Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

Predicting a Distal Outcome Variable From a Latent Growth Model: ML versus Bayesian Estimation

Sanne C. Smid, Sarah Depaoli & Rens Van De Schoot

Structural Equation Modeling: A Multidisciplinary Journal, 27: 169–191, 2020

ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2019.1604140

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Predicting a Distal Outcome Variable From a Latent Growth Model: ML versus Bayesian Estimation

Sanne C. Smid,[1] Sarah Depaoli,[2] and Rens Van De Schoot[1,3]
[1]*Utrecht University*
[2]*University of California*
[3]*North-West University*

Latent growth models (LGMs) with a distal outcome allow researchers to assess longer-term patterns, and to detect the need to start a (preventive) treatment or intervention in an early stage. The aim of the current simulation study is to examine the performance of an LGM with a continuous distal outcome under maximum likelihood (ML) and Bayesian estimation with default and informative priors, under varying sample sizes, effect sizes and slope variance values. We conclude that caution is needed when predicting a distal outcome from an LGM when the: (1) sample size is small; and (2) amount of variation around the latent slope is small, even with a large sample size. We recommend against the use of ML and Bayesian estimation with M*plus* default priors in these situations to avoid severely biased estimates. Recommendations for substantive researchers working with LGMs with distal outcomes are provided based on the simulation results.

Keywords: Simulation study, latent growth model, distal outcome, informative priors

Latent growth models (LGMs) are commonly used to study developmental processes over time (Duncan, Duncan, & Strycker, 2006; Little, 2013, pp. 246–285; McArdle & Nesselroade, 2003; Meredith & Tisak, 1990). LGMs can be extended with a distal (long-term) outcome variable, which refers to a wave of assessment that occurs long after the other waves of assessment in the LGM. By estimating the regression coefficients from the latent intercept and latent slope to the distal outcome variable, researchers can examine whether someone's initial status (latent intercept) or growth rate (latent slope) can predict the distal outcome variable. Examples within the field of public health include predicting: young adult depression from conduct and emotional problems at a younger age (Koukounari, Stringaris, & Maughan, 2017); health-risking sexual behavior among young adults from adolescent substance initiation (Spoth, Clair, & Trudeau, 2014); or reading and writing problems from the development of babies with a family risk of dyslexia (Wijnen, de Bree, van Alphen, de Jong, & van der Leij, 2015).

Another example of an LGM with distal outcomes is from Holgersen, Boe, Klöckner, Weisæth, and Holen (2010), who studied post-traumatic stress caused by an oil rig disaster. An LGM is analyzed with four time points closely after the oil rig disaster (one to three days; four to seven days; two weeks; and three weeks), and two distal outcome variables measured five and 27 years after the disaster. By using this model, Holgersen et al. (2010) were able to investigate whether the participants' initial status or growth rate on post-traumatic stress can predict the levels of stress five and 27 years later.

Hence, LGMs with distal outcomes allow for the assessment of longer-term patterns through the inclusion of distal outcomes. Based on the analysis of LGMs with distal outcomes, a treatment or intervention can be started sooner in order to take preventive actions. Adding a distal outcome variable to an LGM can therefore truly enhance the practical implications of a study.

Correspondence should be addressed to Sanne C. Smid, Department of Methodology and Statistics, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands. E-mail: s.c.smid@uu.nl

In a review of the literature, no methodological or simulation studies were found examining the performance of LGMs with a distal outcome. However, if researchers base their sample size and choice for estimation method on LGM simulation results, then the expected influence of the distal outcome variable is overlooked. Therefore, we see an important need to examine the performance of an LGM with a distal outcome, so that researchers who want to analyze such a model can rely on simulation results suitable for their model of interest. To our knowledge, this is the first simulation study examining the performance of an LGM with a distal outcome.

One important component when examining the performance of LGMs is the estimation method implemented. Much of the literature implementing LGMs represents frequentist estimation (e.g., maximum likelihood).[1] A viable alternative estimation method that is more recently established in the literature is Bayesian estimation (for examples of Bayesian LGM simulation studies see: van de Schoot, Broere, Perryck, Zondervan-Zwijnenburg, & van Loey, 2015; Zhang, Hamagami, Wang, Nesselroade, & Grimm, 2007; Zondervan-Zwijnenburg, Depaoli, Peeters, & van de Schoot, 2018). Within the Bayesian framework, prior information about parameters of the model is combined with the observed data. A Markov chain Monte Carlo (MCMC) estimation algorithm is used to obtain the posterior, which is a compromise between the data and the specified prior distributions.[2] One unique benefit of Bayesian estimation is the inclusion of prior information (e.g., Kruschke, Aguinis, & Joo, 2012; Lee & Wagenmakers, 2014; van de Schoot & Depaoli, 2014). Using informative priors can lead to a decrease in estimation bias and an increase in statistical power compared to the results of frequentist methods, such as maximum likelihood estimation (see, e.g., Miočević, MacKinnon, & Levy, 2017; van de Schoot et al., 2015). These features are especially valuable under instances of small sample sizes. As discussed in – among many others – Gelman et al. (2014) and McNeish (2016a) and echoed in the literature review of Smid, McNeish, Miočević, and van de Schoot (2019), the successful use of Bayesian estimation with small samples requires a thoughtful specification of priors.

[1] Maximum likelihood (ML) estimation is based on asymptotic theory, which implies that large sample sizes are required to meet the assumptions of the estimation method to obtain unbiased parameter estimates. For a conceptual explanation of ML estimation, we refer to Myung (2003), and we refer to Meng and Rubin (1993) for a technical in-depth discussion.

[2] For an elaborative discussion of Bayesian estimation, we refer to, among many others: Depaoli and van de Schoot (2017), Gelman et al. (2014), Kaplan (2014), Kaplan and Depaoli (2013), Kruschke (2015), Lee (2007) and van de Schoot et al. (2014).

Intended goals and organization of the paper

In the current study, we examine the performance of an LGM with a continuous distal outcome. Our interests are specifically on factors that have been shown to be important in the LGM literature, which include: estimator, sample size, the amount of variation around the latent slope, and effect size of the regression coefficients (see, e.g., Hertzog, von Oertzen, Ghisletta, & Lindenberger, 2008; Liu, Zhang, & Grimm, 2016; McNeish, 2016a; van de Schoot et al., 2015; Zondervan-Zwijnenburg et al., 2018). Regarding the estimator, we are particularly interested in comparing Bayesian estimation under various prior specifications (e.g., informative priors versus diffuse default priors) to frequentist maximum likelihood estimation. This investigation will highlight 'best practice' when assessing longitudinal growth in the presence of a distal outcome.

Next, we discuss relevant simulation literature on the model, and then we introduce the (Bayesian) LGM with a distal outcome. We then describe the simulation design and discuss the results. We conclude with a discussion and recommendations for researchers working with LGMs with distal outcomes.

PREVIOUS RESEARCH ON DISTAL OUTCOMES

Within the finite mixture modeling framework, distal outcomes are regularly studied (for empirical studies see, e.g., Eastman, Mitchell, & Putnam-Hornstein, 2016; Hipwell et al., 2016; Jiang et al., 2016; Petras & Masyn, 2010; and for methodological and simulation studies see, e.g., Bakk & Vermunt, 2016; Bray, Lanza, & Tan, 2015; Huang, Brecht, Hara, & Hser, 2010; van de Schoot, Sijbrandij, Winter, Depaoli, & Vermunt, 2017; Vermunt, 2010). Distal outcomes and covariates can impact the latent class structure within mixture models, and several methods are proposed to deal with this (Bakk, Oberski, & Vermunt, 2016; Asparouhov & Muthén, 2014; Bakk & Vermunt, 2016; Lanza, Tan, & Bray, 2013; Vermunt, 2010). Relevant to the current investigation is that adding a distal outcome increases the complexity of the model (Huang et al., 2010). A more complex model has a higher chance of non-convergence during the estimation process (Huang et al., 2010), implying a larger sample size is needed for proper estimation. Additionally, Lanza et al. (2013) discuss another factor that further complicates predicting a distal outcome variable from latent class membership: Namely, the value of the predictor – the true class membership – is not known, but estimated in the model. There are similar concerns for the model under investigation in the current study. Akin to mixture models, the values of the predictors – the true values of the latent intercept and latent slope from the LGM – are unknown and estimated by the growth model. Furthermore, model complexity of LGMs increases when a distal outcome variable is added. We therefore expect that a relatively larger sample size is needed to circumvent convergence problems.

LATENT GROWTH MODELS WITH A DISTAL OUTCOME

There is a rich body of simulation literature examining many different aspects of performance surrounding the LGM (see, e.g., Hertzog et al., 2008; Shin, Davison, & Long, 2017; Tong & Ke, 2016; Ye, 2016). One aspect that is commonly addressed is the performance of LGMs under small sample sizes (see, e.g., McNeish, 2016a, 2016b, 2017; van de Schoot et al., 2015; Zondervan-Zwijnenburg et al., 2018). Different types of LGMs have been examined in this context, however, there is no previous simulation work including distal outcomes. Simulation literature on LGMs (that do *not* include distal outcomes) has shown that small sample sizes in relation to the complexity of the model (i.e., $N < 50$, in an LGM with four time-points and two covariates) can lead to convergence problems when frequentist methods are used (see, e.g., McNeish, 2016a). Furthermore, analyzing LGMs with small sample sizes can lead to biased parameter estimates and low levels of statistical power when frequentist methods are used (e.g., van de Schoot et al., 2015; Zondervan-Zwijnenburg et al., 2018). There are similar concerns for LGMs with distal outcomes. The distal outcome variable can cause higher rates of drop-outs, as the time interval between the different measurement moments is longer than for LGMs without distal outcomes.

The model

Consider a general LGM with a latent intercept and a latent linear slope, as originally described by McArdle (1986), McArdle and Epstein (1987), and Meredith and Tisak (1990).[3] The LGM consists of a measurement model (Equation 1) and structural model (Equation 2):

$$\boldsymbol{y}_{it} = \boldsymbol{\eta}_{\boldsymbol{I}i} + \boldsymbol{\eta}_{\boldsymbol{S}i}\boldsymbol{\lambda}_t + \boldsymbol{\varepsilon}_{it}, \qquad (1)$$

with

$$\boldsymbol{\eta}_{\boldsymbol{I}i} = \alpha_{I0} + \boldsymbol{\xi}_{\boldsymbol{I}i}, \qquad (2)$$

$$\boldsymbol{\eta}_{\boldsymbol{S}i} = \alpha_{S0} + \boldsymbol{\xi}_{\boldsymbol{S}i}\,,$$

where $\boldsymbol{y}_{it}$ is the observed outcome for person i at time t, $\boldsymbol{\eta}_{\boldsymbol{I}i}$ and $\boldsymbol{\eta}_{\boldsymbol{S}i}$ respectively represent the person-specific latent intercept and latent linear slope factors, $\boldsymbol{\lambda}_t$ denotes the time score at time t, and $\boldsymbol{\varepsilon}_{it}$ is the person- and time-specific error term. α_{I0} is the population mean of individual intercept factor values, α_{S0} is the population mean of individual slope factor values, and $\boldsymbol{\xi}_{\boldsymbol{I}i}$ and $\boldsymbol{\xi}_{\boldsymbol{S}i}$ represent the differences between the latent factors ($\boldsymbol{\eta}_{\boldsymbol{I}i}$ and $\boldsymbol{\eta}_{\boldsymbol{S}i}$) and the population means (α_{I0} and α_{S0}).

The LGM can be extended by including a distal outcome variable (see Figure 1). When adding a distal outcome variable, the structural model, as shown in Equation 2, is extended with:

$$\boldsymbol{\eta}_{\boldsymbol{D}i} = a_{D0} + \beta_1\boldsymbol{\eta}_{\boldsymbol{I}i} + \beta_2\boldsymbol{\eta}_{\boldsymbol{S}i} + \boldsymbol{\xi}_{\boldsymbol{D}i}, \qquad (3)$$

where $\boldsymbol{\eta}_{\boldsymbol{D}i}$ is the person-specific latent factor for the distal outcome, a_{D0} is the intercept of the distal outcome; that is, the population mean of the individual distal outcome variable values when $\boldsymbol{\eta}_{\boldsymbol{I}i}$ and $\boldsymbol{\eta}_{\boldsymbol{S}i}$ are zero. β_1 and β_2 are the regression coefficients representing the relations between the LGM and the distal outcome variable, and $\boldsymbol{\xi}_{\boldsymbol{D}i}$ represents the person-specific difference between $\boldsymbol{\eta}_{\boldsymbol{D}i}$ and a_{D0}. A more detailed description of all parameters in this model can be found in Appendix A1.

Bayesian specification of the model

Within the Bayesian framework, prior distributions are specified for all unknown parameters in the model. Hence, for the Bayesian LGM with a distal outcome this contains the following parameters: latent factor means $\boldsymbol{\alpha}$, regression coefficients $\boldsymbol{\beta}$, covariance matrix $\boldsymbol{\Psi}$ containing the latent factor variances and covariance ξ, and matrix $\boldsymbol{\Theta}$ containing the residual variances $\boldsymbol{\varepsilon}$. We refer to these parameters as $\boldsymbol{\theta}$, which represents a vector of the unknown parameters in matrices $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$. Hence, the prior distributions $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Theta})$ are denoted by $p(\boldsymbol{\theta})$.

We followed the discussion in Lee (2007, pp. 95–98), and adjusted the posterior distribution for the inclusion of a distal outcome here. In the posterior analysis, the observed data $\mathbf{Y}$ $(y_{it}, \ldots, y_{nt})$ is augmented with the matrix of latent variables $\boldsymbol{\eta}$, resulting in the joint posterior distribution $[\boldsymbol{\theta}, \boldsymbol{\eta}|\mathbf{Y}]$. The unknown parameters in $\boldsymbol{\theta}$ can be divided into two groups: $\boldsymbol{\theta}_{\boldsymbol{y}}$, the unknown parameters in $\boldsymbol{\Theta}$ associated with the measurement model; and $\boldsymbol{\theta}_{\boldsymbol{w}}$, the unknown parameters in $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\Psi}$ associated with the structural model. The prior distributions of the measurement model are assumed to be independent of the prior distributions of the structural model, and can therefore be seen as two different sets of prior distributions: $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_y)p(\boldsymbol{\theta}_w)$. Hence, the

[3] For an introduction into LGMs, we refer to, among many others: Curran, Obeidat, and Losardo (2010), Duncan et al. (2006), McArdle (2012), Little (2013), and Stoel, van den Wittenboer and Hox (2004).

likelihood is expressed by $p(\mathbf{Y}|\boldsymbol{\eta},\ \boldsymbol{\theta}) = p(\mathbf{Y}|\boldsymbol{\eta},\ \boldsymbol{\theta}_y)$ and $p(\boldsymbol{\eta}|\boldsymbol{\theta}) = p(\boldsymbol{\eta}|\boldsymbol{\theta}_w)$. Accordingly, the posterior distribution of the LGM with a distal outcome is given by $p(\boldsymbol{\theta}_y,\ \boldsymbol{\theta}_w|\mathbf{Y}, \boldsymbol{\eta}) \propto [p(\mathbf{Y}|\boldsymbol{\eta},\ \boldsymbol{\theta}_y)p(\boldsymbol{\theta}_y)]\ [p(\boldsymbol{\eta}|\boldsymbol{\theta}_w)p(\boldsymbol{\theta}_w)]$.

SIMULATION DESIGN

The model of interest in the current simulation study, is an LGM with a latent intercept, latent linear slope, four time points, and one continuous distal outcome variable, as represented by Equations 1–3, and shown in Figure 1. The population values for this model are based on McNeish (2016a). Data sets were generated and analyzed using M*plus* version 8 (Muthén & Muthén, 2017), and R version 3.4.4 via the package MplusAutomation version 0.7 (Hallquist & Wiley, 2017; R Core Team, 2018). The following data generation conditions were varied: sample size (3 levels), effect size (2 levels), and population values for the slope variance parameter (2 levels). These three conditions were fully crossed with each other, resulting in 12 different settings for data generation. For each of these 12 settings, 1,000 data sets were generated, and we analyzed these datasets using eight different estimation methods: maximum likelihood (ML) estimation; Bayesian estimation with M*plus* default priors; Bayesian estimation with weak, medium, and strong informative priors centered at the population values; and Bayesian estimation with weak, medium, and strong priors deviating from the population values. Accordingly, the simulation design includes: 3 (sample sizes) x 2 (effect sizes) x 2 (slope variance values) x 8 (estimation methods) = 96 cells. An overview of the simulation design can be found in Table 1, and the varying conditions are detailed below.

TABLE 1
Overview of the Simulation Design

3 sample sizes: 26 (very small), 52 (small), 325 (large)
2 effect sizes for β_1 and β_2: 0.20 (small), 0.80 (large)
2 population values for slope variance: 0.10 (small), 1.00 (large)
8 estimation settings:
• Maximum likelihood estimation (ML)
• Bayesian estimation with only M*plus* default priors (BayesDefault)
• Bayesian estimation with six sets of informative priors:
• weak, medium and strong priors centered at population values (Info Weak, Info Medium, Info Strong)
• weak, medium and strong priors deviating from population values (Deviating Weak, Deviating Medium, Deviating Strong)

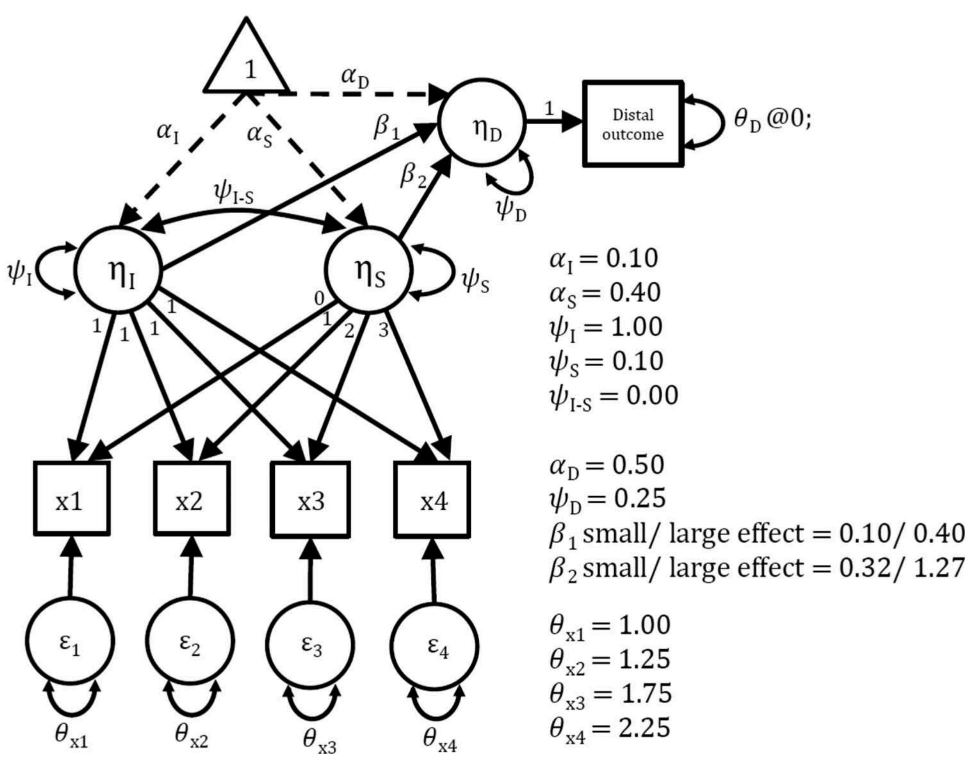

FIGURE 1 The model and population values used in the current simulation study. Note that in this figure, population values are given for a small latent slope variance ψ_s (0.10). For a large latent slope variance (1.00), the regression coefficients for β_2 are adjusted to 0.10/0.40 to still represent a small/large effect.

Conditions simulation design

Sample size

Sample size was computed as a factor of the number of unknown parameters, given by: $n = d*a$, where a denotes the number of unknown parameters in the model of interest, $d = 2$ represents a very small sample, and $d = 4$ represents a small sample (as discussed in Lee & Song, 2004, p. 660). In the current study, the number of unknown parameters in the model, a, is 13. Therefore, $n = 26$ represents a *very small* sample size, and $n = 52$ a *small* sample size. We also included $n = 325$ to see how the various estimation methods perform under a *large* sample size.

Effect size

Two different effect sizes are investigated for β_1 and β_2: a small effect size, represented by a standardized regression coefficient of 0.20; and a large effect size, represented by a standardized regression coefficient of 0.80. Supplementary file S1 shows the computation of the corresponding unstandardized regression coefficients which are used for data generation (all Supplementary files are available on the Open Science Framework: https://osf.io/ycfvg/).

Slope variance

We investigated two levels of variation around the latent slope to examine whether this influences the prediction of the distal outcome variable in an LGM. In empirical studies, the ratio of the intercept and slope variance is often small to moderate (Hertzog et al., 2008; Liu et al., 2016), and the slope variance is usually less than 1/4 of the intercept variance (Ke & Wang, 2015). Small ratios are regularly studied in the context of simulation studies (see, e.g., Bauer & Curran, 2003; Liu et al., 2016; McNeish, 2016a; Muthén & Muthén, 2002). However, larger ratios are also important to examine, as empirical research can produce any values along that continuum. In the current simulation study, the intercept variance was fixed at 1.00, and the two following levels of slope variation were examined: a small slope variation of 0.10 (ratio of 1/10), and a large slope variation of 1.00 (ratio of 1/1).

Estimation methods

To investigate the impact of various estimation methods on the results, we compared ML to seven levels of Bayesian estimation. For ML, all M*plus* default settings were used regarding convergence (see Muthén & Muthén, 2017). For the Bayesian analyses, the median point estimate of the posterior was saved, no thinning was used (i.e. thinning interval = 1), and two Markov chains were specified for each model parameter. The Gelman–Rubin potential scale reduction (PSR) factor (see, e.g., Gelman et al., 2014; Gelman & Rubin, 1992) was used to assess convergence. The convergence criterion was set to 0.01 instead of the 0.05 M*plus* default, to request a stricter criterion and ensure convergence was obtained. The minimum and maximum number of iterations per chain were increased and set at 50,000 and 150,000, respectively. The first half of the iterations within each chain was discarded as the burn-in phase, and the remaining iterations defined the posterior. Aside from the PSR factor, convergence was also visually examined for two randomly selected data sets for each of the 12 data generating conditions. These randomly selected data sets were analyzed using the different Bayesian estimation conditions, and then trace plots for all estimated parameters were visually examined for fluctuations or other signs of non-convergence.

Prior Specifications for Bayesian Estimation. Here, we discuss the prior specifications for the seven Bayesian estimation settings. First, Bayesian estimation with only diffuse M*plus* default priors was used (i.e. BayesDefault); default priors can be found in Appendix A2. The other six levels of Bayesian estimation contain informative prior distributions for five parameters in the model to mimic an achievable applied data situation, where the researcher would have information about some model parameters but not all of them. The model parameters with informed priors were: the mean of the latent intercept; the mean of the latent slope; the intercept of the distal outcome; and the two regression coefficients. M*plus* default priors were used for the remaining model parameters.

Two main types of prior distributions were specified: prior distributions that contained information similar to the population values *(informative prior conditions)*, and distributions that contained information that was deviating from the population values *(deviating prior conditions)*. By investigating these two types of priors, we were able to examine the upper-bound performance (i.e. when prior distributions are centered at the population values), as well as a scenario that is probably more realistic in practice (i.e., when the location of the prior distributions deviates from the population values). Under these two main categories of prior *location* (centered at the population value, and deviating from it), we investigated varying degrees of precision in the prior distributions.

Specifically, we investigated weak, medium, and strong levels of certainty by manipulating the variance hyperparameter of the prior. In order to set up these conditions, we used the information we gained from the results of the condition implementing all M*plus* default prior settings. Upon obtaining the results from the default prior settings, we logged the posterior standard deviation (SD) for the five model parameters in question. We then used this value to help us set three different degrees of (un)certainty within the prior setting. Specifically, for all five parameters, a normal distribution was specified, N(μ, σ^2), where σ^2 = 150%, 100%, and 50% of the (posterior

SD)^2 of the default prior setting results; these three settings created weak, medium, and strong prior distributions, respectively. The weakly informative prior had a variance hyperparameter of *1.50*(posterior SD)*^{2}, indicating it contained relatively more uncertainty (i.e. more variation). The medium informative prior used the posterior standard deviation from the default prior analysis: The variance hyperparameter was *1.0*(posterior SD)*^{2}. Finally, the strong informative prior was computed as *0.50*(posterior SD)*^{2}, indicating it contained the most certainty out of the three conditions.

For the informative prior distributions, the mean hyperparameter (μ) of the prior distribution was set to the population value in order to center the bulk of the prior over the population value. For the deviating prior distribution conditions, the mean hyperparameter (μ) was computed such that it deviated from the true population value. Specifically, μ was specified in order that there was a 5% overlap between the informative and deviating prior distributions. For example, in the case of parameter β_2 with n =325, small slope variance and small effect size, the population value was 0.32. The corresponding posterior SD produced by BayesDefault was 0.4086. Accordingly, the weakly informative prior was N(0.32, *(0.4086^2 * 1.5 =)* 0.25), the medium informative prior was N(0.32, *(0.4086^2 * 1 =)* 0.167), and the strong informative prior was N(0.32, *(0.4086^2 * 0.5 =)* 0.083). Then, the deviating prior distributions were fixed to overlap with these distributions by 5% (see Figure 2). Consequently, the weakly deviating prior was N(−1.642, 0.25), the medium deviating prior was N (−1.282, 0.167), and the strong deviating prior was N (−0.813, 0.083). This allowed us to assess the impact of priors that were slightly deviating from the population, potentially representing a setting more realistic to applied inquiries where the truth of the population is unknown.

Note that in the informative conditions, the mean hyperparameters were the same in the weak, medium, and strong distributions. While in the deviating prior conditions, the means differed for the weak, medium, and strong conditions to maintain the 5% overlap between the informative and deviating prior distributions. The variance hyperparameters were the same in the informative and deviating weak conditions; the informative and deviating medium conditions; and the informative and deviating strong conditions. Appendix A3 shows all of these prior conditions. For more information on the varying conditions in the simulation design, we refer to Supplementary file S1.

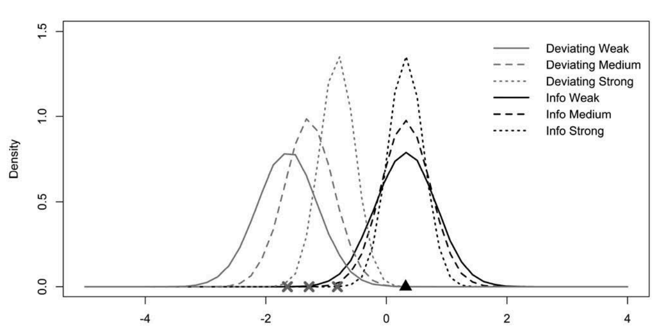

FIGURE 2 Prior distributions for regression coefficient β_2, when n = 325, the slope variance is small (0.10), and the effect size is small (0.20). The three grey lines represent deviating prior distributions, which contain weak, medium and strong amounts of information, respectively. The three black lines represent the weak, medium and strong prior distributions centered at population values. The triangle shows the specified population value, and the three crosses the mean hyperparameters of the three deviating prior distributions. Note that the mean hyperparameters of the three deviating priors differ, to maintain the 5% overlap with the three informative prior distributions.

Evaluation criteria

With small samples or complex models, convergence problems, warnings, and inadmissible parameter solutions can occur. Therefore, the number of completed and non-completed replications (which were highlighted by warning messages) was examined for each of the cells of the simulation design. Furthermore, for all parameters in the model, the following evaluation criteria were examined: relative mean bias, mean squared error (MSE), and coverage.

Relative mean bias was computed by $\left[{(\bar{\theta} - \theta)}/{\theta}\right] * 100$, where $\bar{\theta}$ denotes the average estimate across replications, and θ denotes the specified population value (Muthén & Muthén, 2017). Because the population value of the covariance of the intercept and slope was zero, the relative bias could not be computed. The absolute bias (computed by $(\bar{\theta} - \theta)*100$) is therefore reported for the covariance parameter. In interpreting parameter bias, the cutoff value of ±10% was used as suggested by Hoogland and Boomsma (1998). Values outside this interval represented problematic levels of bias.

The MSE was computed by $(SD)^2 + (\bar{\theta} - \theta)^2$, where SD denotes the standard deviation across replications, $\bar{\theta}$ denotes the average estimate across replications, and θ denotes the specified population value (Muthén & Muthén, 2017). The MSE takes the relative bias and variability across replications into account. Therefore, the smaller the MSE, the closer the estimated value is to the population value, across replications.

Coverage was denoted by the proportion of replications for which the 95% confidence or credibility interval contains the population value. Values between 0.925 and 0.975 are considered to represent good parameter coverage (Bradley, 1978). Values outside the interval could suggest biased standard error estimates.

Finally, for the two regression coefficients β_1 and β_2, statistical power was reported, that is, the proportion of estimates across replications that differs significantly from zero (Muthén & Muthén, 2017). The preferred value for power is considered to be 0.80 (Muthén & Muthén, 2002).

TABLE 2
Overview of the Simulation Process: the Number of Completed Replications and Warnings When Using Maximum Likelihood (ML) Estimation, and the Occurrence of Spikes When Using Bayesian Estimation with M*plus* Default Priors, for the Varying Simulation Conditions

		Small slope variance			*Large slope variance*		
Sample Size	*Effect Size*	*Completed replications for ML*	*Warnings for ML*	*Spikes for Bayes Default*	*Completed replications for ML*	*Warnings for ML*	*Spikes for Bayes Default*
26	Small	830[a]	Total = 345, θ = 45, Ψ = 300	Yes	914	Total = 150, θ = 92, Ψ = 58	Yes
	Large	758	Total = 332, θ = 38, Ψ = 294	Yes	908	Total = 162, θ = 79, Ψ = 83	Yes
52	Small	885[c]	Total = 217, θ = 3, Ψ = 214	Yes	963[a]	Total = 31, θ = 17, Ψ = 14	Yes
	Large	787[b]	Total = 185, θ = 3, Ψ = 182	Yes	962	Total = 41, θ = 12, Ψ = 29	Yes
325	Small	986[b]	Total = 11, θ = 0, Ψ = 11	Yes	1,000	No warnings	No
	Large	968	Total = 22, θ = 22, Ψ = 0	No	1,000	No warnings	No

Note. For each of the cells, 1,000 replications were requested. The [a,] [b,] and [c] denote the occurrence of one, two and three non-completed replication(s), respectively, because the standard errors could not be computed. All other non-completed replications were caused by non-convergence. 'Total' shows the total number of warnings from the completed replications, reported in the M*plus* output when ML estimation was used, θ denotes the number of warnings related to the residual covariance matrix theta, and Ψ the number of problems with the latent variable covariance matrix psi. The detection of spikes (yes/no) for BayesDefault is based on the visual assessment of traceplots for all parameters for 2 randomly selected data sets per data generation condition. For more details on (non-)convergence and spikes, we refer to Supplementary file S1.

RESULTS

In this section, we focus extensively on the results of the parameters related to the distal outcome, and briefly discuss results of the other LGM parameters. The results of the medium prior distributions are very similar to either the weak or strong prior distributions, and have therefore been moved to Supplementary file S2 to conserve space.

Convergence and warnings

The ML analyses did encounter convergence problems in 8.6% of the cases, when taking all 12 data generation conditions into account (out of the 12,000 requested replications, 1,030 of them had convergence problems). Furthermore, standard errors could not be computed in 0.075% of the cases (9 out of 12,000 replications). The total number of completed replications under ML is shown in Table 2. The amount of non-completed replications when the slope variance was small is 3.11 times as high as when the slope variance was large (786 non-completed replications versus 253). Besides non-convergence errors, warnings related to the latent covariance matrix psi, and residual covariance matrix theta were given when ML was used, see Table 2. The total number of warnings was also higher when the slope variance was small: 2.89 times as high than when the variance of the slope was large (total of 1,112 warnings versus 384). Warnings related to the residual covariance matrix theta were present 1.80 times more often when the slope variance was large, while warnings related to the latent variable covariance matrix psi occurred 5.44 times more often when the slope variance was small. For more information on convergence and warnings, we refer to Supplementary file S1.

From the 1,000 requested replications, the Bayesian analyses produced a 100% convergence rate, without any reported warnings. However, when visually examining trace plots for all parameters for 2 randomly selected data sets per data generation condition, to inspect if the multiple chains truly reached convergence, spikes were detected under the M*plus* default priors when small sample sizes were implemented. Spikes are extreme values sampled during MCMC, which cannot always be identified by the potential scale reduction (PSR) factor if they are happening uniformly across the duration of the chain. For instance, for regression coefficient β_2, the trace plot showed spikes with estimates up to 4500 and down to −2000, while the population value for this parameter was 1.27.[4] With a larger sample size and large slope variance, no spikes were observed in the trace plots. Interestingly, no spikes were detected when informative and deviating priors were used in the analyses, even in the conditions with the smallest sample size. The appearance of spikes (yes/no) for the varying simulation conditions when Bayesian estimation with

[4] The values of the extreme spikes for parameter β_2 correspond to the following data generation conditions: small slope variance, large effect size, n =26. The replication number of the data set is 30. The spikes appeared when Bayesian estimation with default priors was used. For more information, see Supplementary file S1.

default priors was used, is reported in Table 2. Interested readers are referred to Supplementary file S1 for the trace plots with spikes for one of the examined data sets, and more details on the visual convergence checks we performed.

Relative bias

We have organized this section into subsections to promote clarity and highlight the most important patterns that emerged. First, the results of the LGM are discussed, followed by an extensive discussion of the results of the distal outcome. The results of the LGM parameters are very similar to findings from previous simulation studies and are therefore only briefly discussed in the main text.

Relative bias in the LGM

Results of the LGM parameters can be found in Supplementary file S3. The most problematic levels of bias are found for the variance parameters of the intercept and slope, which is in line with previous LGM simulation results (see, e.g., McNeish, 2016a, 2016b; van de Schoot et al., 2015). The highest levels of bias for both parameters are reported when the slope variance was small, although the impact of the slope variance was more extreme for the variance parameter of the latent slope. Unexpected was the deterioration of both variance parameter estimates when informative priors were specified for other parameters in the model in combination with M*plus* default priors for the variance parameters (i.e. weak and strong informative prior conditions), and the improvement of the variance of the intercept parameter when deviating priors were specified for other parameters in the model (i.e. weak and strong deviating prior conditions) in comparison to the Bayesian default priors condition and ML results. For the variance parameters of the intercept and slope, ML estimation resulted in the median closest to the population value when samples were small, followed by Bayesian estimation with default priors, Bayesian estimation with informative and deviating priors (for more information, see Figures 3–4 in Supplementary file S4). This might indicate that the specified prior distributions for the variance parameters were not suitable for the current study, in combination with informative priors for other parameters in the model. This issue will be further explored and discussed in the 'Additional Exploration Priors on Variance Parameters' section, and covered in more detail in the Discussion section.

Relative bias in the distal outcome

We start this subsection with a discussion of the two regression coefficients, as these will often be the main parameters of interest in substantive studies. The section will be continued by the description of the results of the intercept and variance of the distal outcome.

Relative Bias for Regression Coefficients β_1 and β_2. With a small slope variance, problematic levels of bias were found when ML and the Bayesian condition with default prior settings (BayesDefault) were used, even in combination with a large sample size. Furthermore, higher levels of bias were also produced with smaller sample sizes. The specification of informative priors led to considerable improvements of the regression coefficient estimates, as can be seen in Figure 3–4. Deviating prior distributions resulted in extremely high levels of bias for the two regression coefficients. Higher levels of bias with deviating priors were associated with a small effect size. Furthermore, a counterintuitive pattern was visible for both coefficients when the slope variance was small and BayesDefault and ML were used. As can be seen in Figure 4a,b, the estimate of β_2 with BayesDefault was negatively biased when n= 26 and positively biased when n = 325, and vice versa for β_1 (see Figure 3a for β_1). The pattern disappeared when informative priors were specified (Figure 4a,b), and did return when deviating priors were used, when the slope variance and effect size were small (Figure 4e). When the slope variance was large, we did not encounter this pattern and results looked more sensible: The amount of bias decreased when the sample size increased. Note that the results for the smaller sample sizes should be interpreted with caution. When inspecting the distribution of estimates across replications, outliers were detected under small sample sizes, as well as when BayesDefault or ML was used. It is reasonable to assume that these outliers influenced the relative mean bias estimates and could have caused the counterintuitive patterns shown in Figure 4a,b. Therefore, boxplots are presented in Figures 5 and 6 to show the entire distribution of estimates across replications. Hence, the figures showing the relative mean bias should be interpreted in combination with the boxplots in which the outliers are clearly visible.

In Figures 5 and 6 it can be seen that when the sample size increases, the amount of outliers decreases and the distributions of estimates are closer to the true population values. Furthermore, the distributions of estimates based on Bayesian estimation with informative priors (weak and strong) were closer to the population values than when ML and BayesDefault were used. The distributions of estimates based on the deviating priors (weak and strong) were clearly deviated from the population values, although less outliers did occur compared to ML and BayesDefault. Comparing β_1 and β_2, more extreme outliers were present for the estimation of β_2 (especially when the slope variance was small, see Figure 6a,b). For β_1, the estimation of the four data generation scenarios led to more or less similar distributions and amount of outliers.

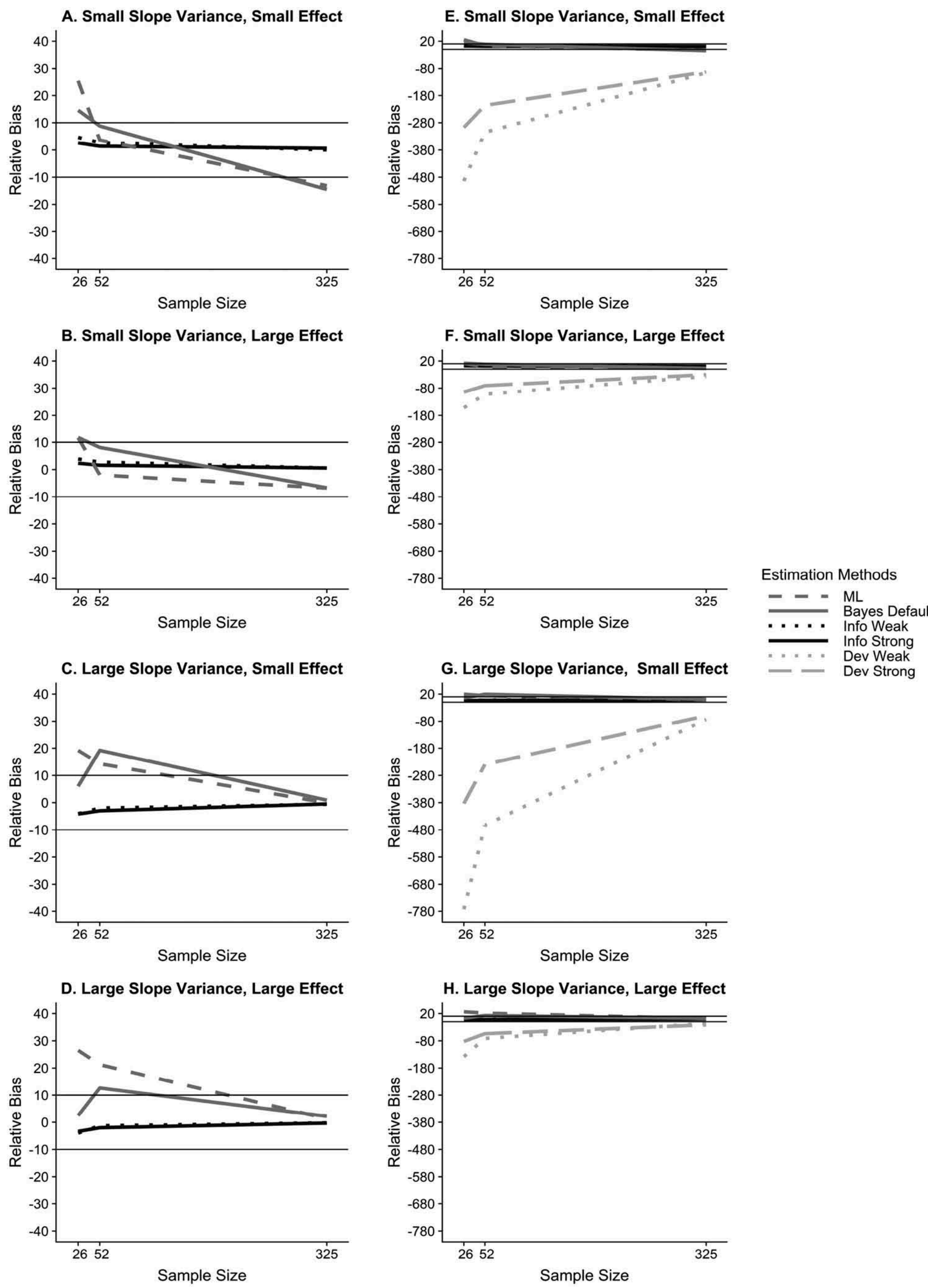

FIGURE 3 Relative bias for regression coefficient β_1, under varying sample sizes, effect sizes, slope variance values, and estimation methods. The static black horizontal lines represent the ±10% interval. Subfigures at the left (A-D) present a smaller range of the y-axis to show the performance close to the ±10% boundaries, and therefore deviating prior conditions are not included here. Subfigures at the right (E-H) represent corresponding data generation conditions as A-D, but also include the deviating conditions. Note that therefore the y-axes of the A-D graphs differ from the y-axes of the E-H graphs.

Relative Bias for Intercept α_D. When ML and BayesDefault were used, too high levels of bias were only reported for the intercept of the distal outcome when the effect size was large and the slope variance was small (see Figure 7). The estimates were biased for all sample sizes, and the counterintuitive pattern that was observed for the regression coefficients was present when the slope variance was small and the effect size was large (see Figure 7b). As expected, the use of informative priors improved the estimates. With a large slope variance (Figure 7c,d) – regardless of the effect size – the bias of the distal outcome intercept was close to 0% when using ML, BayesDefault and the two informative prior conditions. The specification of deviating priors led to extremely biased estimates when sample sizes were small.

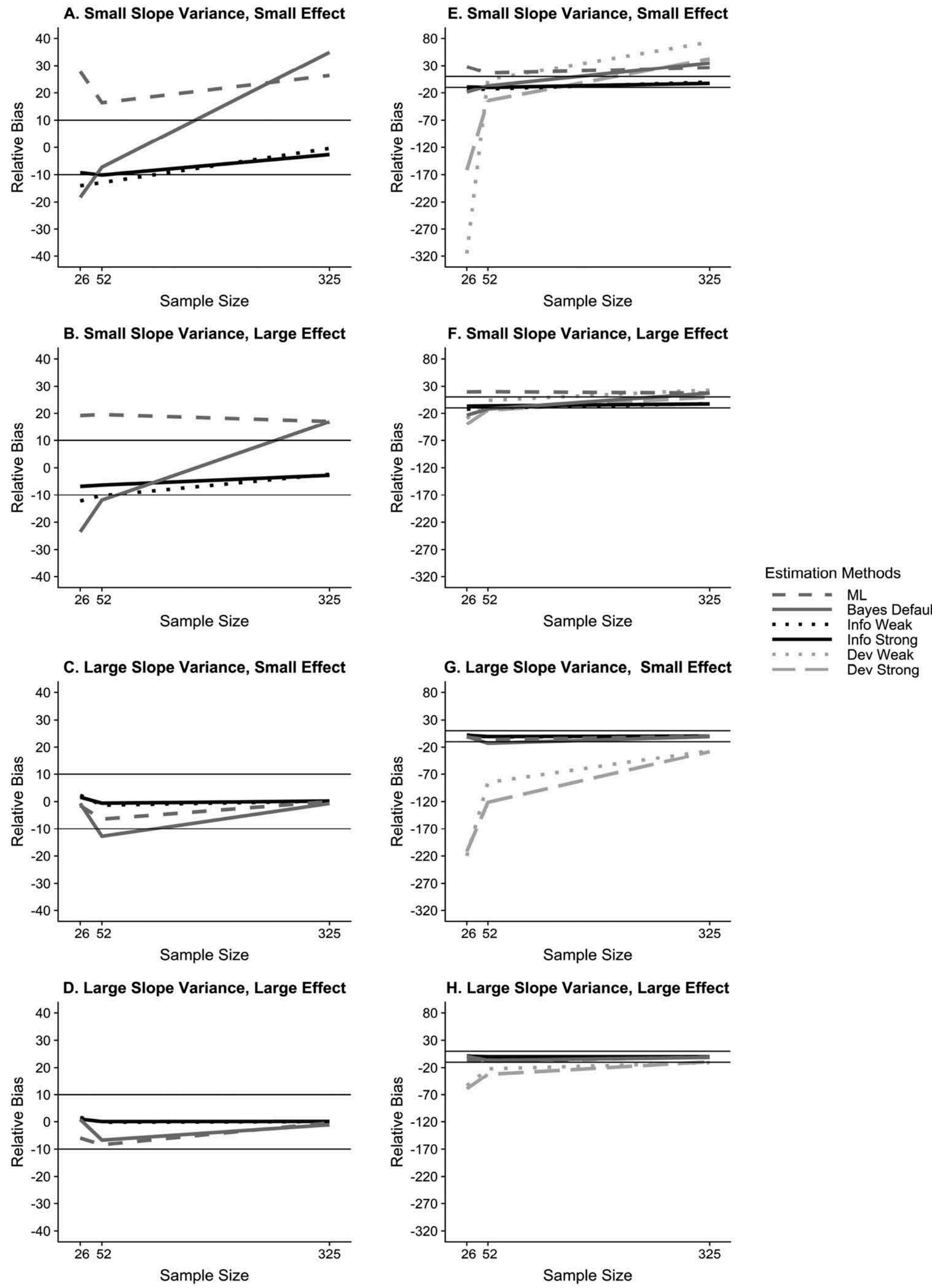

FIGURE 4 Relative bias for regression coefficient β_2, under varying sample sizes, effect sizes, slope variance values, and estimation methods. The static black horizontal lines represent the ±10% interval. Subfigures at the left (A-D) present a smaller range of the y-axis to show the performance close to the ±10% boundaries, and therefore deviating prior conditions are not included here. Subfigures at the right (E-H) represent corresponding data generation conditions as A-D, but also include the deviating conditions. Note that therefore the y-axes of the A-D graphs differ from the y-axes of the E-H graphs.

The boxplots in Figure 8 show larger ranges of distributions and more outliers when the slope variance was small (Figure 8a,b) compared to when the slope variance was large (Figure 8c, d). This indicates that the results were more stable across replications when the slope variance was large.

Relative Bias for Variance Parameter ψ_D. For the variance of the distal outcome, biased estimates were reported for ML when the sample size was small, and this also occurred for BayesDefault when the sample size was small in combination with a small slope variance and small effect (see Figure 9). The estimation of the variance

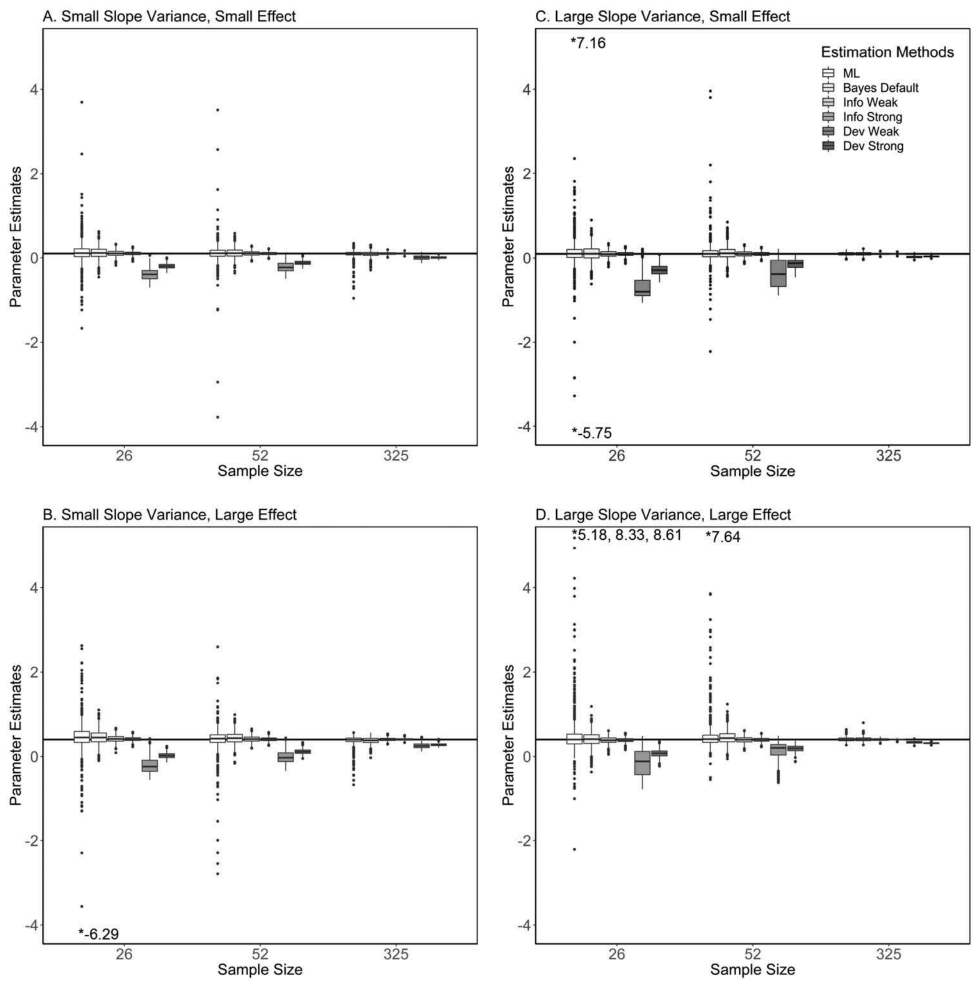

FIGURE 5 Distribution of the estimates for parameter β_1 across completed replications, under varying sample sizes, effect sizes, slope variance values, and estimation methods. The static black horizontal line denotes the true population value for β_1. Outliers are displayed as black circles. Outliers outside the interval [−4; 5] only occurred for ML, and are denoted by *.

parameter improved when informative priors were specified for other parameters in the model and resulted in unbiased estimates. There was only one exception: When n = 26, the slope variance was small, and the effect was large (Figure 9b), the estimate was slightly biased. Deviating prior distributions, specified for other model parameters, led to increased levels of bias in comparison to ML, BayesDefault, and the informative prior conditions.

In the boxplots in Figure 10, it can be seen that the highest outliers were associated with ML estimation and deviating priors, when samples were small (n = 26, 52); and when the effect size was large.

Mean squared error

In Supplementary file S3, the mean squared error (MSE) values are shown for the varying parameters, sample sizes, effect sizes, and slope variance values. For all parameters, higher levels of MSE were associated with smaller sample sizes, the use of Bayesian estimation with weak and/or strong deviating priors, and/or ML estimation. As the MSE took into account both variability and bias of the estimates, the MSE values showed a similar pattern as the distributions of estimates shown in the boxplots in Figures 5, 6, 8, 10, and in Supplementary file S4.

Coverage

Results in terms of coverage can be found in Supplementary file S3. When ML was used, 21.15% of the cases showed under-coverage, but only in 5.13% of all values the coverage values were below 0.90.[5] With BayesDefault, under-coverage only occurred in 3.21%, and values below 0.90 were not

[5] These values represent the percentage of cases that showed under- or over-coverage from a total of 156 cases. 156 is computed as follows: 3 (sample sizes) x 2 (effect sizes) x 2 (slope variance values) x 13 (parameters) = 156.

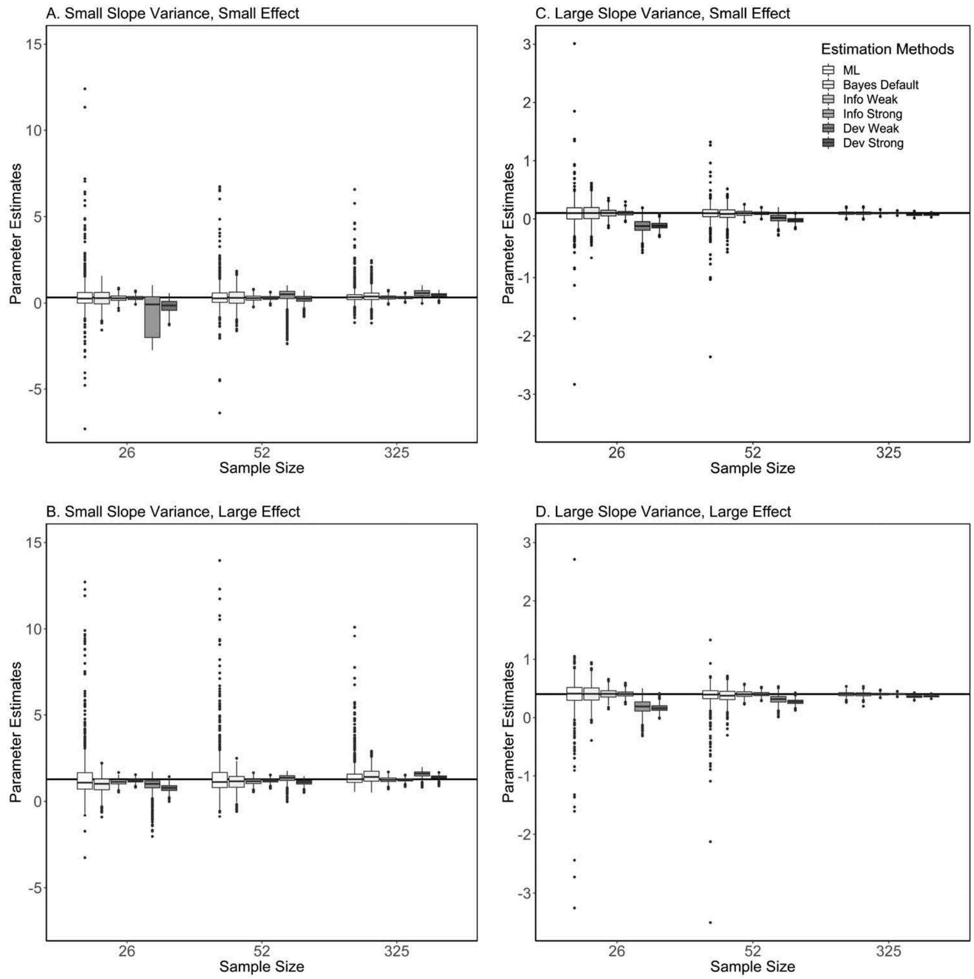

FIGURE 6 Distribution of the estimates for parameter β_2 across completed replications, under varying sample sizes, effect sizes, slope variance values, and estimation methods. The static black horizontal line denotes the true population value for β_2. Outliers are displayed as black circles. Note that the range of the y-axes of subfigures A and B differs from the range of the y-axes of subfigures C and D.

obtained. Under-coverage was especially associated with smaller sample sizes. The use of informative priors never resulted in under-coverage rates, while the use of deviating priors often led to extremely low coverage values – especially for the five parameters for which deviating priors were specified. Under-coverage was found in 60.90% of the cases for weak deviating distributions, and 55.13% of the cases for the strong deviating distributions, and dramatically low coverage values were obtained down to 0.002.

ML resulted in 3.21% of the cases in over-coverage rates, while the use of BayesDefault led to over-coverage rates in 25% of the cases. The use of informative priors resulted in over-coverage rates for all situations for the five parameters for which informative priors were specified. For the other eight parameters, over-coverage occurred in 13.54% of the cases for the weak informative priors, and in 9.38% for the strong informative priors.[6] Deviating priors hardly ever resulted in coverage rates that were too high; 2.56% of the cases for weak deviating priors yielded over-coverage, as well as 3.85% of the cases for strong deviating priors.

Power of the regression coefficients β_1 and β_2

In Table 3, the power rates of β_1 and β_2 are presented. With a small sample size, it was impossible to detect a small effect when ML *or* BayesDefault were used. Only in 16.7%, the power levels of ML and BayesDefault were at or above 0.80.[7] Higher levels of

[6] Here, we used 96 cases instead of 156, because we were interested in the remaining parameters for which no informative and deviating prior distributions were specified: 3 (sample sizes) x 2 (effect sizes) x 2 (slope variance values) x 8 (parameters) = 96.

[7] Here, we used 48 cases instead of 156, because we were interested in the two regression coefficient parameters: 3 (sample sizes) x 2 (effect sizes) x 2 (slope variance values) x 2 (estimation methods) x 2 (parameters) = 48.

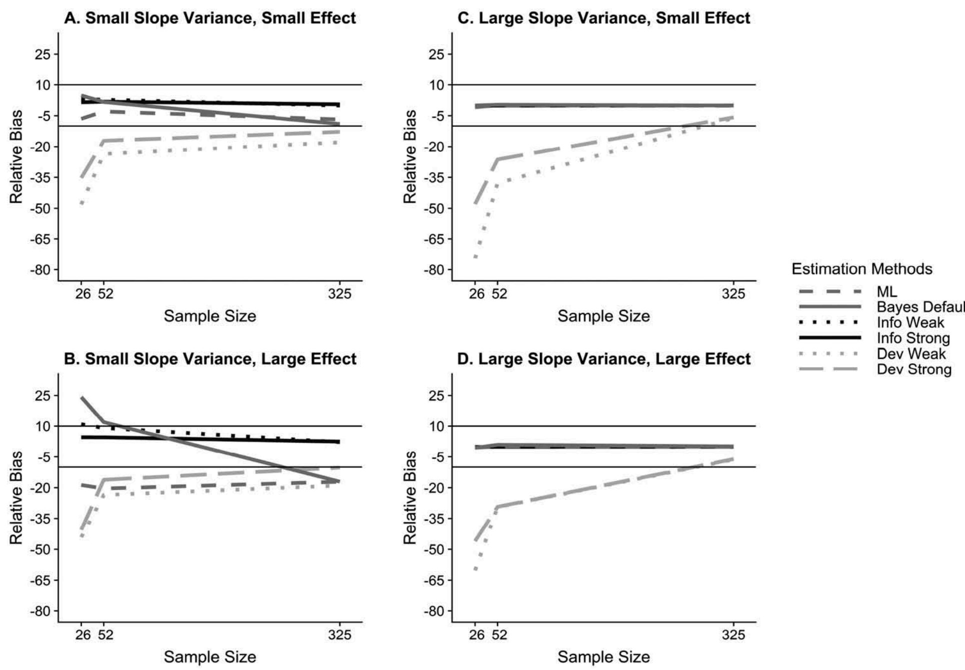

FIGURE 7 Relative bias for the intercept of the distal outcome, under varying sample sizes, effect sizes, slope variance values and estimation methods. The static black horizontal lines represent the ±10% interval.

power were associated with a larger sample size, a large effect in the simulated data and a large slope variance. With the use of informative priors, the large effect was detected with all sample sizes for both small and large slope variances. Additionally, it became possible to detect the small effect when the largest sample size was used under informative priors. Although high levels of power were reported for the deviating priors when the effect and/or sample size was large, the coverage was dramatically low for the corresponding estimates, especially for β_1.

Additional exploration priors on variance parameters

In all simulation conditions, M*plus* default priors were specified for the variance parameters (see Appendix A2). After finding some unexpected results as discussed in the section: ‘Relative bias in the LGM’, we explored alternate prior distributions for the variance parameters. The Inverse Wishart distribution is the default prior distribution in M*plus* for the covariance matrix of a multivariate normal distribution, which means that one prior distribution is specified for all elements in the covariance matrix (Muthén & Muthén, 2017). Consequently, all elements in the covariance matrix are assigned an equal level of informativeness (e.g., Asparouhov & Muthén, 2010). For a comprehensive discussion of the Inverse Wishart prior distribution, we refer to Schuurman, Grasman, and Hamaker (2016) and Liu et al. (2016).

As suggested by Liu et al. (2016), another option is to specify separate priors for the varying parts of the covariance matrix. This type of prior allows for separate prior distributions for each variance and covariance parameter. We followed the suggestions of Liu et al. (2016) when specifying the prior distributions for the additional exploration. An Inverse Gamma prior: IG (0.001, 0.001) was specified for the variance parameters of the intercept, slope, and distal outcome. In turn, a Uniform prior, U [−1, 1], was specified for the covariance of the intercept and slope. For one of the worst-case cells in the design: n =26, small slope variance, small effect size, we ran 1,000 replications for the informative and deviating prior conditions including the separate priors for the variance parameters and compared those to the existing results.

The specification of the Inverse Gamma and Uniform priors for the variance parameters resulted in an improvement of the estimates and led to sensible findings (see Supplementary file S5 for the results in terms of relative bias, and Supplementary file S6 for the boxplots of the distribution of estimates across replications). Informative priors led to a decrease in bias compared to the BayesDefault condition. In turn, deviating priors led to increased levels of bias for all four variance parameters compared to the results of informative prior conditions. For the variance of the slope and variance of the distal outcome, higher levels of bias were found for the deviating prior results compared to BayesDefault results. While for the variance of the intercept, and the covariance of the intercept and slope, the deviating prior condition showed

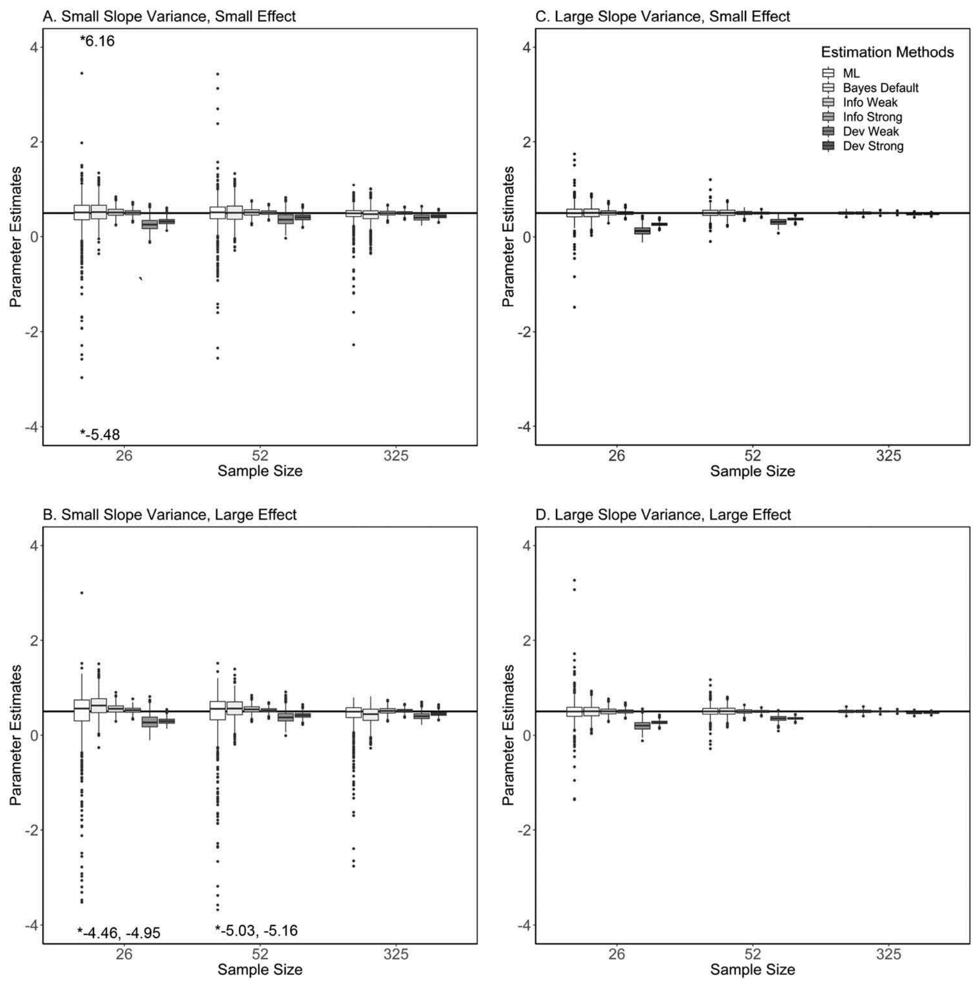

FIGURE 8 Distribution of the estimates for the intercept of the distal outcome across completed replications, under varying sample sizes, effect sizes, slope variance values and estimation methods. The static black horizontal line denotes the true population of 0.50 for the intercept of the distal outcome. Outliers are displayed as black circles. Outliers outside the interval [−4; 4] only occurred for ML, and are denoted by *.

an improvement over BayesDefault results in terms of bias. The distribution of estimates across replications (see boxplots in Supplementary file S6), showed similar patterns. The informative prior conditions resulted in distributions closer to the true population values, and less variable than the results of ML and BayesDefault. The estimates resulting from the specification of deviating priors showed more variable distributions, and their medians were further away from the population value compared to the informative prior condition.

DISCUSSION

The aim of the current study was to examine the performance of an LGM with a continuous distal outcome, under varying estimation methods, sample sizes, effect sizes, and variation around the latent slope. Caution is needed when predicting a distal outcome from an LGM when the sample size is small, or when the slope variance is small – regardless of the sample size. The use of Bayesian estimation with informative priors did improve the estimates in terms of relative bias, MSE, coverage, and power. On the other hand, the specification of priors that deviated from the population values deteriorated the results, especially when sample sizes were small.

Predicting a distal outcome variable can completely fail when there is almost no variation around the latent slope. Even a sample size of 325 is not large enough to yield unbiased regression coefficients when maximum likelihood or Bayesian estimation with default priors are used. Additionally, the prediction of the distal outcome from the latent intercept is also negatively impacted by a small variation around the latent slope, although it is less impactful when the effect size increases. Furthermore, Liu et al. (2016) associated more variation around the latent slope with higher levels of power to identify individual differences around the latent slope. A similar result was found in the

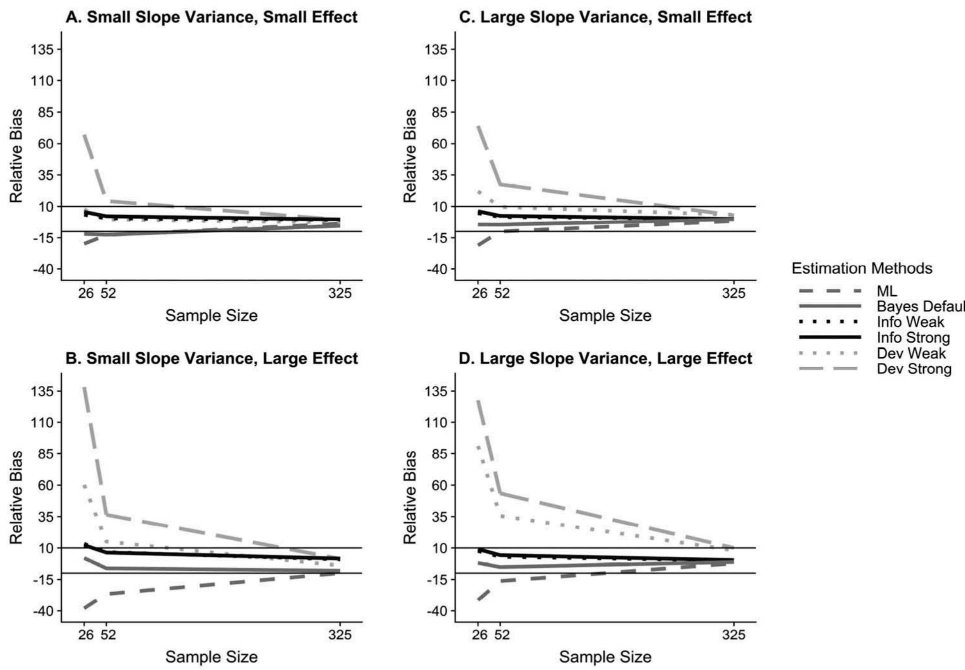

FIGURE 9 Relative bias for the variance of the distal outcome, under varying sample sizes, effect sizes, slope variance values and estimation methods. The static black horizontal lines represent the ±10% interval.

current study: Higher levels of power were reported for the two regression coefficients in the large slope variance condition in comparison to the small slope variance condition.

The results of Bayesian estimation with M*plus* default priors (BayesDefault condition) in the current study, are in line with the many recent simulation studies, claiming Bayesian estimation with default priors is not preferred when sample sizes are small (see, e.g., Depaoli & Clifton, 2015; Holtmann, Koch, Lochner, & Eid, 2016; McNeish, 2016a, 2016b; Shi & Tong, 2017). Spikes were detected when default priors were used in two conditions: (1) when sample sizes were small, and (2) when the slope variance and effect size were small in combination with all examined sample sizes. When prior information was incorporated, by specifying either informative or deviating prior distributions, no spikes were detected. Note that the *weakly informative* and *weakly deviating* prior conditions were still relatively informative distributions in comparison to the default diffuse priors implemented in M*plus*. It is plausible to assume that the M*plus* default prior distributions were not informed enough to prevent spikes for the parameter estimation in the current model and are therefore likely to be the cause of high levels of bias when using Bayesian estimation with default priors.

Findings of the LGM parameters of the model were in line with the existing simulation literature on LGMs; that is, the most problematic levels of bias were detected for the variance parameters of the intercept and slope (e.g., McNeish, 2016a, 2016b; van de Schoot et al., 2015). In McNeish (2016a), an LGM with two covariates was examined with population values and sample sizes comparable to the current study. These similarities allowed us to explore the differences in results. For instance, McNeish (2016a) showed that when using Bayesian estimation with M*plus* default priors, a sample size of 50 was sufficient to obtain an unbiased estimate for the variance parameter of the intercept. While in the current study, with an LGM with a distal outcome, a sample size of 52 still led to a biased estimate for the same parameter.

The additional exploration of the variance parameters in the model indicated that the use of separate priors for the variance components (as suggested by Liu et al., 2016) led to sensible results, whereas the M*plus* default prior distributions for variance parameters did not. Schuurman et al. (2016) examined various specifications of the Inverse Wishart prior for the covariance matrix, and concluded that the prior settings can negatively impact the parameter estimates when the variances are close to zero. However, we decided to implement this prior in the current study based on findings of previous studies. Based on a systematic literature review, Smid et al. (2019) indicated that informative priors for other parameters in the model could improve the estimates of variance parameters, when default priors were specified for the variance parameters. Depaoli (2012), Depaoli and Clifton (2015), and Holtmann et al. (2016) reported similar findings: priors on parameters in one part of the model impacted results for parameters in another part of the model. Further research is needed to examine the exact conditions under which this finding

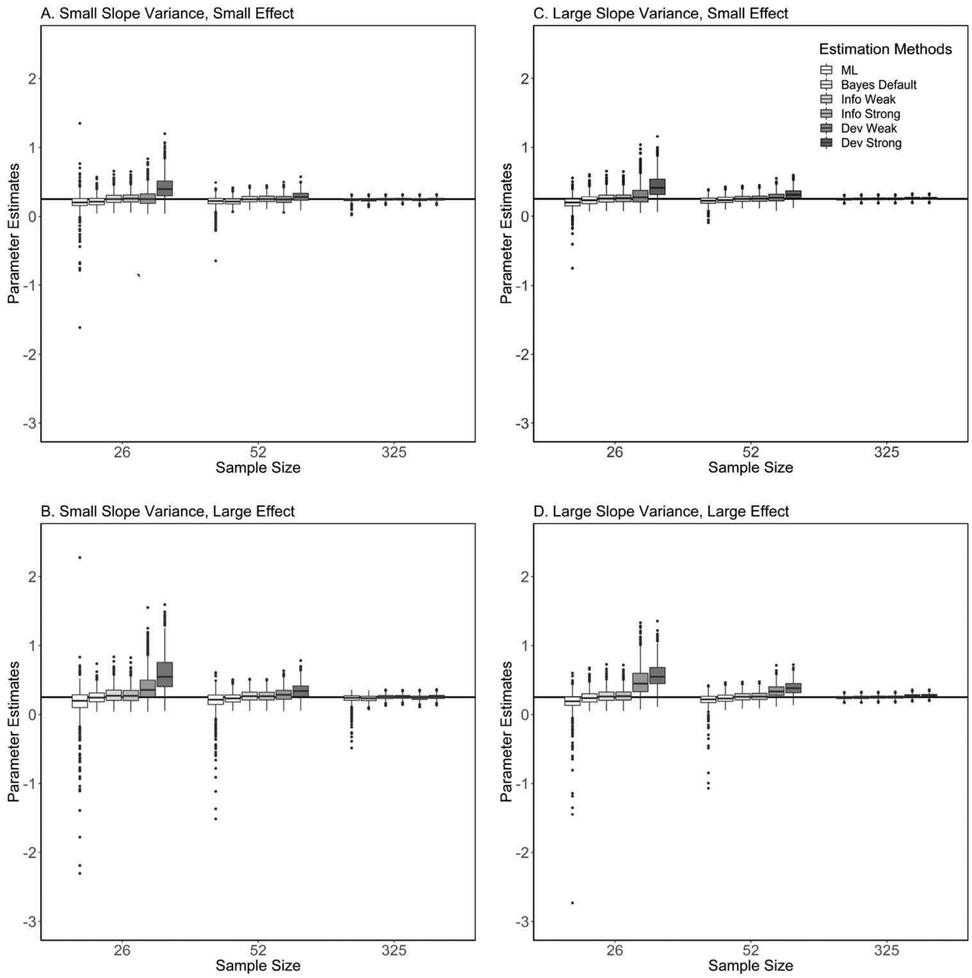

FIGURE 10 Distribution of the estimates for the variance of the distal outcome across completed replications, under varying sample sizes, effect sizes, slope variance values and estimation methods. The static black horizontal line denotes the true population of 0.25 for the variance of the distal outcome. Outliers are displayed as black circles.

holds. Options to further explore the behavior of the variance parameters under varying prior distributions include the use of the half-Cauchy prior as suggested by Gelman (2006), or the use of reference priors as specified in Tsai and Hsiao (2008). Another option is the use of data-dependent priors (Darnieder, 2011; McNeish, 2016b), in which frequentist parameter estimates are implemented in prior distributions. One criticism of data-dependent priors is that data are used twice: first to obtain frequentist parameter estimates, and second when the data are analyzed by using the data-dependent priors (Darnieder, 2011). One way to avoid ‘double-dipping’ is the use of data-splitting techniques. For instance, in the first step, 50% of the data is analyzed using frequentist estimation, and in step two the results of step one are incorporated in prior distributions to analyze the other 50% of the data using Bayesian estimation. On the other side, as the data set needs to be split into two parts, this method is not ideal when the sample size is already small. Hence, there is no clear-cut solution as it depends on the model and the specific situation. However, based on the results of the current study, we conclude that the specification of priors for the variance parameters is of importance – regardless of the use of informative prior distributions for other parameters in the model. It is therefore necessary to further assess the Inverse Wishart (or Inverse Gamma, depending on the model) prior distribution under varying conditions in the future.

Finally, aside from the factors varied in the current simulation design, the number of time points in an LGM can also influence the performance since the number of time points directly impacts the amount of data points. Future research should therefore consider examining the potential influence of the number of time points in an LGM with a distal outcome. Another factor that should be examined is the potential impact of a categorical distal outcome instead of a continuous distal outcome. Also, the impact of adding a quadratic and/or cubic slope to the LGM could be of interest since trajectory shape is also likely to influence the impact of prior settings. Including

TABLE 3
Power of the Regression Coefficients β_1 and β_2

Effect Size	*Slope Variance*	*Sample Size*	*ML*	*Bayes Default*	*Info Weak*	*Info Strong*	*Dev Weak*	*Dev Strong*
Parameter: β_1								
small	small	26	**0.108**	**0.005**	**0.061**	**0.100**	**0.629**	**0.548**
		52	**0.159**	**0.016**	**0.145**	**0.274**	**0.292**	**0.253**
		325	**0.594**	**0.151**	**0.726**	0.977	**0.040**	**0.018**
	large	26	**0.089**	**0.035**	**0.053**	**0.074**	**0.642**	**0.610**
		52	**0.132**	**0.096**	**0.126**	**0.174**	**0.356**	**0.245**
		325	**0.735**	**0.710**	0.979	1.000	**0.113**	**0.324**
large	small	26	**0.534**	**0.068**	0.944	1.000	**0.085**	**0.001**
		52	**0.659**	**0.149**	0.997	1.000	**0.021**	**0.083**
		325	0.867	**0.422**	1.000	1.000	0.904	1.000
	large	26	**0.569**	**0.257**	0.937	0.998	**0.142**	**0.039**
		52	**0.766**	**0.596**	0.993	1.000	**0.332**	**0.485**
		325	0.999	1.000	1.000	1.000	1.000	1.000
Parameter: β_2								
small	small	26	**0.051**	**0.013**	**0.058**	**0.126**	**0.222**	**0.085**
		52	**0.050**	**0.035**	**0.087**	**0.174**	**0.317**	**0.168**
		325	**0.290**	**0.258**	**0.438**	**0.710**	0.811	0.916
	large	26	**0.161**	**0.020**	**0.098**	**0.154**	**0.058**	**0.106**
		52	**0.212**	**0.073**	**0.195**	**0.364**	**0.015**	**0.009**
		325	0.849	**0.794**	0.995	1.000	0.849	0.968
large	small	26	**0.294**	**0.068**	0.996	1.000	**0.592**	**0.726**
		52	**0.366**	**0.157**	0.998	1.000	0.944	0.996
		325	0.810	**0.798**	1.000	1.000	1.000	1.000
	large	26	**0.657**	**0.189**	0.986	1.000	**0.161**	**0.236**
		52	0.839	**0.359**	1.000	1.000	0.913	0.998
		325	0.999	0.994	1.000	1.000	1.000	1.000

Note. Bold values represent power rates below 0.80. ML refers to maximum likelihood estimation; BayesDefault refers to Bayesian estimation using M*plus* default priors; Info Weak and Info Strong refer to the weakly and strongly informative prior settings in the simulation design, and Dev Weak and Dev Strong to the weakly and strongly deviating prior settings.

additional slopes to the growth model increases model complexity, as there could be three to four latent factors that could be used to predict the distal outcome.

RECOMMENDATIONS FOR SUBSTANTIVE RESEARCHERS

The sample size needed to analyze an LGM with a distal outcome depends on the following four items: (1) the parameter of interest, (2) the amount of variation around the latent intercept and slope, (3) the effect size, and (4) the amount of prior information that a researcher can (or wants to) specify. For example, to predict a distal outcome from the participants' growth rate, a sample size of 26 would be sufficient to obtain an unbiased parameter estimate for the regression coefficient β_2 when using ML, BayesDefault, or Bayesian estimation with informative priors when the effect size and slope variance are large (although the statistical power was extremely low in this condition when BayesDefault was used). In contrast, a small effect and slope variance linked to a sample size of 325 would not be sufficient to obtain an unbiased estimate for β_2 when using ML or BayesDefault. In this case, only the use of informative priors could lead to an unbiased estimate when $n = 325$.

The specification of informative priors improved the estimates in terms of bias, MSE, coverage, and power. Accordingly, Bayesian estimation with informative priors can be used with a smaller sample size or slope variance, and it could therefore be a solution for the analysis of data with such characteristics. However, note that informative priors represent the upper-bound performance of Bayesian estimation and not necessarily the practical application of Bayesian estimation in applied research settings. The specification of priors that deviated from the population values deteriorated the results, especially when sample sizes were small. Specifically, the deviating priors negatively influenced the estimation of the parameters for which deviating information was included, but also parameters for which no deviating information was specified.

In real life applications, it is likely to have prior distributions that at least slightly deviate from the data. One might therefore opt to choose BayesDefault instead of risking the specification of deviating priors (when comparing BayesDefault to situations when the prior deviates from the population in the current study). However, as discussed earlier, BayesDefault can lead to severely biased estimates when samples are small (see, e.g., Depaoli & Clifton, 2015; Holtmann et al., 2016; McNeish, 2016a, 2016b; Shi & Tong, 2017), and is therefore hard to recommend as a viable approach.

Hence, we recommend researchers take the most careful approach possible, which entails: (1) carefully constructing prior distributions; and (2) assessing the impact and robustness of the specified priors through an extensive sensitivity analysis. For more information on how to elicit prior information (e.g., based on previous studies, meta-analyses, or knowledge of experts in the field), we refer to: O'Hagan et al. (2006); Zondervan-Zwijnenburg, Peeters, Depaoli, and van de Schoot (2017); Veen, Stoel, Zondervan-Zwijnenburg, and van de Schoot (2017); Bolsinova, Hoijtink, Vermeulen, and Béguin (2017); and van de Schoot et al. (2018). We also refer to Kruschke (2015, pp. 721–725) for an overview of items that should always be reported when Bayesian estimation is used, including reporting details on prior specifications. For information on how to perform a sensitivity analysis, we refer to Depaoli and van de Schoot (2017), and van Erp, Mulder, and Oberski (2018). An example of a sensitivity analysis in an empirical setting can be found in van de Schoot et al. (2018).

The results of the current study further emphasize the importance of inspecting trace plots for all parameters for the appearance of spikes when using Bayesian estimation (see also Depaoli & Clifton, 2015; van de Schoot et al., 2015). The inspection of trace plots should be a standard procedure when Bayesian estimation is used to assess whether the different chains have truly converged (see, e.g., Gelman et al., 2014; Kaplan, 2014; Lynch, 2007) – note that convergence criteria cannot always identify spikes, as we saw in the current investigation.

Although further research is needed to completely examine the performance of the Inverse Wishart prior distribution, researchers should be cautious with the use of the Inverse Wishart default prior in M*plus* for the covariance matrix. Caution is especially needed when the individual variance parameters are expected to be small (as shown by Schuurman et al., 2016). In such a situation, researchers should preferably specify separate priors as suggested by Liu et al. (2016). However, extreme caution is needed if adapting this approach, as one could easily end up with a non-positive definite matrix.

To conclude, LGMs with a distal outcome are useful to assess longer-term patterns, and to detect the need to start a (preventive) treatment or intervention in an early stage. The results of the current study showed that when predicting a distal outcome from an LGM, prudence is called for when: (1) the sample size is small; and (2) the variance of the slope is (expected to be) small. ML and Bayesian estimation with M*plus* default prior settings should not be used in these situations to avoid severely biased estimates. A larger sample size or the specification of informative priors can help to improve the results. Note that the smaller the sample size, the larger the impact of prior distributions on the posterior, and therefore deliberate decisions about prior distributions are necessary. It is our hope that these findings help to uncover the important estimation issues tied to properly assessing the impact of distal outcomes on final model results.

ACKNOWLEDGMENTS

The authors would like to thank Gerbrich Ferdinands for her assistance in preparing the manuscript for resubmission. In addition, the authors would like to thank the anonymous reviewers for their helpful and constructive comments on the manuscript.

FUNDING

This research was supported by a grant from the Netherlands organization for scientific research: NWO-VIDI-452-14-006.

SUPPLEMENTARY MATERIAL

Supplemental data can be accessed here: https://osf.io/ycfvg/.

ORCID

Sanne C. Smid http://orcid.org/0000-0001-6451-202X
Sarah Depaoli http://orcid.org/0000-0002-1277-0462
Rens Van De Schoot http://orcid.org/0000-0001-7736-2091

REFERENCES

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis of latent variable models using Mplus. Retrieved from http://www.statmodel.com/download/BayesAdvantages18.pdf

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using M*plus*. *Structural Equation Modeling*, *21*, 329–341. doi:10.1080/10705511.2014.915181

Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2016). Relating latent class membership to continuous distal outcomes: Improving the LTB approach and a modified three-step implementation. *Structural Equation Modeling*, *23*, 278–289. doi:10.1080/10705511.2015.1049698

Bakk, Z., & Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling*, *23*, 20–31. doi:10.1080/10705511.2014.955104

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338–363. doi:10.1037/1082-989X.8.3.338

Bolsinova, M., Hoijtink, H., Vermeulen, J. A., & Béguin, A. (2017). Using expert knowledge for test linking. *Psychological Methods*, *22*, 705–724. doi:10.1037/met0000124

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*, 144–152. doi:10.1111/bmsp.1978.31.issue-2

Bray, B. C., Lanza, S. T., & Tan, X. (2015). Eliminating Bias in classify-analyze approaches for latent class analysis. *Structural Equation Modeling*, *22*, 1–11. doi:10.1080/10705511.2014.935265

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*, 121–136. doi:10.1080/15248371003699969

Darnieder, W. F. (2011). Bayesian methods for data-dependent priors (Doctoral dissertation).

Depaoli, S. (2012). Measurement and structural model class separation in mixture CFA: ML/EM Versus MCMC. *Structural Equation Modeling*, *19*, 178–203. doi:10.1080/10705511.2012.659614

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling*, *22*, 327–351. doi:10.1080/10705511.2014.937849

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, *22*, 240–261. doi:10.1037/met0000065

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and application* (Second ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Eastman, A. L., Mitchell, M. N., & Putnam-Hornstein, E. (2016). Risk of re-report: A latent class analysis of infants reported for maltreatment. *Child Abuse & Neglect*, *55*, 22–31. doi:10.1016/j.chiabu.2016.03.002

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). *Bayesian Analysis*, *1*, 515–534. doi:10.1214/06-BA117A

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

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

Hallquist, M., & Wiley, J. (2017). MplusAutomation: Automating mplus model estimation and interpretation. (Version R package version 0.7). Retrieved from https://CRAN.R-project.org/package=MplusAutomation

Hertzog, C., von Oertzen, T., Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling*, *15*, 541–563. doi:10.1080/10705510802338983

Hipwell, A. E., Stepp, S. D., Moses-Kolko, E. L., Xiong, S., Paul, E., Merrick, N., … Keenan, K. (2016). Predicting adolescent postpartum caregiving from trajectories of depression and anxiety prior to childbirth: A 5-year prospective study. *Archives of Women's Mental Health*, *19*, 871–882. doi:10.1007/s00737-016-0627-3

Holgersen, K. H., Boe, H. J., Klöckner, C. A., Weisæth, L., & Holen, A. (2010). Initial stress responses in relation to outcome after three decades. *The Journal of Nervous and Mental Disease*, *198*, 230–233. doi:10.1097/NMD.0b013e3181d106a9

Holtmann, J., Koch, T., Lochner, K., & Eid, M. (2016). A comparison of ML, WLSMV, and Bayesian methods for multilevel structural equation models in small samples: A simulation study. *Multivariate Behavioral Research*, *51*, 661–680. doi:10.1080/00273171.2016.1208074

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, *26*, 329–367. doi:10.1177/0049124198026003003

Huang, D., Brecht, M.-L., Hara, M., & Hser, Y.-I. (2010). Influences of a covariate on growth mixture modeling. *Journal of Drug Issues*, *40*, 173–194. doi:10.1177/002204261004000110

Jiang, L., Chen, S., Zhang, B., Beals, J., Mitchell, C. M., Manson, S. M., & Roubideaux, Y. (2016). Longitudinal patterns of stages of change for exercise and lifestyle intervention outcomes: An application of latent class analysis with distal outcomes. *Prevention Science*, *17*, 398–409. doi:10.1007/s11121-015-0599-y

Kaplan, D., & Depaoli, S. (2013). Bayesian statistical methods. In T. D. Little (Ed.), *Oxford handbook of quantitative methods* (pp. 407–437). Oxford, UK: Oxford University Press.

Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York, NY: The Guilford Press.

Ke, Z., & Wang, L. (2015). Detecting individual differences in change: Methods and comparisons. *Structural Equation Modeling*, *22*, 382–400. doi:10.1080/10705511.2014.936096

Koukounari, A., Stringaris, A., & Maughan, B. (2017). Pathways from maternal depression to young adult offspring depression: An exploratory longitudinal mediation analysis. *International Journal of Methods in Psychiatric Research*, *26*, e1520. doi:10.1002/mpr.1520

Kruschke, J. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). London, UK: Academic Press.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*, 722–752. doi:10.1177/1094428112457829

Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling*, *20*, 1–26. doi:10.1080/10705511.2013.742377

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Chichester, UK: John Wiley & Sons.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge, UK: Cambridge University Press.

Lee, S.-Y., & Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, *39*, 653–686. doi:10.1207/s15327906mbr3904_4

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Liu, H., Zhang, Z., & Grimm, K. J. (2016). Comparison of inverse Wishart and separation-strategy priors for Bayesian estimation of covariance parameter matrix in growth curve analysis. *Structural Equation Modeling*, *23*, 354–367. doi:10.1080/10705511.2015.1057285

Lynch, S. M. (2007). *Introduction to applied bayesian statistics and estimation for social scientists*. New York, NY: Springer Science & Business Media.

Masyn, K., Petras, H., & Lu, W. (2014). Growth curve models with categorical outcomes. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 2013–2025). New York, NY: Springer.

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, *16*, 163–200. doi:10.1007/BF01065485

McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analysis in contemporary psychological research. In J. A. Schinka, W. F. Velicer, & I. B. Weiner (Eds.), *Handbook of psychology: Volume 2 research methods in psychology* (pp. 447–480). Hoboken, NJ: Wiley.

McArdle, J. J. (2012). Latent curve modeling of longitudinal growth data. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 547–571). New York, NY: The Guilford Press.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110–133. doi:10.2307/1130295

McNeish, D. (2016a). On using Bayesian methods to address small sample problems. *Structural Equation Modeling*, *23*, 750–773. doi:10.1080/10705511.2016.1186549

McNeish, D. (2016b). Using data-dependent priors to mitigate small sample bias in latent growth models a discussion and illustration using mplus. *Journal of Educational and Behavioral Statistics*, *41*, 27–56. doi:10.3102/1076998615621299

McNeish, D. (2017). Brief research report: Growth models with small samples and missing data. *The Journal of Experimental Education*, 1–12. doi:10.1080/00220973.2017.1369384

Meng, X.-L., & Rubin, D. B. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, *80*, 267–278. doi:10.2307/2337198

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122. doi:10.1007/BF02294746

Miočević, M., MacKinnon, D. P., & Levy, R. (2017). Power in Bayesian mediation analysis for small sample research. *Structural Equation Modeling*, *24*, 666–683. doi:10.1080/10705511.2017.1312407

Muthén, L. K., & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*, 599–620. doi:10.1207/S15328007SEM0904_8

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's guide* (Eighth ed.). Los Angeles, CA: Muthén & Muthén.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100. doi:10.1016/S0022-2496(02)00028-7

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., … Rakow, T. (2006). *Uncertain judgements: eliciting experts' probabilities*. West-Sussex, England: Wiley.

Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 69–100). New York, NY: Springer. doi:10.1007/978-0-387-77650-7_5

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

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*, 185–206. doi:10.1080/00273171.2015.1065398

Shi, D., & Tong, X. (2017). The impact of prior information on Bayesian latent basis growth model estimation. *SAGE Open*, *7*. doi:10.1177/2158244017727039

Shin, T., Davison, M. L., & Long, J. D. (2017). Maximum likelihood versus multiple imputation for missing data in small longitudinal samples with nonnormality. *Psychological Methods*, *22*, 426–449. doi:10.1037/met0000094

Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2019). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Structural Equation Modeling*. doi:10.1080/10705511.2019.1577140

Spoth, R., Clair, S., & Trudeau, L. (2014). Universal family-focused intervention with young adolescents: Effects on health-risking sexual behaviors and STDs among young adults. *Prevention Science*, *15*, 47–58. doi:10.1007/s11121-012-0321-2

Stoel, R. D., van den Wittenboer, G., & Hox, J. J. (2004). Methodological issues in the application of the latent growth curve model. In van Montfort, K., Oud, J., & Satorra, A. (Eds.), *Recent developments on structural equation models: Theory and applications* (pp. 241–261). Dordrecht, NL: Springer.

Tong, X., & Ke, Z. (2016). Growth curve modeling for nonnormal data: A two-stage robust approach versus a semiparametric Bayesian approach. *Quantitative Psychology Research*, *167*, 229–241. doi:10.1007/978-3-319-38759-8_17

Tsai, M.-Y., & Hsiao, C. K. (2008). Computation of reference Bayesian inference for variance components in longitudinal studies. *Computational Statistics*, *23*, 587–604. doi:10.1007/s00180-007-0100-x

van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology*, *6*, 1. doi:10.3402/ejpt.v6.25216

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist*, *16*, 75–84.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*, 842–860. doi:10.1111/cdev.12169

van de Schoot, R., Sijbrandij, M., Depaoli, S., Winter, S. D., Olff, M., & van Loey, N. E. (2018). Bayesian PTSD-Trajectory analysis with informed priors based on a systematic literature search and expert elicitation. *Multivariate Behavioral Research*, *53*, 267–291. doi:10.1080/00273171.2017.1412293

van de Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2017). The GRoLTS-Checklist: Guidelines for reporting on latent trajectory studies. *Structural Equation Modeling*, *24*, 451-467. doi:10.1080/10705511.2016.1247646

van Erp, S. J., Mulder, J., & Oberski, D. L. (2018). Prior sensitivity analysis in default Bayesian structural equation modeling. *Psychological Methods*, *23*, 363–388. doi:10.1037/met0000162

Veen, D., Stoel, D., Zondervan-Zwijnenburg, M., & van de Schoot, R. (2017). Proposal for a five-step method to elicit expert judgment. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.02110

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*, 450–469. doi:10.1093/pan/mpq025

Wijnen, F., de Bree, E., van Alphen, P., de Jong, J., & van der Leij, A. (2015). Comparing SLI and dyslexia: Developmental language profiles and reading outcomes. In S. Stavrakaki (Ed.), *Specific Language Impairment: Current trends in research* (pp. 89–112). Amsterdam, NL: John Benjamins Publishing Company.

Ye, F. (2016). Latent growth curve analysis with dichotomous items: Comparing four approaches. *British Journal of Mathematical and Statistical Psychology*, *69*, 43–61. doi:10.1111/bmsp.12058

Zhang, Z., Hamagami, F., Wang, L., Nesselroade, J. R., & Grimm, K. J. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, *31*, 374–383. doi:10.1177/0165025407077764

Zondervan-Zwijnenburg, M., Depaoli, S., Peeters, M., & van de Schoot, R. (2018). Pushing the Limits: The performance of maximum likelihood and bayesian estimation with small and unbalanced samples in a latent growth model. *Methodology*, *1*, 1–13. doi:10.1027/1614-2241/a000162

Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S., & van de Schoot, R. (2017). Where do priors come from? applying guidelines to construct informative priors in small sample research. *Research in Human Development*, *14*, 305–320. doi:10.1080/15427609.2017.1370966

APPENDIX A1. MODEL MATRICES AND A DESCRIPTION OF THE PARAMETERS IN THE INVESTIGATED LGM WITH A DISTAL OUTCOME

Measurement model matrices:

$$\boldsymbol{\nu} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\Theta} = \begin{bmatrix} \theta_{x1} & & & & \\ 0 & \theta_{x2} & & & \\ 0 & 0 & \theta_{x3} & & \\ 0 & 0 & 0 & \theta_{x4} & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This leads to the measurement model, as given in Equation A1:

$$\boldsymbol{y}_{it} = \boldsymbol{\eta}_{\boldsymbol{I}i} + \boldsymbol{\eta}_{\boldsymbol{S}i}\boldsymbol{\lambda}_t + \boldsymbol{\varepsilon}_{it}. \tag{A1}$$

Structural model matrices:

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{I0} \\ \alpha_{S0} \\ \alpha_{D0} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_1 & \beta_2 & 0 \end{bmatrix}$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_I & & \\ \psi_S & \psi_{I-S} & \\ 0 & 0 & \psi_D \end{bmatrix}$$

This leads to the structural model, as shown in Equations A2 and A3:

$$\boldsymbol{\eta}_{\boldsymbol{I}i} = a_{I0} + \boldsymbol{\xi}_{\boldsymbol{I}i,} \tag{A2}$$

$$\boldsymbol{\eta}_{\boldsymbol{S}i} = a_{S0} + \boldsymbol{\xi}_{\boldsymbol{S}i,}$$

$$\boldsymbol{\eta}_{\boldsymbol{D}i} = a_{D0} + \ \beta_1\boldsymbol{\eta}_{\boldsymbol{I}i} + \ \beta_2\boldsymbol{\eta}_{\boldsymbol{S}i} + \boldsymbol{\xi}_{\boldsymbol{D}i,} \tag{A3}$$

where
$\boldsymbol{y}_{it}$ = observed outcome y for person *i* (i = 1, …, n) at time *t* (in simulation design: 0, 1, 2, 3).
$\boldsymbol{\eta}_{\boldsymbol{I}i}$ = random intercept factor: the expected outcome on y for person *i* at time score $\boldsymbol{\lambda}_t = 0$.
$\boldsymbol{\eta}_{\boldsymbol{S}i}$ = random linear slope factor: the expected outcome on y for person *i* for 1 unit increase in time, on the scale of $\boldsymbol{\lambda}_t$.
$\boldsymbol{\lambda}_t$ = time score at time *t*: 0, 1, 2, 3.
$\boldsymbol{\varepsilon}_{it}$ = represent individual and identically distributed measurement and time-specific errors on $\boldsymbol{y}_{it}$ for person *i* at time *t*, and dε_{it} are usually assumed to be uncorrelated over time.

$\boldsymbol{\eta}_{\boldsymbol{I}i}$ = random intercept factor: the expected outcome on y (here measured by x1-x4) for person *i* at time score $\boldsymbol{\lambda}_t = 0$.
$\boldsymbol{\eta}_{\boldsymbol{S}i}$ = random linear slope factor: the expected outcome on y (here measured by x1-x4) for person *i* for one unit increase in time, on the scale of $\boldsymbol{\lambda}_t$.

$\boldsymbol{\eta}_{Di}$ = random distal outcome factor: the expected outcome on d (here measured by the distal outcome variable) for person *i*, when taking the predictions of the latent intercept and latent slope into account.

a_{I0} = population mean of individual intercept factor values.
a_{S0} = population mean of individual slope factor values.
a_{D0} = population mean of individual distal outcome variable values when $\boldsymbol{\eta}_{Ii}$ and $\boldsymbol{\eta}_{Si}$ are zero, that is, the intercept of distal outcome variable.

$\boldsymbol{\xi}_{Ii}$ = deviation of $\boldsymbol{\eta}_{Ii}$ from α_{I0}
$\boldsymbol{\xi}_{Si}$ = deviation of $\boldsymbol{\eta}_{Si}$ from α_{S0}
$\boldsymbol{\xi}_{Di}$ = deviation of $\boldsymbol{\eta}_{Di}$ from α_{D0}
β_1 = difference in the mean of the distal outcome factor corresponding to the one unit difference in the latent intercept factor; regression coefficient; distal outcome is regressed on latent intercept.
β_2 = difference in the mean of the distal outcome factor corresponding to the one unit difference in the latent slope factor; regression coefficient; distal outcome is regressed on latent slope.

Equations and interpretation are based on Masyn, Petras, and Lu (2014) and Duncan, Duncan and Strycker (2006, pp. 56–62).

APPENDIX A2. M*PLUS* DEFAULT PRIORS (MUTHÉN & MUTHÉN, 2017) AS USED IN THE BAYESDEFAULT SETTING

- Mean latent intercept, mean latent slope, and intercept distal outcome: N (0, 10^{10})
- Regression coefficients: N (0, 10^{10})
- Variances Intercept, Slope, and Covariance Intercept-Slope: IW (0, −3)
- Variance Distal outcome: IG (−1, 0)
- Residual variances: IG (−1, 0)

APPENDIX A3. OVERVIEW OF THE WEAK, MEDIUM AND STRONG INFORMATIVE AND DEVIATING PRIOR DISTRIBUTIONS USED IN THE SIMULATION STUDY

		Small effect size							
Parameter	*Sample Size*	*Population Value*	*Posterior SD*	*Informative Weak*	*Informative Medium*	*Informative Strong*	*Deviating Weak*	*Deviating Medium*	*Deviating Strong*
Small slope variance									
β_1	26	0.1	0.1464	N(0.1,0.032)	N(0.1,0.021)	N(0.1,0.011)	N(−0.603,0.032)	N(−0.474,0.021)	N(−0.306,0.011)
β_2	26	0.32	0.4728	N(0.32,0.335)	N(0.32,0.224)	N(0.32,0.112)	N(−1.95,0.335)	N(−1.533,0.224)	N(−0.991,0.112)
α_D	26	0.5	0.222	N(0.5,0.074)	N(0.5,0.049)	N(0.5,0.025)	N(−0.566,0.074)	N(−0.37,0.049)	N(−0.115,0.025)
α_I	26	0.1	0.2645	N(0.1,0.105)	N(0.1,0.07)	N(0.1,0.035)	N(−1.17,0.105)	N(−0.937,0.07)	N(−0.633,0.035)
α_S	26	0.4	0.1241	N(0.4,0.023)	N(0.4,0.015)	N(0.4,0.008)	N(−0.196,0.023)	N(−0.086,0.015)	N(0.056,0.008)
β_1	52	0.1	0.1205	N(0.1,0.022)	N(0.1,0.015)	N(0.1,0.007)	N(−0.479,0.022)	N(−0.372,0.015)	N(−0.234,0.007)
β_2	52	0.32	0.5101	N(0.32,0.39)	N(0.32,0.26)	N(0.32,0.13)	N(−2.129,0.39)	N(−1.68,0.26)	N(−1.094,0.13)
α_D	52	0.5	0.2149	N(0.5,0.069)	N(0.5,0.046)	N(0.5,0.023)	N(−0.532,0.069)	N(−0.342,0.046)	N(−0.096,0.023)
α_I	52	0.1	0.1856	N(0.1,0.052)	N(0.1,0.034)	N(0.1,0.017)	N(−0.791,0.052)	N(−0.628,0.034)	N(−0.414,0.017)
α_S	52	0.4	0.0878	N(0.4,0.012)	N(0.4,0.008)	N(0.4,0.004)	N(−0.022,0.012)	N(0.056,0.008)	N(0.157,0.004)
β_1	325	0.1	0.0679	N(0.1,0.007)	N(0.1,0.005)	N(0.1,0.002)	N(−0.226,0.007)	N(−0.166,0.005)	N(−0.088,0.002)
β_2	325	0.32	0.4086	N(0.32,0.25)	N(0.32,0.167)	N(0.32,0.083)	N(−1.642,0.25)	N(−1.282,0.167)	N(−0.813,0.083)
α_D	325	0.5	0.1629	N(0.5,0.04)	N(0.5,0.027)	N(0.5,0.013)	N(−0.282,0.04)	N(−0.139,0.027)	N(0.048,0.013)
α_I	325	0.1	0.0717	N(0.1,0.008)	N(0.1,0.005)	N(0.1,0.003)	N(−0.244,0.008)	N(−0.181,0.005)	N(−0.099,0.003)
α_S	325	0.4	0.0347	N(0.4,0.002)	N(0.4,0.001)	N(0.4,0.001)	N(0.233,0.002)	N(0.264,0.001)	N(0.304,0.001)
Large slope variance									
β_1	26	0.1	0.18	N(0.1,0.049)	N(0.1,0.032)	N(0.1,0.016)	N(−0.764,0.049)	N(−0.606,0.032)	N(−0.399,0.016)
β_2	26	0.1	0.1502	N(0.1,0.034)	N(0.1,0.023)	N(0.1,0.011)	N(−0.621,0.034)	N(−0.489,0.023)	N(−0.316,0.011)
α_D	26	0.5	0.1293	N(0.5,0.025)	N(0.5,0.017)	N(0.5,0.008)	N(−0.121,0.025)	N(−0.007,0.017)	N(0.142,0.008)
α_I	26	0.1	0.2639	N(0.1,0.104)	N(0.1,0.07)	N(0.1,0.035)	N(−1.167,0.104)	N(−0.934,0.07)	N(−0.631,0.035)
α_S	26	0.4	0.2224	N(0.4,0.074)	N(0.4,0.049)	N(0.4,0.025)	N(−0.668,0.074)	N(−0.472,0.049)	N(−0.216,0.025)
β_1	52	0.1	0.1588	N(0.1,0.038)	N(0.1,0.025)	N(0.1,0.013)	N(−0.662,0.038)	N(−0.522,0.025)	N(−0.34,0.013)
β_2	52	0.1	0.1113	N(0.1,0.019)	N(0.1,0.012)	N(0.1,0.006)	N(−0.434,0.019)	N(−0.336,0.012)	N(−0.209,0.006)
α_D	52	0.5	0.0851	N(0.5,0.011)	N(0.5,0.007)	N(0.5,0.004)	N(0.091,0.011)	N(0.166,0.007)	N(0.264,0.004)
α_I	52	0.1	0.1861	N(0.1,0.052)	N(0.1,0.035)	N(0.1,0.017)	N(−0.793,0.052)	N(−0.629,0.035)	N(−0.416,0.017)
α_S	52	0.4	0.1583	N(0.4,0.038)	N(0.4,0.025)	N(0.4,0.013)	N(−0.36,0.038)	N(−0.221,0.025)	N(−0.039,0.013)
β_1	325	0.1	0.0406	N(0.1,0.002)	N(0.1,0.002)	N(0.1,0.001)	N(−0.095,0.002)	N(−0.059,0.002)	N(−0.013,0.001)
β_2	325	0.1	0.0334	N(0.1,0.002)	N(0.1,0.001)	N(0.1,0.001)	N(−0.06,0.002)	N(−0.031,0.001)	N(0.007,0.001)
α_D	325	0.5	0.0317	N(0.5,0.002)	N(0.5,0.001)	N(0.5,0.001)	N(0.348,0.002)	N(0.376,0.001)	N(0.412,0.001)
α_I	325	0.1	0.0714	N(0.1,0.008)	N(0.1,0.005)	N(0.1,0.003)	N(−0.243,0.008)	N(−0.18,0.005)	N(−0.098,0.003)
α_S	325	0.4	0.0628	N(0.4,0.006)	N(0.4,0.004)	N(0.4,0.002)	N(0.099,0.006)	N(0.154,0.004)	N(0.226,0.002)

		Large effect size							
Parameter	*Sample Size*	*Population Value*	*Posterior SD*	*Informative Weak*	*Informative Medium*	*Informative Strong*	*Deviating Weak*	*Deviating Medium*	*Deviating Strong*
Small slope variance									
β_1	26	0.4	0.1682	N(0.4,0.042)	N(0.4,0.028)	N(0.4,0.014)	N(−0.408,0.042)	N(−0.259,0.028)	N(−0.066,0.014)
β_2	26	1.27	0.465	N(1.27,0.324)	N(1.27,0.216)	N(1.27,0.108)	N(−0.962,0.324)	N(−0.553,0.216)	N(−0.019,0.108)
α_D	26	0.5	0.2314	N(0.5,0.08)	N(0.5,0.054)	N(0.5,0.027)	N(−0.611,0.08)	N(−0.407,0.054)	N(−0.141,0.027)
α_I	26	0.1	0.2646	N(0.1,0.105)	N(0.1,0.07)	N(0.1,0.035)	N(−1.17,0.105)	N(−0.937,0.07)	N(−0.633,0.035)
α_S	26	0.4	0.124	N(0.4,0.023)	N(0.4,0.015)	N(0.4,0.008)	N(−0.195,0.023)	N(−0.086,0.015)	N(0.056,0.008)
β_1	52	0.4	0.1424	N(0.4,0.03)	N(0.4,0.02)	N(0.4,0.01)	N(−0.284,0.03)	N(−0.158,0.02)	N(0.005,0.01)
β_2	52	1.27	0.4847	N(1.27,0.352)	N(1.27,0.235)	N(1.27,0.117)	N(−1.057,0.352)	N(−0.63,0.235)	N(−0.073,0.117)
α_D	52	0.5	0.2101	N(0.5,0.066)	N(0.5,0.044)	N(0.5,0.022)	N(−0.509,0.066)	N(−0.324,0.044)	N(−0.082,0.022)
α_I	52	0.1	0.1855	N(0.1,0.052)	N(0.1,0.034)	N(0.1,0.017)	N(−0.791,0.052)	N(−0.627,0.034)	N(−0.414,0.017)
α_S	52	0.4	0.0878	N(0.4,0.012)	N(0.4,0.008)	N(0.4,0.004)	N(−0.022,0.012)	N(0.056,0.008)	N(0.157,0.004)
β_1	325	0.4	0.078	N(0.4,0.009)	N(0.4,0.006)	N(0.4,0.003)	N(0.026,0.009)	N(0.094,0.006)	N(0.184,0.003)
β_2	325	1.27	0.4344	N(1.27,0.283)	N(1.27,0.189)	N(1.27,0.094)	N(−0.816,0.283)	N(−0.433,0.189)	N(0.066,0.094)
α_D	325	0.5	0.1759	N(0.5,0.046)	N(0.5,0.031)	N(0.5,0.015)	N(−0.344,0.046)	N(−0.19,0.031)	N(0.012,0.015)
α_I	325	0.1	0.0717	N(0.1,0.008)	N(0.1,0.005)	N(0.1,0.003)	N(−0.244,0.008)	N(−0.181,0.005)	N(−0.099,0.003)
α_S	325	0.4	0.0346	N(0.4,0.002)	N(0.4,0.001)	N(0.4,0.001)	N(0.234,0.002)	N(0.264,0.001)	N(0.304,0.001)
Large slope variance									
β_1	26	0.4	0.1765	N(0.4,0.047)	N(0.4,0.031)	N(0.4,0.016)	N(−0.447,0.047)	N(−0.292,0.031)	N(−0.089,0.016)
β_2	26	0.4	0.157	N(0.4,0.037)	N(0.4,0.025)	N(0.4,0.012)	N(−0.354,0.037)	N(−0.215,0.025)	N(−0.035,0.012)
α_D	26	0.5	0.1394	N(0.5,0.029)	N(0.5,0.019)	N(0.5,0.01)	N(−0.169,0.029)	N(−0.046,0.019)	N(0.114,0.01)
α_I	26	0.1	0.2639	N(0.1,0.104)	N(0.1,0.07)	N(0.1,0.035)	N(−1.167,0.104)	N(−0.934,0.07)	N(−0.631,0.035)
α_S	26	0.4	0.2223	N(0.4,0.074)	N(0.4,0.049)	N(0.4,0.025)	N(−0.667,0.074)	N(−0.471,0.049)	N(−0.216,0.025)
β_1	52	0.4	0.1596	N(0.4,0.038)	N(0.4,0.025)	N(0.4,0.013)	N(−0.366,0.038)	N(−0.226,0.025)	N(−0.042,0.013)
β_2	52	0.4	0.1173	N(0.4,0.021)	N(0.4,0.014)	N(0.4,0.007)	N(−0.163,0.021)	N(−0.06,0.014)	N(0.075,0.007)
α_D	52	0.5	0.0946	N(0.5,0.013)	N(0.5,0.009)	N(0.5,0.004)	N(0.046,0.013)	N(0.129,0.009)	N(0.238,0.004)
α_I	52	0.1	0.1858	N(0.1,0.052)	N(0.1,0.035)	N(0.1,0.017)	N(−0.792,0.052)	N(−0.628,0.035)	N(−0.415,0.017)
α_S	52	0.4	0.1582	N(0.4,0.038)	N(0.4,0.025)	N(0.4,0.013)	N(−0.36,0.038)	N(−0.22,0.025)	N(−0.038,0.013)
β_1	325	0.4	0.0526	N(0.4,0.004)	N(0.4,0.003)	N(0.4,0.001)	N(0.147,0.004)	N(0.194,0.003)	N(0.254,0.001)
β_2	325	0.4	0.0391	N(0.4,0.002)	N(0.4,0.002)	N(0.4,0.001)	N(0.212,0.002)	N(0.247,0.002)	N(0.292,0.001)
α_D	325	0.5	0.0351	N(0.5,0.002)	N(0.5,0.001)	N(0.5,0.001)	N(0.331,0.002)	N(0.362,0.001)	N(0.403,0.001)
α_I	325	0.1	0.0714	N(0.1,0.008)	N(0.1,0.005)	N(0.1,0.003)	N(−0.243,0.008)	N(−0.18,0.005)	N(−0.098,0.003)
α_S	325	0.4	0.0628	N(0.4,0.006)	N(0.4,0.004)	N(0.4,0.002)	N(0.099,0.006)	N(0.154,0.004)	N(0.226,0.002)

Note. Parameters β_1 and β_2 refer to the two regression coefficients, α_D refers to the intercept of the distal outcome, and α_I and α_S refer to the means of the latent intercept and latent linear slope (see also Figure 1).