Structural Equation Modeling, 15:541–563, 2008

ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705510802338983

Psychology Press
Taylor & Francis Group

Evaluating the Power of Latent Growth Curve Models to Detect Individual Differences in Change

Christopher Hertzog
School of Psychology
Georgia Institute of Technology

Timo von Oertzen
Center for Lifespan Psychology
Max Planck Institute for Human Development and
Department of Mathematics
Saarland University

Paolo Ghisletta
Center for Interdisciplinary Gerontology and Faculty of Psychology and
Educational Sciences
University of Geneva

Ulman Lindenberger
Center for Lifespan Psychology
Max Planck Institute for Human Development and
School of Psychology
Saarland University

We evaluated the statistical power of single-indicator latent growth curve models to detect individual differences in change (variances of latent slopes) as a function of sample size, number of longitudinal measurement occasions, and growth curve

Correspondence should be addressed to Christopher Hertzog, School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170. E-mail: christopher.hertzog@psych.gatech.edu

reliability. We recommend the 2 degree-of-freedom generalized test assessing loss of fit when both slope-related random effects, the slope variance and intercept-slope covariance, are fixed to 0. Statistical power to detect individual differences in change is low to moderate unless the residual error variance is low, sample size is large, and there are more than four measurement occasions. The generalized test has greater power than a specific test isolating the hypothesis of zero slope variance, except when the true slope variance is close to 0, and has uniformly superior power to a Wald test based on the estimated slope variance.

A major goal of longitudinal research is the direct identification of interindividual differences in intraindividual change (Baltes & Nesselroade, 1979; Hertzog & Nesselroade, 2003; Wohlwill, 1991). Latent growth curve models (LGCMs) expand on traditional repeated-measures analysis of variance by allowing one to simultaneously model change in the means (fixed effects) and in the variance and covariance of initial level and change (random effects; Bryk & Raudenbush, 2001; Duncan, Duncan, Strycker, Li, & Alpert, 1999; Laird & Ware, 1982; McArdle & Epstein, 1987; Raykov, 1993; Rogosa & Willett, 1985; J. D. Singer & Willett, 2003). Individual differences in rates of change are manifested as reliable random effects in LGCM slopes.

There have been relatively few studies of the properties of LGCM significance tests to detect individual differences in change, and the available evidence is limited in scope (e.g., Pinheiro & Bates, 2000). Simulation studies of LGCMs have typically focused on other questions, such as detecting mean slope differences in multiple groups (Fan, 2003; Kim, 2005; Muthén & Curran, 1997). Aware of the lack of criteria describing the limitations of LGCMs in assessing random effects, we (Hertzog, Lindenberger, Ghisletta, & Oertzen, 2006) recently evaluated the statistical power of bivariate LGCM to detect correlations of slopes between two variables. Our simulation, which used the Satorra and Saris (1985) approximation method, indicated that the power to detect slope covariances was relatively low under a number of conditions, especially when the growth curve reliability (GCR) was less than .90.

The present Monte Carlo simulation evaluates different methods of testing for reliable slope variance in univariate LGCMs. We explicitly evaluate two likelihood ratio tests, a 1 *df* specific variance test and a 2 *df* generalized variance test. We also evaluate the behavior of the standard Wald test (the estimated slope variance divided by its standard error of the estimate).

STATISTICAL MODEL

The simulation study is based on a simple linear LGCM for one variable, y, measured longitudinally over time, $t = 0, \ldots, T$ on $i = 1, \ldots, N$ persons,

generating a data matrix $\mathbf{y_{it}}$. The growth curve model can be written as:

$$\mathbf{y_{it}} = \beta_1 \text{InterceptY}_\text{i} + \beta_{2,t} \text{SlopeY}_\text{i} + \boldsymbol{\varepsilon}_{\mathbf{Yit}}, \tag{1}$$

where $\text{InterceptY}_\text{i}$ and SlopeY_i are latent variables defining the individual intercepts and slopes of the latent growth curve. A growth curve design matrix, $\mathbf{B}$, can be defined to have T rows, one for each occasion, and two columns, β_1 for the intercept and $\beta_{2,t}$ for the slope. Each column of $\mathbf{B}$ is a vector of regression weights establishing the relation of occasions of measurement, t, to the growth curve. For the intercepts in column 1, all values are fixed at 1. Identification of the intercept at origin $t(0)$ is accomplished by fixing the regression weight for the slope parameter at $t(0)$ to 0. For the purposes of this article, the remaining weights for the slope are fixed and define a linear growth curve for each individual on y (see McArdle & Epstein, 1987, and Rovine & Molenaar, 2000, for discussion and alternative scalings of the growth curves).

The model can be used to derive expectations (across individuals i at times t) for $\mathbf{y_{it}}$.

$$\begin{aligned} &\text{Defining E } (\mathbf{y_{it}}) = \mathbf{M_{yt}}, \\ &\mathbf{M_{yt}} = M_{\text{InterceptY}} + \beta_{2,t} M_{\text{SlopeY}}, \end{aligned} \tag{2}$$

where $\beta_{2,t}$ is the corresponding element from the second column of $\mathbf{B}$ (the weights that define the slope of the growth curve), $M_{\text{InterceptY}}$ is the mean population intercept, and M_{SlopeY} is the mean population slope. These two parameters are commonly referred to as fixed effects in multilevel models for LGCM.

Consistent with common practice in empirical applications of LGCM, we assume that the errors, $\boldsymbol{\varepsilon}_{\mathbf{Yit}}$, are distributed normally and are stochastically independent of the latent intercepts and slopes, as well as independent of each other. We also assume the errors have homogeneous variance across occasions, denoted $\sigma^2_{\varepsilon y}(\boldsymbol{\varepsilon}_{\mathbf{Yit}} \approx N(0, \sigma^2_{\varepsilon y}))$. Hence the expectation for the covariance matrix of the observed variables, $\boldsymbol{\Sigma}_\mathbf{y}$, aggregating over individuals, is

$$\boldsymbol{\Sigma}_\mathbf{y} = \mathbf{B}\boldsymbol{\Psi}_\mathbf{IS}\mathbf{B}' + \boldsymbol{\Theta}_\mathbf{y}, \tag{3}$$

where $\boldsymbol{\Psi}_\mathbf{IS}$ is the covariance matrix of the two latent growth curve parameters, [InterceptY, SlopeY] and $\boldsymbol{\Theta}_\mathbf{y}$ is a $T \times T$ diagonal matrix containing the homogenous error variances, $\sigma^2_{\varepsilon y}$, on its diagonal. The parameters in $\boldsymbol{\Psi}_\mathbf{IS}$ and $\boldsymbol{\Theta}_\mathbf{y}$ are termed random effects in multilevel models.

LONGITUDINAL DESIGN MODEL

We used the same longitudinal design model employed by Hertzog et al. (2006), namely, a prospective single-cohort longitudinal design (Baltes, Reese, & Nesselroade, 1988; Schaie, 1977). We assume that a population of adults 50 years of age have been measured on a variable generating linear age decline. We assume simple random sampling from the population, followed by up to 19 longitudinal retest observations on individuals at 2-year intervals.

Statistical power in any context depends on sample size, effect size, and the chosen Type I error rate (Cohen, 1988). Our focus is on detecting slope variance, so we varied the relevant slope-related parameters in $\mathbf{\Psi_{IS}}$ while holding the variance of intercepts constant. We also systematically varied the sample size and the number of longitudinal occasions that had been sampled to address a fundamental longitudinal design question: How long must one measure samples of a given size to have adequate power for detecting variance in change?

It is well known that random measurement error (unreliability) affects statistical power, including tests of means, variances, correlations, and factor loadings (e.g., Bollen, 1989; MacCallum, Widaman, Zhang, & Hong, 1999; Marcoulides, 1996). We expected that GCR would influence the power to detect individual differences in change and correlations in change components. Hertzog et al. (2006) demonstrated a substantial effect of GCR on power to detect nonzero slope covariances in bivariate LGCM. GCR, that is, $[\sigma^2_{yt} - \sigma^2_{\varepsilon y}]/\sigma^2_{yt}$, is defined as the variance determined by the latent growth curve at each time t, divided by the total variance of y. Error variances are assumed to be homogenous over time, but the variance of y varies over t, because it carries the effects of individual differences in slopes that vary with t. We scale the GCR in our simulation design at time $t(0)$, study onset. The GCR has two components: random measurement error in each variable and variability of the residuals for the true scores of y around the linear regression functions of the latent growth curve. Our simulation systematically varied the residual variance, $\sigma^2_{\varepsilon y}$, as the method of varying GCR.

Our LGCMs were scaled in proportional metric (McArdle, 1988; see Rovine & Molenaar, 2000) so that change is expressed as a function of proportion of growth from beginning to end of the longitudinal study. In terms of the latent growth curve model, $\beta_{2,0} = 0$ and $\beta_{2,19} = 0.95$, with $\beta(t)$ increasing by .05 units for each unit increase in t. The temporal units can be treated as representing years, so that can we conceive of the simulation as representing change from mean age 50 to mean age 69. However, the results we report here can be translated into any temporal scale of measurement, as well as for variables that grow instead of decline. Hence the results are generally applicable to any univariate linear LGCM.

SIGNIFICANCE TESTING

In practice, one evaluates the hypothesis of individual differences in change by testing whether the slope variance is reliably greater than zero. Three different standard methods of testing this null hypothesis exist. Often, researchers use a Wald test, in which the estimated variance is divided by its estimated standard error to produce a statistic that is asymptotically a normal deviate (e.g., Bollen, 1989), and can be evaluated against a critical value at a specified Type I error criterion. However, a likelihood ratio (LR) χ^2 test is generally considered superior to the Wald test (e.g., Gonzalez & Griffin, 2001), especially for small samples (Pinheiro & Bates, 2000; Raudenbush & Bryk, 2002).[1] Two different LR tests for the hypothesis of zero slope variance have been illustrated in treatments of LGCM, a *specific* variance test and a *generalized* variance test. The specific test involves isolating the restriction of zero slope variance. The generalized test is based on the argument that any slope-related element of the covariance matrix of the latent variables, $\mathbf{\Psi_{IS}}$, carries information about individual differences in slopes. Hence, the generalized test evaluates whether all slope-related elements of $\mathbf{\Psi_{IS}}$ are equal to zero.

The elements of the covariance matrix of the latent variables, $\mathbf{\Psi_{IS}}$, are the intercept variance, σ_I^2, the covariance of intercept and slope, σ_{IS}, and the variance of slopes, σ_S^2. The specific variance test evaluates the null hypothesis:

$$H_0\text{: } \sigma_S^2 = 0 \text{ with 1 } df;$$

the generalized variance test instead tests the following null hypothesis:

$$H_0\text{: } \sigma_S^2 = 0 \text{ and } \sigma_{IS} = 0 \text{ with 2 } df.$$

Hypothesis testing is accomplished by computing LR tests based on nested models that isolate the restrictions of interest in each null hypothesis. The LR χ^2 test statistic generated from model comparisons on the data is evaluated against a standard central χ^2 distribution, adopting some Type I error criterion. Here we assume that the Type I error criterion is set at .05. There are three nested models of interest here. M1 is a model with freely estimated mean intercept and slope and with all three parameters in Ψ_{IS} freely estimated. M2 is a model that restricts $\sigma_{IS} = 0$, and M3 is a model that restricts both $\sigma_S^2 = 0$ and $\sigma_{IS} = 0$. Then the difference in fit (-2LL) between M3 and M2 is a 1 *df* LR test of the specific hypothesis that $\sigma_S^2 = 0$, whereas the difference in fit between M3

[1]Our LR tests use full information maximum likelihood estimation. In small samples, restricted maximum likelihood estimation may yield better accuracy in LR tests (e.g., Verbeke & Molenberghs, 2000), but we do not consider that estimation approach in this article.

and M1 is a 2 *df* LR test of the generalized hypothesis that both slope-related random components are zero. The generalized test uses the information in the covariance of intercept and slope, but ignores the information carried in σ_{I}^2 because it is determined by stable individual differences and not by variance in change. Although not all treatments of slope variance tests adopt this approach, the generalized test is common in the mixed model literature (e.g., Pinheiro & Bates, 2000), if only because specifying random effects for the LGCM slopes and intercepts automatically generates a random effects covariance matrix involving all three elements in some software packages (e.g., SAS PROC MIXED; Littell Milliken, Stroup, & Wolfinger, 2006; see Verbeke & Molenberghs, 2000). Then, computing differences in −2LL between the model with random slopes and intercepts and a model with random effects for intercepts alone produces the generalized test. On the other hand, the specific test is possible in structural equation modeling (SEM) approaches to LGCM or in more flexible model specifications using mixed model software. Because we believe the generalized test to be the preferred test for slope variances, we emphasize its behavior in this article, but also explicitly compare it with the specific variance test, especially when varying the σ_{IS} parameter.

One reason for considering the generalized over the specific variance test is that the distinction between the slope variance and the intercept-slope covariance, although conceptually meaningful, is to an extent statistically arbitrary (Rovine & Molenaar, 1998). Any choice of occasion for fixing a 0 basis element of $\boldsymbol{\beta}_{2,\mathbf{t}}$ to identify the solution will redefine the intercept and slope without a loss of fit to the model, but with a major change in the estimated intercept and slope parameters and their interpretation. Given this scale equivalence or arbitrariness, both slope variances and slope covariances can carry information about variance in change. Moreover, onset of a study at $t(0)$ does not conform to onset of the true growth process (as in the present example, where maturity prior to age-related decline is achieved much earlier), arguing that the intercept carries growth-related variance.

It is instructive to examine the expected values for the variances of y as a function of LGCM parameters. At study onset, the variance of y at $t(0)$ is composed of level variance and error variance; that is, $\sigma_{y0}^2 = \sigma_{\mathrm{I}}^2 + \sigma_{\varepsilon y}^2$. These two components are assumed invariant with respect to t. In LGCMs, changes in variance over time are assumed to derive from σ_{S}^2 and σ_{IS}. At any given time t,

$$\sigma_{yt}^2 = \sigma_{\mathrm{I}}^2 + \sigma_{\varepsilon y}^2 + (t/T)^2\sigma_{\mathrm{S}}^2 + 2(t/T)\sigma_{\mathrm{IS}}. \tag{4}$$

Likewise, the covariance of y at any two points in time t and t' is

$$\sigma_{yt,yt'} = \sigma_{\mathrm{I}}^2 + 2[(tt')/T]\sigma_{\mathrm{IS}} + 2[tt'/T]\sigma_{\mathrm{S}}^2. \tag{5}$$

The model predicts a quadratic expansion of the variance of σ^2_{yt} over time constrained by the covariance of intercepts and slopes. Egression from the mean will result for individuals if the covariance of initial level and change is nonnegative (Nesselroade, Stigler, & Baltes, 1980; Raykov, 1993), creating a fan-spread pattern of change. Conversely, a negative σ_{IS} will tend to reduce the increase in observed variance. Note that in Model M2, specifying $\sigma_{IS} = 0$, the expectations are simplified by dropping the covariance terms. For Model M3, the expected values for σ^2_{yt} and $\sigma_{yt,yt'}$ for all t reduce to $\sigma^2_{I} + \sigma^2_{\varepsilon y}$ and σ^2_{I}, respectively. Hence the power to detect nonzero slope variance for both the specific and generalized tests depends on the degree to which the associated LR test can detect a loss of fit to all sample variances and covariances by the restricted model in M3 that constrains all variances and covariances of y to be stationary over time, versus the unrestricted LGCM models that do not.

STATISTICAL POWER

Power in this context is the likelihood of rejecting the null hypothesis of zero slope variance when the variance is in fact different from zero. Of course, power varies as a function of effect size and sample size (which determine the separation of the noncentral χ^2 distribution under the alternative hypothesis from the central χ^2 under the null hypothesis) and the selected Type I error criterion (which defines the area under the noncentral χ^2 distribution in the region of rejection for the null hypothesis). Empirical power estimates in our study were obtained by setting the Type I error criterion at .05, and then computing the proportion of samples in which the observed LR χ^2 test statistic exceeded the critical value of 3.84 for a 1 *df* test and the critical value of 5.99 for a 2 *df* test.

BOUNDARY CONDITIONS

Recently, Stoel, Garre, Dolan, and Wittenboer (2006) argued that special attention is needed for LR tests of boundary parameters, such as testing the null hypothesis of zero slope variance in an LGCM. Typically, the LR test estimates the probability that the sample data could have been drawn from a population with a true zero slope variance, using a standard central χ^2 distribution under the null hypothesis. Stoel et al. argued that individuals testing the hypothesis of zero slope variance will treat inadmissible solutions (Heywood cases of negative estimated slope variance) as instances counting for the null hypothesis, assuming the true value is really zero. They therefore advocated an explicit mixture distribution approach (Stram & Lee, 1994) to generate accurate critical values

for the LR χ^2 test statistic. In the case of the generalized variance test, their assumption was that the true distribution under the null hypothesis involved a mixture ratio of 50:50 (admissible to inadmissible variance estimates), leading to a mixture of equally weighted χ^2 distributions with 1 and 2 *df* (see also Verbeke & Molenberghs, 2000). The practical implication is that one would use a critical value of 5.14 versus the standard 2 *df* χ^2 distribution with a critical value of 5.99 for the generalized variance test. Note that the practical consequence of using the mixture distribution is slightly higher power to correctly reject the null hypothesis of no random effects in slopes.

Because the same adjustment in critical value would be used for all LR tests in our simulation, use of the mixture distribution's critical value would result in a consistent increase in power (for LR tests not at ceiling or floor levels of power). Hence use of the standard versus the mixture distribution approaches would not alter conclusions about the effects of variations in LGCM parameter values (e.g., GCR) on power of the generalized variance test. Hence we have opted to report only empirical power estimates for the standard LR test, not one based on the mixture distribution.

SIMULATION DESIGN

We used Monte Carlo methods to simulate tests of the null hypothesis of zero slope variance with the specific and generalized variance tests while manipulating several variables we suspected would influence power. Our previous study of the power of LGCM to detect slope covariances used the method of Satorra and Saris (1985) to generate approximate power curves, and we showed these approximations were very or quite satisfactory when validated against Monte Carlo simulation results (Hertzog et al., 2006). In the present circumstances (tests of slope variance) we evaluated the Satorra–Saris approximation. There were errors of approximation for the Satorra–Saris approximation of the specific variance test under the boundary restrictions created by Models M2 and M3 that we deemed unacceptable.[2] Hence we use only Monte Carlo simulation in this article.

Table 1 summarizes the variables in our simulation design. Given the developmental frame of reference, our interest focused on evaluating power as the number of longitudinal occasions builds. We generated simulated data points at all 20 time points in the temporal epoch defined earlier, but we assumed that

[2]This is not an implicit criticism of Satorra and Saris (1985), who made it clear that their method is an approximation that may result in significant errors of approximation under specific conditions. In this case, the specific variance test requires −2LL comparisons between two models, M2 and M3, that are both misspecified with respect to the true alternative hypothesis, which creates distortion in the approximate noncentrality parameter estimate generated by their method.

TABLE 1
Simulation Design

1 Intercept mean ($M_{\text{InterceptY}}$)	50					
2 Slope mean (M_{SlopeY})	−20					
3 Intercept variance (σ^2_{I})	100					
4 Correlation of intercept and slope (ρ_{IS})	−0.5	−0.25	0	0.25	0.5	
5 Sample size	100	200	500	1000		
6 Error variance ($\sigma^2_{e_Y}$)	1	10	25	100		
7 Slope variance (σ^2_{S})	50	25	0			
8 Occasions of measurement[a]	0, 2, 4	0, 2, 4, 6	0, 2, 4, 6, 8	0, 2, 4, . . . , 10	0, 2, 4, . . . , 18	0 . . . 19

Note. The simulation design crossed sample size, variance of error (growth curve reliability), slope variance, intercept–slope correlation, and occasions of measurement. The condition of slope variance of 0 was only crossed with zero covariance of intercept and slope.

[a]Treated as a within-subjects factor.

the scientist would actually conduct a longitudinal assessment at every other time point (i.e., $t = 0, 2, 4, \ldots$). We started with three-occasion data (0, 2, 4) and continued upward, and we also evaluated the results if all occasions of measurement (including odd-numbered occasions) had been measured as a best possible scenario for power for the temporal design.

Mean intercepts, mean slopes, and the intercept variance were held constant across facets of the simulation design. Error variance was treated as homogenous across all T occasions. Our general approach was to simulate the data with 3,000 replicates for each cell in the design in Table 1. We also conducted additional, focused simulations that systematically varied certain parameter values in small increments (σ^2_{S}, GCR, and σ_{IS}) at selected values of sample size or other variables. The more fine-grained manipulation of slope variance effect sizes allowed us to determine the shape of standard power curves and when those curves reached a typical criterion for sufficient power (.80 according to the convention suggested by Cohen, 1988). The manipulation of GCR checked on whether the magnitude of error variance affects power, as it did for slope covariances (Hertzog et al., 2006). The manipulation of intercept–slope correlations explored differences in power of the specific and generalized variance test when nonzero values of σ_{IS} contributed information to the generalized test. We also included a condition with $\sigma^2_{\text{S}} = 0$ and $\sigma_{\text{IS}} = 0$ to evaluate Type I error for the three types of tests.

The variables y_t were scaled as T scores ($M = 50$, $SD = 10$) at $t(0)$, and the linear growth curve parameters were scaled so as to be psychologically plausible, based on prior longitudinal studies of adult cognitive development, and statistically possible. Empirical studies indicate that variance in change is small to moderate relative to variance in initial level (e.g., Hultsch, Hertzog, Dixon, & Small, 1998; Lövdén et al., 2004; Rabbitt, Diggle, Smith, Holland, & McInnes, 2001; Schaie, 2005; T. Singer, Verhaeghen, Ghisletta, Lindenberger,

& Baltes, 2003). We therefore scaled change variance to be either 50 or 25 at $t(19)$, relative to the intercept variance of 100, to arrive at ratios of total change over intercept variance of 1:2 and 1:4, respectively, values also used by Hertzog et al. (2006). Observed variance ratios are generally smaller than 1:4, making it even more difficult to detect individual differences in change. For this reason, one of our supplemental simulations focuses on a continuous manipulation of slope variance effect size.

Error variance, $\sigma^2_{\varepsilon y}$, was set to 100, 25, 10, and 1, yielding GCRs of .50, .80, .91, and .99, respectively at study onset, $t(0)$. Close-to-perfect GCR (i.e., .99) was included in the simulation to examine power under optimal measurement conditions when deviations from linearity were also at a minimum.

METHOD

Monte Carlo Simulation

We developed a Monte Carlo simulation engine to evaluate the contribution of the factors listed in Table 1 to the estimation and testing of parameters from the LGCM (Oertzen & Ghisletta, 2007). The engine generates individual scores based on LGCM population parameters by first generating samples of two independent variates from a standard normal distribution. These variates are then scaled, using the intercept and slope means and a Cholesky decomposition of the covariance matrix of each variable's intercept and slope scores. Taking the predicted scores from the LGCM for each occasion of measurement and adding stochastic error terms generated the full data matrix. Errors were computed by sampling random normal deviates and rescaling to expected population error variances. We simulated longitudinal sampling by selecting occasions of measurement from this full set of 20 possible occasions, assuming 2-year test intervals.

The simulation engine takes the raw data matrix for a given longitudinal sampling frame and estimates the latent growth curve parameters using both least-squares and full information maximum likelihood (FIML) estimation procedures. The FIML algorithm was programmed using first and second derivative expressions for the raw data vector likelihood (with respect to the LGCM parameters) obtained from Longford (1987) and Lange, Westlake, and Spence (1976). We checked the simulation engine's estimation algorithms by analyzing the data for a few randomly selected samples with Mx (Neale, Boker, Xie, & Maes, 1999), M*plus* (L. K. Muthén & Muthén, 1998–2004), and our engine. The results agreed in all cases.

For each model estimated, we computed and stored the −2 log likelihood (−2LL) of the FIML solution for the freely estimated LGCM (M1) and the

two restricted models (M2 and M3). Finally, at each replication we used the information matrix to compute asymptotic standard errors for the parameter estimates. In particular, we formed a Wald test for the hypothesis of zero slope variance as

$$z = \text{estimated } \sigma_S^2 / \text{estimated var } (\sigma_S^2)^{.5},$$

where var (σ_S^2) is the asymptotic variance of estimate for σ_S^2 taken from the information matrix of the converged solution.

RESULTS

We first present detailed power curves for the generalized variance test. We then report comparative results with the three standard tests (generalized variance test, specific variance test, and Wald test). Finally, we consider results from alternative critical values generated by a mixture distribution approach (Stoel et al., 2006).

Statistical Power of the Generalized Variance Test

Figure 1 shows Monte Carlo power curves for the 2-*df* generalized variance test as a function of varying the correlation of intercepts and slopes, ρ_{IS}, from $-.995$ to $+.995$ for the larger variance effect size. The models estimated the unstandardized random effects, including the covariance of intercepts and slopes, but we show the correlation of slopes in Figure 1 for ease of interpretation. Curves for the smaller variance effect size were similar, with lower power, of course, and hence are not shown here.

Each Figure 1 depicts a family of curves for different values of GCR at $t(0)$: (.50, .80, 91, .99). Power is a roughly quadratic function of ρ_{IS}, with lower power when the correlation is low and increasing as $|\rho_{IS}|$ increases. The curves appear to be symmetrical, but around approximately $\rho_{IS} = -.10$, not $\rho_{IS} = 0$. Obviously, manipulating GCR had a dramatic effect on power to detect the slope variances, generating curves that were fairly widely spread apart. When GCR was .99, estimated power to detect variances in change is uniformly perfect, even with four occasions of measurement. However, power drops dramatically as GCR decreases. With four longitudinal occasions, GCR = .50, and the smallest sample size ($N = 100$), power is .20 or below. Power exceeds the often-mentioned benchmark of .80 (Cohen, 1988) for four occasions only when GCR was .91 and $N = 500$. Power does increase monotonically as the number of longitudinal occasions increases, as expected, and is higher for larger sample sizes. Nevertheless, the relatively low power of the 2 *df* tests under many conditions, especially when ρ_{IS} was close to 0 and GCR $< .91$, is striking.

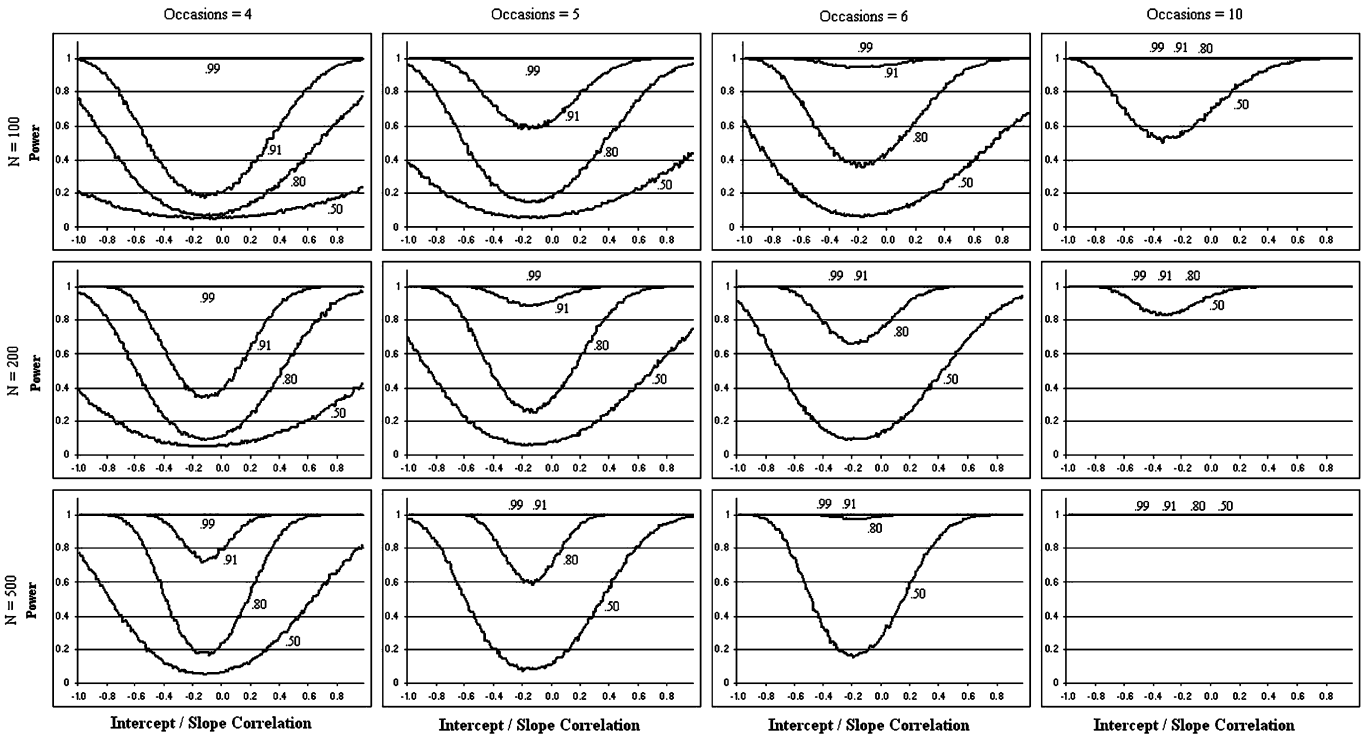

FIGURE 1 Power curves for the generalized variance test to correctly reject the null hypothesis of zero slope variance as function of σ_{IS} (varying on x-axis from $-.995$ to $+.995$). Data are plotted for three sample sizes (rows corresponding to $N = 100$, 200, and 500) and four longitudinal designs (columns corresponding to 4, 5, 6, and 10 occasions of measurement). Separate curves within each panel correspond to GCR of .50, .80, .91, and .99.

Figure 2 presents a different look at the power of the generalized variance test, plotting manipulated GCR from .50 to .995 on the x-axis, with different numbers of longitudinal occasions of measurement generating the family of curves within each plot. These curves demonstrate the profound attenuating effect of low GCR on power to detect slope variances. The middle column in Figure 2, with $\rho_{IS} = 0$, clearly generated lower power than the other two columns with $\rho_{IS} \neq 0$, as the criterion of .80 was reached only when GCR was relatively high, or when the number of longitudinal occasions was relatively large. When $N = 100$, power exceeds .80 only when GCR $> .90$ unless there are six or more longitudinal occasions. The picture is better when $N = 500$, but even then the power to reject the hypothesis of zero slope variance is only .80 with four occasions and GCR of .91. When intercept–slope correlations are $-.50$ or $+.50$, the situation improves at least somewhat.

Figure 3 plots power curves for a continuous range of slope variance effect sizes, σ_S^2, from near 0 to 100. The latter value represents a situation of equal

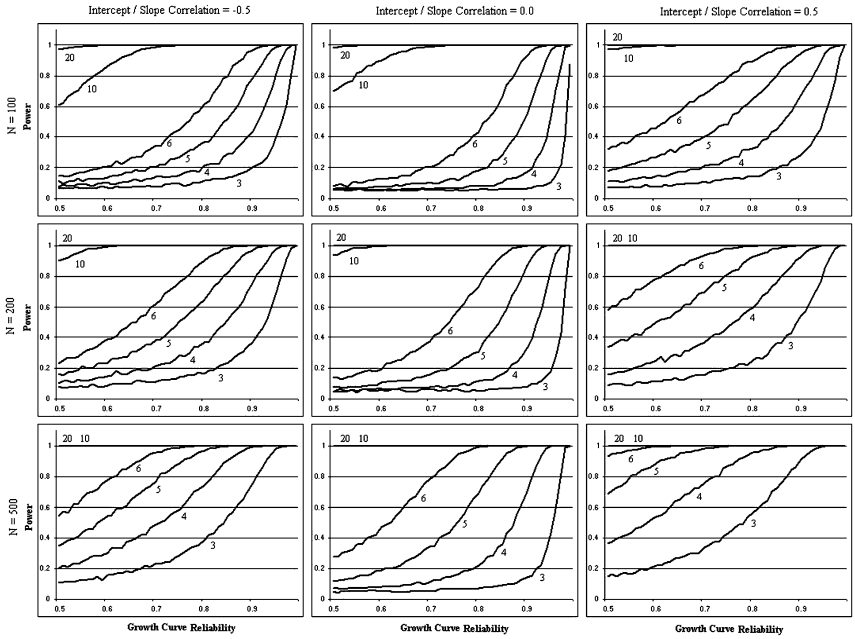

FIGURE 2 Power curves for the generalized variance test to correctly reject the null hypothesis of zero slope variance as a function of GCR (varying on x-axis from .50 to .995). Data are plotted for three sample sizes ($N = 100$, 200, and 500) and three slope-intercept correlations ($-.5$, 0, .5). Separate curves within each panel correspond to number of longitudinal occasions.

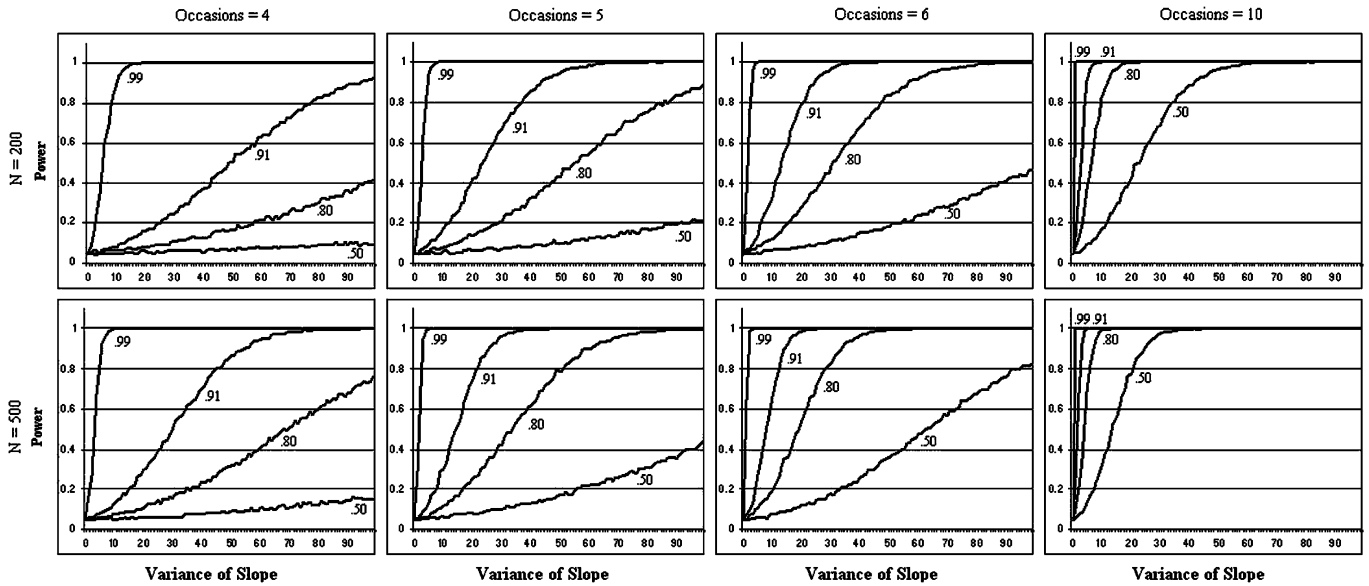

FIGURE 3 Power curves for the generalized variance test to correctly reject the null hypothesis of zero slope variance as a function of slope variance, σ_{S2} (varying on x-axis from 0.05 to 100). Data are plotted for three sample sizes ($N = 100$, 200, and 500) and three slope-intercept correlations (−.5, 0, .5). Separate curves within each panel correspond to GCR of .50, .80, .91, and .99.

slope variance and intercept variance, which from our practical experience in cognitive aging research would represent an extremely large effect size. The families of curves within a panel represent different GCR values. When GCR is near perfect, the power to detect a slope variance rises quickly above the benchmark of .80 power with relatively small variance effect sizes (< 10, even for only four occasions of measurement). However, when GCR is low (.50), power to detect the slope variance is poor even at the largest effect sizes, unless the number of occasions exceeds typical values in longitudinal designs (six or more occasions of measurement).

Comparisons of Generalized, Specific, and Wald Tests

Type I error. We evaluated the behavior of the slope variance tests when $\sigma_S^2 = 0$ and $\rho_{IS} = 0$ using Monte Carlo simulation. Figure 4 shows the aggregate behavior of the 2 *df* generalized variance test and the Wald test when the null hypothesis of no individual differences in change is true, for GCR = .91 and $N = 100$ (aggregated over all longitudinal occasions). The empirical distribution of the LR test closely approximates a theoretical central χ^2 variable (cf. Pinheiro & Bates, 2000). It did so because we did not impose any boundary restrictions on the estimated slope variance parameters, which were allowed to be negative. Results with the specific variance test were similar to the generalized test and are not shown. The estimated Type I error rates in the simulation cells were close to the specified .05 level for the two LR tests ($M = .051$). The Wald test deviated slightly more from its expected value, making it somewhat liberal; 8% of the cases would have been rejected at a theoretical Type I error level of 5%. We were using a two-tailed criterion and allowing rejection of the null

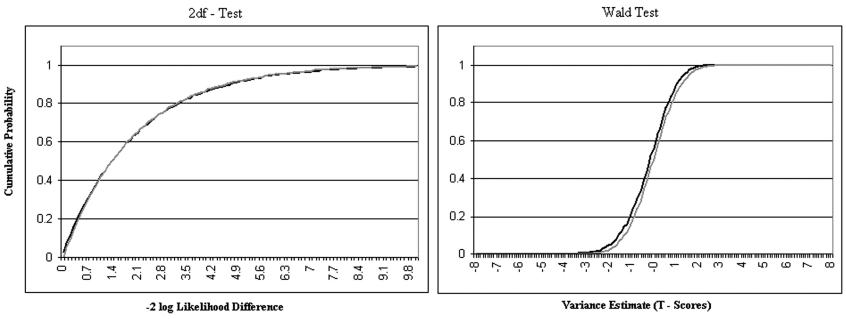

FIGURE 4 Comparison of generalized (2 *df*) variance test to its theoretical parent cumulative χ^2 distribution (left panel) and of the Wald test of zero variance to its theoretical parent, a cumulative normal distribution.

hypothesis for extreme negative variances. Note that the proportion of negative slope variance estimates (plotted on the x-axis of the Wald test plot) was about .50, as expected.

Power. The substantial effect of intercept–slope correlations on power shows indirectly that the 2 *df* test benefits from using both slope-related parameters. However, this is not the full story. Our simulation also indicated that manipulating ρ_{IS} has an effect on the 1 *df* specific variance test.

Table 2 reports the average power for the three tests for $\sigma^2_S = 25$ (smaller variance effect size) across different values of ρ_{IS} with four and five longitudinal occasions, averaging over GCR and N. These are the cells in which power was most consistently not at floor or ceiling. There is a clear power advantage for the generalized test, and inferior performance by the Wald test. Figure 5 plots the power of the generalized variance (2 *df*) test, the specific variance (1 *df*) test, and the Wald test for nine selected simulation cells to illustrate the patterns in the data. The Wald test has lower power, overall, relative to the two LR tests. Note also that the Wald test is not affected by the magnitude of ρ_{IS}, whereas the specific LR variance test is strongly affected by it. The specific variance test shows roughly monotonic increases in power as ρ_{IS} moves from -0.5 to $+0.5$. The specific variance test actually manifests *lower* power than the Wald test when $\rho_{IS} = -0.5$ as long as power is not near ceiling or floor, and has lower power when $\rho_{IS} = -0.25$ for some cells. When ρ_{IS} is equal or greater than 0, the specific variance test has a considerable power advantage over the Wald test. In contrast, the generalized variance test shows the slightly asymmetric U-shaped power curve revealed in Figure 5, and has superior power to the Wald test for all simulated values of ρ_{IS}. The specific variance test is slightly superior in power to the generalized variance test when ρ_{IS} was very close to 0, but not otherwise.

TABLE 2
Average Power of Three Different Tests to Reject Null Hypothesis of Zero Slope Variance for $\sigma^2_S = 25$ (Small Variance Effect Size)

ρ_{IS}	*Generalized Test*	*Specific Test*	*Wald Test*
−0.5	0.60	0.33	0.40
−0.25	0.44	0.36	0.40
0	0.41	0.44	0.41
0.25	0.54	0.52	0.41
0.5	0.68	0.60	0.40

Note. Data are taken from four-occasion and five-occasion cells, averaged over GCR growth curve reliability and sample size.

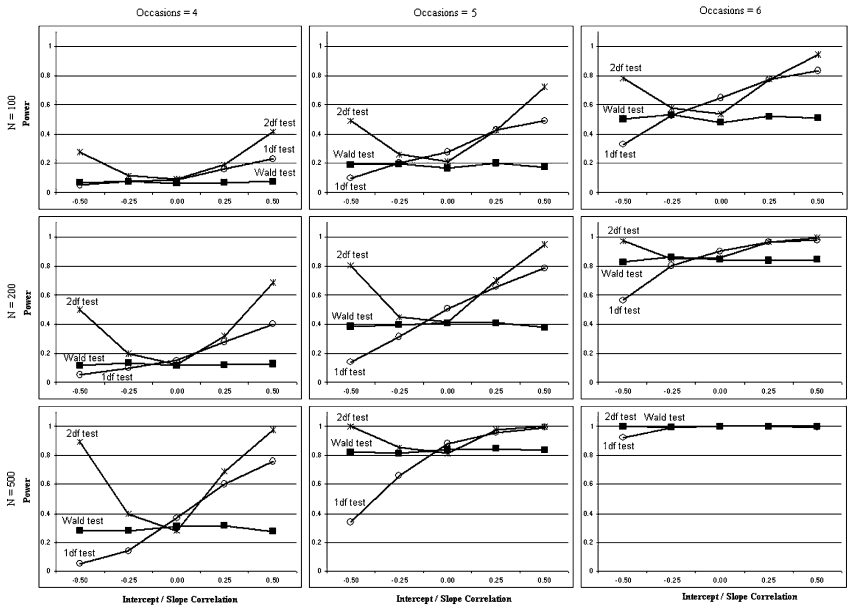

FIGURE 5 Comparison of statistical power for the generalized variance (2 *df*) test, specific variance (1 *df*) test, and Wald test of the null hypothesis of zero slope variance as a function of slope-intercept correlation for two sample sizes ($N = 100$ and 200) and two different longitudinal epochs (four and six occasions).

DISCUSSION

Effects of GCR on Statistical Power

A major result from this simulation is the profound effect of GCR on the power of all of the variance tests, including the generalized test. These results closely resemble effects of GCR on the power to detect slope correlations in a bivariate LGCM (Hertzog et al., 2006). When GCR was essentially perfect, the power to detect slope variance was excellent with only three longitudinal occasions of measurement. When GCR was only .50, power was not satisfactory until one had 10 longitudinal occasions of measurement for the smaller sample sizes. The variance effect size plots in Figure 3 demonstrated conclusively that power was low to detect large slope variances when GCR was below .91, even with a sample size of 500 and five occasions of measurement. We believe that most applied longitudinal researchers would be surprised to learn of the low power to detect slope variances in LGCMs under the conditions we simulated here,

particularly because GCRs have not been emphasized as an important feature of LGCM in typical applications.

LGCM parameter estimates are disattenuated for random measurement error (Bollen, 1989).[3] However, this does not imply elimination of the influence of random measurement error—as one component of GCR—on statistical inferences regarding the LGCM parameters. Multiple-indicator LGCM (McArdle, 1988; Sayer & Cumsille, 2001) should in principle minimize effects of measurement error on statistical power to detect slope variances. Multiple indicator models remove stochastic measurement errors in the observed variables from GCR, estimating them in separate variance components. The GCR in a multiple-indicator model reflects only deviations from linearity in the growth curve for latent variables (latent residual variance around the linear regression line). Given the poor statistical power to detect slope variances with GCR $< .91$, these results suggest that use of multiple indicators could in principle greatly enhance statistical power to detect individual differences in change when they do exist.

Longitudinal studies often involve small samples, limited numbers of occasions, less-than-perfect GCR, and nonrandom sample attrition over time. We believe, therefore, that the results reported here indicate that investigators will often fail to detect slope variances when they exist. Furthermore, this simulation treats the basic LGCM assumptions as true (i.e., there is a universal linear functional form of growth, errors are homoscedastic, statistically independent of latent change, and uncorrelated with each other over time, virtually zero attrition, etc.). The effects of violating these statistical assumptions on power to detect individual differences in change are currently unknown. It is possible that power is even lower under certain conditions (e.g., failure to properly specify a nonlinear functional form of growth).

The Generalized Test Yields Better Power

The results of this simulation indicate that the 2 *df* generalized variance test is clearly preferable to either the specific variance test or the Wald test. Although the specific variance test has slightly better power than the generalized test when $\rho_{IS} = 0$, it has lower power otherwise and is influenced by the magnitude of ρ_{IS}—even though the specific test nominally accounts for the covariance by nested LR tests that isolate σ_S^2. Given that it is generally unlikely that the slope–intercept correlation is very close to zero in a given population, the generalized test will usually give better results, in terms of power to detect the nonzero slope variance. Moreover, the generalized test avoids the problem of shift in the

[3]We do not report these results to save space, but the mean parameter estimates in our simulation were good approximations to the true parameter values, implying appropriate corrections for measurement error by the LGCM.

magnitudes of σ_{IS} and σ_S^2 as a function of choice of scaling the latent growth curve (Rovine & Molenaar, 1998). The Wald test had lower power than the generalized LR test of slope variance. This finding reinforces arguments in the mixed model literature that the generalized LR test, not the Wald test, should be used for hypothesis testing on random effects (e.g., J. D. Singer & Willett, 2003; Snijders & Boskers, 1999). Ironically, the apparent benefit of the Wald test in this simulation—that it is invariant with respect to the manipulation of ρ_{IS}—is offset by the fact that it can be affected by placement of identification constraints in a given model, including scaling of the growth curve (Gonzalez & Griffin, 2001). However, the 1 *df* specific variance test actually had lower power than the Wald test when $\rho_{IS} < -.25$. Considering all these factors, only the generalized variance test can be recommended as a routine approach to testing for individual differences in change.

Boundary Conditions and Mixture Distributions

As noted earlier, we evaluated results without using mixture distribution critical values (Stoel et al., 2006). Elsewhere, we (Oertzen, Ghisletta, Lindenberger, & Hertzog, 2007) have suggested that the mixture distribution recommended by Stoel et al. (2006) represents only one of the alternatives to the standard LR test, with the others being governed by alternative rationales for handling Heywood cases. Stoel et al. did not explicitly evaluate the degree of benefit on statistical power from using the mixture distribution approach. Oertzen et al.'s (2007) simulation results indicate that use of modified Type I error criteria through mixture distributions provides a modest power advantage over the standard tests reviewed here, in keeping with the fact that the only difference between the standard and mixture distribution tests is a more liberal Type I error criterion (smaller critical value for the χ^2 test statistic) invoked by the mixture distribution approach. Most important for the results of this simulation, the power benefit of the mixture distribution approach for LR tests was additive to the factors evaluated in this study (e.g., GCR). Hence our conclusions about factors influencing power are unaffected by which type of Type I error criterion one uses, the standard LR critical value or a mixture-distribution adjusted value.

A practical problem with a mixture distribution method is that one cannot necessarily rely on a priori assumptions about mixture distribution weights because sample sizes in longitudinal studies may be sufficiently far from asymptotic as to affect the empirical likelihood of inadmissible solutions. Instead one needs to simulate the expected frequency of negative slope variance estimates when the null hypothesis is true for a given longitudinal design (Oertzen et al., 2007; Stoel et al., 2006). Many users, therefore, may opt to avoid the simulation and simply use the unadjusted LR tests we evaluate here, even though greater power could be achieved through a mixture distribution approach. Given the low power

to detect slope variances in many conditions, our results suggest that the pain of conducting the extra simulation to generate the proper mixture-distribution critical value may be worth the small gain in power one realizes in the process.

Limitations

This simulation has a number of limitations. We did not evaluate violations of statistical assumptions, nor did we fully explore the universe of possible combinations of LGCM parameter values. For instance, we did not evaluate the consequences of choice of scaling on power to detect slope variances. It is often the case that mixed model analysts prefer to center the time (or age) variable in an LGCM rather than to define it to be zero at time $t(0)$ as was done in this simulation. We do not yet know the consequences of such scaling choices for the statistical power of the 2 *df* generalized LR test.

Furthermore, there are a number of other change-related models and applications we have not evaluated, such as the introduction of exogenous covariates that predict LGCM random effects, or more complicated models that build off the basic LGCM but add dynamic regression coefficients, such as McArdle's bivariate dual-change score model (e.g., McArdle et al., 2002). Ghisletta, McArdle, Oertzen, Hertzog, and Lindenberger (2005) reported preliminary results suggesting good power to detect dynamic lagged regression effects in the bivariate dual-change score model (which are defined as fixed, not random, effects). Hence, one should not overgeneralize our findings to the statistical power of other classes of developmental models for detecting change-related parameters. Such issues remain open empirical questions that can and should be explored in new studies.

The linear LGCM we have simulated is an increasingly common application in developmental psychology (e.g., Duncan et al., 1999; J. D. Singer & Willett, 2003). Users of this technique should be wary of accepting the null hypothesis of no random effects in change unless they conduct post hoc power analyses to ensure their study had adequate power to detect plausible variance effect sizes in a given content domain. At minimum, researchers should calculate estimates of GCR in their study and evaluate whether it is sufficiently low to raise concerns about power to detect random effects, which could be done to a crude approximation from the simulation results provided in this study. Generically, our simulation indicates that GCR values under .90 are potentially problematic. Longitudinal design decisions can and should be based on a Monte Carlo simulation (which is readily available in M*plus*; see also L. K. Muthén & Muthén, 2002) to generate a priori power estimates to guide longitudinal design decisions. Such simulations are also critical for determining weights for use of different mixture distributions testing more complex hypotheses as well (Stoel et al., 2006). Our study reinforces the importance of explicitly considering the

statistical power of tests for LGCM parameters when designing longitudinal studies focused on individual differences in change.

ACKNOWLEDGMENTS

This work was partially supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) to Ulman Lindenberger in the context of the Collaborative Research Center (SFB) 378, and by an internal developmental leave award from the Georgia Institute of Technology to Christopher Hertzog. We thank Martin J. Sliwinski for advice regarding random effects estimation, Florian Schmiedek for assistance in simulation data processing, and J. Jack McArdle for helpful comments and discussion.

REFERENCES

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1988). *Life-span developmental psychology: Introduction to research methods.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bollen, K. A. (1989). *Structural equation models with latent variables.* New York: Wiley.

Bryk, A. S., & Raudenbush, S. W. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Fan, X. (2003). Power of latent growth modeling for detecting group differences in linear growth trajectory parameters. *Structural Equation Modeling, 10,* 380–400.

Ghisletta, P., McArdle, J. J., Oertzen, T. v., Hertzog, C., & Lindenberger, U. (2005, November). *Power to identify lead–lag relationships in bidimensional dynamic systems based on latent difference score models.* Paper presented at the 58th Annual Scientific Meeting of the Gerontological Society of America, Orlando, FL.

Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every 'one' matters. *Psychological Methods, 6,* 258–269.

Hertzog, C., Lindenberger, U., Ghisletta, P., & Oertzen, T. v. (2006). On the power of latent growth curve models to detect correlated change. *Psychological Methods, 11,* 244–252.

Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging, 18,* 639–657.

Hultsch, D. F., Hertzog, C., Dixon, R. A., & Small, B. J. (1998). *Memory change in the aged.* New York: Cambridge University Press.

Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling, 12,* 368–390.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38* 963–974.

Lange, K., Westlake, J., & Spence, M. A. (1976) Extensions to pedigree analysis: III. Variance component by the scoring method. *Annals of Human Genetics, 39,* 485–491.

Littell, R. C, Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (2006). *SAS system for mixed models* (2nd ed.). Cary, NC: SAS Institute.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation for unbalanced mixed models with nested random effects. *Biometrika, 74,* 817–827.

Lövdén, M., Rönnlund, M., Wahlin, Å., Bäckman, L., Nyberg, L., & Nilsson, L-G. (2004). The extent of stability and change in episodic and semantic memory in old age: Demographic predictors of level and change. *Journal of Gerontology: Psychological Sciences, 59B,* P130–P134.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84–99.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 561–614). New York: Plenum.

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

McArdle, J. J., Ferrer-Caja, E., & Hamagami, E. Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology, 38,* 115–142.

Marcoulides, G. A. (1996). Estimating variance components in generalizability theory: The covariance structure analysis approach. *Structural Equation Modeling, 3,* 290–299.

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2,* 371–402.

Muthén, L. K., & Muthén, B. O. (1998–2004). *M*plus *user's guide* (3rd ed.). Los Angeles: Muthén & Muthén.

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,* 599–260.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (1999). *Mx: Statistical modeling* (5th ed.). Richmond: Medical College of Virginia.

Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin, 88,* 622–637.

Oertzen, T. v., & Ghisletta, P. (2007). *A simulation engine for structural equation modeling.* Unpublished manuscript. Max Planck Institute for Human Development, Berlin, Germany.

Oertzen, T. v., Ghisletta, P. G., Lindenberger, U., & Hertzog, C. (2007). *Strategies for hypothesis testing when parameters are at or near a boundary.* Unpublished manuscript. Max Planck Institute for Human Development, Berlin, Germany.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effect models in S and S-PLUS.* New York: Springer.

Rabbitt, P. M. A., Diggle, P., Smith, D., Holland, F., & McInnes, L. (2001). Identifying and separating the effects of practice and of cognitive aging during a large longitudinal study of elderly community residents. *Neuropsychologia, 39,* 532–543.

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

Raykov, T. (1993). A structural equation model for measuring residualized change and discerning patterns of growth or decline. *Applied Psychological Measurement, 17,* 53–71.

Rogosa, D., & Willett, J. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203–228.

Rovine, M. J., & Molenaar, P. C. M. (1998). The covariance between level and shape in the latent growth curve model with estimated basis vector coefficients. *Methods of Psychological Research Online, 3*(2). Retrieved May 4, 2001, from http://www.pabst-publications.de/mpr/

Rovine, M. J., & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research, 35,* 51–88.

Satorra, A., & Saris, W. E. (1985). The power of the LR test in covariance structure analysis. *Psychometrika, 50,* 83–90.

Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 177–200). Washington, DC: American Psychological Association.

Schaie, K. W. (1977). Quasi-experimental designs in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 39–58). New York: Van Nostrand Reinhold.

Schaie, K. W. (2005). *Intellectual development in adulthood: The Seattle Longitudinal Study.* New York: Cambridge University Press.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis.* Oxford, UK: Oxford University Press.

Singer, T., Verhaeghen, P., Ghisletta, P., Lindenberger, U., & Baltes, P. B. (2003). The fate of cognition in very old age: Six-year longitudinal findings in the Berlin Aging Study (BASE). *Psychology and Aging, 18,* 318–331.

Snijders, T. A. B., & Boskers, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Stoel, R. D., Garre, F. G., Dolan, C., & Wittenboer, G. v. d. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods, 11,* 439–455.

Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics, 50,* 1171–1177.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer-Verlag.

Wohlwill, J. F. (1991). The partial isomorphism between developmental theory and methods. *Annals of Theoretical Psychology, 6,* 1–43.