

On the Power of Multivariate Latent Growth Curve Models to Detect Correlated Change

Christopher Hertzog

Georgia Institute of Technology

Ulman Lindenberger

Max Planck Institute for Human Development
and Saarland University

Paolo Ghisletta

University of Geneva

Timo von Oertzen

Max Planck Institute for Human Development
and Saarland University

We evaluated the statistical power of single-indicator latent growth curve models (LGCMs) to detect correlated change between two variables (covariance of slopes) as a function of sample size, number of longitudinal measurement occasions, and reliability (measurement error variance). Power approximations following the method of Satorra and Saris (1985) were used to evaluate the power to detect slope covariances. Even with large samples ($N = 500$) and several longitudinal occasions (4 or 5), statistical power to detect covariance of slopes was moderate to low unless growth curve reliability at study onset was above .90. Studies using LGCMs may fail to detect slope correlations because of low power rather than a lack of relationship of change between variables. The present findings allow researchers to make more informed design decisions when planning a longitudinal study and aid in interpreting LGCM results regarding correlated interindividual differences in rates of development.

Keywords: statistical tests, latent growth curves, power analysis, aging, structural equation models

Assessing individual differences in change with appropriate statistical models is critical for empirical studies of life span development (Baltes & Nesselroade, 1979; Hertzog & Schaie, 1986; Hofer & Sliwinski, 2006; McArdle, 1988;

Wohlwill, 1991). When developmental change can be characterized as progressive or cumulative (e.g., cognitive development from childhood to early maturity or cognitive aging in older adults), latent growth curve models (LGCMs) provide an attractive means of assessing both average developmental function and individual differences in rates of developmental change (Bryk & Raudenbush, 1987; Duncan, Duncan, Strycker, Li, & Alpert, 1999; Laird & Ware, 1982; McArdle & Epstein, 1987; Raykov, 1993; Rogosa & Willett, 1985). LGCMs can be used to test a critical developmental question, namely, whether there are covariances in change between variables. Whether change is associated across two or more variables is a hallmark of theories about life span cognitive development (Hertzog, 1985; Hultsch, Hertzog, Dixon, & Small, 1998; Lindenberger & Baltes, 1994; Rabbitt, 1993).

Christopher Hertzog, School of Psychology, Georgia Institute of Technology; Ulman Lindenberger, Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany, and School of Psychology, Saarland University, Saarbrücken, Germany; Paolo Ghisletta, Center for Interdisciplinary Gerontology and Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland; Timo von Oertzen, Center for Lifespan Psychology, Max Planck Institute for Human Development, and Department of Mathematics, Saarland University, Saarbrücken, Germany.

This work was partially supported by a grant from the Deutsche Forschungsgemeinschaft to Ulman Lindenberger in the context of Collaborative Research Center 378 and by an internal developmental leave award from the Georgia Institute of Technology to Christopher Hertzog.

Correspondence concerning this article should be addressed to Christopher Hertzog, School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170. E-mail: christopher.hertzog@psych.gatech.edu, lindenberger@mpib-berlin.mpg.de, paolo.ghisletta@cig.unige.ch, or vonoertzen@mpib-berlin.mpg.de

Currently, we lack good information about the capability of LGCMs to detect a covariance in slopes between two variables. Existing simulation studies of LGCMs have focused on other issues, including the power of LGCMs to detect differences in average slopes between different groups of persons (e.g., Fan, 2003; Muthén and Curran, 1997). Consequently, psychologists seeking to design prospective longitudinal studies of psychological development have no simulation-based guidelines regarding sample size,

spacing of temporal retest intervals, the temporal length of studies needed to detect different effect sizes of variance in change, and so on. A major risk, then, is that researchers may design studies that do not have a realistic chance of detecting relationships in change between variables because their choices regarding sample size, occasions of measurement, or other relevant design features led to insufficient statistical power to detect potentially existing correlations between two or more change processes.

Statistical Model

Given our interest in evaluating individual differences in adult cognitive development (e.g., Ghisletta & Lindenberger, 2003; Hertzog, Dixon, Hultsch, & MacDonald, 2003; Lindenberger & Oertzen, 2006), we frame the LGCM issues around the problem of detecting reliable individual differences in rates of cognitive decline and reliable associations of age-related decline in two variables. However, the results we report are actually arbitrary with respect to (a) epoch of time, (b) direction of change (growth or decline), and (c) choice of psychological constructs.

Consider a simple bivariate linear LGCM for two variables, x and y , measured longitudinally over time $t = 0, \dots, T$, on $i = 1, \dots, N$ persons, generating data matrices \mathbf{x}_{it} and \mathbf{y}_{it} . The LGCM can be written as

$$\begin{aligned} \mathbf{x}_{it} &= \beta_{x1} \cdot \text{Intercept}X_i + \beta_{x2t} \cdot \text{Slope}X_i + \boldsymbol{\epsilon}_{xit} \\ \mathbf{y}_{it} &= \beta_{y1} \cdot \text{Intercept}Y_i + \beta_{y2t} \cdot \text{Slope}Y_i + \boldsymbol{\epsilon}_{yit}, \end{aligned} \quad (1)$$

where $\text{Intercept}X_i$, $\text{Slope}X_i$, $\text{Intercept}Y_i$, and $\text{Slope}Y_i$ are latent variables defining the intercepts and slopes of the latent growth curves for x and y . For x and y , a growth curve design matrix, \mathbf{B} , is defined with $2T$ rows for occasions of measurement for x and y and four columns for the intercept and slope of x and y . Each column of \mathbf{B} is a vector of regression weights establishing the relationship of occasions of measurement to the growth curve. For intercepts, all values are fixed at 1. The regression weights for slopes are increased from 0 to 1 in steps of $1/T$. (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 (E) on the basis of the mean growth curves across individuals for \mathbf{x}_{it} and \mathbf{y}_{it} :

$$\begin{aligned} E(\mathbf{x}_{it}) &= M_{\text{Intercept}X} + \boldsymbol{\beta}_{2t} \cdot M_{\text{Slope}X} \\ E(\mathbf{y}_{it}) &= M_{\text{Intercept}Y} + \boldsymbol{\beta}_{2t} \cdot M_{\text{Slope}Y}, \end{aligned} \quad (2)$$

where $\boldsymbol{\beta}_{2t}$ is the column vector of weights defining the linear growth curve slope across T occasions, $M_{\text{Intercept}X}$ and $M_{\text{Intercept}Y}$ are the population mean intercepts and $M_{\text{Slope}X}$ and $M_{\text{Slope}Y}$ are the population mean slopes.

Consistent with common practice in empirical applications of LGCMs, we assume that the errors, $\boldsymbol{\epsilon}_{xit}$ and $\boldsymbol{\epsilon}_{yit}$, 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 of measurement, denoted $\sigma_{\boldsymbol{\epsilon}_{ex}}^2$ and $\sigma_{\boldsymbol{\epsilon}_{ey}}^2$. Hence the expectation for the covariance matrix of the observed variables, $\boldsymbol{\Sigma}_{xy}$, aggregating over individuals, is

$$\boldsymbol{\Sigma}_{xy} = \mathbf{B}_{pt} \boldsymbol{\Sigma}_{\text{IS}} \mathbf{B}_{pt}' + \boldsymbol{\Theta}_{xy}, \quad (3)$$

where p is an index for each observed variable, $\boldsymbol{\Sigma}_{\text{IS}}$ is the covariance matrix of Intercept X , Intercept Y , Slope X , and Slope Y and $\boldsymbol{\Theta}_{xy}$ is a diagonal matrix containing the error variances $\sigma_{\boldsymbol{\epsilon}_{ex}}^2$ and $\sigma_{\boldsymbol{\epsilon}_{ey}}^2$. The parameters in $\boldsymbol{\Sigma}_{\text{IS}}$ and $\boldsymbol{\Theta}_{xy}$ are termed random effects in multilevel models. Our interest in this study is on the power to detect nonzero slope covariances, namely, $\sigma_{\text{Slope}X\text{Slope}Y}$, one element of $\boldsymbol{\Sigma}_{\text{IS}}$.

Longitudinal Design Model

The setting for this power study is a prospective single-cohort longitudinal design (Baltes, Reese, & Nesselroade, 1988; Schaie, 1977). We assume that a population of adults of homogeneous initial age (i.e., 50 years) has been measured on two variables that produce linear age decline from age 50 years through age 69 years. We assume that the scientist draws a random sample of individuals from the general population and then measures these individuals longitudinally. To enable comparisons across different temporal epochs, we framed the linear growth process as change over 20 occasions of measurement. For relevance to studies of adult cognitive development, the temporal unit is years, but other substantive problems would dictate other time scales (e.g., days or weeks for prosocial development in preschool). The variables x and y were scaled as T scores ($M = 50$, $SD = 10$) at $t(0)$, and the LGCM parameters were scaled to be psychologically plausible, on the basis of prior longitudinal studies of adult cognitive development, and statistically possible. Data from longitudinal studies have demonstrated that variance in change is small to moderate relative to variance in initial level (e.g., Hertzog & Schaie, 1986; Hultsch et al., 1998; Lövdén, Ghisletta, & Lindenberger, 2004; Lövdén, Rönnlund, et al., 2004; Rabbitt, Diggle, Smith, Holland, & McInnes, 2001; Schaie, 1996). On the basis of these empirical considerations, we scaled change variance to be either 50 or 25 at $t(19)$, relative to the intercept variance of 100, to arrive at a ratio of total change over intercept variance of 1:2 or 1:4. Note that estimated ratios reported in the literature are generally smaller, in all likelihood making it even more difficult to detect interindividual differences in change. Both variables were treated as having equal magnitudes of mean parameters for intercept and slope, intercept variance and covariance parameters,

and slope variance (i.e., we did not simulate the situation of differential change in the two variables). Mean intercepts, mean slopes, intercept variances, and intercept covariances were held constant. Also, covariances between intercepts and slopes were fixed to 0. Error variance was treated as homogenous across all T occasions.

One important issue is the number of occasions of measurement needed to reliably detect a covariance between the slopes of x and y . Hence we evaluated changes in statistical power as a function of increasing the number of longitudinal occasions of measurement, as if we were conducting an actual longitudinal study.

A second issue is the effect of growth curve reliability (GCR) on the power to detect covariance between the slopes. Define the total variance of x at time t as σ_{xt}^2 . This variance can be decomposed into two components, (a) that due to individual differences in latent intercepts and slopes, namely $\sigma_{lx}^2 + \beta_{2t}^2 \cdot \sigma_{sx}^2$ (where σ_{lx}^2 and σ_{sx}^2 are the variances of the intercepts and slopes for x , respectively), and (b) that due to error (σ_{ex}^2). Then GCR is defined as $(\sigma_{lx}^2 + \beta_{2t}^2 \cdot \sigma_{sx}^2) / \sigma_{xt}^2$ or $(\sigma_{lx}^2 - \sigma_{ex}^2) / \sigma_{xt}^2$; that is, the ratio of variance is determined by the latent growth curve to total variance. The same expressions apply to variable y . The error variance has two components: random measurement error in each variable and variability of the residuals for the true scores of x and y around the linear regression functions of the growth curves. At $t(0)$, GCR reduces to $\sigma_{lx}^2 / (\sigma_{lx}^2 + \sigma_{ex}^2)$ for x and $\sigma_{ly}^2 / (\sigma_{ly}^2 + \sigma_{ey}^2)$ for y (namely, variance in latent intercepts divided by total variance), because by scaling constraint, individual differences in slopes do not contribute to the reliable variance at $t(0)$. These expressions are, of course, closely related to reliability as defined by classical test theory, differing only in the growth curve basis for estimating the residual variances σ_{ex}^2 and σ_{ey}^2 and hence the resulting absorption of regression residual variance into the error term. In general, GCR varies with t , so we refer to GCR as scaled at $t(0)$ for the power curves we report in this article.

On the basis of the well-known effects of random measurement error (unreliability) on change scores (e.g., Rogosa, Brandt, & Zimowski, 1982), we also expected that GCR would influence the power of LGCMs to detect slope covariances. Moreover, LGCMs can be framed as an instantiation of restricted common factor analysis. MacCallum, Widaman, Zhang, and Hong (1999) demonstrated that power to detect unstandardized factor loadings depends on the magnitude of residual variances. Note that although LGCMs define latent factors of intercept and change, each of these latent factors is typically based on just one indicator per construct measured over time. We suspected that the statistical power of such models may be more susceptible to measurement error than is commonly assumed.

In practice, researchers evaluate the null hypothesis of zero slope covariance with a likelihood ratio (LR) chi-square test (e.g., Steiger, Shapiro, & Browne, 1985), com-

puted as the difference in fit between a model that fixes the slope covariance to 0 versus one that estimates it freely. Power is defined as the probability of rejecting the null hypothesis when it is false, that is, of detecting a covariance in change when it actually exists. We used the method developed by Satorra and Saris (1985) to evaluate the power to detect slope covariances of varying magnitudes (effect sizes). Satorra and Saris showed that if the true model is in a sufficiently close neighborhood to the model of the null hypothesis, then the LR of a set of data generated by the true model tested against the null hypothesis will be distributed as a noncentral chi-square variate. The noncentrality parameter of this distribution is then the optimal $-2 \log$ likelihood index for the true model against the null hypothesis model, multiplied by the number of observed data points. Power can then be estimated in turn by comparing the overlap of the noncentral chi-square with the central chi-square generated for the LR test (Saris & Satorra, 1993; Satorra & Saris, 1985).

Results

We first traced power curves as a function of number of measurement occasions and reliability for six conditions: slope correlations of .25, .50, and .75 crossed with sample sizes of 200 and 500. Figure 1 plots these curves for each number of longitudinal occasions (from left to right, 20, 10, 6, 5, 4, and 3). The curves for the larger sample size ($N = 500$) are shown in the top row; the curves for the smaller sample size ($N = 200$) are shown in the bottom row. The curves were generated by computing Satorra–Saris power estimates by stepping through GCRs—scaled for occasion $t(0)$, the onset of the longitudinal study—from .50 to .99 in .005 increments.

Power was generally low for GCRs below .90, although this depended on sample size and effect size, as expected. For the slope correlation of .25 (small effect size), the steepest rise in the power curves occurred for reliabilities greater than .91, and the plots reveal the rapidly ascending behavior of the sigmoidal power curves in this region. With a sample size of 200, power did not exceed .80 until reliability was nearly perfect ($>.96$) for longitudinal designs with 3, 4, 5, or 6 occasions of measurement and achieved .80 when reliability was approximately .90 for 10 occasions of measurement. When sample size increased to 500 for this effect size, the picture improved somewhat, with a 6-occasion longitudinal design achieving .80 power with a GCR of about .92. The increase in power as a function of effect size is seen by going from left to right within a row of the figure. With a sample size of 500 for the moderate effect size, power of .80 was achieved with GCR of .82 for 6 occasions of measurement and with GCR of .93 with 4 occasions of measurement. With the largest effect size (slope correlation of .75), power for a sample of 500 persons rose to .80 with

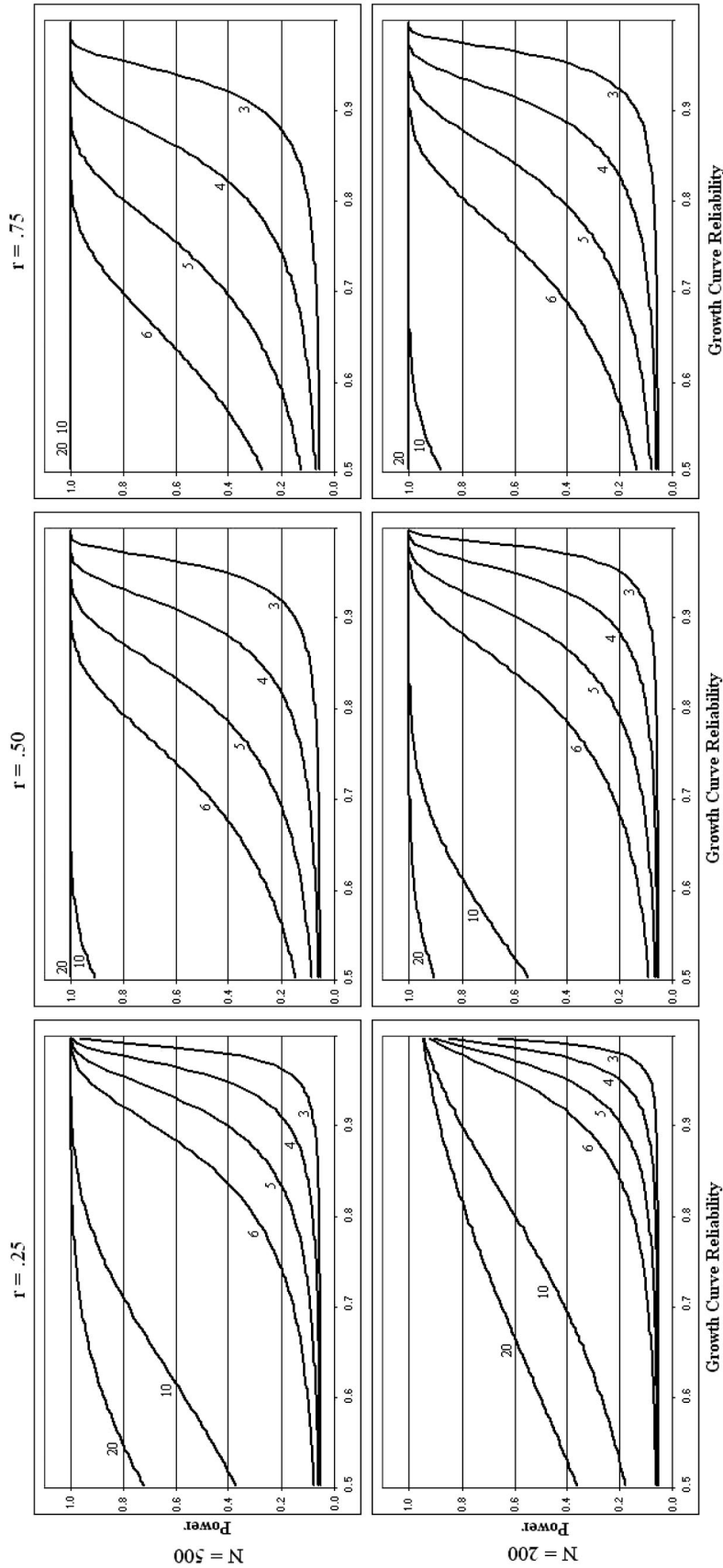


Figure 1. Satorra–Saris power curves mapping reliability (x-axis) against power (y-axis) for different numbers of longitudinal occasions of measurement (organized from left to right as 20, 10, 6, 5, 4, and 3 occasions, as numbered). The top row corresponds to a sample size of 500, the bottom row to a sample size of 200. The columns represent slope correlation effect sizes (from left to right) of .25, .50, and .75.

4 occasions of measurement and reliability of just under .90. Clearly, use of a bivariate LGCM requires sample sizes greater than 500 and good GCRs to achieve acceptable power to detect slope covariances with only 3 or 4 occasions of measurement, assuming all other assumptions hold.

To gain further understanding of the power to detect slope covariances, we estimated a second set of Satorra–Saris power curves that systematically varied the magnitude of the implied slope correlation from 0.0 to 1.0 for a sample of 500 persons. Figure 2 presents these curves. Each panel reports a different number of longitudinal occasions of measurement, varying from 3 to 10. Separate curves are plotted for GCRs at $t(0)$ of .75, .80, .85, .90, and .95 within each panel.

When the number of longitudinal occasions of measurement is 3, power remains relatively low in all cases except when GCR is .99. As the number of occasions increases, the power curves improve, reaching the benchmark of .80 with smaller effect sizes and GCRs. With 4 occasions of measurement and a GCR of .91, better than .80 power is achieved to detect a slope correlation of approximately .70. With 5 occasions of measurement and a GCR of .91, .80 power is found for a slope correlation of .40. As the number of occasions increases, power for lower levels of GCRs becomes more satisfactory, and with 10 occasions of measurement, the curves for the different levels of GCRs lie relatively close to one another. Nevertheless, it is clear that with 5 or fewer longitudinal occasions (e.g., the norm for actual longitudinal studies in the cognitive aging literature), power remains quite low for even high levels of GCRs until effect size becomes very large.

Discussion

The results of this study show that the power to detect individual differences in change as well as correlations among changes between two variables is often disappointingly low. We had anticipated that sample size and the number of longitudinal occasions of measurement would affect power, as they did. What was unexpected was the degree of influence of GCR on power in LGCMs. LR tests of slope covariance were affected by measurement error to a degree that we did not foresee.

It is well-known that arguing for the null hypothesis is problematic when statistical power is low (e.g., Cohen, 1994). Our results demonstrate that failures to reject the null hypotheses of no covariance in change often do not permit any firm conclusions about the absence of such effects in reality. Even with moderate or large change-related effect sizes, LGCMs failed to detect these effects under many combinations of sample size, number of measurement occasions, and GCR. When one is evaluating theoretical propositions about the degree of commonality of change in various attributes within and across domains of functioning,

this vexing fact of low statistical power needs to be kept in mind.

In cognitive aging research, for instance, LGCM analyses of longitudinal studies with relatively large sample sizes, extended longitudinal observation periods, and high measurement reliability have tended to yield significant estimates of variance and covariance in change. For instance, Wilson et al. (2002) detected significant correlations among LGCM slopes in a sample of over 600 persons and with up to 6 years of longitudinal data, using composite variables of multiple psychometric tests for each construct that undoubtedly had high measurement reliability. Be that as it may, studies that conclude that there is a lack of individual differences in slopes or that there are no correlations among slopes between variables should be treated with skepticism unless the power of the study design to detect such effects is evaluated.

The results indicated that statistical power to detect covariance in change is generally most satisfactory when GCR is above .90. The two components of GCR are variation in each person's scores around his or her intraindividual regression line and random measurement error. The former is a characteristic of persons developing in contexts, whereas the latter is a function of the scientist's measurement choices and hence can be manipulated to improve GCR. One can speculate that if multiple indicators of a construct are measured so that a multiple-indicator measurement model can be specified at each occasion, then an LGCM on the latent rather than manifest variables (see McArdle, 1988; Sayer & Cumsille, 2001; see also Christensen et al., 2004) would produce improved statistical power for a given sample size and number of occasions of measurement. In this case, one should, in effect, have corrected for random measurement error, leaving only latent regression residuals to influence the power of the test for correlated slopes.

If our single-indicator results generalize to multiple-indicator LGCMs, these results may explain an apparent discrepancy in the literature on cognitive development in adulthood. In contrast to some null results with single-indicator LGCM studies, latent difference score models (using multiple indicators to estimate differences in latent constructs; see Hertzog & Nesselroade, 2003; McArdle & Nesselroade, 1994) have detected reliable correlations of change in cognitive latent variables over relatively short time periods (e.g., 6 years) with samples under 500 persons (Hertzog et al., 2003; Hultsch et al., 1998; Lövdén, Rönnlund, et al., 2004). With the same approach, Raz et al. (2005) were able to detect reliable variance and covariance in brain volume changes among 72 adults measured over 5 years. These results are consistent with the hypothesis that multiple-indicator LGCMs should make it possible to detect covariance in change among constructs over shorter periods of time and with smaller samples.

The results of this study can be helpful in making longi-

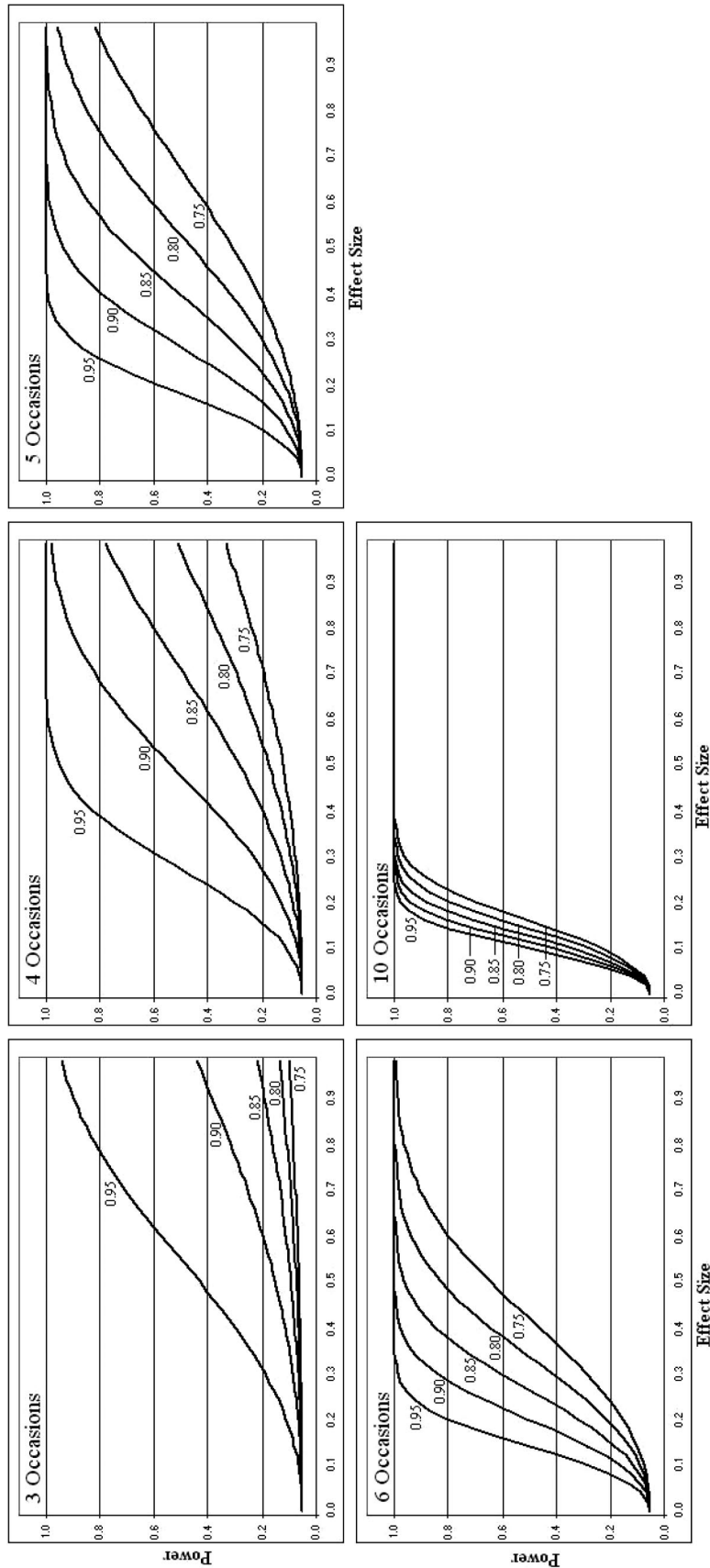


Figure 2. Satorra-Saris power curves mapping effect size (x-axis), represented by slope correlation ranging from .00 to 1.00, against power (y-axis) for different numbers of longitudinal occasions of measurement, organized from left to right and top to bottom as 3, 4, 5, 6, and 10 occasions, respectively. In each panel, the different power curves correspond to growth curve reliability at $t(0)$ (from left to right) of .95, .90, .85, .80, and .75, respectively (as numbered).

tudinal design decisions. The problem, in part, is how to trade off sample size for study duration. If the researcher desires a reasonable opportunity to detect correlations of slopes with fewer than 10 occasions of measurement (in our arbitrary example metric, 18 years of elapsed time), then the power curves suggest that a large sample size ($N = 500$ or greater) is critical, especially if the true correlation of change is modest (.25) rather than large (.75). Most developmental psychologists would shudder at needing this long a temporal epoch to have the power to detect correlations among slopes. Increasing the sample size would, in turn, greatly reduce the number of longitudinal occasions needed to have excellent power to detect correlations among slopes.

Of course, this conclusion is strongly moderated by the GCR of the variables. The unexpectedly strong influence of GCR on the power curves suggests that one must have measures with, at a minimum, a GCR of .91 or greater to have adequate power with even 10 occasions of measurement. In fact, the effect of GCR is so potent that a longitudinal researcher might even be better off investing several potential retest cycles (e.g., 2 to 4 years) in refining measurements and enhancing their reliability before the inception of a longitudinal panel study (see Nesselroade, 1988, for further discussion of longitudinal design decisions from a selection perspective). There are boundaries on how much reliability can be increased for different kinds of measures, however. Hence, increasing GCR through the use of multiple-indicator models could save many cases and several temporal epochs of longitudinal sampling to achieve adequate power to detect correlations of change. In addition, the use of multiple indicators also enhances construct validity in the sense that it increases the likelihood of identifying the center of the construct domain in question (e.g., Little, Lindenberger, & Nesselroade, 1999).

One problem for using the present results to inform about power is that in actual longitudinal research, nonrandom attrition from the sample reduces the size of the remaining sample as a function of the number of occasions of measurement (e.g., Schaie, 1977). The present results cannot speak to effects of nonrandom attrition. However, one can compensate for sample size reductions, assuming random attrition, in a crude sense: One can assume a particular attrition rate and base power estimates on the complete sample available at the end of the study. So, for example, if one assumes that a 50% attrition rate occurs over the desired sampling epoch, then one needs to double the estimated sample size at the inception of the study.

The power curves we reported here can be used to illustrate provisional longitudinal design decisions (but see Kim, 2005). For example, say that a researcher wishes to have .80 or better power to detect a slope correlation of .50. Figure 1 indicates that a sample of 500 persons would need to be collected over four occasions of measurement (6 years elapsed time) on variables generating GCR of .93 to have

.80 power to detect this correlation. With five occasions of measurement, .80 power would be achieved with GCR of .88. However, if GCR were .99, Figure 1 indicates that a sample size of 200 persons could be assessed over four occasions of measurement (6 years elapsed time) with power of 1.0 to detect the same slope correlation.

Likewise, Figure 2 enables a more fine-grained evaluation of statistical power for an expected effect size (slope correlation). Say that the investigator is interested in detecting a slope correlation of .5 with .80 power in a study with sample size of 500 but is only willing to assume a GCR of .80. The curves indicate that the investigator will not be able to achieve this level of power until measurements are collected on five or more occasions. If GCR is .91, then this level of power would be achieved with four occasions of measurement.

Researchers contemplating longitudinal designs can and should use their own specific conjectures on parameter values by the Satorra–Saris approximation (Kim, 2005) and by Monte Carlo simulation (which is readily available in Mplus; see Muthén & Muthén, 2002). This will allow them to generate a priori power estimates to guide longitudinal design decisions.

Given that longitudinal studies often involve small samples, limited numbers of measurement occasions, and less-than-perfect reliability, the present results are sobering. Perhaps even more sobering is the likelihood that the present simulation could present a best-case scenario for power, because it treats the basic LGCM assumptions as true (e.g., there is a universal linear functional form of growth; errors are homoscedastic, statistically independent of latent change, and uncorrelated with themselves and each other over time). Further evaluation of the effects of assumption violations on LGCM power is needed. The present study has a number of limitations. We did not fully explore the universe of possible combinations of LGCM parameter values and change-related effect sizes. Furthermore, there are a number of other LGCM models and applications that we have not evaluated, such as the introduction of exogenous covariates that predict variance in intercept and change 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, Ferrer-Caja, Hamagami, & Woodcock, 2002). Hence one should not overgeneralize our findings to the statistical power of other classes of developmental models for detecting relations among change processes. Such issues remain open empirical questions that can and should be explored in future research.

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 Press.
- Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1988). *Life-span developmental psychology: Introduction to research methods*. Hillsdale, NJ: Erlbaum.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147–158.
- Christensen, H., Mackinnon, A., Jorm, A. F., Korten, A., Jacomb, P., Hofer, S. M., & Henderson, S. (2004). The Canberra Longitudinal Study: Design, aims, methodology, outcomes, and recent empirical investigations. *Aging, Neuropsychology, and Cognition*, *11*, 169–195.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997–1003.
- 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: Erlbaum.
- 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., & Lindenberger, U. (2003). Age-based structural dynamics between perceptual speed and knowledge in the Berlin Aging Study: Direct evidence for ability dedifferentiation in old age. *Psychology and Aging*, *18*, 696–713.
- Hertzog, C. (1985). An individual differences perspective: Implications for cognitive research in gerontology. *Research on Aging*, *7*, 7–45.
- Hertzog, C., Dixon, R. A., Hulstsch, D. F., & MacDonald, S. W. S. (2003). Latent change models of adult cognition: Are changes in processing speed and working memory associated with changes in episodic memory? *Psychology and Aging*, *18*, 755–769.
- Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging*, *18*, 639–657.
- Hertzog, C., & Schaie, K. W. (1986). Stability and change in adult intelligence: 1. Analysis of longitudinal covariance structures. *Psychology and Aging*, *1*, 159–171.
- Hofer, S. M., & Sliwinski, M. J. (2006). Design and analysis of longitudinal studies of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (6th ed., pp. 15–37). San Diego, CA: Academic Press.
- Hulstsch, 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.
- Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging*, *9*, 339–355.
- Lindenberger, U., & Oertzen, T. von (2006). Variability in cognitive aging: From taxonomy to theory. In F. I. M. Craik & E. Bialystok (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 297–314). Oxford, UK: Oxford University Press.
- Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When “good” indicators are bad and “bad” indicators are good. *Psychological Methods*, *4*, 192–211.
- Lövdén, M., Ghisletta, P., & Lindenberger, U. (2004). Cognition in the Berlin Aging Study (BASE): The first 10 years. *Aging, Neuropsychology, and Cognition*, *11*, 104–133.
- 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. *Journals 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 Press.
- 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, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*, 115–142.
- McArdle, J. J., & Nesselroade, J. R. (1994). Structuring data to study development and change. In S. H. Cohen & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological contributions* (pp. 223–268). Hillsdale, NJ: Erlbaum.
- 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. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*, 599–620.
- Nesselroade, J. R. (1988). Sampling and generalizability: Adult development and aging research issues examined within the general methodological framework of selection. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 13–42). New York: Springer.
- Rabbitt, P. (1993). Does it all go together when it goes? The 19th Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *46(A)*, 385–434.
- Rabbitt, P. M. A., Diggle, P., Smith, D., Holland, F., & McInnes, L. (2001). Identifying and separating the effects of practice and of cognitive ageing during a large longitudinal study of elderly community residents. *Neuropsychologia*, *39*, 532–543.
- Raykov, T. (1993). A structural equation model for measuring

- residualized change and discerning patterns of growth or decline. *Applied Psychological Measurement*, *17*, 53–71.
- Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., et al. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex*, *15*, 1676–1689.
- Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92*, 726–748.
- 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. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research*, *35*, 51–88.
- Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 181–204). Newbury Park, CA: Sage.
- Satorra, A., & Saris, W. E. (1985). The power of the likelihood ratio 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. (1996). *Intellectual development in adulthood: The Seattle Longitudinal Study*. New York: Cambridge University Press.
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square tests. *Psychometrika*, *50*, 253–264.
- Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging*, *17*, 179–193.
- Wohlwill, J. F. (1991). The partial isomorphism between developmental theory and methods. *Annals of Theoretical Psychology*, *6*, 1–43.

Received December 13, 2004

Revision received May 10, 2006

Accepted May 17, 2006 ■