MULTIVARIATE BEHAVIORAL RESEARCH
2025, VOL. 60, NO. 5, 930–953

Empirical Bayes Priors for MCMC Estimation of the Multivariate Social Relations Model

Aditi M. Bhangale[a,b]* and Terrence D. Jorgensen[a]

[a]Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands; [b]Methodology and Statistics Department, Institute of Psychology, Leiden University, Leiden, The Netherlands

ABSTRACT

The social relations model (SRM) is a linear random-effects model applied to examine dyadic round-robin data within social networks. Such data have a unique multilevel structure in that dyads are cross-classified within individuals who may be nested within different social networks. The SRM decomposes perceptual or behavioral measures into multiple components: case-level random effects (in-coming and out-going effects) and dyad-level residuals (relationship effects), the associations among which are often of substantive interest. Multivariate SRM analyses are increasingly common, requiring more sophisticated estimation algorithms. This article evaluates Markov chain Monte Carlo (MCMC) estimation of multivariate-SRM parameters, compares MCMC to maximum-likelihood estimation, and introduces two methods to reduce bias in MCMC point estimates using empirical-Bayes priors. Four simulation studies are presented, two of which reveal dependency of small-group results on priors by manipulating location and precision hyperparameters, respectively. The third simulation study explores the impact of sampling more small groups on prior sensitivity. The fourth simulation study explores how Bayesian model averaging might compensate for underestimated variance due to empirical-Bayes priors. Finally, recommendations for future research are made and extensions of the SRM are discussed.

KEYWORDS

Social network; social relations model; maximum likelihood estimation; Markov chain Monte Carlo estimation; empirical Bayes priors

Introduction

The social relations model (SRM) is a linear random-effects model applied to examine dyadic data within social networks. Dyadic network data occur when perceptual or behavioral ratings are available for multiple pairs (or dyads) in a sample and each individual participates in more than one pair. The round-robin design (Gleason & Halperin, 1975), a common dyadic design (Kenny et al., 2006, Chapters 8, 9, & 11), is typically a multiple-group reciprocal approach wherein a participant i interacts with or provides perceptual ratings of every other member j of their group g (where $i \neq j \in 1, \ldots, n_g$ and $g \in 1, \ldots, G$). In a group of size n_g, each member participates in $n_g - 1$ dyads, resulting in $N_g = n_g \times (n_g - 1)$ interactions per group, and $N_g \times G$ interactions overall, given equal n_g across groups. Hence, every interaction within a dyad $\{ij\}$ yields two observations—i's perception of or behavior toward j and vice versa—stored in a vector $\mathbf{y}_{\{ij\}} = \begin{bmatrix} y_{ij} \\ y_{ji} \end{bmatrix}$. When there is no meaningful difference between dyad members—for example, on the basis of sex or age—dyads in a group are considered to be indistinguishable (e.g., same-sex group members are not distinguishable based on sex). We use braces $\{ij\}$ to indicate that the distinction between persons i and j is arbitrary.

Dyadic data have a complex nesting structure as an observation $y_{\{ij\}}$ is cross-classified within data from all dyads of which i is a member *and* data from all dyads of which j is a member. Multivariate data from dyadic designs enable quantifying the degree to which the total (co)variance in dyadic variables—for example, perceptual measures of liking or behavioral ratings of social mimicry during an interaction—is due to case-level differences among participants (such as their individual tendency to like [or be liked by] others), unique dyad-level characteristics (such as whether i particularly mimics [or is mimicked by] j beyond their individual tendencies to mimic [or be mimicked by] others), and group-level variation in the interactions (Kenny et al., 2006, p. 186–187).

ANOVA-based method-of-moments (Warner et al., 1979), restricted maximum likelihood (REML; Nestler, 2016), and full-information maximum likelihood (FIML; Nestler, 2018) estimators have been proposed to estimate the decomposed (co)variances of SRM components. Markov chain Monte Carlo (MCMC) estimation has also been proposed (Lüdtke et al., 2013), which enables Bayesian inference and incorporation of prior information. However, MCMC has only been explored for univariate and bivariate SRM, mainly to accommodate regressing SRM components on explanatory covariates (e.g., Jorgensen et al., 2018; Koster & Leckie, 2014; Lüdtke et al., 2018). Although Bayesian estimation can avoid some computational difficulties associated with frequentist estimation (e.g., analytic derivation of *SE*s), the accuracy of MCMC estimates can depend heavily on the specified prior distributions per parameter (McElreath, 2018, p. 31). This is particularly the case with small-samples, which are common in SRM research.

Our goal with this article is to demonstrate the impact of manipulating the location (i.e., accuracy) and scale (i.e., precision) of MCMC prior distributions on estimates of multivariate SRM (co)variance components. We also explore two methods to obtain empirically informed hyperparameters for MCMC priors of SRM (co)variances—specifically, choosing hyperparameters based on ANOVA-based method-of-moments or FIML estimates of SRM parameters. We begin with a brief overview of the SRM and its estimators. Then, we present the results of four simulation studies evaluating various MCMC prior specifications. We conclude with some considerations when conducting (Bayesian) SRM analyses and provide suggestions for future research.

Social relations models

Perceptual or behavioral ratings for a dyad $\{ij\}$ in a group g may be decomposed in the following random-effects model (Warner et al., 1979):

$$\mathbf{y}_{g\{ij\}} = \begin{bmatrix} y_{gij} \\ y_{gji} \end{bmatrix} = \mu_g + \begin{bmatrix} E_{gi} + A_{gj} + R_{gij} \\ E_{gj} + A_{gi} + R_{gji} \end{bmatrix}, \tag{1}$$

where μ_g is the mean of the dyadic variable (e.g., average social mimicry) in a group g. Usually, group-level differences are not of primary interest in social relations analyses. In this paper, we therefore only consider group-mean centered data, so μ_g need not be included in the SRM:

$$\mathbf{y}_{\{ij\}} = \begin{bmatrix} y_{ij} \\ y_{ji} \end{bmatrix} = \begin{bmatrix} E_i + A_j + R_{ij} \\ E_j + A_i + R_{ji} \end{bmatrix}. \tag{2}$$

Jorgensen et al. (2024) provided more details about the multivariate SRM with group-level random effects and (co)variance components.

The case level in most SRM applications is typically the person level, but SRM can also be applied to networks of households (Koster & Leckie, 2014) or countries (Dorff & Ward, 2013). Measurements $\mathbf{y}_{\{ij\}}$ are composed of ego (E) and alter (A) effects (also called actor and partner effects, perceiver and target effects, or sender and receiver effects) at the case level. An ego effect E_i is an out-going effect, interpreted as person i's the general perception of or behavior toward others. A person i's ego effect of liking, for instance, represents how much they generally like others. Likewise, A_i is an in-coming effect indicative of others' general perception of or behavior toward person i. A person i's alter effect of liking is then representative of how much they are generally liked by others. At the dyad level (also known as the relationship level), the relationship effects $\mathbf{R}_{\{ij\}}$ are residual effects composed of measurement error and i and j's unique perceptions of or behavior toward one another beyond their case-level tendencies. In the context of the liking example, relationship effects indicate the extent to which a pair i and j like one another beyond their individual tendencies to like (and be liked by) others.

Case-level effects are uncorrelated across individuals, but a particular i's ego effect E_i and alter effect A_i are assumed to be bivariate normally distributed with location 0 and a covariance matrix Σ_{EA}:

$$\begin{bmatrix} E_i \\ A_i \end{bmatrix} \sim \mathcal{MVN}\left(\mu_{EA} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma_{EA} = \begin{bmatrix} \sigma_E^2 & \\ \sigma_{EA} & \sigma_A^2 \end{bmatrix} \right), \tag{3}$$

where σ_E^2 and σ_A^2 are ego and alter variances, and σ_{EA} is a generalized covariance (generalized reciprocity ρ_{EA} when standardized) of the ego and alter effects for case i (Kenny et al., 2006, ch. 8). A positive σ_{EA}

implies that individuals with a greater ego effect also have a greater alter effect—that is, individuals who like others more are also generally liked more by others. A negative σ_{EA} means that individuals with a greater ego effect have a lower alter effect—that is, individuals who like others more are generally liked *less* by others.

Similarly, relationship effects R_{ij} and R_{ji} at the dyad level are also considered to be bivariate normally distributed:

$$\begin{bmatrix} R_{ij} \\ R_{ji} \end{bmatrix} \sim \mathcal{MVN}\left(\mu_R = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma_R = \begin{bmatrix} \sigma^2_R & \\ \sigma^2_R \rho_R & \sigma^2_R \end{bmatrix} \right), \tag{4}$$

where the relationship variances $\sigma^2_{R_{ij}}$ and $\sigma^2_{R_{ji}}$ are constrained to equality ($\sigma^2_{R_{ij}} = \sigma^2_{R_{ji}} = \sigma^2_R$) when dyads are indistinguishable. The R_{ij} and R_{ji} effects are assumed to be uncorrelated between dyads, but the correlation between R_{ij} and R_{ji} per dyad is labeled the dyadic reciprocity ρ_R (Kenny et al., 2006, ch. 8). A positive ρ_R indicates that, within a dyad $\{ij\}$, an increase in *i*'s rating of or behavior toward *j* is associated with an increase in *j*'s rating of or behavior toward *i*. For example, if *i* particularly likes *j*, then *j* also particularly likes *i*, beyond their case-level tendencies to like (or be liked by) others. A negative ρ_R means that an increase in *i*'s rating of or behavior toward *j* is associated in a decrease in *j*'s rating of or behavior toward *i*. That is, if *i* particularly likes *j*, then *j* likes *i* particularly *less* than their individual-level tendencies to like others and be liked by others.

Multivariate SRM

The SRM may also be used when multiple dyadic variables are measured. In a trivariate case, for example, the vector of SRM equations expands as follows:

$$\begin{bmatrix} \mathbf{y}_{1,\{ij\}} \\ \mathbf{y}_{2,\{ij\}} \\ \mathbf{y}_{3,\{ij\}} \end{bmatrix} = \begin{bmatrix} y_{1,ij} \\ y_{1,ji} \\ y_{2,ij} \\ y_{2,ji} \\ y_{3,ij} \\ y_{3,ji} \end{bmatrix} = \begin{bmatrix} E_{1,i} \\ E_{1,j} \\ E_{2,i} \\ E_{2,j} \\ E_{3,i} \\ E_{3,j} \end{bmatrix} + \begin{bmatrix} A_{1,j} \\ A_{1,i} \\ A_{2,j} \\ A_{2,i} \\ A_{3,j} \\ A_{3,i} \end{bmatrix} + \begin{bmatrix} R_{1,ij} \\ R_{1,ji} \\ R_{2,ij} \\ R_{2,ji} \\ R_{3,ij} \\ R_{3,ji} \end{bmatrix}, \tag{5}$$

where $\mathbf{y}_{1,\{ij\}}$, $\mathbf{y}_{2,\{ij\}}$, and $\mathbf{y}_{3,\{ij\}}$ are variables measured with a round-robin design—such as reports of liking at first impression, behavioral ratings of mimicry during an interaction, and subsequently reported post-interaction liking (Salazar Kämpf et al., 2018).

As with the univariate SRM, case-level effects for an individual *i* are assumed to be multivariate normally distributed:

$$\begin{bmatrix} E_{1,i} \\ A_{1,i} \\ E_{2,i} \\ A_{2,i} \\ E_{3,i} \\ A_{3,i} \end{bmatrix} \sim \mathcal{MVN}\left(\mu_{EA} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_{EA} = \begin{bmatrix} \sigma^2_{E_1} & & & & & \\ \sigma_{A_1,E_1} & \sigma^2_{A_1} & & & & \\ \sigma_{E_2,E_1} & \sigma_{E_2,A_1} & \sigma^2_{E_2} & & & \\ \sigma_{A_2,E_1} & \sigma_{A_2,A_1} & \sigma_{A_2,E_2} & \sigma^2_{A_2} & & \\ \sigma_{E_3,E_1} & \sigma_{E_3,A_1} & \sigma_{E_3,E_2} & \sigma_{E_3,A_2} & \sigma^2_{E_3} & \\ \sigma_{A_3,E_1} & \sigma_{A_3,A_1} & \sigma_{A_3,E_2} & \sigma_{A_3,A_2} & \sigma_{A_3,E_3} & \sigma^2_{A_3} \end{bmatrix} \right). \tag{6}$$

In Equation 6, the covariances between the case-level effects of the three variables can be estimated. For example, σ_{A_2,E_1} is an ego–alter covariance that can be used to investigate whether individuals who generally like others more at first impression are also mimicked more during a subsequent interaction. Likewise, σ_{E_3,E_2} is an ego–ego covariance that estimates whether individuals who mimic others more during a social interaction subsequently display a greater liking toward others post-interaction. Alter–alter covariances (e.g., σ_{A_3,A_1}) can be interpreted in a similar manner.

The equality constraints at the dyad level extend to the multivariate case, following from the assumption that dyad members are indistinguishable:

$$\begin{bmatrix} R_{1,ij} \\ R_{1,ji} \\ R_{2,ij} \\ R_{2,ji} \\ R_{3,ij} \\ R_{3,ji} \end{bmatrix} \sim \mathcal{MVN}\left(\mu_R = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_R = \begin{bmatrix} \sigma^2_{R_1} & & & & & \\ \sigma^2_{R_1}\rho_{R_1} & \sigma^2_{R_1} & & & & \\ \sigma^{\text{intra}}_{R_2,R_1} & \sigma^{\text{inter}}_{R_2,R_1} & \sigma^2_{R_2} & & & \\ \sigma^{\text{inter}}_{R_2,R_1} & \sigma^{\text{intra}}_{R_2,R_1} & \sigma^2_{R_2}\rho_{R_2} & \sigma^2_{R_2} & & \\ \sigma^{\text{intra}}_{R_3,R_1} & \sigma^{\text{inter}}_{R_3,R_1} & \sigma^{\text{intra}}_{R_3,R_2} & \sigma^{\text{inter}}_{R_3,R_2} & \sigma^2_{R_3} & \\ \sigma^{\text{inter}}_{R_3,R_1} & \sigma^{\text{intra}}_{R_3,R_1} & \sigma^{\text{inter}}_{R_3,R_2} & \sigma^{\text{intra}}_{R_3,R_2} & \sigma^2_{R_3}\rho_{R_3} & \sigma^2_{R_3} \end{bmatrix} \right). \tag{7}$$

In Equation 7, $\sigma^{\text{intra}}_{R_2,R_1}$ is termed an *intra*personal covariance and indicates whether, in dyad $\{ij\}$, *i*'s rating of *j* on the first dyadic variable y_1 is associated with *i*'s rating of *j* on the second dyadic variable y_2. That is, it may be hypothesized that *i*'s unique first-impression liking of *j* is associated with *i*'s subsequent mimicry of *j* during an interaction. An *inter*personal covariance $\sigma^{\text{inter}}_{R_3,R_1}$ indicates whether *i*'s rating of *j* on the first dyadic variable y_1 is associated with *j*'s rating of *i* on the third dyadic variable y_3. In the context of the liking–mimicry example, it is possible to estimate the extent to which *i*'s unique first-impression liking of *j* is associated with *j*'s post-interaction liking of *i*.

Given that dyadic observations are cross-classified within each case/individual, the multivariate SRM decomposes the covariance matrix of dyadic observations Σ_y into case- and dyad-level components (see appendix of Ten Hove et al., 2025, for derivations):

$$\Sigma_y = \Sigma_{EA} + \Sigma_{AE} + \Sigma_R, \tag{8}$$

where Σ_{AE} is simply a rearrangement of the Σ_{EA} displayed in Equation 6, with the order of components as $[A_1, E_1, A_2, E_2, \ldots]$. Jorgensen et al., (2024, Equation 23) provided a decomposition that includes group-level (co)variances.

Estimation of multivariate SRMs

The ANOVA-based method-of-moments estimator (Warner et al., 1979) uses the sample mean and mean-square estimated components of scores on a dyadic variable to compute SRM (co)variances. The method is adequate to estimate SRM (co)variances for complete round-robin designs (i.e., when all group members rate all others). However, existing software—SOREMO (Kenny, 2013) and the R package `TripleR` (Schönbrodt et al., 2012, 2022)—has only been implemented for univariate and bivariate SRMs. Estimating multivariate-SRM parameters using existing software would require modeling multiple bivariate analyses, which is time-consuming and computationally intensive when many round-robin variables are investigated. Additionally, inadmissible solutions are produced when true values of (co)variances are near their border—e.g., near-zero variances and correlations near ±1 (Kenny et al., 2006, p. 212–213)—or when the number of round-robin groups G and group sizes n_g are small (Lüdtke et al., 2013).

FIML estimation (Nestler, 2018; Nestler et al., 2020), implemented in the R package `srm` (Nestler, Robitzsch, et al., 2022b), is also available to estimate multivariate-SRM parameters. Nestler et al. (2020) apply a Fisher-scoring algorithm to derive SRM-(co)variance point and *SE* estimates, which is applicable with unbalanced or incomplete normally distributed data, overcoming some limitations of the method-of-moments estimator above. However, the accuracy of FIML generally depends on sample size (Bhangale & Jorgensen, 2024; Hoff, 2005) and the shape of the SRM variance components' sampling distributions (Lüdtke et al., 2013).

MCMC estimators, for example Gibbs sampling (Gill & Swartz, 2001; Hoff, 2005; Lüdtke et al., 2013) and Hamiltonian Monte Carlo (HMC; Jorgensen et al., 2018, 2024), provide some practical advantages. These approaches handle unbalanced and incomplete designs (Gill & Swartz, 2001) and can incorporate model uncertainty well. Additionally, MCMC estimators possess the added benefit of estimating complex models that might be computationally intractable for FIML. The possibility to specify prior distributions informed by expectations and previous knowledge may result in more accurate estimates of SRM parameters. The mean (expected a posteriori; EAP), median (50th percentile), or mode (maximum a posteriori; MAP) of the empirical posterior distribution may be selected as the point estimate(s) of the SRM parameters. Although EAPs and MAPs provide similar estimates given non-negligible true values of SRM (co)variances (Bhangale & Jorgensen, 2024), MAP estimates can be less biased than EAP estimates when the distribution of a parameter is skewed (i.e., in boundary conditions such as near-zero variances) and less information is available from the sample to estimate parameters (Lüdtke et al., 2013). Thus, EAP and MAP estimates will diverge when variances are expected to be near-zero in, for example, small-group conditions unless informative prior distributions are specified (Lüdtke et al., 2013; Ten Hove et al., 2020).

Recently, Jorgensen et al. (2024) proposed using the No-U-Turn Sampler (NUTS; Hoffman & Gelman, 2014), a modified HMC algorithm, to estimate SRM (co)variances. NUTS is available in the Bayesian modeling R package `rstan` (Stan Development Team, 2023) and has been adapted for the SRM in the R package `lavaan.srm` (Jorgensen, 2023). Whereas Gibbs sampling updates posterior point estimates sequentially, NUTS samples a complete vector of all unknown point estimates simultaneously from the posterior distribution. In addition, NUTS does not require conjugate prior distributions, resulting in greater flexibility for prior specification.

The unknown parameters to be estimated via MCMC in the `lavaan.srm` package include the level-specific random effects, and *SD*s of and correlations among the random effects, as well as means when not analyzing group-mean-centered variables. Jorgensen et al. (2024) provide exhaustive technical details about the MCMC algorithm and its application, and we provide details relevant to our current studies when describing Simulation Study 1. The accuracy and efficiency of NUTS to estimate SRM parameters has been previously investigated by Bhangale and Jorgensen (2024), who found that specifying software-default (diffuse) priors for *SD*s and correlations results in biased point estimates compared to FIML. This is consistent with Smid et al.'s (2020) systematic review, which revealed that Bayesian estimates based on software-default priors display more bias than frequentist estimates in the small samples that are common in SRM research. Smid et al. (2020) instead recommended the use of thoughtful (i.e., theoretically informed) and data-dependent priors to derive more accurate estimates. Given that thoughtful and data-dependent priors are yet to be explored in the context of the SRM, in this article, we conducted

four simulation studies to explore sensitivity of results to various MCMC prior specifications.

In sufficiently large samples, estimated posterior distributions are influenced almost exclusively by information provided in the data (likelihood), yielding results that are less dependent on the specified priors. However, the intensive nature of round-robin data collection frequently motivates research designs that sample from (multiple) small groups. Thus, all our simulations explore a range of small and large round-robin group sizes typically encountered in SRM applications. The article is organized such that the design of each subsequent simulation reflects our learning from the previous results and addresses the questions that arise from them. We summarize the goals of each simulation here, before detailed reports are provided in subsequent sections. First, we ascertained the effect of manipulating the magnitude of prior information for highly accurate priors. Second, for a given magnitude of prior information, we compared theoretically versus empirically informed prior locations. We then sampled larger numbers of small groups, to explore whether this can stabilize estimation of SRM components or minimize prior sensitivity. Finally, we explored Bayesian model averaging as a potential solution for the "double-dipping" problems inherent in using the same data to determine priors as are used to update the priors when estimating posterior distributions (i.e., underestimated uncertainty; Carlin & Louis, 2000b, p. 1287; Zitzmann et al., 2024, p. 3).

Simulation study 1: Prior precision

All of our simulation studies involve the trivariate SRM described in Equations 5, 6, and 7, using parameters described in the following section. We first compared FIML estimation to MCMC using (a) diffuse (current software-default) prior distributions or (b) accurately located priors with expected values equal to the population values. Although the latter are unrealistic because parameters are unknown in practice, it is informative to learn about the quality of MCMC estimates in the "worst-case" (diffuse) and "best-case" scenarios (i.e., accurately located priors supplement minimal information from data). Because the discrepancy between a prior location and a true parameter can introduce bias[1] in the posterior estimate (e.g., EAP or MAP)—particularly when proportionally less information is available from the data—we designed two simulations to separately investigate the effects of prior precision (using accurate prior locations in Simulation 1) and of prior accuracy (by varying prior locations in Simulation 2).

We describe in the Method how we manipulated the precision of the prior information in the best-case scenario (accurately located priors). We refer to the accurately located priors as *prophetic* priors, to acknowledge the need for an uncanny ability to know the unknowable in order to specify such a perfectly accurate prior location.

Method

We used the population values specified by (Nestler et al., 2020, see https://osf.io/9twkm/) to derive the population correlations and *SD*s in Tables 1 and 2. Nestler et al. (2020) chose population values based on previous SRM research such that the majority of a dyadic variable's variance is in the relationship component (compare the *SD* columns of Tables 1 and 2), and that ego effects have greater variance than alter effects (compare odd rows to even rows in Table 2).

Prior distributions

Priors distributions for random effects and residuals are the multivariate normal distributions shown in Equations 6 and 7, respectively. Because the hyperparameters Σ_{EA} and Σ_R are unknown, they are also estimated from the data[2] using hyperprior distributions for the *SD*s and correlations. We describe diffuse priors below, which are the default settings in the `lavaan.srm` package. More informative prior conditions in each study are described in later sections.

Hyperprior distributions for *SD*s in Table 1 were specified as a t distribution (left-truncated at 0) with $\nu = 4$ degrees of freedom:

$$\sigma > 0 \sim t(\nu = 4, \mu, \varsigma), \tag{9}$$

which has been shown to work well for variance-decomposition models in past simulation research (e.g., Ten Hove et al., 2020) and is the default prior for scale parameters in the R package `brms` (Bürkner, 2017). A location parameter μ can be specified to shift the t distribution's mean, and a scaling parameter ς

[1]Prior-induced bias can be compensated by the reduction in sampling variability offered by the informative prior, thus yielding a lower overall mean-squared error (MSE), which combines both (squared) bias and sampling variance. This trade-off between accuracy and precision is well documented (see, e.g., Zitzmann et al., 2021).

[2]This has also been referred to as a fully Bayesian variety of empirical-Bayes estimation—"Bayes empirical Bayes" (Carlin & Louis, 2000b, p. 1286)—which is distinct from the variety we propose and evaluate in Simulation Study 2. Gelman et al. (2013) refer to this variety instead as "hierarchical Bayes" (ch. 5, section 5.2).

Table 1. Dyad-level population *SD* and correlation values for Simulations 1–4.

	SRM component	1	2	3	4	5	6	*SD*
1.	$R_{1,ij}$	–						0.949
2.	$R_{1,ji}$	.167	–					0.949
3.	$R_{2,ij}$	.538	.135	–				0.940
4.	$R_{2,ji}$	.135	.538	.222	–			0.940
5.	$R_{3,ij}$	.705	.176	.569	.142	–		1.255
6.	$R_{3,ji}$	.176	.705	.142	.569	.060	–	1.255

Table 2. Case-level population *SD* and correlation values for Simulations 1–4.

	SRM component	1	2	3	4	5	6	*SD*
1.	$E_{1,i}$	–						0.775
2.	$A_{1,i}$	.236	–					0.548
3.	$E_{2,i}$	.703	.124	–				0.881
4.	$A_{2,i}$	.093	.528	.099	–			0.415
5.	$E_{3,i}$	.574	.102	.606	.080	–		0.629
6.	$A_{3,i}$	.093	.528	.099	.419	.195	–	0.415

can vary its spread. The `lavaan.srm` package specifies diffuse priors by default (as elaborated by Jorgensen et al., 2024), setting both μ and ς as the sample[3] *SD*, which is the maximum value that any SRM component can have.

Hyperprior distributions for correlations were specified as a Beta(x, $\alpha = 1.5, \beta = 1.5$) distribution, which has an expected value $M = .5$ and $SD = .25$. During MCMC estimation, parameters sampled from a Beta distribution lie in the range $\{0, 1\}$, which are rescaled to the $\{-1, +1\}$ range using the transformation $(2x - 1)$, so the rescaled-Beta(1.5, 1.5) prior implies correlations vary with $M = 0$ and $SD = .5$. This is equivalent to placing a prior on transformed correlations:

$$\frac{\rho + 1}{2} \sim \text{Beta}(\alpha = 1.5, \beta = 1.5), \tag{10}$$

which is the same method employed by the R package `blavaan` (Merkle & Rosseel, 2018) for structural equation models.

Simulation conditions

To specify prophetic priors, we centered each SRM parameter's prior distribution at its population value. For each *SD*, the population parameter was specified as the location hyperparameter μ of the prophetic *t* distribution, whose scaling hyperparameters were specified as either $\varsigma = 0.05$ (Pr-0.05), $\varsigma = 0.10$ (Pr-0.1), or $\varsigma = 0.20$ (Pr-0.2). For prophetic Beta priors of SRM correlations, an optimization algorithm[4] was used to identify the α and β hyperparameters that yielded expected values equal to population correlations and prior *SD* equal to 0.05, 0.1, or 0.2. The optimization algorithm simply minimized the sum of two squared discrepancies: (a) the difference between the population correlation and expected value[5] of the rescaled Beta distribution, and (b) the difference between the intended prior *SD* and the expected[6] *SD* of the rescaled Beta distribution. We provide an example in Figure 1 of four different rescaled-Beta priors for $\rho_{A_1,E_1} = .236$ (see Table 2) with hyperparameters α and β chosen to yield the same location (M) but different prior *SD*s for ρ_{A_1,E_1}. Prior distributions for the other correlation parameters may be visualized in a similar manner.

The prior *SD* reflects the degree of uncertainty about the prior expectation. For example, narrow priors (i.e., smaller prior *SD*) imply more precision and, thus, more certainty about the prior expectation. Figure 1 shows that the diffuse prior (the dotted-line distribution) reflects considering any positive or negative correlation to be nearly equally likely, except the most extreme values near ± 1. Informative priors are more restrictive, to varying degrees. The Pr-0.1 prior reflects high certainty that the correlation is positive, with 95% of probability density between 0.036–0.436, whereas the Pr-0.2 prior does not rule out large or even small negative correlations (−.164–0.636). The Pr-0.05 prior, on the other hand, reflects greater certainty that the correlation is in the small-to-medium range (0.136–0.336). Thus, it is logical to expect that Pr-0.05 priors (the solid-line distribution in Figure 1) will provide more accurate estimates with lower sampling variability than Pr-0.1 priors (the dashed-line distribution in Figure 1), which will in turn provide more accurate estimates with lower sampling variability than Pr-0.2 priors (the dash-dotted line distribution in Figure 1) which will furthermore provide more accurate estimates with lower sampling variability than the diffuse priors.

Posterior point estimates (EAPs) of the prophetic and diffuse priors were compared with FIML point estimates, yielding five estimator conditions. We also

[3]This constitutes using the data twice—to specify the prior, which is then updated using the same data—which generally results in posterior distributions with underestimated variance (i.e., underestimated uncertainty; Carlin & Louis 2000b, p. 1287; Zitzmann et al. 2024, p. 3). However, diffuse priors lack sufficient information to substantially impact estimated posteriors (Schuurman et al., 2016).

[4]The code for the optimization algorithm used to compute α and β hyperparameters for this and subsequent simulations can be found in our supplementary material on the Open Science Framework (OSF): https://osf.io/ju4fd/.

[5]The expectation of a Beta(α, β)-distributed random variable is $\bar{x} = \frac{\alpha}{\alpha+\beta}$; which is transformed using the function $2\bar{x} - 1$ to obtain the expected value of the rescaled Beta distribution.

[6]The variance of a Beta(α, β)-distributed random variable is $\text{Var}(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$, so the *SD* of the rescaled Beta distribution is $\sqrt{2 \times \text{Var}(x)}$.

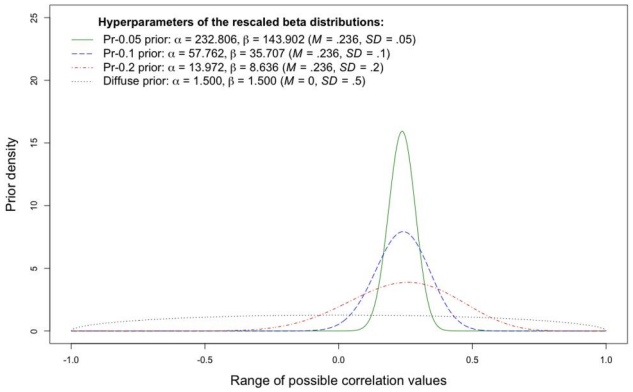

Figure 1. Figure depicting the impact of manipulating prior information (i.e., width of the prior distribution) for a fixed prior location on the shape of the rescaled Beta distribution for correlations. The solid line shows the Pr-0.05 prior, the dashed line shows the Pr-0.1 prior, the dash-dotted line shows the Pr-0.2 prior, and finally, the dotted line shows a diffuse prior (i.e., the current software default).

manipulated the number of round-robin groups ($G =$ 10 or 25) and the size of each group ($n_g = 6$, 8, 10, or 20 per group). For these 5 (estimator/priors) $\times 4$ $(n_g) \times 2$ $(G) = 40$ simulation conditions, we generated $R = 1000$ replications per condition.

Analysis plan

All analyses were conducted in R (R Core Team, 2023). A total of 33 unique correlations and *SD*s were estimated across the case (6 *SD*s and 15 correlations) and dyad (3 *SD*s and 9 correlations) levels.

FIML estimates were obtained using the `srm` package (version 0.5-1). The `srm` package assumes SRM components to be multivariate normally distributed with a mean vector μ and covariance matrix Σ, and can accommodate structural relations between SRM components. However, we fit a saturated model at the case and dyad levels—equivalent to fitting a multivariate SRM—to derive (co)variance estimates between SRM components. Group effects were not treated as fixed in the `srm` package. This is because the algorithm fails to converge for the specified model if group effects are fixed. We do not expect that this limitation will greatly affect our results, as we generated data such that the mean of each variable per group is zero. Furthermore, the `srm` package produces only unstandardized output (i.e., variances and covariances). We calculated the SRM *SD*s, correlations, associated *SE*s, and confidence intervals using the delta method, provided in the `car` package (Fox & Weisberg, 2019). The delta method relies on Taylor series approximation to provide an approximate asymptotic variance—and, by extension, *SE*—of a non-linear transformation of one or more random variables when the expected values of the Taylor polynomial are known up to a certain order of derivatives (provided that the random variables are themselves asymptotically normal). Although higher-order approximations provide more accurate estimates, first-order Taylor approximations are deemed sufficient for sufficiently large samples in psychological research. Note that the `deltaMethod()` function in the `car` package provides only the first-order asymptotic variance approximation of the Taylor polynomial. The `srm` package results were saved and the delta method was applied only if an internal diagnostic tool indicated convergence.

MCMC estimates were obtained from the `lavaan.srm` package (version 0.1-0.0044), using the `mvsrm()` function to estimate a multivariate SRM with priors described in the previous section. For all MCMC analyses, we initialized four Markov chains with random starting values and ran each for 2000 iterations, discarding the first 1000 as burn-in, which yielded 4000 posterior samples to estimate the joint posterior distribution of the SRM parameters. All variables were group-mean centered to remove any group differences due to sampling error. We monitored the multivariate potential scale-reduction factor (mPSRF; Brooks & Gelman, 1998), using values > 1.05 as an indication that the four chains had not yet converged on the same posterior space, in which case the MCMC estimation was then repeated with double the initial number of iterations (i.e., discarding 2000 as burn-in retaining 2000 samples per chain). We also calculated each parameter's effective sample size (ESS) and potential scale-reduction factor (PSRF or $\hat{R}$; Gelman & Rubin, 1992) to make a more informed decision about convergence. $ESS = 100$ and $\hat{R} \approx 1.02$—which is equivalent to $ESS = 100$ when four MCMC chains are run (see Equation 13, Kwon et al., 2025)—were applied as cutoff values. Because any poorly sampled parameter could call convergence into question, a sample was flagged and removed from the final analysis if either $ESS < 100$ or $\hat{R} > 1.02$ for any correlation or *SD* estimate. If $ESS > 100$ and $\hat{R} < 1.02$ for all correlations and *SD*s, we included that sample's EAP estimates of SRM correlations, *SD*s, and (co)variances for analysis[7].

[7]Kwon et al. (2025) and Zitzmann and Hecht (2019) recommend using $ESS = 400$—equivalent to $\hat{R} \approx 1.005$ with four MCMC chains—as a cutoff value to minimize the impact of Monte Carlo error. However, doing so resulted in zero samples eligible for final analysis in the $n_g = 6$ conditions (with $G = 10$ or 25 groups) with diffuse priors. Furthermore, the resulting plots differed only negligibly between criteria ($ESS = 100/\hat{R} \approx 1.02$ vs. $ESS = 400/\hat{R} \approx 1.005$), leading to the same conclusions about the general patterns of results. Thus, we chose to retain $ESS = 100$ and $\hat{R} \approx 1.02$ as our cutoff values in order to display these patterns across all simulated conditions. Interested readers can find a side-by-side comparison of these plots in our OSF project.

Outcome variables

We inspected the robust bias (RB) of point estimates, the bias in *SE* estimates, coverage rate (CR) of interval estimates and the root mean-squared error (RMSE) to assess the accuracy and efficiency of the different prior types and FIML.

The RB of an estimate was computed as the difference between (a) the median of all estimates ($\hat{\theta}_r$, where $r \in 1, \ldots, R$) per condition and (b) the true parameter (θ).

The *SE* bias per simulation condition was computed as the difference between (a) the mean estimated *SE* and (b) the empirically observed *SE* (i.e., the *SD* of point estimates across replications).

CRs of interval estimates were computed as the percentage of samples for which the true value of a parameter was captured by the interval. The `lavaan.srm` package provides central 95% Bayesian credible intervals (BCI) for the MCMC estimates by default. For FIML estimates of (co)variance parameters, we constructed a normal-theory 95% confidence interval (CI) for the transformed parameter (*SD*s and correlations) using delta-method *SE*s.

Finally, consistent with Nestler et al. (2020), the RMSE was computed as the sum of the squared robust bias and the squared median absolute deviation (MAD), which is a robust estimate of the true sampling variability ($RMSE = \sqrt{bias^2 + MAD^2}$). The MAD was computed as $g \times Med(\hat{\theta}_r - Med(\hat{\theta}_r))$ where $g = 1.4826$ (Talloen et al., 2019).

Results

One FIML sample in the smallest sample-size condition ($n_g = 6$, $G = 10$) did not converge. In addition, a total of 108 FIML samples, mostly from the $n_g = 6$, $G = 10$ condition were not considered in the final analysis, given that they produced inadmissible solutions (namely, correlation estimates with absolute value > 1). Table A1 in Appendix A shows the number of converged samples for all MCMC prior types and for FIML.

The results for all 24 correlations and 9 *SD*s are visually summarized in plots. All other results can be found on our OSF page.

For this, and subsequent simulations, we structure our results by Monte Carlo outcome, beginning with accuracy (of point, *SE*, and interval estimates) and then efficiency (using RMSE, which combines both sampling error and bias). The results are presented in multifaceted line plots with error bars that display the range (minimum and maximum) per condition. Each panel displays results for a particular sample-size condition, the prior type (or FIML) appear along the *x*-axis, and the *y*-axis contains a scale of values for the outcome variable. Each panel contains two lines—dotted for the case level and solid for the dyad level—with the median displayed using different symbols per level (● for the dyad level and ▲ for the case level). For the *SE* bias plots, additional ribbons displaying the range of the empirically observed *SE*s per level have been included. For CR plots, a solid horizontal line at 95% to indicate nominal coverage and a dashed horizontal line at 90% to indicate minimally acceptable coverage are added. Readers interested in viewing plots containing the results of individual parameters can find these in our supplementary material on the OSF.

Robust bias

RB results across all simulation conditions are presented in Figure 2.

We found that the diffuse—i.e., software-default—priors could lead to highly biased estimates at the case level, which is consistent with Smid et al.'s (2020) conclusions that software-default MCMC estimates display greater bias than frequentist estimates in small samples. We found that this was particularly the case in the small-group conditions ($n_g = 6$ or 8), wherein information from only $n_g \times G =$ 60, 80, 150, or 200 people is available when estimating the parameters. However, the distribution of bias was closer to 0 with $G = 10$ groups of $n_g = 10$, despite the total number of cases $n_g \times G = 100$ being smaller than $G = 25$ groups of $n_g = 6$ ($n_g \times G = 150$), for which distribution of bias was more negative. This finding reinforces results from Nestler (2018) and Lüdtke et al. (2013), who concluded that group size n_g was more important than the number of groups G to accurately estimate SRM parameters. Larger groups imply more interactions per person and, by extension, more information about each case.

Diffuse priors yielded much lower bias in dyad-level estimates. Even in the $n_g = 6$ conditions, maximum absolute bias value did not exceed approximately 0.5 units. This is due to estimating relationship-level parameters with more dyads ($n_g \times \frac{n_g - 1}{2} \times G$) than cases ($n_g \times G$). Therefore, dyad-level parameters tend to be more accurately and precisely estimated.

Comparing the prophetic-prior conditions reveals the effect of prior information. Estimates of the Pr-0.2 priors were, as expected, more accurate than those of the diffuse priors. However, these estimates yielded the

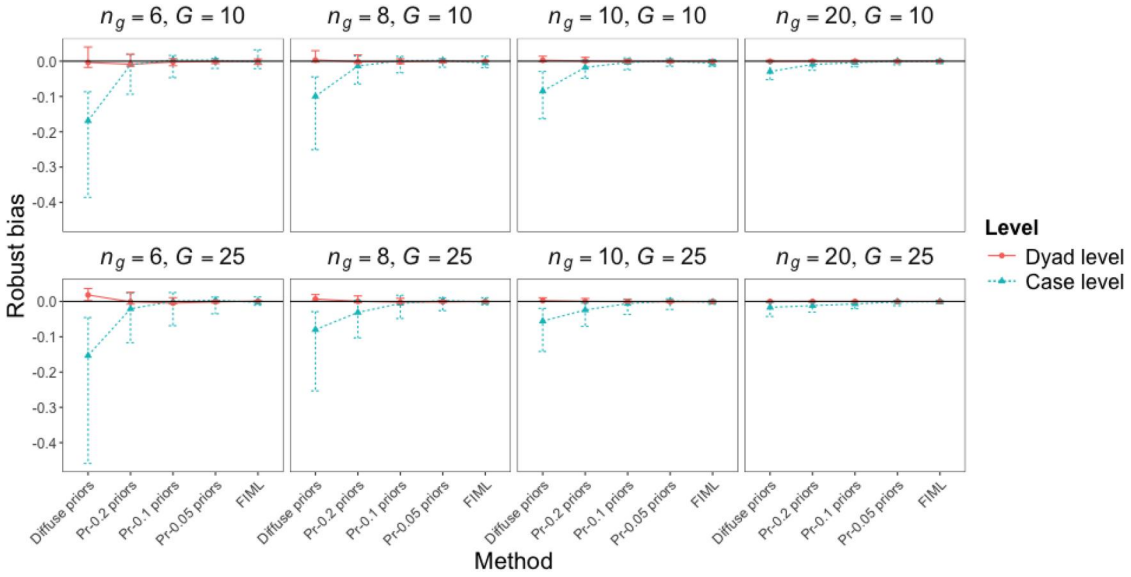

Figure 2. Plot depicting the robust bias for Simulation-1 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the robust bias values. Each facet presents the results for a separate $\{n_g, G\}$ combination. The symbols ▲ (case level) and ● (dyad level) represent the median robust bias across all estimates per level, whereas the error bars extend to the minimum and maximum robust bias per level. Solid lines are used for the dyad level and dotted lines are used for the case level.

most variability in bias around 0 out of the three prophetic prior conditions, at both levels and across all sample sizes. The Pr-0.05 prior estimates had the least variability in bias around 0. Bias still tended to be distributed around 0 at the dyad level, but case-level bias seemed to be distributed more negatively with less prior information (i.e., Pr-0.2 performed worse than Pr-0.1). This result suggests that even if accurately located priors are specified, sufficient prior information (i.e., a sufficiently small prior *SD*) would be required to prevent bias of posterior point estimates.

Finally, FIML produced relatively unbiased estimates at both levels, parallel to results in Nestler (2018); Nestler et al. (2020); Nestler, Lüdtke, et al., 2022a). The distribution of bias was spread very little around 0, even in samples of a few small groups ($n_g = 6, G = 10$). Note, however, that FIML produced 98 samples with at least one out-of-bounds correlation estimate in the $n_g = 6, G = 10$ condition, which were excluded from the results.

Standard-error bias

The *SE* bias is illustrated in Figure 3, wherein the ribbons display the range of empirical *SE*s (i.e., the *SD*s of each parameter's estimate). As evident Figure 3, the empirical *SE*s for FIML were higher than for any prior condition when $n_g < 10$, implying that even with diffuse priors, MCMC can be more efficient than FIML in small samples. Diffuse priors have a similar spread of efficiency as FIML when $n_g \geq 10$.

The median *SE* bias for diffuse priors was close to zero, implying that the median bias in *SE*s was small relative to the actual sampling variability, even in the small-group conditions. However, *SE*s for some parameters at the case level were overestimated in the $n_g = 6$ condition, leading to the error bar for *SE* bias in these conditions to be positively skewed. As with point estimates, *SE* bias at the dyad level was also minimal.

The median *SE* bias for the prophetic priors was also, on average, close to zero across all sample-size conditions. However, note that the Pr-0.2 priors overestimated the *SE*s for some parameters at the case level in the $n_g = 6$ conditions. This is because a lower prior precision (i.e., larger prior *SD*) allows for more variability in sampled estimates, resulting in higher posterior variance than the actual variability in estimated parameters.

For all MCMC conditions, across all n_g at the case level, the overestimation of *SE*s reduces as G increases from 10 to 25. This is because collecting data from a greater number of groups per sample decreases the amount of sampling variability in the estimated parameters.

The *SE*s of FIML estimates displayed low to minimal bias across all conditions, even in the $n_g = 6$ conditions. However, note that the empirically

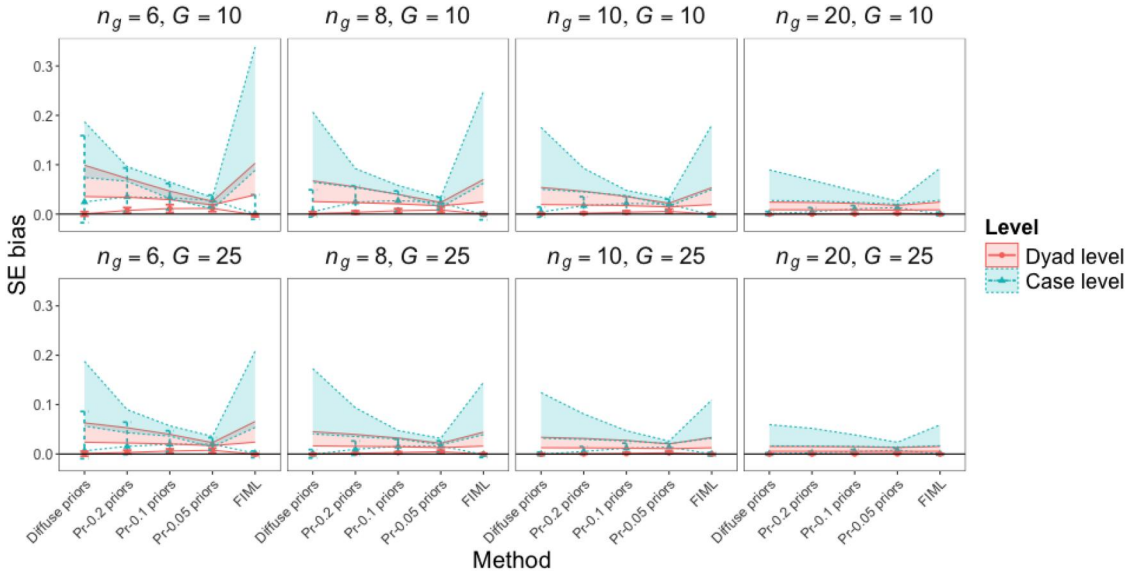

Figure 3. Plot depicting the *SE* bias for Simulation-1 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the *SE* bias values. Each facet presents the results for a separate $\{n_g, G\}$ combination. The symbols ▲ (case level) and ● (dyad level) represent the median *SE* bias across all parameters per level, whereas the error bars extend to the minimum and maximum *SE* bias per level. The ribbons show the range of the empirically observed *SE*s across all parameters per level. Solid lines are used for the dyad level and dotted lines are used for the case level.

observed *SE*s for FIML are the greatest of the methods compared in this simulation, implying that the estimated *SE*s are also large. In addition, out of the 98 samples with at least one out-of-bounds correlation estimate in the $n_g = 6, G = 10$ condition, most had unrealistically high associated *SE*s—for example, the associated *SE* for an out-of-bounds estimate $\hat{\rho}_{A_3,E_3} = -53.426$ (which is impossible for a correlation estimate) was as high as 187,344.504. These samples were excluded from the final analysis.

Coverage rates

The CRs per level and condition are presented in Figure 4.

In the diffuse-prior conditions, case-level interval estimates had coverage rates that were lower than nominal in small-to-moderate samples. This shows how greatly case-level parameters were underestimated, given that their overestimated sampling variability (high *SE* bias) would make the interval estimates too wide. Coverage at the dyad level was much better, but not always nominal (e.g., between 80 and 90% when $n_g = 6$ or 8).

Out of the three prophetic-prior conditions, Pr-0.05 priors capture the true value of the parameter almost 100% of the time, whereas Pr-0.1 and Pr-0.2 priors had lower CRs. This is to be expected, given that not only do Pr-0.1 and Pr-0.2 estimates display a greater magnitude of bias than Pr-0.05 estimates, but also larger sampling variability. Whereas the width of the Pr-0.05 prior is narrower with greater probability density concentrated at the expected value, Pr-0.1 and Pr-0.2 are wider priors allowing for estimation of a diverse set of values (see Figure 1), which affects their estimated posterior distributions and, by extension, their CRs.

FIML estimates had nominal coverage at both levels, even in small groups. This follows from the lack of bias in point and *SE* estimates.

RMSE

The previous outcomes were used to compare accuracy of point, *SE*, and interval estimates, but even in the absence of bias, estimates can be inaccurate (unequal to their population parameter) due to sampling error. RMSE incorporates (in)efficiency, making it possible to compare a less biased but less efficient estimator to one that is more efficient but more biased. Our results so far have shown FIML to be quite accurate (even in small samples), but less efficient than MCMC in small samples (even with diffuse priors). The RMSE results presented in Figure 5 allow us to compare MCMC (with varying prior information) to FIML in a way that accounts for the accuracy–precision tradeoff: lower RMSE indicates that estimates are closer to population values.

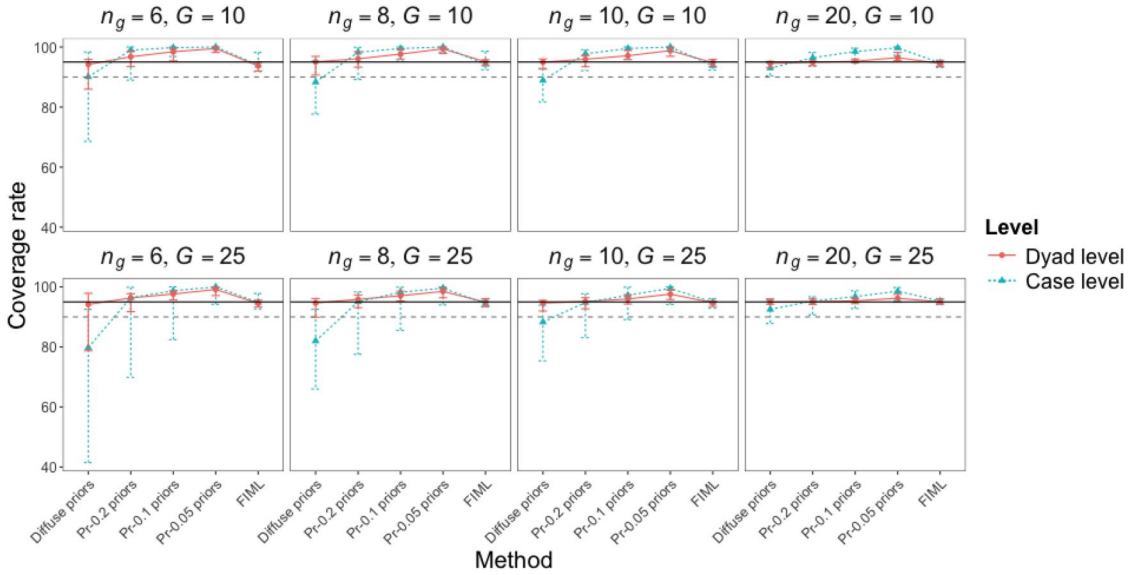

Figure 4. Plot depicting the coverage rate for Simulation-1 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the coverage rates. Each facet presents the results for a separate $\{n_g, G\}$ combination. The symbols ▲ (case level) and ● (dyad level) represent the median coverage rate across all estimates per level, whereas the error bars extend to the minimum and maximum coverage rate per level. Solid lines are used for the dyad level and dotted lines are used for the case level. A solid horizontal line indicates the nominal 95% confidence level, and a dashed horizontal line indicates a minimally acceptable 90% coverage.

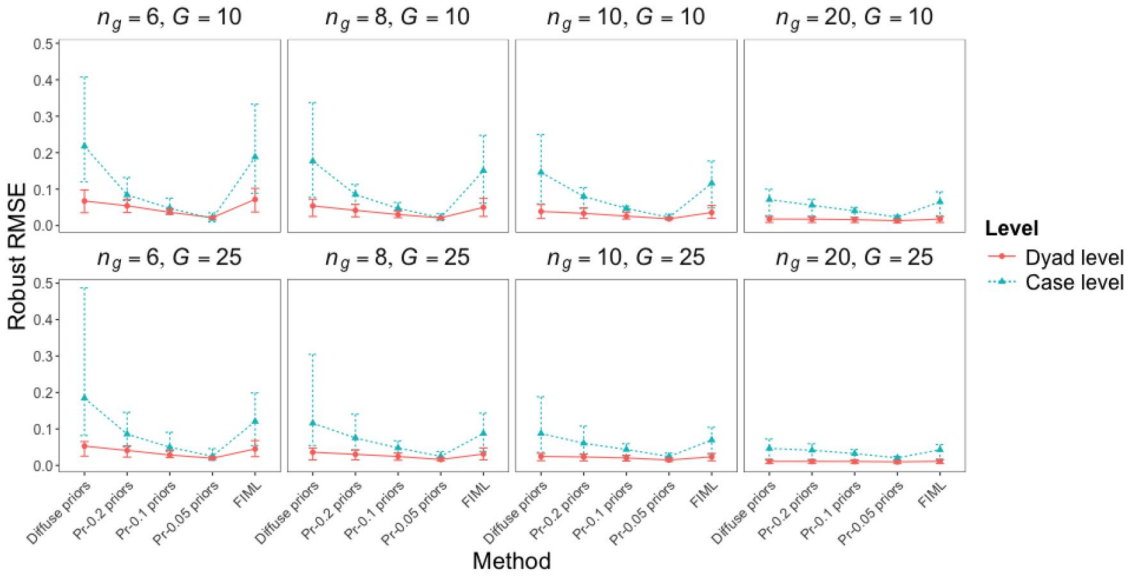

Figure 5. Plot depicting the RMSE for Simulation-1 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the RMSE values. Each facet presents the results for a separate $\{n_g, G\}$ combination. The symbols ▲ (case level) and ● (dyad level) represent the median RMSE across all estimates per level, whereas the error bars extend to the minimum and maximum RMSE per level. Solid lines are used for the dyad level and dotted lines are used for the case level.

On average, across both levels, all approaches are asymptotically comparable, given that both bias and sampling variability decrease as the sample size increases. Dyad-level RMSE was quite low across all sample sizes and estimators. Differences in case-level RMSE were more apparent, particularly in smaller groups with less prior information. Even weakly informative priors—which were not informative enough to minimize bias—yielded notably lower RMSE than FIML or diffuse priors. As expected, the

Pr-0.05 priors had lowest RMSE values compared to the Pr-0.1 and Pr-0.2 priors, owing to the low estimate bias and MAD values.

Discussion

These results show that even with accurately located priors, the accuracy of posterior point, *SE*, and interval estimates of multivariate SRM parameters can depend on prior precision (i.e., the *SD* of the prior distribution). Naturally, more informative priors yielded more efficient estimates, minimizing RMSE due to the priors being accurately located. However, informative priors can overwhelm the data in small samples, and the highest level of precision we considered (Pr-0.05) might not be representative of how confident researchers are likely to be about their expectations (if any) when estimating SRM parameters with MCMC.

Researchers in substantive settings do not have knowledge of the true values of parameters they estimate with MCMC. In the next simulation, we explore the use of less accurate priors—based on theory or data, rather than prophetically knowing the true parameters. We therefore manipulate prior locations, but we hold the magnitude of prior information constant.

Simulation study 2: Prior location

Whereas Simulation 1 investigated the effects of prior precision under the condition of unrealistically accurate location, Simulation 2 compares practical methods to specify priors with approximately accurate locations. Reasonably accurate prior locations can be specified based on theoretical expectations—perhaps also informed by previous research findings—which Smid et al. (2020) referred to as *thoughtful* priors. In the context of analyzing round-robin data, thoughtful priors would incorporate researchers' prior knowledge—for example, from previous studies—about the strength of relations between SRM components (correlations) and the relative contributions of case- and dyad-level components (often expressed as proportions of the total variance).

It is also possible to choose hyperparameters based on preliminary (frequentist) analyses of the round-robin sample data. These are called empirical Bayes (EB) priors, of which we consider two types to determine approximate prior locations of SRM *SD*s and correlations: method-of-moments estimation (EB-MOM prior) and FIML estimation (EB-FIML prior). We designed Simulation 2 to compare thoughtful priors to the two EB prior types.

Simulation conditions

The main difference from Simulation Study 1 is our manipulation of prior locations for *SD* and correlation parameters, which we describe in detail here.

Thoughtful priors were designed to simulate a realistic situation wherein a researcher specifies hyperparameters based on expert knowledge. In this case, we assume the researcher would correctly expect all correlations at both levels to be positive, but to have less confidence about how large the correlations would be. Thus, we specified a Beta $(\alpha = 58.500, \beta = 31.500)$ prior, which after rescaling has an expected value of $M = .30$ (i.e., most likely to be a "medium" correlation; Cohen, 2013) with $SD = .1$. We consider this reasonably located and weakly informative, as approximately 95% of the prior probability mass is located within $\pm .2$ of the specified location. That is, the prior is not so informative as to rule out small (0.1) or large (0.5) correlations as being reasonably likely—a range which covers most of the correlations in Tables 1 and 2. The default location of the prior *t* distributions were retained for *SD*s in the thoughtful condition as the location hyperparameters (μ) are chosen based on the total sample *SD*.

The EB-MOM and EB-FIML prior locations were chosen in two steps. In the first step, a univariate SRM was conducted for each round-robin variable using frequentist estimation (method-of-moments or FIML) to obtain point estimates of *SD*s and correlations (reciprocities). A bivariate SRM was conducted for each pair of round-robin variables to obtain point estimates of between-variable correlations among SRM components. The following section provides more detail about frequentist estimation.

After obtaining point estimates, EB priors were specified with hyperparameters that implied an expected value equal to the point estimate in the second step. For *SD* parameters, this was simply the location parameter, μ, of the prior *t* distributions. For correlation parameters, an optimization algorithm was used to find hyperparameters of the rescaled-Beta priors that yielded expected values equal to the frequentist correlation estimates. As with thoughtful priors, priors *t* distributions were specified with scaling hyperparameter $\varsigma = 0.1$, and rescaled Beta distributions had a prior $SD = .1$.

The EB priors described above involve using the same data both to determine prior hyperparameters,

then updating the prior with the same data to estimated the posterior distribution. This is a contentious practice in Bayesian literature—sometimes called "double dipping"—both philosophically (due to being "not particularly Bayesian" Carlin & Louis, 2000b, p. 1287) and due to its empirical consequences. Especially in small samples, informative (narrow) priors may overwhelm the data to exert greater impact on the estimated posterior's location (Darnieder, 2011). A data-informed prior location will be consistent with the data (re)used in the likelihood, resulting in a similarly located posterior; however, greater prior precision will lead to more precisely estimated posterior. Because the precision is illusory (i.e., prior information is borrowed from the data, ignoring sampling error), uncertainty intervals will be too narrow, leading to lower-than-nominal coverage (documented in, e.g., Kass & Steffey, 1989; Schuurman et al., 2016). Our simulation results clearly demonstrate this problem in the SRM context, and we explore a potential solution in Simulation 4.

As in Simulation Study 1, we specified the size of round-robin groups $n_g = 6$, 8, 10, or 20 and the number of round-robin groups $G = 10$ or 25, resulting in 3 (priors) $\times 4$ $(n_g) \times 2$ $(G) = 24$ simulation conditions. We generated $R = 1000$ samples per condition, saving EAP estimates and calculating the same outcome variables as for Simulation Study 1.

Frequentist analyses for EB priors

The first step of specifying the EB-MOM priors was to fit univariate and bivariate SRMs in the `TripleR` package (version 1.5.4). `TripleR` estimates groupwise variances and covariances for (pairs of) SRM components, the weighted means of which are the resulting method-of-moments estimates. However, the method-of-moments estimator sometimes produces Heywood cases (i.e., negative variances) which are also considered in the weighted mean by default. Following Lüdtke et al. (2013), we assumed groupwise negative variance estimates were an outcome of negligible variances and rescaled these to 0 before computing the weighted mean. The square-roots of the weighted-mean variances were then set as the μ hyperparameters for the prior t distributions. `TripleR` follows a similar procedure for covariance parameters, wherein the weighted mean of the covariance among SRM component pairs is computed. We standardized the weighted means and fixed correlation estimates $> .9$ to .9. The correlation values were then passed through the optimization algorithm, resulting in unique α and β hyperparameters per SRM correlation parameter in each sample.

To specify the EB-FIML priors, we first fit univariate and bivariate SRMs in the `srm` package. FIML may produce negative variance estimates given near-zero population variances and small sample sizes. In such cases, the μ of the t-prior was set to 0. When computing a correlation ρ_{xy} between SRM components x and y, covariances were standardized only if both $\sigma_x^2 > 0$ and $\sigma_y^2 > 0$. Large correlations were fixed to 0.9 and the α and β hyperparameters were chosen based on the FIML estimates. Similar to Simulation 1, an optimization algorithm computed α and β values that minimized the sum of two squared discrepancies—(a) the difference between the FIML correlation estimate and the expectation M of the rescaled Beta distribution, and (b) the difference between the intended prior SD and the expected SD of the rescaled Beta distribution, the code for which can be found in our supplementary materials. If $\sigma_x^2 < 0$ or $\sigma_y^2 < 0$, the α and β hyperparameters for that correlation were fixed to the default value of 1.5. Although it is possible to fit multivariate models in `srm`, we performed all above computations using univariate model estimates for SDs and bivariate model estimates for correlations. This is because Simulation 1 revealed convergence issues that led to unstable or unrealistic estimates when fitting multivariate models in the `srm` package. Fitting bivariate SRMs minimized convergence problems.

Results

As with Simulation 1, MCMC samples were not considered in the final analysis if ESS < 100 or $\hat{R} > 1.02$ for any estimated correlation or SD parameter. Table A1 in Appendix A shows the number of converged samples for thoughtful, EB-MOM, and EB-FIML priors per sample-size condition. We include FIML and Pr-0.1 estimates from Simulation Study 1 as benchmarks for comparison with Simulation-2 results.

Robust bias

Figure 6 presents the (range of) robust bias for the thoughtful, EB-MOM, and EB-FIML priors per level.

Although the median bias in thoughtful-prior estimates lies close to 0, the distribution of the outcome is wide, indicating that thoughtful priors poorly estimate some parameters. As expected, parameters with true values $> .3$ were underestimated, whereas those

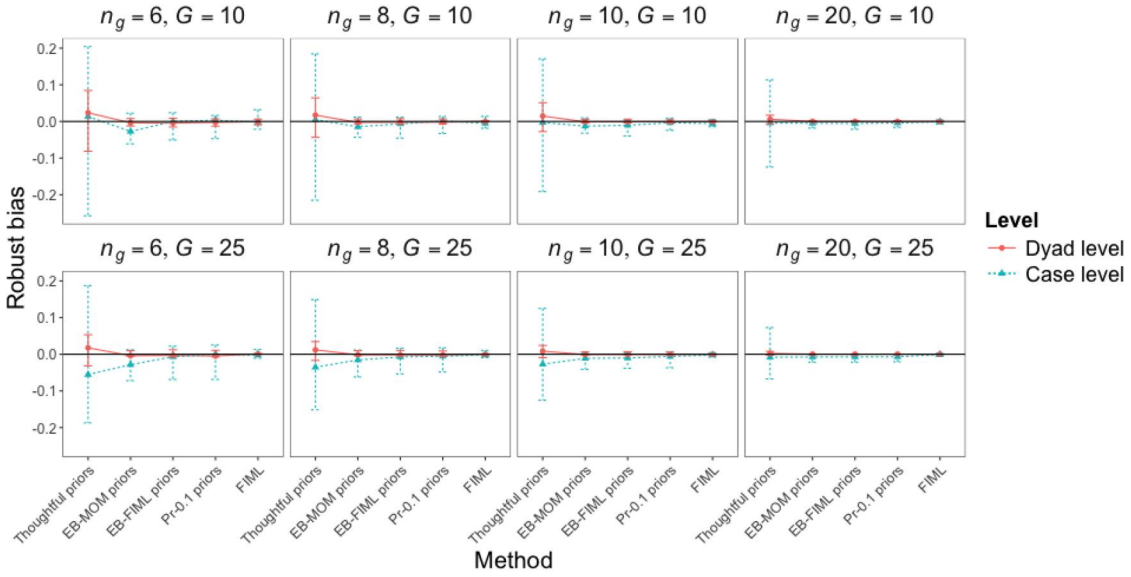

Figure 6. Plot depicting the robust bias for Simulation-2 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the robust bias values. Each facet presents the results for a separate $\{n_g, G\}$ combination. The symbols ▲ (case level) and ● (dyad level) represent the median robust bias across all estimates per level, whereas the error bars extend to the minimum and maximum robust bias per level. Solid lines are used for the dyad level and dotted lines are used for the case level.

with true values $< .3$ were overestimated (interested readers can find figures displaying aggregate results per parameter on our OSF project). Although the magnitude of the bias diminishes in larger groups, the pattern of over/underestimation persists. The overall bias at the dyad level is much lower, but parameters still appear to be over/underestimated in the small-group conditions.

Both the EB priors generally display a near-zero median bias at both the case and dyad levels for all sample-size conditions. Further, the range of bias values is also lower for these prior types. This is sensible, since the frequentist estimates used as the location for the prior distributions are presumably much closer to the true value of the parameter than a fixed location of .3. There does not appear to be a substantial difference between the bias of EB-MOM and EB-FIML prior locations.

Standard-error bias

The *SE* bias is presented in Figure 7. As indicated by the shaded ribbons, both EB priors are as (in)efficient as FIML[8], but the *SE* estimates (i.e., posterior *SD*s) are more biased for the EB priors than FIML. That is, EB priors expect less variability than is actually observed, consistent with past research on EB priors (Carlin & Louis, 2000b; Schuurman et al., 2016).

[8]A plot containing the estimated *SE*s for both EB priors and FIML can be found on our OSF page.

The *SE*s for thoughtful priors are only slightly overestimated. The bias is greatest in the smallest condition, with a maximum of approximately 0.06 units, but reduces as sample size increases. Also note that, across all n_g, *SE* bias reduces from $G = 10$ to 25 since the sampling variability of estimates reduces when more information is available.

Most evident for the case level, *SE*s for the EB priors are underestimated. This implies that, in reality, there is greater variability in estimated values across samples than approximated by the posterior *SD*. As mentioned in Simulation Study 1, the underestimation of sampling variance due to "double dipping" is a common problem when using EB priors (Carlin & Louis, 2000b). We derived estimates for our EB priors by using the complete dataset twice for a hierarchical process: first deriving frequentist estimates of the desired model parameters, then using these as hyperparameters (or computing hyperparameters based on these) before proceeding with Bayesian estimation as if the hyperparameters were known (or believed) a priori (Carlin & Louis, 2000a, ch. 3). Thus, to more accurately estimate sampling variability, it is necessary to account for uncertainty from two sources: (a) the uncertainty in the frequentist estimation of the hyperparameters, and (b) the uncertainty in the Bayesian estimation of SRM parameters. Our EB priors fail to adjust for the uncertainty in the frequentist process used to choose hyperparameters, which leads to underestimating the true sampling variance. However,

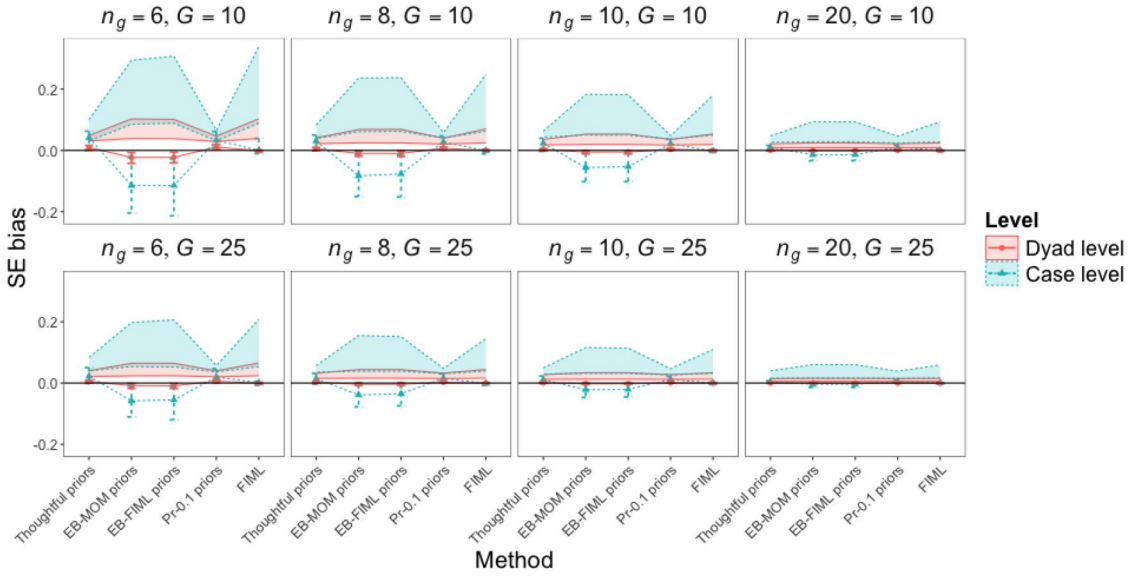

Figure 7. Plot depicting the *SE* bias for Simulation-2 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the *SE* bias values. Each facet presents the results for a separate $\{n_g, G\}$ combination. The symbols ▲ (case level) and • (dyad level) represent the median *SE* bias across all parameters per level, whereas the error bars extend to the minimum and maximum *SE* bias per level. The ribbons show the range of the empirically observed *SE*s across all parameters per level. Solid lines are used for the dyad level and dotted lines are used for the case level.

the *SE* bias is negligible for dyad-level estimates, and even *SE* bias of case-level estimates is attenuated by more data, reducing in conditions with more (G) or larger (n_g) groups.

We conduct 2 additional small-scale simulations, first to explore the effect of sampling more small groups ($n_g = 6$ and 8) in Simulation 3. Because sampling small round-robin groups is so common (to minimize the burden of data collection), it is valuable to discover whether unbiased results for EB priors can be obtained from small groups when G is large. In Simulation 4, we explore a potential solution to double dipping by using only some data to estimate hyperparameters.

Coverage rates

The CR for each analysis type and sample-size condition is presented in Figure 8.

Across all sample-size conditions, thoughtful priors display very low CRs due to their inaccurate point estimates. Figure 6 shows that parameters are either greatly overestimated or greatly underestimated by the thoughtful priors. The BCIs fail to capture the true value of the parameters at nominal levels. This implies inflated Type I error rates when using BCIs to apply traditional null-hypothesis significance tests.

CRs for the EB priors are also (but to a lesser degree than thoughtful priors) below nominal in small-sample conditions, but CRs slightly improve as n_g or G increase (i.e., when data overwhelm the prior). Given that the EB priors underestimate *SE*s, their BCIs are too narrow, which yields low coverage even when point estimates are unbiased—the same consequence of double dipping observed for *SE* bias.

RMSE

The ranges of RMSE values for the thoughtful priors and EB priors were, for the most part, comparable. Figure 9 shows the EB priors also had comparable RMSE values to FIML, indicating a similar accuracy–precision tradeoff. In line with Nestler et al. (2020) results, we also found that FIML more efficiently estimates case-level parameters with fewer large groups (e.g., $n_g \times G = 20 \times 10 = 200$) than with many small groups (e.g., $8 \times 25 = 200$), given the same number of cases (200).

Discussion

Simulation 2 shows how the accuracy of weakly informative prior locations can impact quality of estimation. We found that while the EB priors provided relatively accurate estimates of parameters, they underestimated their sampling variability—especially of case-level estimates in smaller samples. This is due to double dipping, which involves a hierarchical process in computing EB-prior estimates: first obtaining frequentist estimates of hyperparameters, then treating those EB-hyperparameters as a priori knowledge. Posterior *SD*s estimate sampling variability expected if the same model were fitted to new data. But priors

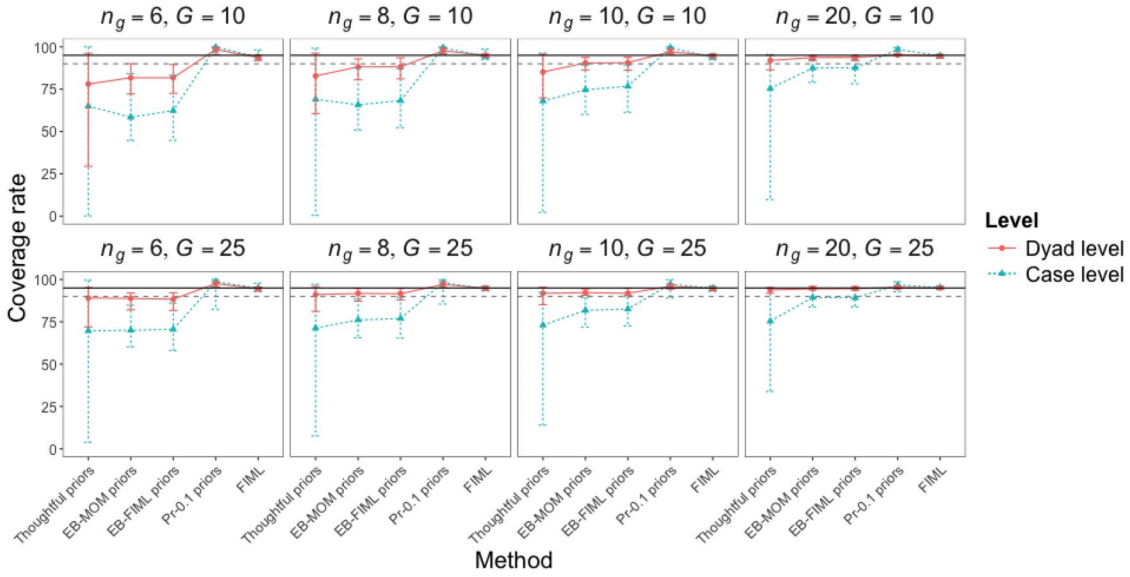

Figure 8. Plot depicting the coverage rate for Simulation-2 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the coverage rates. Each facet presents the results for a separate $\{n_g, G\}$ combination. The symbols ▲ (case level) and ● (dyad level) represent the median coverage rate across all estimates per level, whereas the error bars extend to the minimum and maximum coverage rate per level. Solid lines are used for the dyad level and dotted lines are used for the case level.

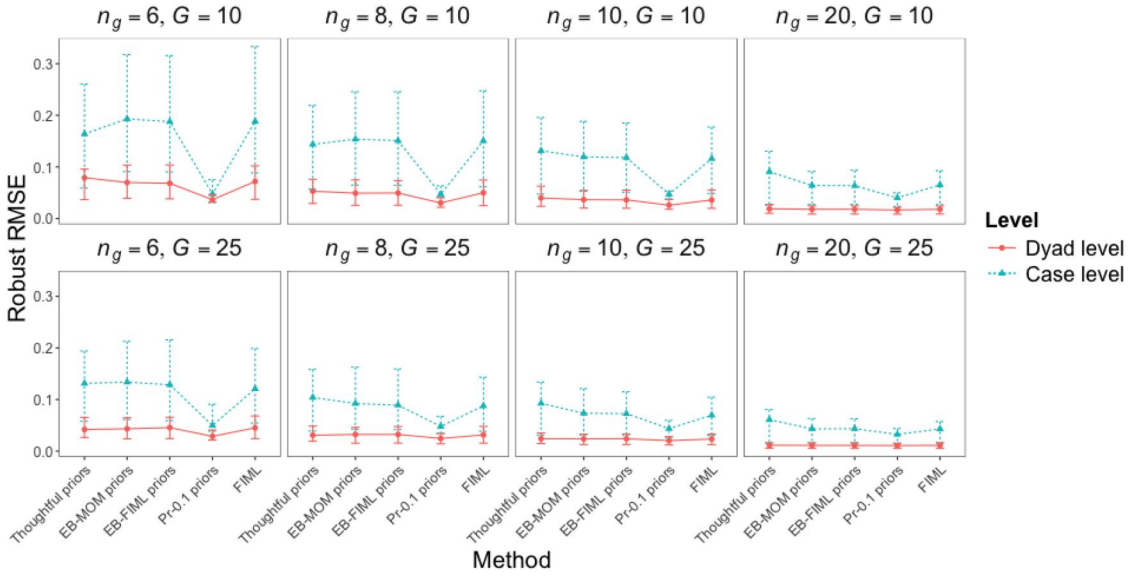

Figure 9. Plot depicting the RMSE for Simulation-2 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the RMSE values. Each facet presents the results for a separate $\{n_g, G\}$ combination. The symbols ▲ (case level) and ● (dyad level) represent the median RMSE across all estimates per level, whereas the error bars extend to the minimum and maximum RMSE per level. Solid lines are used for the dyad level and dotted lines are used for the case level.

are part of the model, and EB priors cause the model to fluctuate from sample to sample. That additional "model variability" is not captured by a single MCMC estimation. In Simulation 4, we explore BMA as a solution to the underestimated-variance problem of EB priors.

We also found that EB-MOM and EB-FIML yielded comparable point and *SE* estimates. Thoughtful priors produced biased estimates even in the large-sample conditions, indicating that the accuracy of a posterior estimate is highly dependent on how accurate the expected value of the prior distribution is. In our thoughtful-prior conditions, we specified the same prior location for all correlations. In practice, a researcher with expert knowledge might be able to specify more accurate thoughtful prior

locations that depend on the correlation—for example, generalized and dyadic reciprocities are likely to be the largest correlations, and one might reasonably expect intrapersonal correlations to be larger than interpersonal correlations (as in Table 1). In such cases, the performance of thoughtful priors might more closely resemble that of EB priors, which vary across correlations by taking the data into account, so our simulation merely shows how poorly thoughtful priors might perform if they are not well informed.

Simulation study 3: More small groups

Simulation Studies 1 and 2 showed that MCMC estimation of SRM effects is inaccurate in small-group ($n_g = 6$ or 8) conditions. For a given group size n_g, sampling more groups reduces the range of estimation bias when using diffuse or thoughtful priors. The EB priors yielded relatively unbiased point estimates in small-group conditions, but they produced biased *SE*s when fewer groups were sampled for a given n_g. The goal of this simulation is to determine whether substantially larger G would yield sufficiently unbiased point estimates for diffuse or thoughtful priors, or sufficiently less biased *SE*s for EB priors.

Simulation conditions

Data with $G = 50$ and 100 groups of size $n_g = 6$ as well as $G = 50$ groups of size $n_g = 8$ were generated and analyzed for this simulation. We compared the software-default diffuse priors with thoughtful and EB-FIML priors. The population values (see Tables 1 and 2) and prior specifications from Simulations 1 and 2 were retained. EB-MOM priors were not specified in this simulation, given that we found EB-FIML and EB-MOM priors perform similarly (see Simulation 2's Results). We simulated $R = 1000$ replications in each of these new conditions.

Results

As in Simulations 1 and 2, only converged samples (using criteria $ESS = 100$ and $\hat{R} = 1.02$) were considered for the final analysis (see Table A2 in Appendix A). We include plots for $n_g = 6$ and 8 with $G = 10$ and 25 (from Simulations 1 and 2) in the figures for comparison. Readers interested in viewing plots based on stricter convergence criteria ($ESS = 400$ and $\hat{R} = 1.005$) can find these in our supplementary materials.

We present results for only RB and *SE* bias, to evaluate whether sampling more G reduces point estimate bias for diffuse and thoughtful priors and reduces *SE* bias for EB-FIML priors.

Robust bias

Figure 10 displays the results for robust bias.

The median bias of diffuse-prior estimates for the $n_g = 6$ conditions at the case level remain consistent as G increases from 10 to 100. Increasing G for a given n_g decreases the sampling variability of estimates, but smaller groups still yield biased estimates. Thus, the inaccurate results are more precise, which yields even worse coverage (see our supplementary materials on the OSF). For the $n_g = 8$ condition, both the median bias and range of bias remain consistent across the various G. At the dyad level, there is a substantial increase in accuracy of estimation from $G = 25$ to $G = 50$ for both $n_g = 6$ and 8.

For both $n_g = 6$ and 8 in the thoughtful-prior conditions, the median bias is approximately constant across all G conditions. The distribution of bias around the median value becomes tighter as G increases. This same pattern persists at the dyad level.

Bias for EB-FIML priors remains stable across G for both levels.

Standard-error bias

SE bias is presented in Figure 11.

For both levels, the *SE* bias for diffuse prior estimates greatly improves when G is increased from 25 to 50. This is sensible, given that information from more groups is available to estimate parameters, thereby reducing the overall sampling variability.

There appears to be no substantial difference in the median *SE* bias and range of (consistently low) *SE* bias values for thoughtful priors at either the case or dyad level.

EB priors greatly underestimated case-level *SE*s of estimated parameters when $n_g = 6$. However, the magnitude of underestimation appeared (in Simulation 2) to reduce when G increased from 10 to 25. Following this pattern, the *SE* bias reduced further when G increased to 50 and then 100. The $n_g = 8$ conditions display a similar pattern. This overall improvement in estimation of *SE*s follows from the prior exerting less relative influence on the posterior when more information is available from data to estimate the parameters.

Despite the improvement in *SE* bias as G increases, the CR for some parameters remains quite low. The CR plot can be found in our supplementary material.

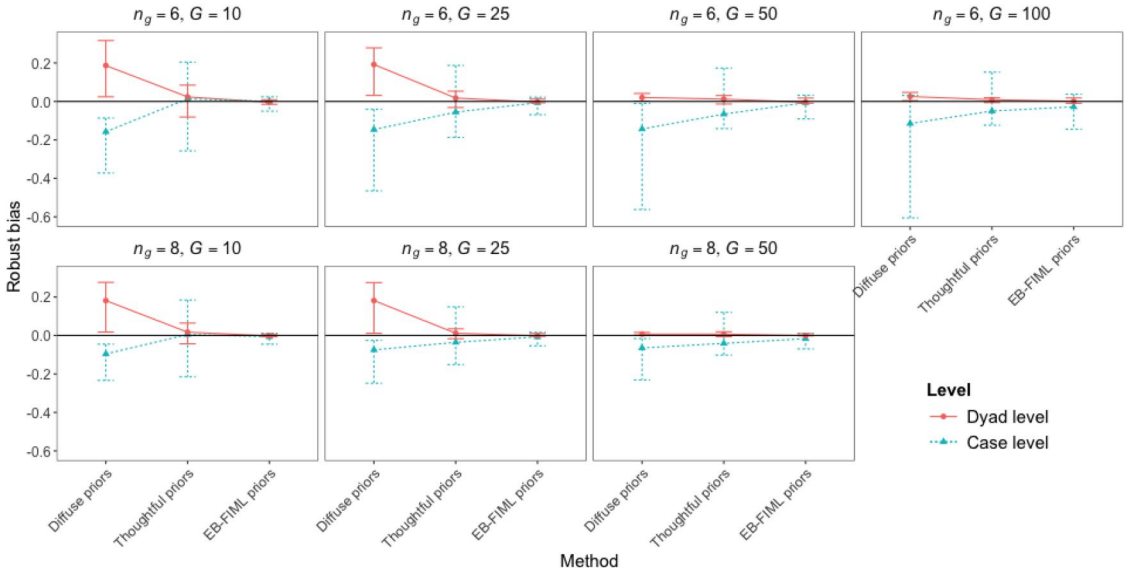

Figure 10. Plot depicting the robust bias for Simulation-3 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the robust bias values. Each facet presents the results for a separate $\{n_g, G\}$ combination. The symbols ▲ (case level) and • (dyad level) represent the median robust bias across all estimates per level, whereas the error bars extend to the minimum and maximum robust bias per level. Solid lines are used for the dyad level and dotted lines are used for the case level.

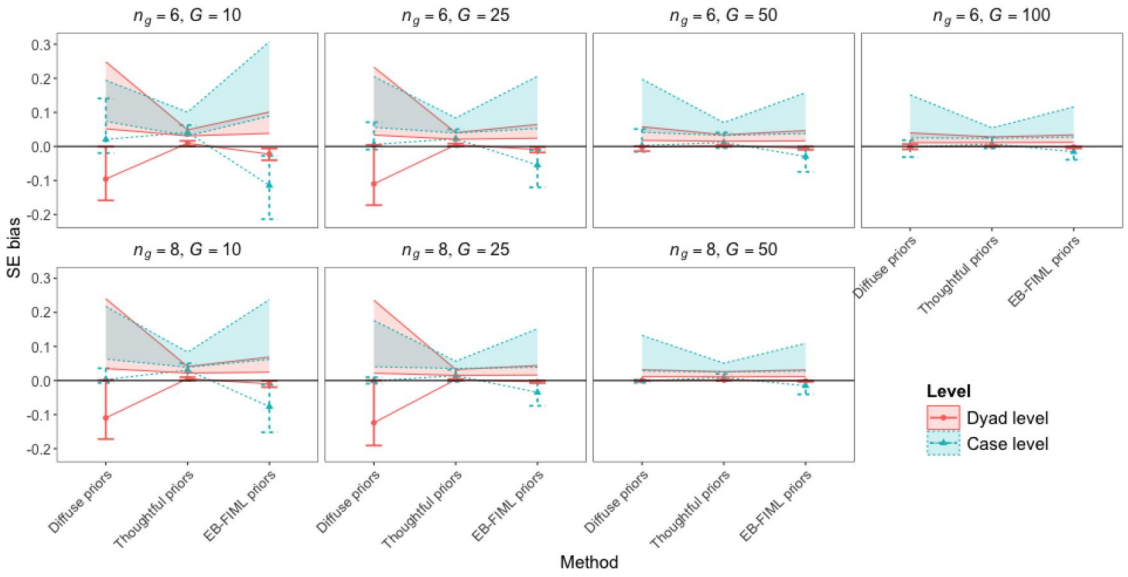

Figure 11. Plot depicting the *SE* bias for Simulation-3 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the *SE* bias values. Each facet presents the results for a separate $\{n_g, G\}$ combination. The symbols ▲ (case level) and • (dyad level) represent the median *SE* bias across all parameters per level, whereas the error bars extend to the minimum and maximum *SE* bias per level. The ribbons show the range of the empirically observed *SE*s across all parameters per level. Solid lines are used for the dyad level and dotted lines are used for the case level.

Discussion

Simulation 3 shows that sampling many groups is insufficient to compensate for the lack of information due to small groups, specifically at the case level, when considering point-estimate bias for diffuse and thoughtful priors. Although bias in *SE*s for EB-FIML priors improves with greater *G*, it is often impractical to sample 100 or even 50 round-robin groups in applied SRM settings.

Simulation 4: Bayesian model averaging

For the additional sampling variability induced by using data-dependent priors to be captured by the posterior, we adapted a method for a similar problem:

capturing uncertainty due to incomplete data. When missing data are imputed multiple times, MCMC estimation can be applied to each completed data set. Given satisfactory convergence criteria per imputed data set, inference can proceed by mixing the posterior samples from all imputations (Gelman et al. 2013, p. 452; Zhou & Reiter 2010).

Similarly, we used FIML to estimate SRM parameters with subsets of round-robin groups, to specify EB priors with different location parameters. The degree to which the estimated posterior distribution varied across these subsets should reflect the uncertainty about the estimates used as location hyperparameters. Because priors are part of the model, merging the separate posterior samples in this scenario is a form of Bayesian model averaging (BMA), although BMA is usually applied when models differ in their covariates or estimated parameters (see Hinne et al., 2020; Wasserman, 2000, for an introduction). We label this prior type as the "BMA-FIML" prior to reflect the two components involved.

Simulation conditions

Adequately investigating this BMA method would warrant its own paper (to review the BMA literature, motivating design factors to manipulate), so this small-scale simulation study serves only as a proof of concept, both to assess its effectiveness in addressing the underestimated posterior variance issue and to determine the computational feasibility of such a method in the SRM context. For this reason, we focused only on a single small-sample condition: $n_g = 6$ with $G = 10$. We compared point and *SE* estimates of FIML and EB-FIML priors to those computed using BMA-FIML priors. As with Simulation 3, we used the same population values and prior specifications detailed previously and simulated $R = 1000$ replications.

Bayesian model averaging

We first created five subsets—containing six groups each—per dataset, resulting in 36 cases and 90 dyads (15 dyads per group) in each subset. EB-FIML priors were constructed for each of the five subsets (see Simulation 2's Method for the procedure). Similar to previous simulations, we specified weakly informative priors and fixed precision to 0.1. Posterior samples were then generated for each of the five subsets.

For each MCMC analysis, we initialized two Markov chains with random starting values and ran each for 2000 iterations, discarding 1000 as burn-in (which returned 2000 posterior samples per subset). This resulted in five separate sets of posterior samples, corresponding to each subset. Then, we computed the mPSRF—using values mPSRF > 1.05 as an indication that the algorithm had not yet converged—for each subset posterior. If mPSRF > 1.05 for any of the five subsets, we repeated the MCMC estimation for all subsets with 10,000 iterations, discarding 5000 as burn-in (i.e., 10,000 posterior samples per subset). Subsets for which mPSRF remained > 1.05 after the second round of MCMC were discarded before merging[9] the remaining posterior samples to estimate the joint posterior distribution of the SRM parameters. EAP estimates of the SRM correlations, *SD*s and (co)variances were saved from this pooled posterior distribution.

Results and discussion

A total of 802 samples were considered for the final analysis in the BMA-FIML condition (see Table A3 in Appendix A). That is, at least two subsets converged with mPSRF < 1.05 for each of these samples.[10] Plots of FIML (from Simulation 1) and EB-FIML (from Simulation 2) are included for comparison.

We evaluate the robust bias, *SE* bias, and coverage rates below. Plots for all three outcome variables can be found in Figure 12. The plot for robust RMSE can be found in our supplementary material.

The left panel in Figure 12 presents the robust bias comparisons for EB-FIML priors, BMA-FIML priors, and FIML. Estimates at the dyad level are comparable across priors, but at the case level, BMA-FIML priors displayed slightly greater absolute median bias than EB-FIML estimates; however, this difference is negligible (approximately 0.02 units). This slight increase in absolute median bias is likely due to the fewer number of cases per subset for BMA-FIML priors (36 cases) compared to EB-FIML priors (60 cases). That is, for BMA-FIML priors, data from only 36 cases is used to estimate FIML estimates of case-level parameters and subsequently compute the final posterior distribution per subset, in contrast to the 60 cases used for EB-FIML priors.

[9]Merging separate chains invalidates the standard calculation of ESS and $\hat{R}$ because additional between-chain variance is expected, not due to lack of convergence, but due to conditioning on different priors (based on different subsets of data). Arguably, the ESS for the merged posterior could be calculated as the sum of the estimated ESS per subset, but exploring that issue is beyond the scope of this article.

[10]We did not consider samples for which only one subset converged with mPSRF < 1.05, as these samples did not undergo mixing of chains from multiple subsets (thus, no model averaging was involved).

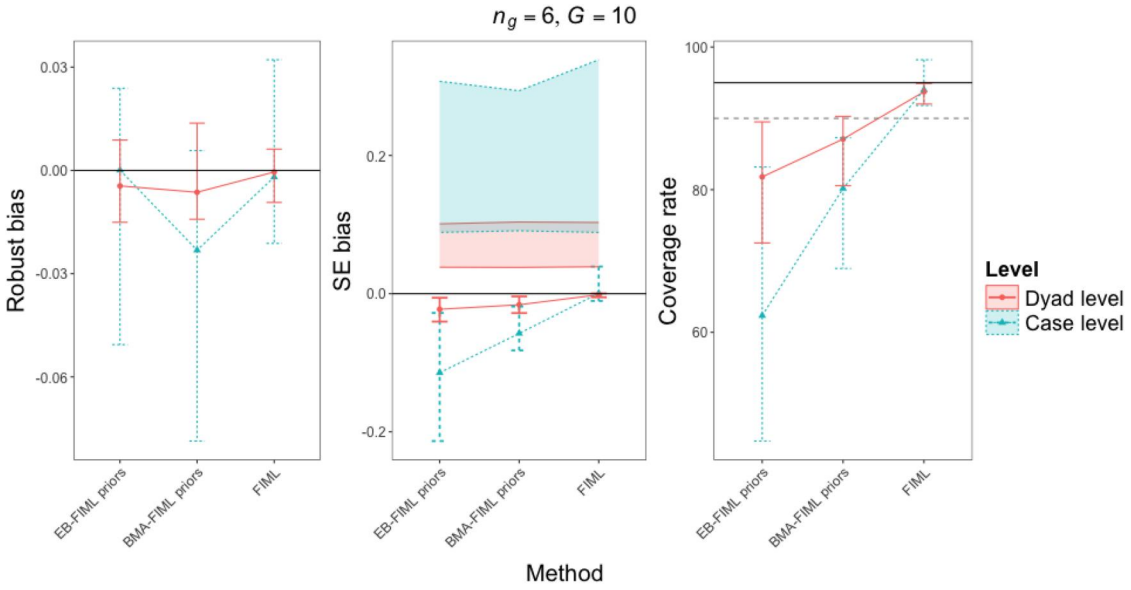

Figure 12. Plot depicting the robust bias (left panel), *SE* bias (center panel), and coverage rate (right panel) for Simulation-4 SRM correlation and *SD* estimates. The estimation methods compared in this simulation are presented in the *x*-axis and the *y*-axis contains the outcome variable values. In all three panels, the symbols ▲ (case level) and ● (dyad level) represent the median value across all estimates per level, whereas the error bars extend to the minimum and maximum per level. Solid lines are used for the dyad level and dotted lines are used for the case level. In the center panel, the ribbons show the range of the empirically observed *SE*s across all parameters per level.

The range of *SE* bias values is presented in the center panel of Figure 12. The range of *SE* bias is already minimal at the dyad level, but is improved with BMA. BMA more noticeably improves case-level *SE* bias, although some bias still remains. The practical impact of the remaining *SE* bias is clear from the range of low CR, displayed in the right panel of Figure 12.

Overall, the results of Simulation 4 show that merging posterior samples from multiple subsets of data attenuates, but does not completely eliminate, *SE* bias. More work is needed to explore whether a more optimal implementation of BMA (e.g., larger or more subsets) can further reduce *SE* bias.

General discussion

In this article, we explored MCMC estimation of the multivariate SRM, comparing it to FIML across a range of commonly employed round-robin design conditions. Bayesian inference for SRM parameters has several theoretical and practical advantages, as demonstrated in several developments and applications for univariate round-robin outcomes (e.g., Jorgensen et al., 2018; Koster & Leckie, 2014; Lüdtke et al., 2013, 2018). An increased demand for multivariate modeling of round-robin data has led to recent developments using FIML estimation (Nestler, 2018; Nestler et al., 2020), and Jorgensen et al. (2024) demonstrated the practical utility of MCMC estimation, acknowledging the need for Monte Carlo research to establish best practices, as well as reveal conditions that warrant caution (e.g., small samples). To this end, we presented four simulation studies. The first three simulations explored the impact of manipulating (a) the precision of prior distributions for a fixed prior location, (b) the accuracy of prior locations for a fixed prior precision, and (c) the number of groups sampled when information from only a few participants per group is available. For (b), we explored two methods for specifying more accurate EB priors to better estimate SRM correlations and *SD*s. The fourth simulation study applied BMA to attenuate a well-documented problem with EB priors.

In Simulation 1, we found that even for "prophetic" prior locations, the accuracy of point estimates depends on prior precision, with more precise priors yielding more accurate posterior estimates (see Figure 5). However, as prophetic prior locations are an unrealistic scenario, we fixed the prior precision in Simulation 2, manipulating prior locations to compare thoughtful and EB priors.

Of the MCMC priors considered in Simulation 2 (diffuse, thoughtful, EB-FIML and EB-ANOVA), the EB priors (which performed comparably) had lower point-estimate bias but higher *SE* bias, resulting in lower but similar RMSE than our minimally thoughtful priors (see Figure 9)—the familiar bias–variance tradeoff. The accuracy of diffuse priors was quite

poor, but improved as n_g increased. Thoughtful priors over/underestimated the SRM parameters depending on whether the true value of the parameter was greater/lesser than the chosen expected value of the rescaled Beta distribution ($M = .30$). We selected a single expected value for all correlations across both levels. In practice, researchers may select separate thoughtful-prior locations for different parameters informed by theory or previous research which may improve the accuracy of the estimation. Though the EB priors displayed low bias, they underestimated sampling variability, leading to low coverage in small groups since the algorithms failed to account for the uncertainty in estimating the Beta hyperparameters.

In Simulation 3, we considered conditions wherein a greater number of small groups are sampled to assess whether (a) point-estimate bias for diffuse and thoughtful priors improves and (b) *SE*s are more accurately estimated for EB-FIML priors. Case-level point-estimate bias for diffuse and thoughtful priors did not show substantial improvement as more groups were sampled. Instead, in the smallest-group condition, bias increased with more groups due to more precision for inaccurate estimates. Dyad-level estimates appeared to improve as G increased from 25 to 50. Additionally, the underestimation of *SE*s for EB-FIML priors improved as more groups were sampled. However, sampling many small round-robin groups may be costly or burdensome; thus, this may not be a practical solution in applied settings.

Use of the same data twice—first to specify prior distributions, and again for the MCMC estimation—can be problematic and results in underestimated posterior variability (Carlin & Louis, 2000b). Simulation 2 demonstrated this issue in the SRM context. In Simulation 4, we attempted to address this issue by merging posterior samples obtained from multiple subsets of the same data, using priors we termed BMA-FIML priors. We found that the *SE*s for BMA-FIML priors were more accurately estimated than those for EB-FIML priors, making the combination of EB priors and BMA a promising topic for future exploration.

Our findings are largely consistent with prior research on univariate SRM with MCMC estimation (Lüdtke et al., 2013) and multivariate SRM with FIML estimation (Nestler, 2018). Generally speaking, dyad-level parameters are more accurately and efficiently estimated than case-level parameters, particularly in small groups. Measuring a participant's multiple interactions within larger groups (i.e., greater n_g) results in more accurate estimates of SRM correlations and *SD*s than smaller groups. Furthermore, increasing G for a given n_g did not substantially improve the estimation of SRM parameters. Thus, if researchers must consider a tradeoff between n_g and G, it is preferable to collect data from fewer large groups than many small groups (Kenny et al., 2006, p. 215). Our third simulation study showed that even 100 small groups (perhaps not even practical) does not substantially reduce the bias of point estimates, and their decreased sampling variability makes BCI coverage even worse.

Recommendations for future research

In Simulation 4, we explored only a single subset condition—specifically, we created five subsets of six groups each for $n_g = 6, G = 10$ data. This combination barely yielded more observations than parameters, given that data from 36 cases were available to estimate 21 parameters at the case level. When more parameters are to be estimated, the number of subsets, size of subsets, or size of groups may have to be increased. It is necessary to reveal the extent to which these factors impact estimates of posterior variability.

In the present simulations, we saved both EAP and MAP posterior estimates of SRM correlations and *SD*s, but found that these negligibly differed. These posterior summary estimates have been previously found to differ for boundary conditions (i.e., near-zero variances or large correlations) in small samples (Lüdtke et al., 2013; Ten Hove et al., 2020). It is interesting for future research to explore whether these estimates differ based on the magnitude of SRM *SD*s and sample size.

Finally, Jorgensen et al. (2024) demonstrated how social-relations structural equation model (SR-SEM) parameters can be estimated, using as input data the level-specific covariances matrices calculated from the correlations and *SD*s estimated with MCMC in this paper. SR-SEMs enable researchers to test measurement and structural hypotheses about round-robin variables, and Nestler et al. (2020) evaluated FIML estimation of SR-SEM parameters. The two-stage method proposed by Jorgensen et al. (2024) remains to be evaluated.

Advice for applied researchers

Our simulations show that informative prior distributions are not a substitute for sufficient data and that larger groups are preferable to derive more accurate estimates. However, we recognize the impracticality of collecting round-robin data from larger groups, which

involve many interactions (e.g., 190 dyads in a group of 20). Researchers may have difficulties setting up studies of this kind, and participants are likely to find it burdensome to interact with or rate many other individuals. In such cases, researchers can plan n_g and G sizes based on their specific hypotheses. For example, if researchers are interested in testing only dyad-level hypotheses, it may be acceptable to collect data from smaller groups, but if they are interested in case-level SRM correlations and *SD*s, collecting data from fewer large groups may be more suitable. It is necessary to note that we provide these recommendations in the context of our analysis of a trivariate SRM. A potentially useful extension in this context may be to explore unequal n_g in the same sample—for instance, whether sampling many small groups in combination with a few large groups may partly solve the issues with accuracy and variability. Another solution could be planned missing-data designs, wherein each of 20 subjects only interacts with (or responds about) a subset of partners (Brunson et al., 2016; Øverup et al., 2021).

Furthermore, our results showed that the choice of prior is highly influential in parameter estimation; hence, it may also be useful—especially in small-group conditions—to conduct a prior sensitivity analysis to evaluate the robustness of estimates across varying prior location and precision. If this reveals results to be sensitive to prior specification, then confidence in any results would be undermined by lacking sufficient data.

Finally, we used ESS, $\hat{R}$, and mPSRF to specify convergence criteria in our simulations, which is more feasible than inspecting visual diagnostics (e.g., traceplots) for 1000 samples per condition. In practice, researchers are advised to use more informative visual diagnostics to evaluate whether the algorithm has converged. Cowles and Carlin (1996) and Roy (2020) provide an overview of convergence diagnostic tools for MCMC. Kwon et al. (2025) and Zitzmann and Hecht (2019) use simulations to compare various convergence criteria.

Conclusion

In summary, the present article explored the impact of manipulating the location and scale of MCMC priors of SRM parameter estimates. We compared MCMC to FIML and found that although FIML produced less biased estimates, they were less efficient than MCMC estimates in small samples. However, MCMC estimates were biased in small samples without prior information, which can be improved using EB priors. RMSE indicated the efficiency gain can outweigh the bias, making MCMC more accurate overall, although the biased estimates of sampling variability led to low coverage, threatening validity of inferences based on interval estimates, particularly in small samples. Thus, MCMC estimation has distinct advantages over FIML estimation of multivariate SRM parameters, but depending on the researcher's priority on valid inference (e.g., nominal BCI coverage), those may be outweighed by bias in small groups. For less biased, more efficient results with either estimator, researchers should prioritize designs with (fewer) large groups rather than (many) small groups.

Article information

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. No human participants were sampled for the research reported in this article.

Funding: This work was supported by the Dutch Research Council (NWO) under project numbers 016.Veni.195.457 and 406.XS.01.078, awarded to Terrence D. Jorgensen.

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

Bhangale, A. M., & Jorgensen, T. D. (2024). Comparing maximum likelihood to Markov chain Monte Carlo estimation of the multivariate social relations model. In H. Hwang, H. Wu, & T. Sweet (Eds.), *Quantitative psychology: 88th annual meeting of the psychometric society*

(pp. 65–75, Vol. 452) Springer. https://doi.org/10.1007/978-3-031-55548-07

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455. https://doi.org/10.1080/10618600.1998.10474787

Brunson, J. A., Øverup, C. S., & Mehta, P. D. (2016). A social relations examination of neuroticism and emotional support. *Journal of Research in Personality*, *63*, 67–71. https://doi.org/10.1016/j.jrp.2016.05.012

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Carlin, B. P., & Louis, T. A. (2000a). *Bayes and empirical Bayes methods for data analysis*. (2nd ed.). Hall/CRC.

Carlin, B. P., & Louis, T. A. (2000b). Empirical Bayes: Past, present and future. *Journal of the American Statistical Association*, *95*(452), 1286–1289. https://doi.org/10.1080/01621459.2000.10474331

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge. https://doi.org/10.4324/9780203771587

Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, *91*(434), 883–904. https://doi.org/10.2307/2291683

Darnieder, W. F. (2011). *Bayesian methods for data-dependent priors* [Doctoral dissertation]. Ohio State University.

Dorff, C., & Ward, M. D. (2013). Networks, dyads, and the social relations model. *Political Science Research and Methods*, *1*(2), 159–178. https://doi.org/10.1017/psrm.2013.21

Fox, J., & Weisberg, S. (2019). An R companion to applied regression. (Third) Sage. https://uk.sagepub.com/en-gb/eur/an-r-companion-to-applied-regression/book246125

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. (3rd ed.) Hall/CRC. https://sites.stat.columbia.edu/gelman/book/

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gill, P. S., & Swartz, T. B. (2001). Statistical analyses for round robin interaction data. *Canadian Journal of Statistics*, *29*(2), 321–331. https://doi.org/10.2307/3316080

Gleason, J. R., & Halperin, S. (1975). A paired compositions model for round-robin experiments. *Psychometrika*, *40*(4), 433–454. https://doi.org/10.1007/BF02291548

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2020). A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, *3*(2), 200–215. https://doi.org/10.1177/2515245919898657

Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, *100*(469), 286–295. https://doi.org/10.1198/016214504000001015

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

Jorgensen, T. D., Bhangale, A. M., & Rosseel, Y. (2024). Two-stage limited-information estimation for structural equation models of round-robin variables. *Stats*, *7*(1), 235–268. https://doi.org/10.3390/stats7010015

Jorgensen, T. D., Forney, K. J., Hall, J. A., & Giles, S. M. (2018). Using modern methods for missing data analysis with the social relations model: A bridge to social network analysis. *Social Networks*, *54*, 26–40. https://doi.org/10.1016/j.socnet.2017.11.002

Jorgensen, T. D. (2023). *lavaan.srm: Fit structural equation models to round-robin data* [R package version 0.1-0.0044]. https://github.com/TDJorgensen/lavaan.srm

Kass, R. E., & Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association*, *84*(407), 717–726. https://doi.org/10.1080/01621459.1989.10478825

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. Guilford Press.

Kenny, D. A. (2013). https://davidakenny.net/srm/srmp.htm

Koster, J. M., & Leckie, G. (2014). Food sharing networks in lowland Nicaragua: An application of the social relations model to count data. *Social Networks*, *38*, 100–110. https://doi.org/10.1016/j.socnet.2014.02.002

Kwon, S., Zhang, S., Köhn, H. F., & Zhang, B. (2025). MCMC stopping rules in latent variable modelling. *The British Journal of Mathematical and Statistical Psychology*, *78*(1), 225–257. https://doi.org/10.1111/bmsp.12357

Lüdtke, O., Robitzsch, A., Kenny, D. A., & Trautwein, U. (2013). A general and flexible approach to estimating the social relations model using Bayesian methods. *Psychological Methods*, *18*(1), 101–119. https://doi.org/10.1037/a0029252

Lüdtke, O., Robitzsch, A., & Trautwein, U. (2018). Integrating covariates into social relations models: A plausible values approach for handling measurement error in perceiver and target effects. *Multivariate Behavioral Research*, *53*(1), 102–124. https://doi.org/10.1080/00273171.2017.1406793

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. (2nd ed.) Chapman & Hall/CRC.

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, *85*(4), 1–30. https://doi.org/10.18637/jss.v085.i04

Nestler, S. (2016). Restricted maximum likelihood estimation for parameters of the social relations model. *Psychometrika*, *81*(4), 1098–1117. https://doi.org/10.1007/s11336-015-9474-9

Nestler, S. (2018). Likelihood estimation of the multivariate social relations model. *Journal of Educational and Behavioral Statistics*, *43*(4), 387–406. https://doi.org/10.3102/1076998617741106

Nestler, S., Lüdtke, O., & Robitzsch, A. (2020). Maximum likelihood estimation of a social relations structural equation model. *Psychometrika*, *85*(4), 870–889. https://doi.org/10.1007/s11336-020-09728-z

Nestler, S., Lüdtke, O., & Robitzsch, A. (2022a). Analyzing longitudinal social relations model data using the social relations structural equation model. *Journal of Educational and Behavioral Statistics*, *47*(2), 231–260. https://doi.org/10.3102/10769986211056541

Nestler, S., Robitzsch, A., Lüdtke, O. (2022b). *srm: Structural equation modeling for the social relations model* [R package version 0.5-1]. https://github.com/alexander-robitzsch/srm

Øverup, C. S., Brunson, J. A., & Mehta, P. D. (2021). A social relations model of need supportiveness. *Journal of Research in Personality*, *94*, 104142. https://doi.org/10.1016/j.jrp.2021.104142

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Roy, V. (2020). Convergence diagnostics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application*, *7*(1), 387–412. https://doi.org/10.1146/annurev-statistics-031219-041300

Salazar Kämpf, M., Liebermann, H., Kerschreiter, R., Krause, S., Nestler, S., & Schmukle, S. C. (2018). Disentangling the sources of mimicry: Social relations analyses of the link between mimicry and liking. *Psychological Science*, *29*(1), 131–138. https://doi.org/10.1177/0956797617727121

Schönbrodt, F. D., Back, M. D., & Schmukle, S. C. (2012). TripleR: An R package for social relations analyses based on round-robin designs. *Behavior Research Methods*, *44*(2), 455–470. https://doi.org/10.3758/s13428-011-0150-4

Schönbrodt, F. D., Back, M. D., Schmukle, S. C. (2022). *TripleR: Social Relation Model (SRM) analyses for single or multiple groups* [R package version 1.5.4]. https://CRAN.R-project.org/package=TripleR

Schuurman, N. K., Grasman, R. P. P. P., & Hamaker, E. L. (2016). A comparison of inverse-Wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research*, *51*(2-3), 185–206. https://doi.org/10.1080/00273171.2015.1065398

Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2020). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 131–161. https://doi.org/10.1080/10705511.2019.1577140

Stan Development Team (2023). *RStan: The R interface to Stan* [R package version 2.26.23]. https://mc-stan.org/

Talloen, W., Loeys, T., & Moerkerke, B. (2019). Consequences of unreliability of cluster means and unmeasured confounding on causal effects in multilevel mediation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(2), 191–211. https://doi.org/10.1080/10705511.2018.1533406

Ten Hove, D., Jorgensen, T. D., van der Ark, L. A. (2020). Comparing hyperprior distributions to estimate variance components for interrater reliability coefficients. *Quantitative Psychology: 84th Annual Meeting of the Psychometric Society*, (pp. 79–93). https://doi.org/10.1007/978-3-030-43469-47

Ten Hove, D., Jorgensen, T. D., & van der Ark, L. A. (2025). Interrater reliability for interdependent social network data: A generalizability theory approach. *Multivariate Behavioral Research, 60*(3), 444–459. https://doi.org/10.1080/00273171.2024.2444940

Warner, R. M., Kenny, D. A., & Stoto, M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology*, *37*(10), 1742–1757. https://doi.org/10.1037/0022-3514.37.10.1742

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, *44*(1), 92–107. https://doi.org/10.1006/jmps.1999.1278

Zhou, X., & Reiter, J. P. (2010). A note on Bayesian inference after multiple imputation. *The American Statistician*, *64*(2), 159–163. https://doi.org/10.1198/tast.2010.09109

Zitzmann, S., & Hecht, M. (2019). Going beyond convergence in Bayesian estimation: Why precision matters too and how to assess it. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(4), 646–661. https://doi.org/10.1080/10705511.2018.1545232

Zitzmann, S., Lindner, C., & Hecht, M. (2024). A straightforward and valid correction to Nathoo et al.'s Bayesian within-subject credible interval. *Journal of Mathematical Psychology*, *122*, 102873. https://doi.org/10.1016/j.jmp.2024.102873

Zitzmann, S., Lüdtke, O., Robitzsch, A., & Hecht, M. (2021). On the performance of Bayesian approaches in small samples: A comment on Smid, McNeish, Miocevic, and van de Schoot (2020). *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(1), 40–50. https://doi.org/10.1080/10705511.2020.1752216

Appendix A: Converged samples across all conditions per simulation

Table A1. Number of converged samples (out of 1000) per condition in simulations 1 and 2.

		Simulation condition							
		$G = 10$				$G = 25$			
Study	Prior type	6	8	10	20	6	8	10	20
1	Default priors	121	385	747	1000	94	594	983	1000
1	Pr-0.05 priors	998	1000	1000	1000	1000	1000	1000	902
1	Pr-0.1 priors	981	993	1000	1000	991	1000	1000	986
1	Pr-0.2 priors	678	907	980	1000	667	967	1000	1000
1	None (FIML)	901	994	999	1000	997	1000	1000	1000
2	Thoughtful priors	991	995	999	1000	961	993	999	1000
2	EB-MOM priors	734	932	994	999	917	997	1000	989
2	EB-FIML priors	648	910	980	1000	889	994	1000	990

Table A2. Number of converged samples (out of 1000) per condition in simulation 3.

	Simulation condition		
	$G = 50$		$G = 100$
Analysis type	6	8	6
Default priors	48	849	24
Thoughtful priors	879	999	813
EB-FIML priors	946	1000	931

Table A3. Number of samples in simulation 4 that merged converged posterior samples from 2–5 subsets of data.

Converged subsets	Number of samples
2	270
3	252
4	200
5	80