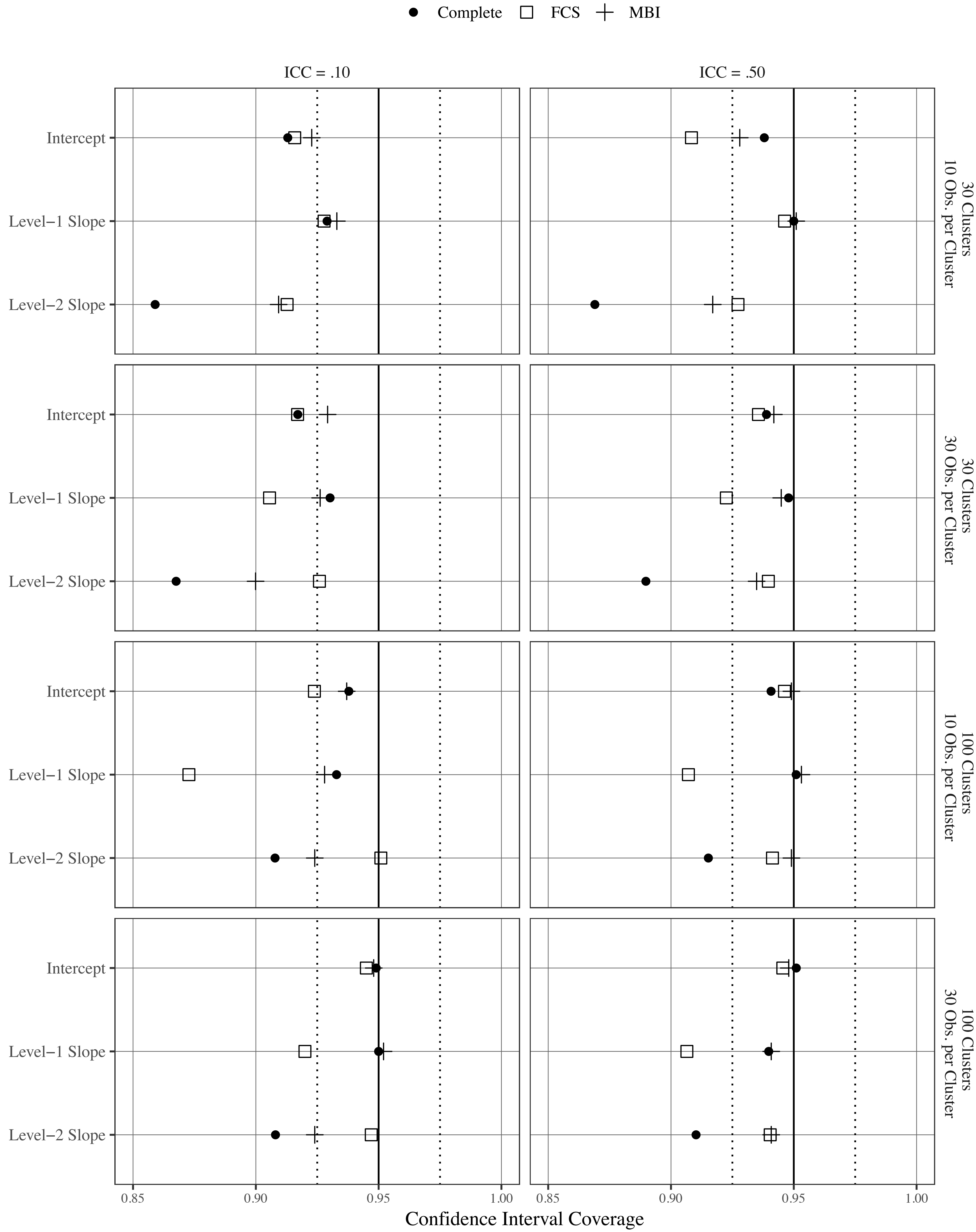
Simulation 1
Interval Coverage: Normal Distribution, 25% Missingness

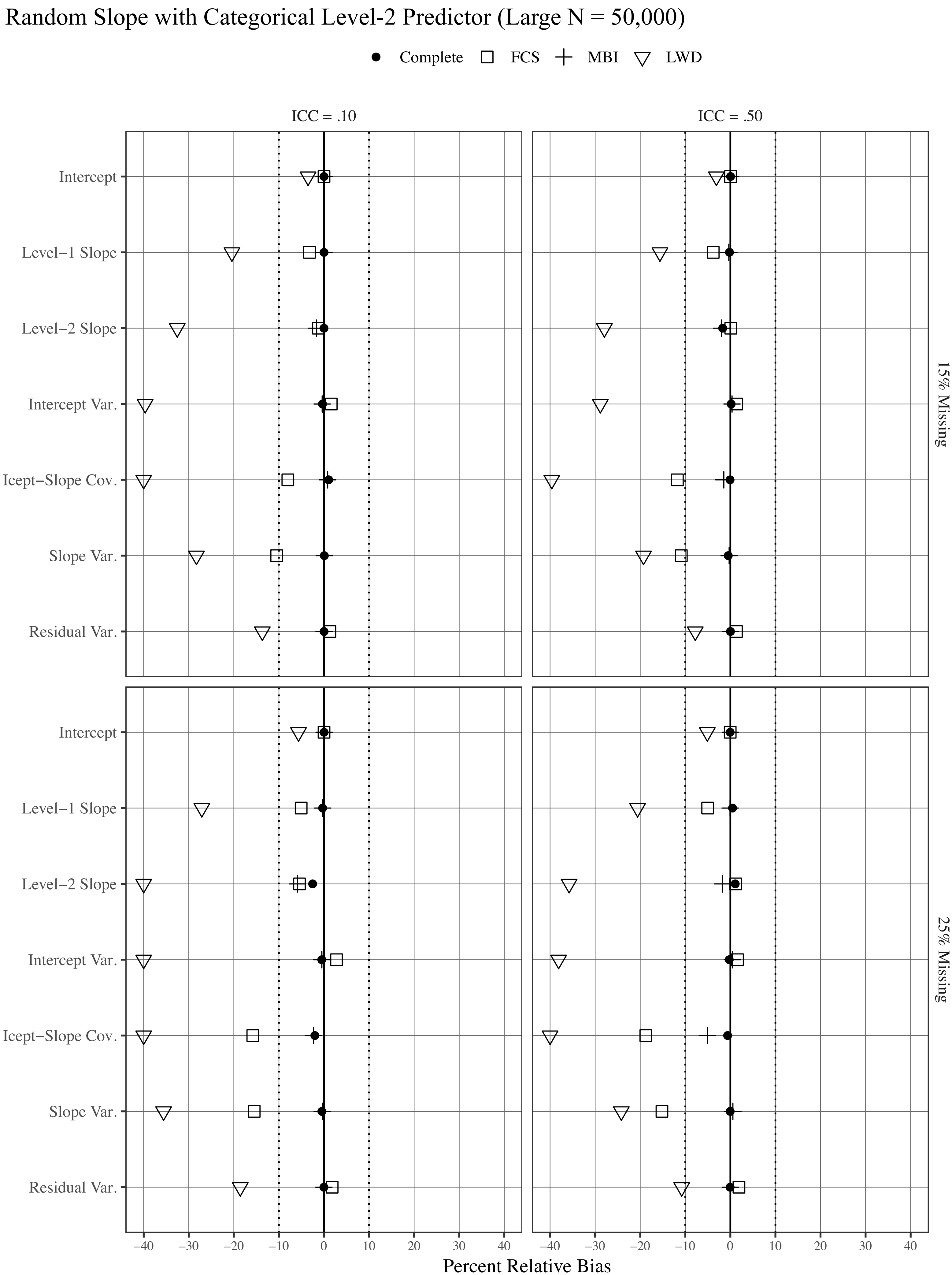


Simulation 1
Interval Coverage: Skewed Distribution, 15% Missingness

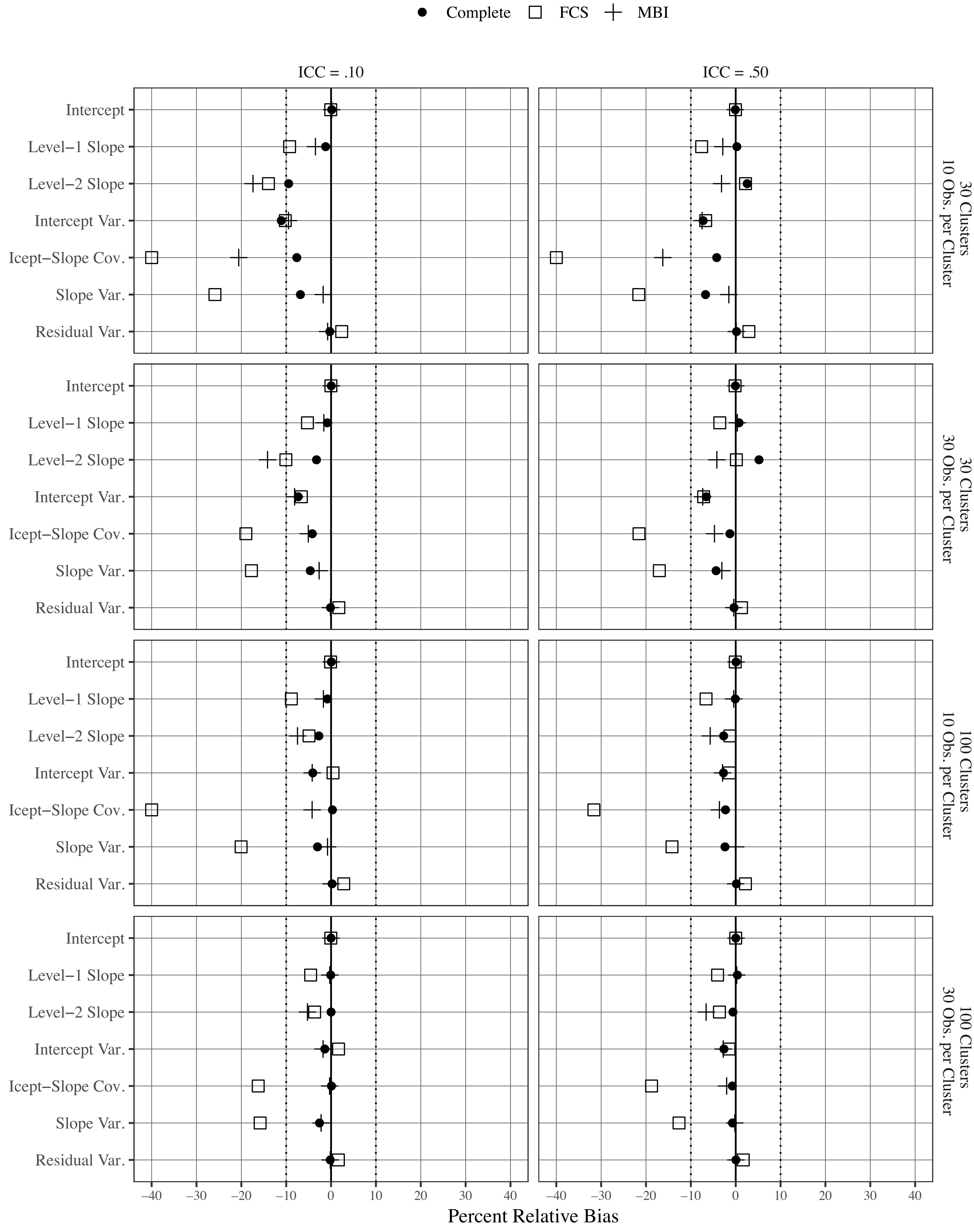


Simulation 1
 Interval Coverage: Skewed Distribution, 25% Missingness

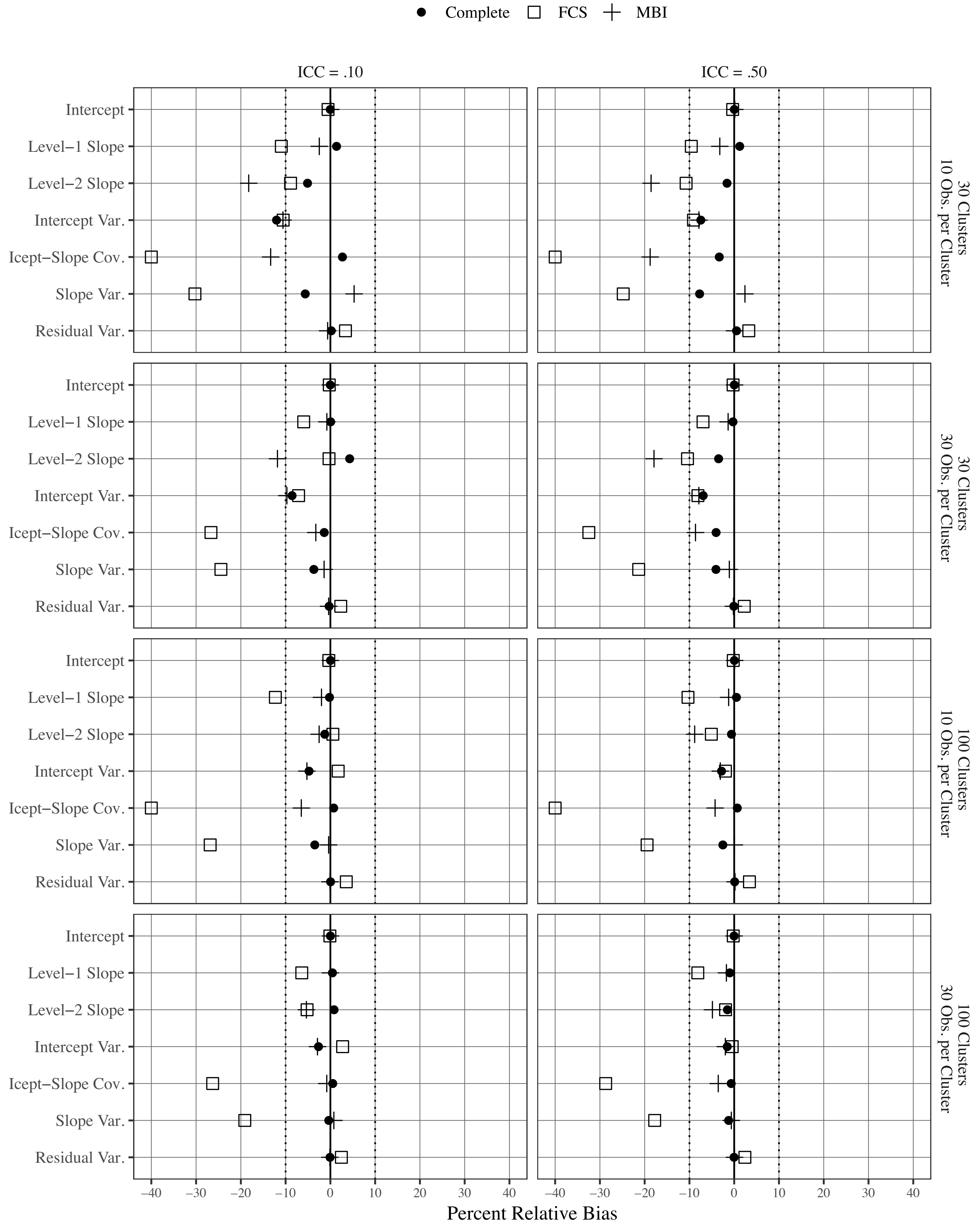



Random Slope with Categorical Level-2 Predictor (Large N = 50,000)
Complete
FCS
MBI
LWD
ICC = .10
ICC = .50
Intercept
Level−1 Slope
Level−2 Slope
Intercept Var.
Icept−Slope Cov.
Slope Var.
Residual Var.
15% Missing
25% Missing
−40
−30
−20
−10
0
10
20
30
40
Percent Relative Bias

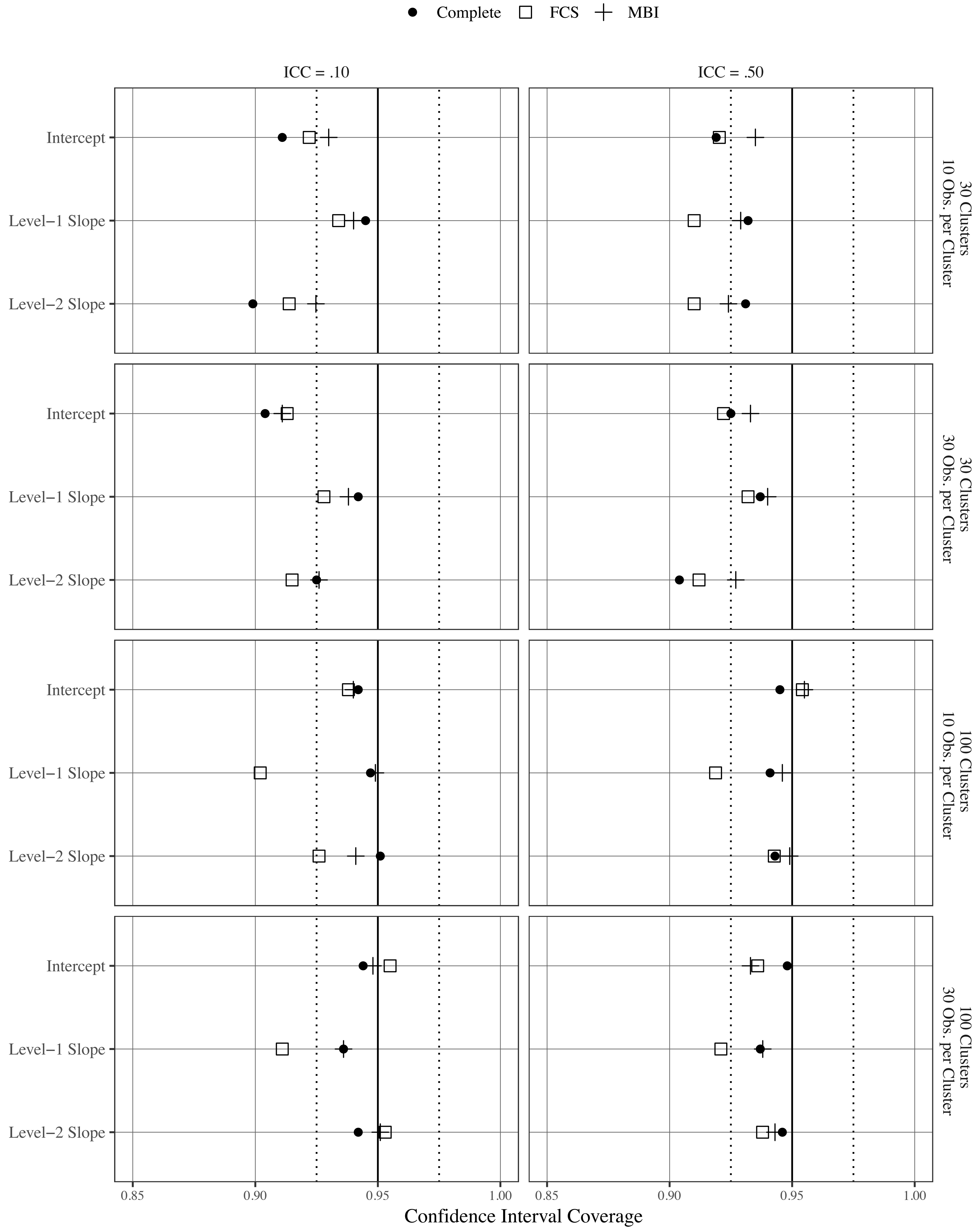
Simulation 2
Relative Bias: 15% Missingness



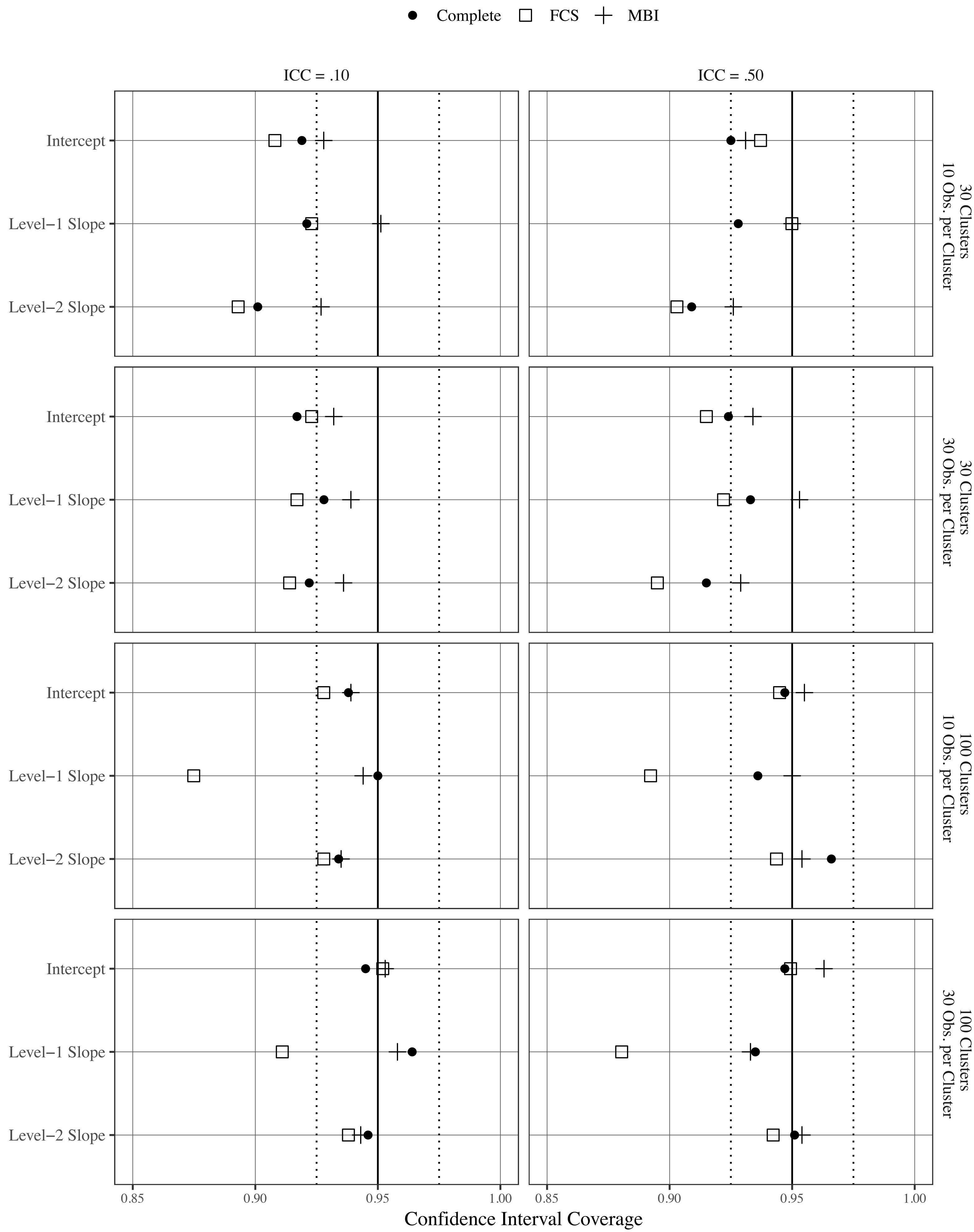
Simulation 2
Relative Bias: 25% Missingness



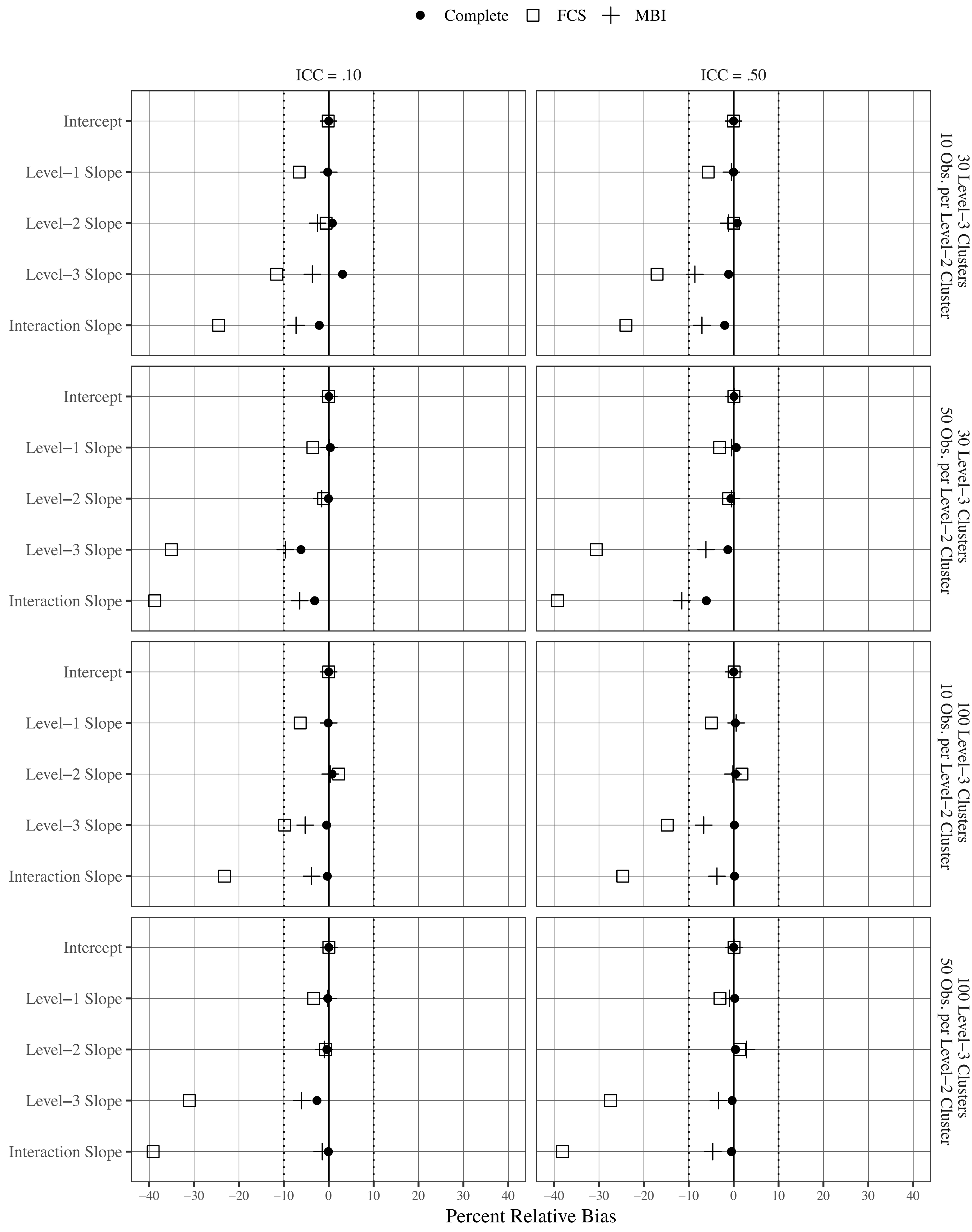
Simulation 2
Interval Coverage: 15% Missingness

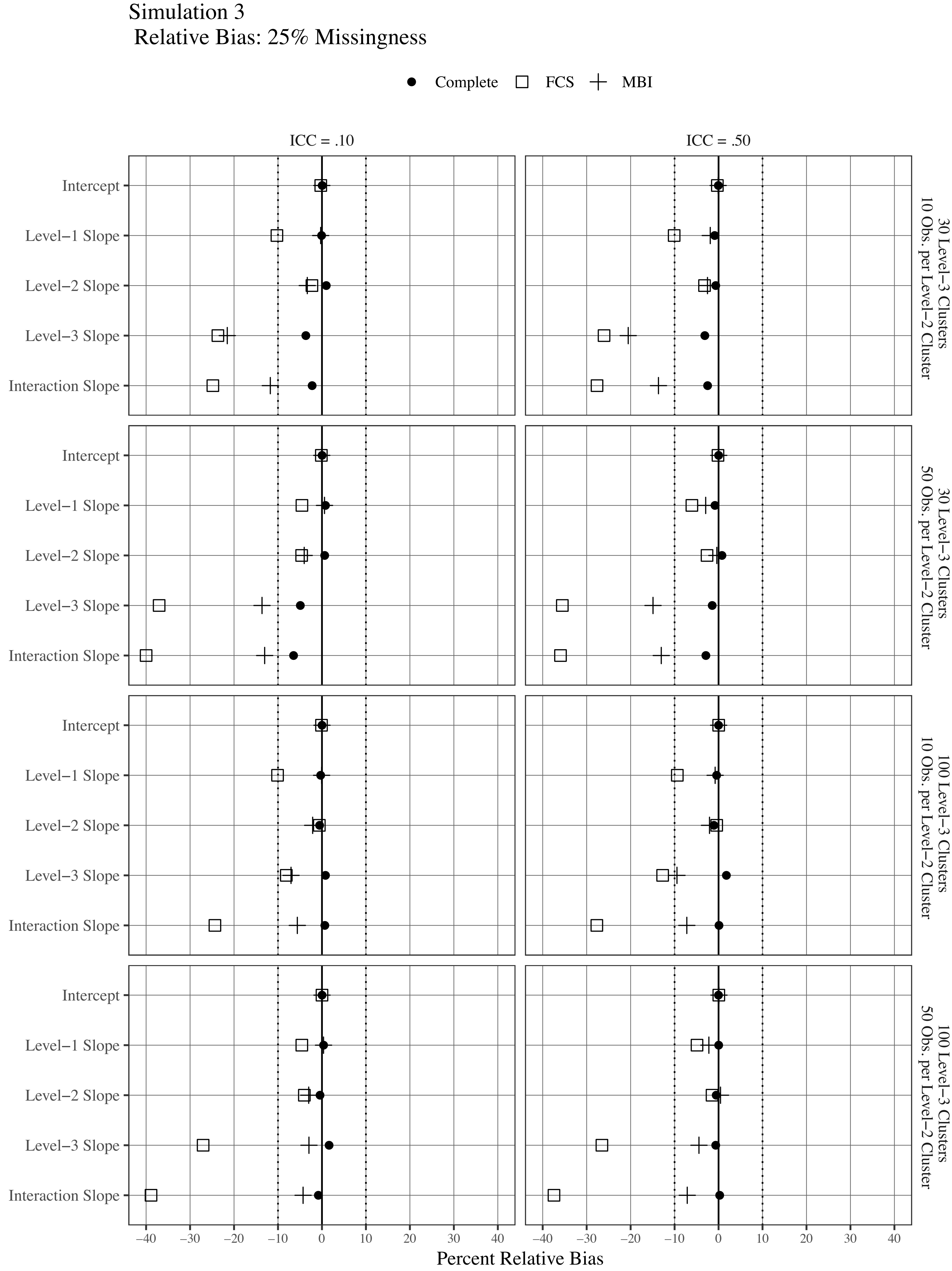


Simulation 2
Interval Coverage: 25% Missingness

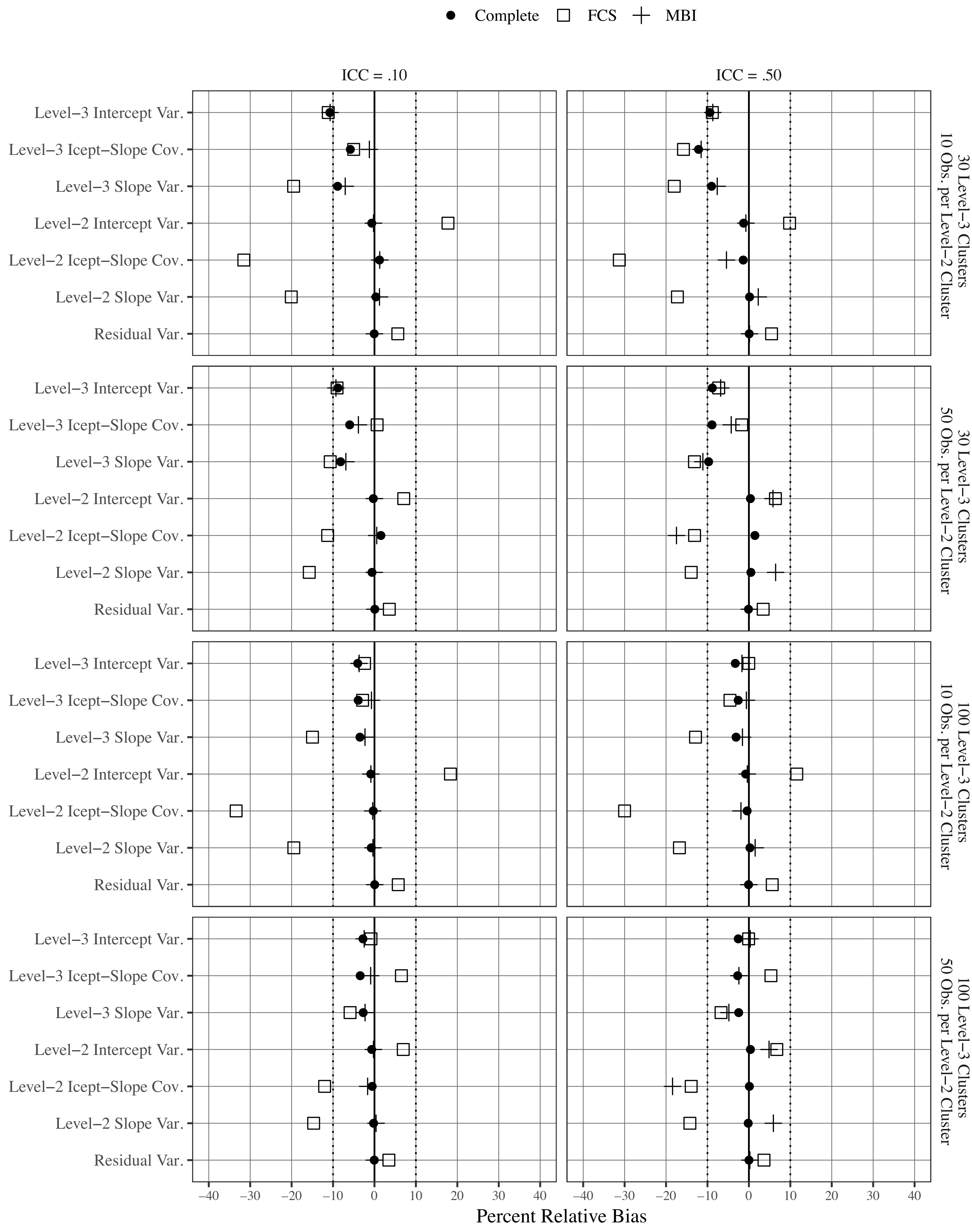


Simulation 3
Relative Bias: 15% Missingness

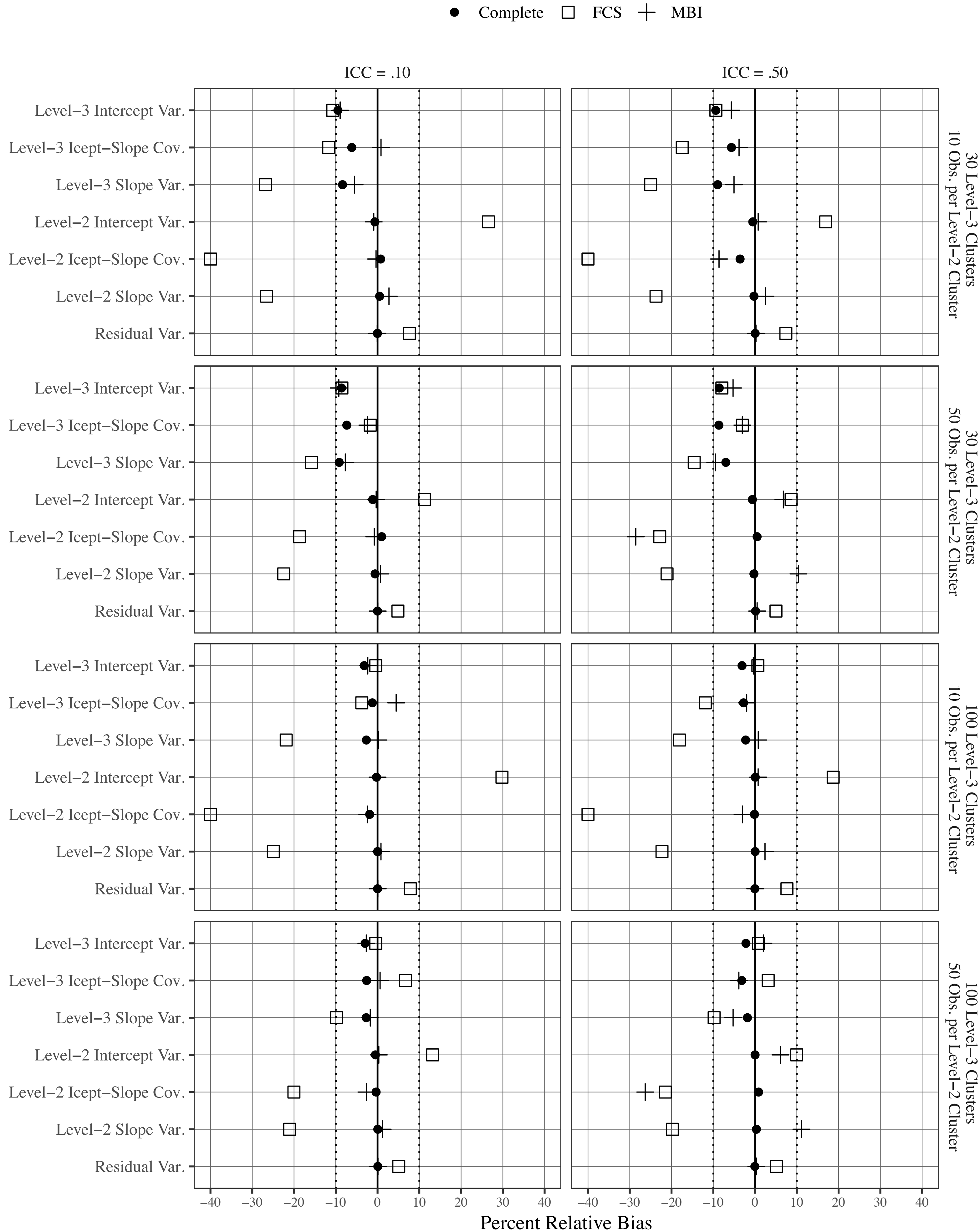



Simulation 3
Relative Bias: 25% Missingness
Complete
FCS
MBI
ICC = .10
ICC = .50
Intercept
Level−1 Slope
Level−2 Slope
Level−3 Slope
Interaction Slope
30 Level−3 Clusters
10 Obs. per Level−2 Cluster
30 Level−3 Clusters
50 Obs. per Level−2 Cluster
100 Level−3 Clusters
10 Obs. per Level−2 Cluster
100 Level−3 Clusters
50 Obs. per Level−2 Cluster
−40
−30
−20
−10
0
10
20
30
40
Percent Relative Bias

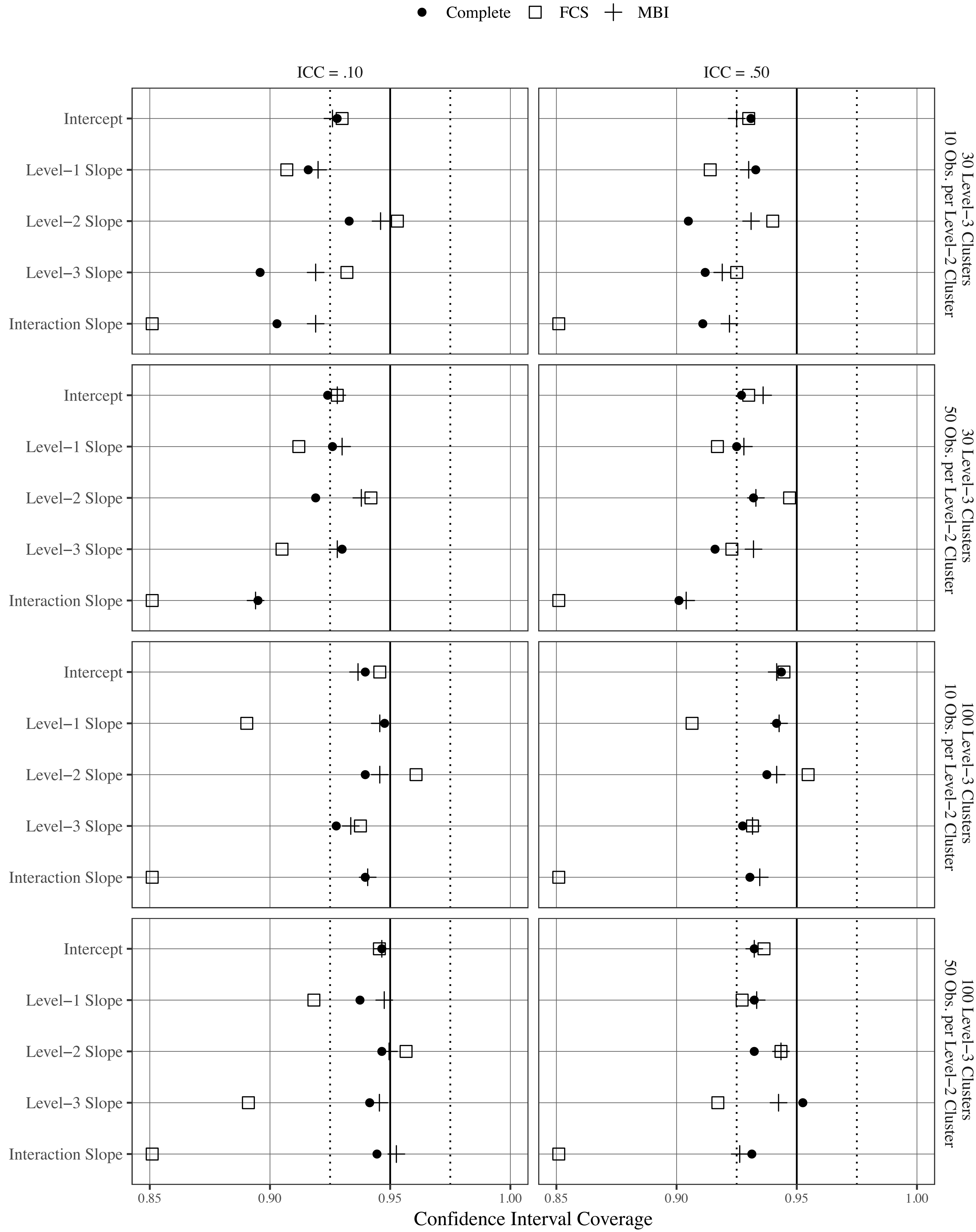
Simulation 3
Relative Bias: 15% Missingness



Simulation 3
Relative Bias: 25% Missingness



Simulation 3
Interval Coverage: 15% Missingness



Simulation 3
Interval Coverage: 25% Missingness