Research Report

Department of Statistics

No. 2007:2

Slope Differences in Latent Growth Curve Models under General Measurement Schemes

Linda Wänström

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Slope Differences in Latent Growth Curve Models under General Measurement Schemes

Linda Wänström
Stockholm University

An algebraic expression for the variance of the estimate of slope differences in latent growth curve models was presented in Wänström (2007). This expression can be used in formulas to determine needed sample sizes. In this paper, we present more general algebraic variance expressions that allow researchers to evaluate factors such as spacing of occasions, indicator reliabilities, factor variances and covariances, as well as attrition. Not surprisingly, sample sizes decrease with indicator reliabilities and number of indicators, and increase with loss of observations, e.g with attrition. Observations close to the end points are more important than middle observations, and later observations are more important than earlier observations in cases of equal baseline levels in groups. Correlations between baseline levels and growth as well as between indicator residuals may either increase or decrease needed sample size.

Key words: sample size; latent growth curve models; slope differences; sample size calculations.

1. Introduction

Researchers are often interested in the nature of change in a population, the average individual growth or decline of one or several variables, the variation around this change, as well as differences in change between groups. Consider for example the IQ scores of a group of children followed over time. One may want to examine the children's average change in IQ per year, look at variations in change across all children, or at differences between groups such as boys or girls, or children randomized to a treatment group and a control group.

There are advantages to conducting longitudinal studies as opposed to cross-sectional studies when the aim is to study change. Assessing units, e.g. individuals, repeatedly allows us to distinguish between variation between individuals at one point in time as well as variation within individuals over time (e.g. Diggle, Heagerty, Liang, & Zeger, 2002). Longitudinal studies can be more expensive, however, because we assess individuals several times. Observing a small number of individuals, on the other hand, can result in poor power and meaningful effects may be neglected (see e.g. discussion in Tran, 1997). Sample size calculations can be used to aid in determining the number of individuals to observe as well as the number of occasions that they should be observed.

In this paper, we will present algebraic variance expressions that can be used in sample size calculations when the aim is to examine group differences in slope means using latent growth curve models. These expressions can be used for multiple indicator designs, for designs in which we assume equal baseline levels in the groups, as well as for designs with individually varying times of observation. We will illustrate the expressions in sample size calculations and discuss various effects on sample size. Before looking at the formulas, however, we will provide some background on LGC models and sample size estimation.

2. Background

2.1 Latent growth curve (LGC) models

Latent growth curve (LGC) models (McArdle, 1988; Meredith & Tisak, 1990; Rao, 1958; Tucker, 1958), also referred to as latent trajectory models, have been widely used in the study of average change as well as variation around the change. They can be thought of as confirmatory factor analysis models where individual growth or decline is captured by latent factors with (usually) non-zero means and some fixed loadings (Meredith & Tisak, 1990; Muthén & Curran, 1997). Observed scores at repeated occasions are thought to reflect the underlying latent variables. For example, assuming linear growth, an individual's score at each occasion is the result of his or her latent level, latent growth, and latent "error".

LGC models are closely related to multilevel models, hierarchical linear models, random effects and other similar models (for descriptions of these models, see e.g. Goldstein, 2003; Bryk & Raudenbush, 1992; Longford, 1993) and the factor means in the LGC models correspond to the fixed effects, whereas the factor variances and covariances correspond to the random effects in these models. Comparisons among LGC and multilevel models can be found e.g. in Hox and Stoel (2005), Stoel, van der Wittenboer and Hox (2003), and Raudenbush (2002). LGC model parameters can be estimated through structural equation modeling methods (SEM: Bollen, 1989; Jöreskog & Sörbom, 1979, 1993; Loehlin, 1992) and SEM software can be used (e.g. LISREL, Mplus, Amos, Mx etc.). Parameters can also be estimated through multilevel software (e.g. HLM and MLwiN).

When growth in latent constructs is considered, second order LGC models (Duncan & Duncan, 1996; McArdle, 1988) can be employed. A construct, such as cognitive ability, is then measured by one or several indicators at each occasion. Assuming linear growth, an individual's score on each indicator is the result of an underlying (first order) factor as well as an error term. In turn, his or her score on the underlying (first order) factor is the result of an underlying level and growth factor as well as an error term. These models enable researchers to estimate reliabilities of the indicators, and also to test for measurement equalities over time and across groups (see e.g. Sayer and Cumsille (2002) for a description of factorial invariance testing over time for second order LGC models). They decompose the residual variance at each occasion into variance around the growth curve and variance specific to measurement.

A description of a second order LGC model will be given below, however for more thorough descriptions of first order models, see e.g. Duncan, Duncan, Strycker, Li, and Alpert (1999), McArdle (1988), McArdle and Bell (2000), Raykov (2000), and Willet and Sayer (1994). Also, see Sayer and Cumsille (2002) for descriptions of second order models.

2.2. Sample size

We might be interested in knowing the number of individuals needed to detect group differences in levels or growth. We might also be interested in knowing how the number of measurement occasions and their spacing, the number of indicators and their reliabilities, expected attrition of participants over time etc. will affect the total number of individuals needed. Power and sample sizes in LGC models have been studied, for example, by Curran and Muthén (1999), Fan and Fan (2005), Fan (2003), Hertzog, Ghisletta, Lindenberger and Oertzen (2006), Muthén and Curran (1997), and Wänström (2007). Commonly used techniques for doing this include Monte Carlo simulations and approximations using Satorra and Saris' (1985) technique. Descriptions of Satorra and Saris' technique can be found in e.g. Duncan et al. (1999) and Muthén and Curran (1997).

Researchers have also developed sample size formulas to aid in decisions about sample size for various models. Sample size formulas for cross-sectional designs are widely developed (see e.g. Desu & Raghavarao, 1990). Algebraic formulas have also been provided for longitudinal models (e.g. Diggle, Heagerty, Liang, & Zeger, 2002; Liu & Liang, 1997; Liu, Boyett & Xiong, 2000; Liu, Shih, & Gehan, 2002; Rochon, 1991). Raudenbush and Liu (2000; 2001) developed variance formulas for group differences in means and trends in hierarchical linear models that they used to study the power of these models. Their formulas allow researchers to explore the effects of duration of the study and frequency of observations, among other factors, on sample size. Wänström (2007) extended their variance formulas for slope differences to account for growth in latent constructs. In addition, the formulas in Wänström (2007) can be used in situations where baseline levels can be assumed equal for the groups. These formulas are useful for designs in which participants are observed at equal and equidistant occasions. In this paper we will present more general algebraic variance formulas that can be used for arbitrary measurement occasions, and for arbitrary factor and residual variances.

Measuring the construct of interest with high reliability usually results in lower needed sample sizes. For example, Wänström (2007) noted that needed sample sizes decreased for higher reliabilities and that more indicators can make up some for lower reliabilities. The variance formulas presented in this paper allow for unequal reliabilities of indicators and we can explore effects of adding indicators with various reliabilities to models.

An intervention may affect the slope mean for the treatment group, however it may also affect the variance of the slope or the covariance between the intercept and the slope (e.g. Muthén & Curran, 1997). A treatment may for example work differently for different individuals and increase the variance in the treatment group, or it may work differently for individuals at different levels, producing different covariances between intercepts and slopes for treatment and control. Our formula allows for general and unequal factor variances and covariances.

In second order LGC models, the indicator residuals (measurement errors) may be correlated over time such that the residual for indicator one at the first occasion is correlated with the residual for indicator one at the second occasion etc. We will look at a simple case of autocorrelated residuals. Results for longitudinal models in which the covariance structure of the repeated measures is compound symmetric have observed that variance and sample size decreases for higher correlations between the repeated measures (e.g. Diggle, Heagerty, Liang, & Zeger, 2002). Results for models with AR(1) structures, however, have found that the relationship is more complex. For example, Yi & Panzarella (2002) noted that variance decreased for increases in the correlation for models with few occasions, and that the variance first increases with increases in the correlation, and then decreases, for models with more occasions, the maximum being obtained later for more occasions.

Not surprisingly, power increases (or equivalently, sample size decreases) as intermittent observations are added (e.g. Maxwell, 1998; Raudenbush & Liu, 2001; Wänström, 2007; Yi & Panzarella, 2002). Power and sample size will also depend on the spacing of the added observations. The formulas presented in this paper allow for exploration of effects of spacing on sample size for LGC models. We will look at some illustrations and examine designs with several different spacings.

All participants may not be observed at all occasions, by design or because of dropout for instance. Effects of attrition on power and sample size have been studied in longitudinal models (e.g. Rochon, 1998; Yi & Panzarella, 2002; Zucker & Denne, 2002). For example, Muthén and Muthén (2002) showed how to use simulations to study effects of attrition on power and sample size requirements for latent variable models. Rochon (1998) presented matrix variance formulas based on GEE methodology and showed how to incorporate expected attrition into sample size formulas. Yi and Panzarella (2002) studied effects of attrition on sample size requirements for random effect models, and they also presented matrix variance formulas. They noted that attrition increases the required sample sizes, and simulations also suggested that missing observations at the end of the study have larger effects on sample size. Our formula allows for an unequal number of measurements for individuals, and we can explore effects of attrition on sample size in LGC models. In addition, we can evaluate some simple rules of thumb for sample size adjustment in cases of attrition, as well as evaluate effects of listwise deletion on sample size.

3. The model

3.1 One group

We described a second order LGC model in Wänström (2007), however we assumed that all individuals were observed on all occasions. Let us now assume that we will observe individuals on occasions 1, …, *T*, but that all individuals may not be observed on all occasions. We observe individual *i*, *i* = 1, …, *n*, with *K* indicators on T_i occasions. Let x_{ti} be the elapsed time since the start of the study for individual *i* at occasion *t*. Further, let y_{itk} denote the observed value for individual *i* on indicator *k* at occasion *t*.

A second order LGC model consists of two types of latent variables: *T* first order factors $\eta_h=(\eta_1, \ldots, \eta_T)'$, and *J* second order factors $\Pi=(\pi_1, \ldots, \pi_J)'$. The model may be written in terms of a structural model that describes the structure among the latent variables, and a measurement model that relates the measurements to the latent variables. A structural model, where Π and η_h are stacked above each other, may look as follows for individual *i*

$$\begin{pmatrix}\eta_h \\ \Pi\end{pmatrix} = \begin{pmatrix}\alpha_h \\ \alpha_p\end{pmatrix} + \begin{pmatrix}0 & B_h \\ 0 & 0\end{pmatrix}\begin{pmatrix}\eta_h \\ \Pi\end{pmatrix} + \begin{pmatrix}\varsigma_h \\ \varsigma_p\end{pmatrix} \quad (1)$$

where η_h is a column vector with the individual's latent level at the T_i measurement occasions, α_h is an intercept vector, B_h is a matrix with T_ix*J* factor loadings and ζ_h is a random residual vector with mean 0 and covariance matrix $\Sigma_{\eta i}$ that may be different between individuals because of different measurement occasions. The *J* latent growth factors in Π are assumed to come from a normal distribution with mean vector $\alpha_p=(\alpha_1, \ldots, \alpha_J)'$ and covariance matrix Σ_π containing the variances and covariances of the growth factors. Since we cannot estimate both α_h and α_p separately, we will set $\alpha_h = 0$.

We can write (1) as[1]

$$\eta = \alpha + B\eta + \zeta \qquad (2)$$

where the column vector η contains the T_i+J latent factors, α is a T_i+J vector of factor means, B is a $(T_i+J)\times(T_i+J)$ matrix of factor loadings and ζ is a T_i+J vector of latent residuals which is normal with zero mean and covariance matrix Ψ_i.

A measurement model relates the measurements to the latent variables

$$Y = \tau + \Lambda\eta + \varepsilon, \qquad (3)$$

where Y is the column vector containing the KT_i observed measurements, τ is the column vector containing the corresponding KT_i measurement intercepts and Λ is a $KT_i\times(T_i+J)$ matrix containing the factor loadings λ relating the measurements to the factors. We must set the scale of the factors, and one way of doing this is to fix one factor loading for each latent factor to 1. The corresponding intercept in τ can be fixed to 0 for identification purposes. The column vector ε contains KT_i normal error terms with zero mean and covariance matrix Θ_i. We assume that the covariance matrices Θ_i are identical for individuals except for differences caused by different times of measurements.

3.1 An example

Let us consider the simple case where all individuals have been observed on $T = 4$ equidistant occasions[2]. Consider the path diagram in Figure 1 of a second order LGC model. The two latent growth factors, an intercept and a slope, describe linear growth: $\eta_t = \pi_1 + \pi_2 x_t$.

[1] See also Sayer and Cumsille (2002).
[2] This case was also considered in Wänströn (2007).

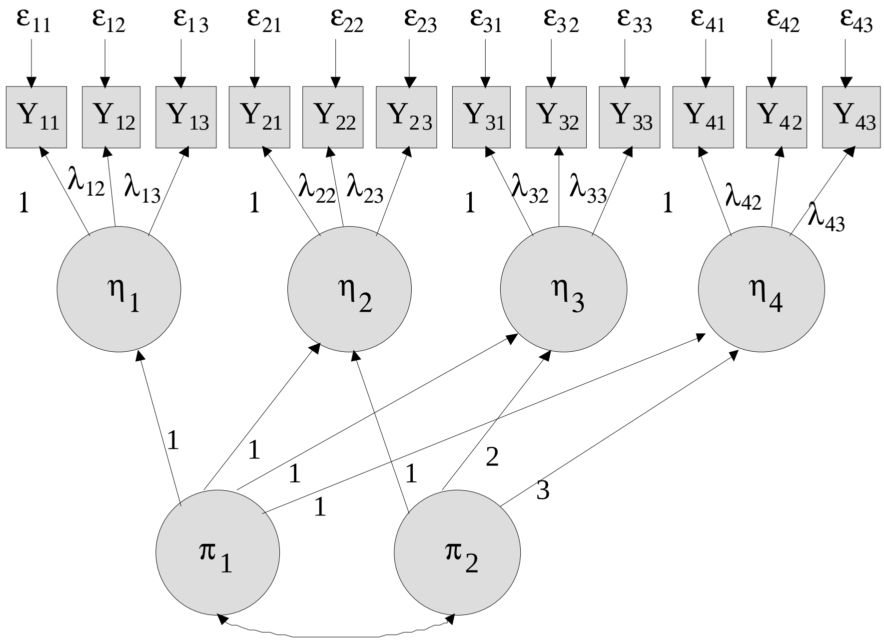

Figure 1. Path diagram of a second order LGC model. Circles indicate latent variables, squares indicate observed variables, one-headed arrows indicate regression coefficients, and double headed arrows indicate covariances.

The structural model corresponding to Figure 1 can be written

$$\begin{bmatrix}\eta_1\\\eta_2\\\eta_3\\\eta_4\\\pi_1\\\pi_2\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\\\alpha_1\\\alpha_2\end{bmatrix}+\begin{bmatrix}0&0&0&0&1&0\\0&0&0&0&1&1\\0&0&0&0&1&2\\0&0&0&0&1&3\\0&0&0&0&0&0\\0&0&0&0&0&0\end{bmatrix}\times\begin{bmatrix}\eta_1\\\eta_2\\\eta_3\\\eta_4\\\pi_1\\\pi_2\end{bmatrix}+\begin{bmatrix}\zeta_{\eta_1}\\\zeta_{\eta_2}\\\zeta_{\eta_3}\\\zeta_{\eta_4}\\\zeta_{\pi_1}\\\zeta_{\pi_2}\end{bmatrix}$$

where

$$\mathrm{cov}(\zeta) = \Psi = \begin{bmatrix} \sigma^2_{\eta_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{\eta_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{\eta_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{\eta_4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\pi_1} & \sigma_{\pi_1\pi_2} \\ 0 & 0 & 0 & 0 & \sigma_{\pi_1\pi_2} & \sigma^2_{\pi_1} \end{bmatrix}.$$

The measurement model can be written

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{41} \\ Y_{42} \\ Y_{43} \end{bmatrix} = \begin{bmatrix} 0 \\ \tau_{12} \\ \tau_{13} \\ 0 \\ \tau_{22} \\ \tau_{23} \\ 0 \\ \tau_{32} \\ \tau_{33} \\ 0 \\ \tau_{42} \\ \tau_{43} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{12} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{13} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{22} & 0 & 0 & 0 & 0 \\ 0 & \lambda_{23} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{32} & 0 & 0 & 0 \\ 0 & 0 & \lambda_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{42} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{43} & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \pi_1 \\ \pi_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \end{bmatrix}$$

where cov(ε) = Θ.

3.3 Indicator reliability

As mentioned previously, second order LGC models can model change in latent constructs. The variance in Y_{tk} can be divided into two parts; the variance explained by the structural model and error variance. The reliability R_{tk} of indicator k at time t can be defined as the variance in Y_{tk} that is explained by the structural model divided by the total variance, which corresponds to reliability in classical test theory (e.g. McDonald, 1999).

$$R_{tk} = \lambda^2_{tk} V(\eta_t) / \left(\lambda^2_{tk} V(\eta_t) + \sigma^2_{\varepsilon tk}\right) \tag{4}$$

where λ_{tk}^2 is the indicator factor loading, $V(\eta_t)$ is the first order factor variance at time t, and $\sigma_{\varepsilon tk}^2$ is the residual variance of indicator k at time t.

3.4 Two or more groups

Sometimes we want to compare the growth of different populations, such as boys and girls, individuals living in urban and rural areas, or different treatment groups in an experimental setting. We extend our notations by adding a superscript (g), g = 1, …, G, denoting group. Equation (2) becomes[3]

$$\eta^{(g)} = \alpha^{(g)} + B^{(g)}\eta^{(g)} + \zeta^{(g)} \tag{5}$$

and equation (3) becomes

$$Y^{(g)} = \tau^{(g)} + \Lambda^{(g)}\eta^{(g)} + \varepsilon^{(g)} \tag{6}$$

where vectors and matrices containing parameters that may differ between groups are indexed by g. Multiple group analyses may be conducted and parameters may be estimated simultaneously for the groups (e.g. Jöreskog & Sörbom, 1979).

3.5 Analysis of the LGC model

The expected value of η from (2) is $E(\eta) = (I - B)^{-1}\,\alpha$ and the variance is $V(\eta) = (I - B)^{-1}\,\Psi\,(I - B)^{-1'}$. The expected value of the observed values is thus from (3)

$$E(Y) = \tau + \Lambda(I - B)^{-1}\alpha \tag{7}$$

and the covariance structure is

$$V(Y) = \Lambda(I - B)^{-1}\Psi(I - B)^{-1'}\Lambda' + \Theta. \tag{8}$$

Since the last J columns of Λ are zero vectors, we may write (7) as

$$E(Y) = \tau + \Lambda_h B_h \alpha_p. \tag{9}$$

Further, (9) may be rewritten $E(Y) = \Lambda_0(\alpha_p,\tau_0)'$ where $\Lambda_0 = (\Lambda_h B_h, I)$.

We may also write (8) as

$$V(Y) = \Lambda_h B_h \Sigma_\Pi B_h' \Lambda_h' + \Lambda_h \Sigma_\eta \Lambda_h' + \Theta. \tag{10}$$

---

3 see e.g. Muthén & Curran (1997).

In Wänström (2007), we presented well known generalized least squares estimates of parameters along with variances of the estimates. When individuals have been measured on unequal occasions, these will not be the same. For each individual, the generalized least square estimate of the parameter $(\alpha_p,\tau_0)'$ is $\left(\Lambda_{0i}' V_i^{-1} \Lambda_{0i}\right)^{-1} \Lambda_{0i}' V_i^{-1} Y_i$ with variance $\left(\Lambda_{0i}' V_i^{-1} \Lambda_{0i}\right)^{-1}$ where V_i is the variance in (8), assuming that all parameters are estimable and that V_i has full rank. For n individuals, the estimate is $\left(\sum_{i=1}^{n} \Lambda_{0i}' V_i^{-1} \Lambda_{0i}\right)^{-1} \sum_{i=1}^{n} \Lambda_{0i}' V_i^{-1} Y_i$. The variance of the estimate is

$$Var\left(\hat{\alpha}_p, \hat{\tau}_0\right)' = \left(\sum_{i=1}^{n} \Lambda_{0i}' V_i^{-1} \Lambda_{0i}\right)^{-1} \tag{11}$$

which simplifies to $\left(\Lambda_0' V^{-1} \Lambda_0\right)^{-1} / n$ if all matrices are equal between individuals.

When we have two or more groups, we might be interested in differences between group parameters. When we have two groups with no common parameters, the estimate of the difference between their parameters is the difference between the corresponding group estimates and the variance is the sum of the variances in the groups

$$\begin{gathered} Var\left[\left(\hat{\alpha}_p^{(1)}, \hat{\tau}_0^{(1)}\right)' - \left(\hat{\alpha}_p^{(2)}, \hat{\tau}_0^{(2)}\right)'\right] = \\ \left(\sum_{i=1}^{n^{(1)}} \Lambda_{0i}^{(1)\prime} V_i^{(1)-1} \Lambda_{0i}^{(1)}\right)^{-1} + \left(\sum_{i=1}^{n^{(2)}} \Lambda_{0i}^{(2)\prime} V_i^{(2)-1} \Lambda_{0i}^{(2)}\right)^{-1} \end{gathered} \tag{12}$$

which simplifies to $\left(\Lambda_0^{(1)\prime} V^{(1)-1} \Lambda_0^{(1)}\right)^{-1} / n^{(1)} + \left(\Lambda_0^{(2)\prime} V^{(2)-1} \Lambda_0^{(2)}\right)^{-1} / n^{(2)}$ if all matrices are equal for individuals within groups and to $2\left(\Lambda_0' V^{-1} \Lambda_0\right)^{-1} / n$ if all matrices and sample sizes are equal in the two groups as well.

In some cases, we may assume that some parameters are equal in the groups. This is often the case in experimental situations when the groups should be equal at baseline, t = 1. Assume, for example, that we have two groups. Let α_c be the common parameters and $\alpha^{(1)}$ and $\alpha^{(2)}$ be the group-specific parameters. In each group $\left(\hat{\alpha}_c, \hat{\alpha}^{(g)}\right)'$ is normal with mean $(\alpha_c, \alpha^{(g)})'$ and variance $\left(\sum_{i=1}^{n^{(g)}} \Lambda_{0i}^{(g)\prime} V_i^{(g)-1} \Lambda_{0i}^{(g)}\right)^{-1}$. We can stack these expressions above each other and we have that $\left(\hat{\alpha}_c, \hat{\alpha}^{(1)}, \hat{\alpha}_c, \hat{\alpha}^{(2)}\right)'$ is normal with mean $(\alpha_c, \alpha^{(1)}, \alpha_c, \alpha^{(2)})'$ and the covariance matrix is block diagonal with blocks

$\left(\sum_{i=1}^{n^{(g)}} \Lambda_{0i}^{(g)'} V_i^{(g)-1} \Lambda_{0i}^{(g)}\right)^{-1}$. We can combine these to form a least squares estimate of the parameters and to find its variance (see also Wänström, 2007).

4. Sample size, variance, and standardized effect size

4.1 Sample size

In case of normality and known covariance matrix, the usual sample size formula[4] can be used to determine the group sample size needed for tests

$$n = \frac{\left(\lambda_{\alpha/2} + \lambda_{\beta}\right)^2 \sigma^2}{e.s.^2} \quad (13)$$

where n is the sample size per group, α is the significance level, 1-β the power, $\lambda_{\alpha/2}$ and λ_{β} quantiles of the normal curve, σ^2 / n the variance of the estimate, and *e.s.* the effect size or the difference between the alternative and the null hypotheses.

If we want to find the group sample size needed for a certain length of a confidence interval, *h*, formula (13) is changed to

$$n = \frac{\left(2\lambda_{\alpha/2}\right)^2 \sigma^2}{h^2}. \quad (14)$$

Multiple group analyses may be conducted for two or more groups. However, from now on, we will concentrate on two groups where the structural part is linear, $E(\eta_t^{(g)}) = \alpha_1^{(g)} + \alpha_2^{(g)} x_t$. Our focus will be on finding the sample size needed to detect differences in slopes between two groups. The factor of interest will be $\pi_3 = \pi_2^{(2)} - \pi_2^{(1)}$, with mean $\alpha_3 = \alpha_2^{(2)} - \alpha_2^{(1)}$ in a model with two growth factors, an intercept and a slope, such as the model in Figure 1.

4.2 Variance

Several researchers have presented algebraic expressions for the variance of the estimate of slope differences between two groups that applies to first order LGC models when we have equally spaced measurements for all individuals, equal variances and sample sizes in the groups, and no assumptions

[4] This formula can be found in many statistics books, see e.g. Desu & Raghavarao (1990) for an overview of sample size formulas.

on baseline level equalities (Friedman, Furberg, & DeMets, 1985; Raudenbush & Liu, 2001). Wänström (2007) presented an algebraic expression that applies to second order models as well. Assuming that measurements are obtained at times 0, 1, 2, …, T-1, the variance of the estimate is

$$Var(\hat{\pi}_3) = \frac{2}{n}\left(V_2 + \sigma^2_{\pi_2}\right) \tag{15}$$

where

$$V_2 = \frac{\sigma^2_\eta + \sigma^2_\varepsilon / K}{\left(T^3 - T\right)/12}.$$

The first term inside the brackets, V_2, is the conditional variance of the least squares estimate of the individual slope, and the second term inside the bracket is variance between individuals. K is the number of indicators, T is the number of equally spaced measurements for all individuals, $\sigma^2{}_\eta$ is the common first order factor residual variance, and $\sigma^2{}_\varepsilon$ is the common indicator residual variance.

Wänström (2007) also presented an expression that can be used if we have reason to assume that the groups have the same baseline levels (intercept means) and the two factors π_1 and π_2 are independent. In that case, the variance of the estimate is

$$Var(\hat{\pi}_3) = \frac{2}{n}\left(\frac{1}{\frac{1}{V_2} + \frac{\left((T-1)/2\right)^2}{V_1 + \sigma^2_{\pi_1}}} + \sigma^2_{\pi_2}\right) \tag{16}$$

where

$$V_1 = \frac{\sigma^2_\eta + \sigma^2_\varepsilon / K}{T}.$$

The numerators in formulas (15) and (16) can be used in (13) or (14) to find the required sample sizes. These formulas were explored in Wänström (2007) and required sample sizes decreased for larger effect sizes, indicator reliabilities, number of indicators, frequency of observation, and duration of study. Sample sizes were also larger in designs with no assumptions on equality of baseline levels.

4.3 Standardized effect size

When our interest is in slope differences between groups, the effect size is the mean difference in slopes between groups, α_3. Sometimes it is valuable to standardize the effect. Cohen (1988) suggested that a standardized effects size, *d* = .2, is small, whereas *d* = .5 and *d* = .8 correspond to medium and large respectively. There are different ways in which effects can be standardized in longitudinal models. For example, Raudenbush and Liu (2001) used the group slope difference divided by the population standard deviation of the slope

$$d = \alpha_3 \Big/ \sqrt{\sigma^2_{\pi_2}} \tag{17}$$

whereas Muthén & Curran (1997) and Wänström (2007) used the difference between groups at the last time point divided by the standard deviation at the last time point

$$d = \frac{\alpha_3 D}{\sqrt{\sigma^2_{\eta_t} + \sigma^2_{\pi_1} + D^2\sigma^2_{\pi_2} + 2D\sigma_{\pi_1\pi_2}}} \tag{18}$$

where *D* is the intended duration of the study. This will be $T - 1$ if we have measurements at times 0, 1, …, $T-1$.

4.4 Numerical example

We will now use variance formulas (15) and (16) to examine the sample sizes needed for detecting a small slope difference with significance level .05 and power[5] .8. Consider again Figure 1. We will base this as well as coming numerical illustrations on this basic second order LGC model. We will assume $K = 3$ indicators, all with reliability $R = .9$ at $t = 1$ (we assume that $\sigma_\varepsilon^2 = 1/9$ and $\lambda = 1$). In addition, we will use $T = 4$ occasions, and we will assume equal first order factor residual variances $\sigma^2_\eta = 0.5$, and second order factor variances $\sigma^2_{\pi_1} = 0.5$, $\sigma^2_{\pi_2} = 0.1$, and $\sigma_{\pi_1\pi_2} = 0$ (which provides a commonly seen intercept/slope variance ratio (Muthén & Muthén, 2002)). We will use effect size, *e.s.* = $\alpha_3 = 0.092$, obtained from assuming a small standardized effect size *d* = .2 in (18). This corresponds to *d* = .29 using the definition in (17). These parameter values were also used in numerical illustrations in Wänström (2007).

[5] a power level recommended by Cohen (1988) for the social sciences.

With no assumptions on equality in baseline levels, we substitute the numerator in (15) for σ^2 in (13)

$$n = \frac{(\lambda_{\alpha/2} + \lambda_\beta)^2 \sigma^2}{e.s.^2} = \frac{(1.96 + 0.8416)^2 \times 0.4148}{(0.092)^2} = 384.6585$$

where $\sigma^2 = 2(V_2 + \sigma^2_{\pi_2}) = 2(0.1074 + 0.1) = 0.4148$

and $V_2 = \frac{0.5 + 0.111/3}{(4^3 - 4)/12} = 0.1074$.

Assuming equal baseline levels, we substitute the numerator in (16) for σ^2,

$$n = \frac{(\lambda_{\alpha/2} + \lambda_\beta)^2 \sigma^2}{e.s.^2} = \frac{(1.96 + 0.8416)^2 \times 0.3556}{(.092)^2} = 329.7603$$

where

$$\sigma^2 = \frac{2}{\frac{1}{V_2} + \frac{((T-1)/2)^2}{V_1 + \sigma^2_{\pi_1}}} + 2\sigma^2_{\pi_2} = \frac{2}{\frac{1}{0.1074} + \frac{1.5^2}{0.1343 + 0.5}} + 2 \times 0.1 =$$

$= 0.3556$,

$$V_2 = \frac{0.5 + 0.111/3}{(4^3 - 4)/12} = 0.1074,$$

and

$$V_1 = \frac{0.5 + 0.111/3}{4} = 0.1343.$$

As noted, sample sizes are smaller if we can assume equal baseline levels, in which case we will need 330 individuals per group compared to 385.

As mentioned previously, formulas (15) and (16) can be used in situations with equally spaced measurements for all individuals as well as equal residual variances at each time point and for each indicator. We will now extend these to situations with arbitrary factor and/or residual variances and covariances or situations with varying measurement occasions. Our formulas and illustrations will be based on the model in Figure 1 but we will make different changes in the assumptions, such as allowing for unequal residual variances or varying measurement occasions.

5. Arbitrary factor variances and covariances

5.1 Arbitrary reliabilities and residual variances

When the reliabilities of different indicators are unequal, $\sigma_{\varepsilon}^2 / K$ in (15) and (16) should be replaced by $\left(\sum_{k=1}^{K} 1/\sigma_{\varepsilon kt}^2\right)^{-1}$. When first order factor residual variances $\sigma_{\eta_t}^2$ are unequal at each time point as well, formula (16) extends to

$$Var(\hat{\pi}_3) = \frac{2}{n}\left(\frac{1}{\sum_{t=1}^{T}\left(\frac{(x_t - \bar{x})^2}{\sigma_{\hat{Y}_t}^2}\right) + \frac{\bar{x}^2}{\sigma_{\pi_1}^2 + \frac{1}{\sum_{t=1}^{T}\frac{1}{\sigma_{\hat{Y}_t}^2}}}} + \sigma_{\pi_2}^2 \right) \qquad (19)$$

where $\sigma_{\hat{Y}_t}^2 = \sigma_{\eta_t}^2 + \frac{1}{\sum_{k=1}^{K}\frac{1}{\sigma_{\varepsilon kt}^2}}$,

$\bar{x} = \sum_{t=1}^{T}\left(x_t / \sigma_{\hat{Y}_t}^2\right) / \sum_{t=1}^{T}\left(1/\sigma_{\hat{Y}_t}^2\right)$,

x_t is the elapsed time since the start of the study, and *n* is the common sample size per group. If we cannot assume equal baseline levels in groups, the second term in the denominator inside the bracket disappears and the variance increases.

Let us illustrate this formula by examining sample sizes for designs with different number of indicators and indicator reliabilities. Using the parameter values from section 4.4, Figure 2 shows the required group sample sizes when adding one or three indicators of varying reliabilities to a one-indicator model with reliability .9 or .6. The bottom horizontal line (K = 1, $R(1)$ = .9) shows the sample sizes for a one-indicator model with reliability .9. The bottom curves (K = 2 and K = 4) show the sample sizes needed when adding one or three indicators of varying reliabilities, keeping the first at R = .9. As shown, sample sizes do not decrease much when one indicator of low reliability is added to this model. Adding three indicators results in a larger reduction in needed sample sizes, although the decrease is still not large for low reliabilities.

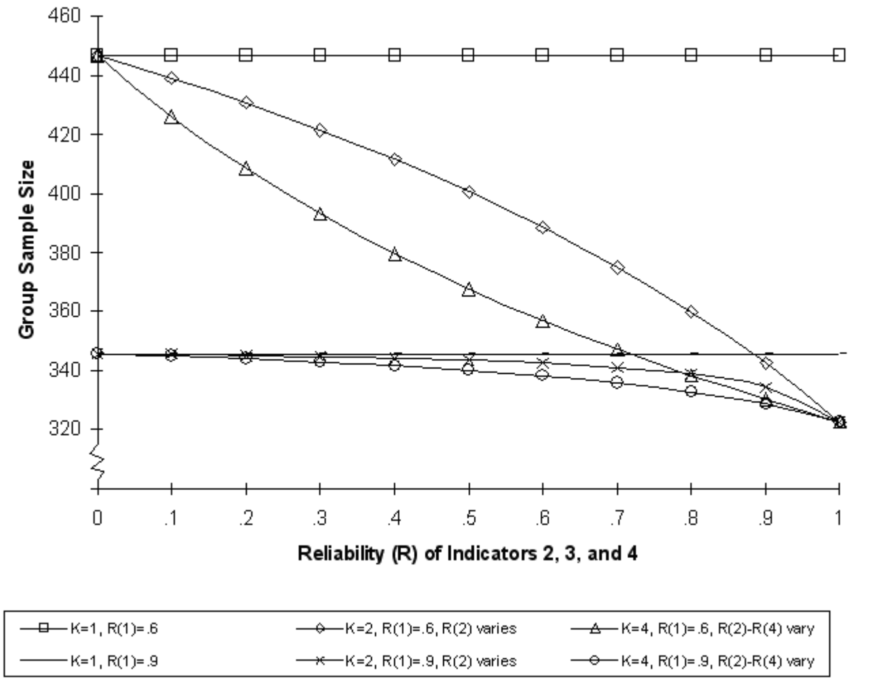

Figure 2. Required group sample sizes for models with different number of indicators of varying reliabilities. The horizontal lines show sample sizes for a one-indicator (K = 1) model with reliability, R, of .6 or .9. The curves show sample sizes for the same models where one or three indicators with varying reliabilities have been added.

The top horizontal line shows sample sizes for a one-indicator model with reliability .6. Not surprisingly, the reduction in sample sizes are larger when one or three indicators are added, even with small reliabilities. We can also see that all curves intersect at one point for reliabilities of 1. We only need one indicator if it has perfect reliability.

We can easily check the number of indicators and reliabilities that are needed for a certain sample size. Models in which $1\Big/\left(\sum_{k=1}^{K} 1\Big/\sigma^2_{\varepsilon kt}\right)$ is the same will give the same sample size. For example, a model with one indicator with reliability .9 (and thus $\sigma^2_{\varepsilon kt}$ = 1/9) will yield the same sample size as a model with 6 indicators with .6 reliability each: (2/(3×6))= 1/9.

5.2 Arbitrary intercept-slope covariances and unequal variances in groups

Often, there is reason to believe that the covariance between the intercept and slope is positive, indicating that those with higher than average scores at baseline also have a higher than average growth. When $\sigma_{\pi_1\pi_2} \neq 0$, the variance formula (16) should be replaced by (see also Wänström, 2007)

$$Var(\hat{\pi}_3) = \frac{2}{n}\left(V_2 + \sigma^2_{\pi_2} - \frac{\left(V_2\left((T-1)/2\right) - \sigma_{\pi_1\pi_2}\right)^2}{V_1 + \sigma^2_{\pi_1} + V_2\left((T-1)/2\right)^2}\right) \qquad (20)$$

where V_1 and V_2 are given after formulas (15) and (16).

Figure 3 shows required group sample sizes for a model with varying indicator reliabilities and intercept – slope correlations ($r = \sigma_{\pi_1\pi_2}\Big/\sqrt{\sigma^2_{\pi_1}\sigma^2_{\pi_2}}$) using parameter values from section 4.4. As seen, sample sizes increase for positive correlations, especially for models with low reliability indicators. For large reliabilities, the sample sizes first increase and then decrease some.

We can see from (20) that the maximum sample size is needed when $V_2(T-1)/2 = \sigma_{\pi_1\pi_2}$, as was also noted in Wänström (2007). For example, when the reliability R = 1, this occurs for $\sigma_{\pi_1\pi_2} = .15$ which corresponds to r = .67. When R is smaller, the maximum occurs for larger covariances and correlations. A minimum sample size is reached for r = -1.

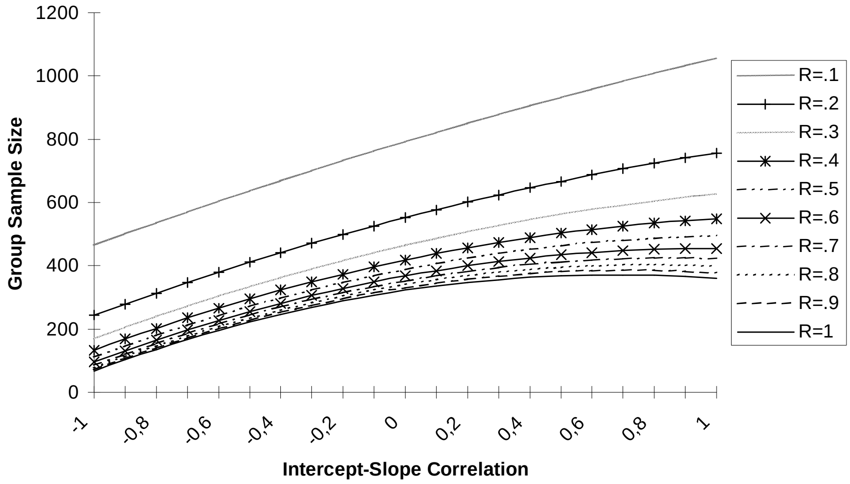

Figure 3. Required group sample sizes for varying, but equal for groups, intercept- and slope correlations, *r*, and varying reliabilities, *R*.

Formula (20) can be used when we assume equal variances and covariances in the groups, thus the factor "2" in the denominator. However variances may differ between groups, such as when the treatment produces variance increases. In addition, the intercept-slope covariance may differ between groups. A treatment might work very well for the already above average individuals. We might then have unequal covariances between the intercept and slope in the groups. We can easily change the formula to handle unequal variances or covariances. A formula in which we assume that the slope variances as well as the intercept-slope covariances differ between groups can be written

$$Var(\hat{\pi}_3) = \frac{1}{n}\left(2V_2 + \sigma_{\pi_2}^{(1)2} + \sigma_{\pi_2}^{(2)2} - \frac{\left(2V_2\left((T-1)/2\right) - \left(\sigma_{\pi_1\pi_2}^{(1)} + \sigma_{\pi_1\pi_2}^{(2)}\right)\right)^2}{2V_1 + 2\sigma_{\pi_1}^2 + 2V_2\left((T-1)/2\right)^2}\right) \quad (21)$$

where the terms are as defined previously and the superscript in parenthesis refers to group one or two.

Figure 4 shows group sample sizes for different intercept – slope correlations for the groups using parameter values from section 4.4 as previously. The middle curve shows situations where group two has zero correlation, whereas the correlation varies for group one. Because only the sum of the covariances matters, all else being equal (see (21)), this middle curve

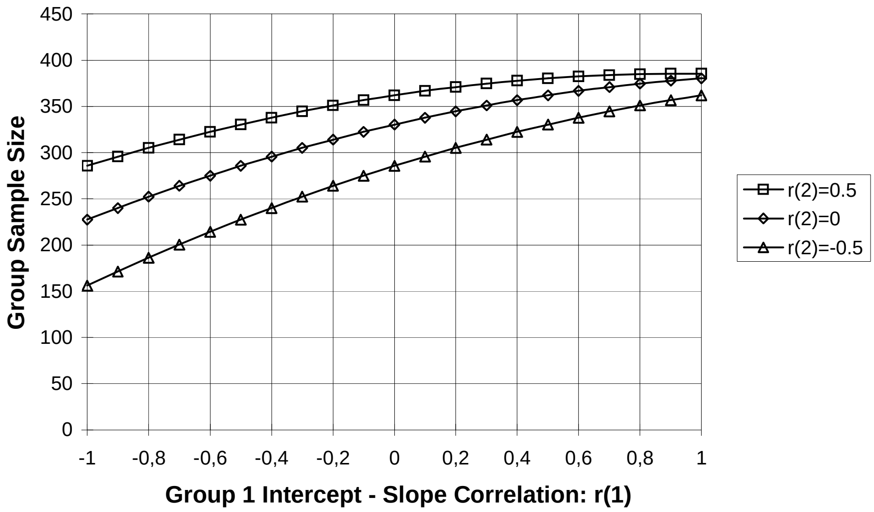

Figure 4. Required group sample sizes for varying intercept- and slope correlations in the groups. Group 1 has intercept – slope correlations r(1) varying between –1 and 1, and group 2 has intercept – slope correlations r(2) = -0.5 (bottom curve), r(2) = 0 (middle curve), and r(2) = 0.5 (top curve).

corresponds to the R = .9 curve in Figure 3 for the interval $-0.5 \le r \le 0.5$. For example, a group one correlation of -1, together with a group two correlation of 0 results in sample sizes 227. This can be found either from Figure 3 for a common intercept – slope correlation of -0.5 and R = .9 (the sum of the correlations is -1), or from the middle curve in Figure 4 for r(1)= -1 (r(2)=0), or the bottom curve in Figure 4 for r(1) = -0.5 (r(2)= -0.5).

5.3 Unequal allocation to groups

Sometimes we want the sample sizes in the groups to be unequal, for cost- or other reasons. Several authors have discussed sample sizes in cases of differential allocation to groups (e.g. Liu, Boyett, & Xiong, 2000; Rochon, 1998). If $n^{(1)}$ and $n^{(2)}$ denote the required number of individuals in group one and two respectively, and if $n^{(1)} = cn^{(2)}$ then the required sample size $n^{(1)}$ can be calculated by the formulas in this paper but the factor *2* is replaced by 1 + *c*.

If the variances in the groups are not equal, we may sample more individuals from the group with the larger variance. For example, Muthén and Curran (1997) examined power for LGC models, and noted that it was greatest for designs that were nearly balanced but with slightly more participants in the group with the larger variance. Using Neyman allocation[6], and given the total sample size, n, we should choose $n^{(1)} = n\sigma^{(1)} /(\sigma^{(1)} + \sigma^{(2)})$, and $n^{(2)} = n\sigma^{(2)} /(\sigma^{(1)} + \sigma^{(2)})$ for optimal allocation, where $\sigma^{(g)}$ is the standard deviation in group g.

5.4 Correlated errors of measurement

Indicator residuals (errors of measurement) may be correlated over time. The variance of the slope difference estimate can be found as element (3,3) in the matrix

$$Var(\hat{\pi}_3) = \frac{1}{n}\left[Z^{'} R^{-1} Z\right]^{-1} \tag{22}$$

where

$$Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1/2 & -1/2 \end{bmatrix}$$

and R is a block diagonal matrix with blocks containing the covariance matrix of the parameter estimates in each group (see Wänström (2007)for more details).

Any error covariance structure can be specified and used with (22). We will illustrate the formula in a simple case where the residual of indicator one at time one is correlated with the residual of indicator one at time two etc. Assume that $K = 1$, that $\sigma_\eta^2 = 0$, and that $\sigma_\varepsilon^2 = 1$. Also assume that an autocorrelation coefficient can be written $\rho^{|x_j - x_l|}$, where $|x_j-x_l|$ is the distance between time points j and l. For equally spaced measurements at times 0, 1, …, T-1, the error covariance matrix is

$$\Theta = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ . & . & . & \cdots & . \\ \rho^{T-1} & . & . & \cdots & 1 \end{bmatrix}.$$

[6] see most statistics books on sampling methodology, e.g. Scheaffer, Mendenhall and Ott (1996).

Figure 5 shows the required group sample sizes for varying number of occasions, T, and varying correlations between the indicator residuals, ρ. When T = 2, the maximum sample size is reached for $\rho = 0$ (not shown in the figure). When $T > 2$, the sample size first increases and then decreases for increasing values of ρ. The maximum is reached for larger values of ρ the larger T is. This agrees with results for models with AR(1) structures (e.g. Yi & Panzarella, 2002). We can also see that the minimum sample size is reached for $\rho = 1$, and that the curves intersect at that point. This corresponds to a situation in which we perfectly measure the slope means.

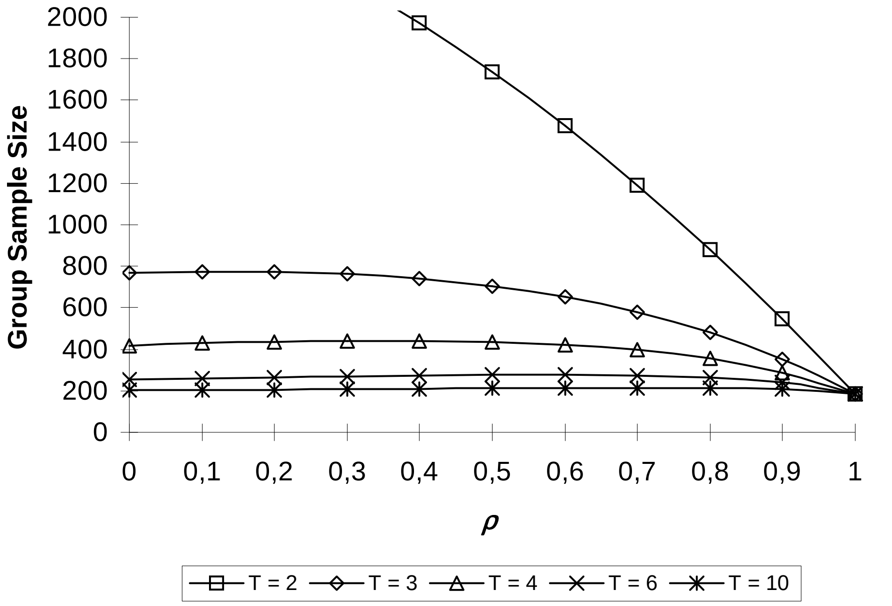

Figure 5. Required group sample sizes for autocorrelated residuals with number of occasions, T, and autocorrelation coefficient ρ for a one indicator model with σ_{ε}^{2} = 1 and σ_{η}^{2} = 0.

Although not shown in Figure 5, the sample sizes will be 330, 330, and 326 for ρ = .1, .5, and .9 respectively for the model in Figure 1 (with K = 3 indicators, reliability R = .9, residual variances σ_{η}^{2} = .5 at each time point, and autocorrelated errors). The required sample sizes are 472, 486, and 404 respectively for R = .3. These may be compared to n = 330 for R = .9 and ρ = 0, and 464 for R = .3 and ρ = 0.

As mentioned, (22) covers any error covariance structure. A special case is when the indicator residuals are correlated with correlation ρ within the same time point but independent between time points. In that case, σ_ε^2 / K in the formulas from the previous sections is replaced by $\left(\sigma_\varepsilon^2 + (K-1)\rho\sigma_\varepsilon^2\right)/K$. More generally, if the covariance matrix of the indicators at time *t* is $\Sigma_{\varepsilon kt}$ one should use $\left(1'\Sigma_{\varepsilon tk}^{-1}1\right)^{-1}$.

6. General measurement occasions

6.1 General and equal measurement occasions

When we have generally spaced measurement occasions which occur at the same time points for all individuals, the variance formula (16) extends to

$$Var(\hat{\pi}_3) = \frac{2}{n}\left(\frac{1}{\frac{1}{V_2^*} + \frac{\bar{x}^{**2}}{V_1 + \sigma_{\pi_1}^2}} + \sigma_{\pi_2}^2\right) \qquad (23)$$

where $V_2^* = \dfrac{\sigma_\eta^2 + \sigma_\varepsilon^2 / K}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$,

$$\bar{x}^{**} = \frac{1}{T}\sum_{t=1}^{T} x_t ,$$

x_t is the elapsed time since the start of the study, and V_1 is as defined previously. If there are no assumptions of equal baselines in groups, the second term in the denominator inside the bracket disappears.

It is easily seen that the variance (23) decreases with increases in $\sum_{t=1}^{T}(x_t - \bar{x})^2$ and $(\sum_{t=1}^{T} x_t)^2$. The first expression indicates that it might be beneficial to place measurement occasions closer to the beginning and end. The second expression indicates that the later the measurement occasion the better. The optimal plan is a compromise between these two factors. If it is unknown whether the groups start at the same level, there is no $(\sum_{t=1}^{T} x_t)^2$ term in the denominator and only the first term remains.

Figure 6 shows required sample sizes where values of x_t (the elapsed time since the start of the study) are 0, .5, 1, 1.5, 2, 2.5, and 3. The top bar shows the required sample size for these 7 equally spaced measurements. The seven “x’s” (“xxxxxxx”) illustrate this situation. The second bar from above shows the case where we remove the second measurement occasion and thus have measurements at times 0, 1, 1.5, 2, 2.5, and 3. The notation “xoxxxxx”

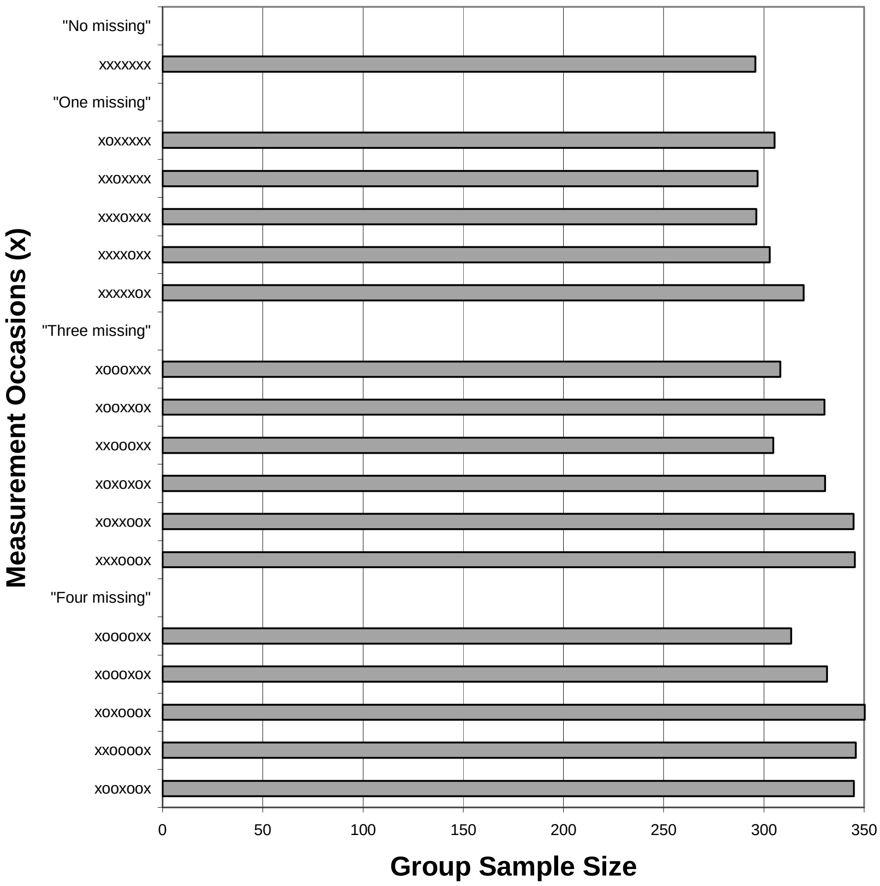

Figure 6. Required group sample sizes for unequally spaced occasions. The top bar shows the “starting model” with 7 equally spaced occasions (where “x” indicates a measurement), and the remaining bars show situations where some occasions have been removed (“o” indicates no measurement).

illustrates this situation where “o” indicates no measurement. In the same manner, the remaining bars show different situations where we have removed different number of occasions at varying places.

In Figure 6, we can see that removing measurement occasions close to the middle does not affect sample size much, whereas the effect is somewhat larger when removing measurement occasions close to the start point. It is worst to remove measurement occasions close to the end point. We can also see, however, that the differences between required sample sizes are not large, most sample sizes are between 300 and 350.

6.2 General and unequal measurement occasions

The formulas above deal, in some sense, with ideal situations where all participants have been observed on all occasions. Using the theory from section 3.5, the variance of the estimate of the slope difference can be calculated when we have individually varying occasions. This is done in the appendix, and the result is

$$Var(\hat{\pi}_3) = 2\left(\sum_{i=1}^{n}\left(\frac{\bar{x}_i^{*2}}{\sigma_{z_i}^2} + \frac{1}{\sigma_{r_i}^2}\right)\right)^{-1} \quad (24)$$

where n is the group sample size, x_{ti} is the elapsed time since the start of the study for individual i at time t,

$\bar{x}_i^* = \bar{x}_i - \sigma_{\bar{Y}_i r_i} / \sigma_{r_i}^2$,

$\bar{x}_i = \sum_{t=1}^{T_i}\left(x_{ti} / \sigma_{\hat{Y}_{ti}}^2\right) \Big/ \sum_{t=1}^{T_i}\left(1 / \sigma_{\hat{Y}_{ti}}^2\right)$,

$\sigma_{\bar{Y}_i r_i} = \sigma_{\pi_1 \pi_2} + \sigma_{\pi_2}^2 \bar{x}_i$,

$\sigma_{ri}^2 = V_{ri} + \sigma_{\pi_2}^2$,

$V_{ri} = \sum_{t=1}^{T_i} \sigma_{\hat{Y}_{ti}}^2 \Big/ \left(x_{ti} - \bar{x}_i\right)^2$,

$\sigma_{\hat{Y}_{ti}}^2 = \sigma_{\eta_{ti}}^2 + 1 \Big/ \sum_{k=1}^{K} 1 / \sigma_{\varepsilon_{kti}}^2$,

$\sigma_{z_i}^2 = \sigma_{\bar{Y}_i}^2 - \sigma_{\bar{Y}_i r_i}^2 / \sigma_{r_i}^2$,

$\sigma_{\bar{Y}_i}^2 = V_{\bar{Y}_i} + \sigma_{\pi_1}^2 + \bar{x}_i^2 \sigma_{\pi_2}^2 + 2\bar{x}_i \sigma_{\pi_1 \pi_2}$, and $V_{\bar{Y}_i} = 1 \Big/ \sum_{t=1}^{T_i} \left(1 \big/ \sigma^2_{\hat{Y}_{ti}} \right)$.

Variance formula (24) also includes arbitrary factor variances and covariances within groups, and arbitrary residual variances over time as well as arbitrary indicator residual variances. It can be shown that formulas (15) and (16) presented in Wänström (2007), as well as formulas (19), (20), and (23) are special cases of (24).

Formula (24) can for instance be used to find the needed sample sizes when we expect attrition during the duration of the study.

6.3 Attrition

Even though we plan to observe individuals on all occasions, we may expect that some individuals will be lost during the study due to attrition. For example, we might expect that individuals will provide a complete set of measurements with a certain probability. We might also expect others to drop out at different time points with certain probabilities.

Assume that the elapsed time since the start of the study is x_t, t = 1, …, T. We also assume that individuals that drop out are lost to further study. In each group, we will then have one subgroup of individuals that has been observed on occasion 1, one that has been observed on occasions 1 and 2, one that has been observed on occasions 1, 2, 3 etc. and one subgroup that has been observed on occasions 1, 2, …, T. Let p_m be the probability that individuals are observed exactly the first m occasions where $\sum_{m=1}^{T} p_m = 1$ (see also Rochon, 1998). For simplicity, we also assume that attrition occurs equally in both groups, and that attrition is MAR (see Little & Rubin, 1987). We also assume that variances are equal in the groups and that the intercept-covariances are zero. For groups with m measurements, the variance of the estimate is

$$Var(\hat{\pi}_3)_m = \frac{2}{n_m} \left(\left(\bar{x}_m^{*2} / \sigma^2_{z_m} \right) + \left(1 / \sigma^2_{r_m} \right) \right)^{-1} \qquad (25)$$

where $\bar{x}_m^* = \bar{x}_m - \bar{x}_m \sigma^2_{\pi_2} / \sigma^2_{r_m}$,

$\sigma^2_{z_m} = V_{\bar{Y}_m} + \sigma^2_{\pi_1} + \sigma^2_{\pi_2} \bar{x}^2_m - \left(\sigma^2_{\pi_2} \bar{x}^2_m \right)^2 / \sigma^2_{r_m}$,

$\sigma^2_{r_m} = V_{r_m} + \sigma^2_{\pi_2}$,

$\bar{x}_m = \sum_{t=1}^{m} x_t / m$,

$V_{\bar{Y}_m} = \left(\sigma_\eta^2 + \sigma_\varepsilon^2 / K\right) / m$,

and $V_{r_m} = \left(\sigma_\eta^2 + \sigma_\varepsilon^2 / K\right) / \sum_{t=1}^{m} \left(x_t - \bar{x}_m\right)$.

We will define $Var\left(\hat{\pi}_3\right)_m = \sigma_m^2 / n_m$. The total variance is then

$$Var\left(\hat{\pi}_3\right) = \frac{1}{n}\left(\sum_{m=2}^{T} p_m / \sigma_m^2\right)^{-1} \qquad (26)$$

Figure 7 shows required group sample sizes for our usual situation with T = 4 equally spaced occasions. The bottom bar shows a situation with no attrition (a response rate of 100% throughout the study), and the remaining bars show some situations with different expectations on attrition. With a 100% response rate, the required sample size per group is 330 as we have found previously.

If we define an observation to be a measurement on an individual at one occasion, bars 2 to 5 from below show situations where we expect 7.5% of the observations to be lost throughout the study but at varying occasions. The second bar from below shows the case where 30% of the individuals drop out at the fourth occasion (we lose 7.5% of the observations at the fourth occasion). The third bar from below shows the case where 15% of the individuals drop out at the third occasion (we lose 3.75% of the observations at the third occasion and 3.75% at the fourth). In the same manner, the fourth bar from below shows the case where 10% drop out at the second occasion, and the fifth bar shows the case where 5% drop out at occasions two, three, and four respectively. We can see that the required sample size increases with attrition, and that the increase depends somewhat on when and how the attrition occurs. For example, loosing all 7.5% of the observations at the last time point (second bar from below) results in a sample size of 375, whereas loosing 2.5% at occasions 2, 3, and 4 respectively (fourth bar from below) results in a sample size of 367. Bars 6 to 9 from below show situations where we expect 15% of the observations to be lost. The same pattern is seen here.

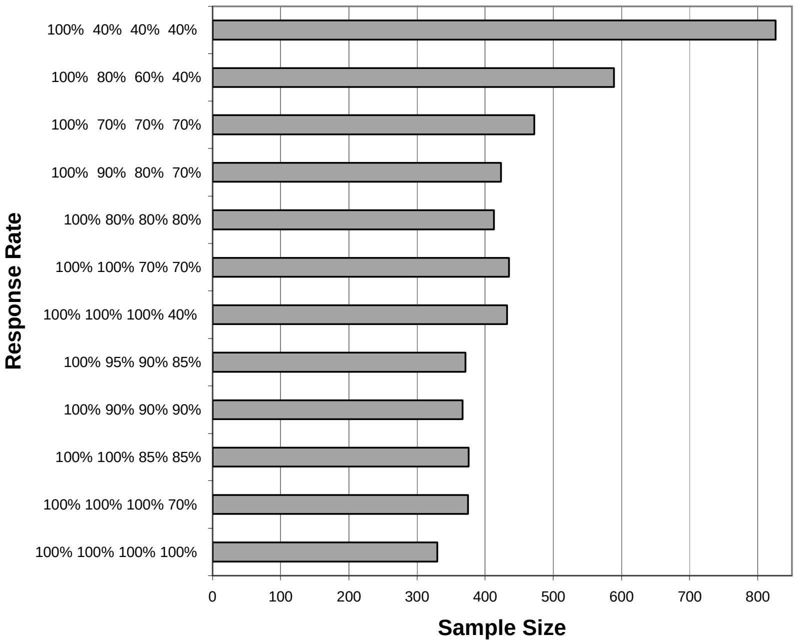

Figure 7. Required group sample sizes for studies with attrition. The X-axis shows the percentage of participants expected to stay in the study at each of four occasions (response rate) and the Y-axis shows the required group sample size.

We can also see, from (26), that individuals who only participate at the first time point do not add any precision to the estimation of differences in slope means across groups. The fourth bar from below, with a 90% response rate at occasions 2, 3, and 4, shows the same sample size as a response rate of 90% at all four occasions: $330/.9 = 366.67$. This can be understood by realizing that participants who are measured at one occasion only add precision to the estimation of the intercept and not to the slope. Even if we could measure the intercept perfectly in the two groups, this would not help us in estimating the difference in slopes. This bar, and the third bar from above, thus also show the required sample size if we were to use listwise deletion, i.e. only keep the 90% or 70% of the individuals respectively who were observed on all occasions. The third bar from above can be compared to the second bar from below where we keep all available individuals. As shown, the difference in required sample size is fairly large, and the required sample size increases substantially if we plan to use only individuals with a complete set of measurements. The increase in sample size is not as large if we use listwise deletion in situations where we loose participants earlier on in the study (compare e.g. bars three and four from above).

Formula (26) is quite complex, especially when the number of measurement occasions, T, is large. Overall, Shobaki, Shivakumar, and Steele (1998) suggested using a simple rule of thumb to adjust the originally computed required sample size for expected attrition: $n^* = n\,(1 + \text{DRP})$, where n^* is the adjusted sample size, n is the sample size obtained from calculations based on zero attrition, and DRP is the dropout rate expected to occur randomly across measurements. Using this rule of thumb we obtain $n = 330$ for the case with no attrition. Adjusting this expecting 15% attrition scattered randomly throughout the measurement occasions give $n^* = 330\ (1 + .15) = 379.5$ individuals per group. This can be compared to $n = 371$ obtained from (26) – see the fifth bar from below. In the same manner, we have $n^* = 429$ (compared to $n = 423$) for 30% attrition – see the fourth bar from above, and $n^* = 528$ (compared to $n = 589$) for 60% attrition – see the second bar from above. As noticed, there are some discrepancies, especially for the case with 60% attrition in which the rule of thumb technique underestimates the required sample size by more than 10%. Overall, Tonidandel, and Starbuck (2006) recommended that this rule of thumb be used when $T \geq 5$ equally spaced measurements, and when expected attrition is no more than 40%, however.

If attrition occurs uniformly throughout the study, we suggest to use a correction where $n^* = n/(1\text{-}\ c\ \text{DRP})$. This is based on a theoretical consideration and assumes that attrition occurs uniformly. If the aim is to compare two group level means, we should choose $c = 1/2$. If the aim is to compare two slope means with no assumptions on equal baseline levels, we should choose $c = 3/4$. In cases where we assume equality of baseline levels, c should be between 2/3 and 3/4. Using $c = 3/4$ to obtain conservative sample sizes, we obtain $n^* = 330\ /(1 - 3\times.15/4) = 371.8$ for 15% attrition, $n^* = 425.8$ for 30% attrition, and $n^* = 600$ for 60% attrition, and these sample sizes are very close to those obtained by the formula (371, 423, and 589).

7. Discussion

7.1 Summary and conclusions

We presented algebraic expressions for the variance of the estimate of slope differences between two groups for LGC models. They allow researchers to evaluate effects of the number of measurement occasions, effect size, number of indicators and their reliabilities, factor variances and covariances, correlated errors, and attrition on needed sample size. Results from numerical illustrations agree with suggestions from Yi and Panzarella (2002) that researchers should focus on obtaining measurements at the end of the study. As long as we have measurements at the end points, dropping intermittent occasions does not make a big difference. Although this is of course dependent on the variance around the individual regression lines.

If groups can be assumed to have equal baseline levels, observations at the end are more important than early observations. In cases of attrition, we lose observations at the end, and attrition can therefore increase sample sizes needed substantially. Individuals who are only measured at the start of the study do not add to the precision of measuring the slope difference between groups. Using listwise deletion, i.e. using only those individuals with a complete set of measurements, can substantially increase the required sample size, and more so if the attrition occurs primarily at the end. Overall et al. (2006) recommended a rule of thumb that can be used to correct the original computed sample size for expected dropout. It appeared to work fairly well in our examples although it underestimated sample sizes in some cases. They recommended it for designs with more than four occasions and a dropout rate of less than 40% however. We suggested using another correction if attrition occurs uniformly, and it appeared to work for our examples. With only 4 occasions, as in our examples, attrition does not occur uniformly, however, and this correction should work better the more occasions we have.

Even though observations at the beginning and end are more important than middle observations, there are several reasons to use designs with intermittent observations. If we only obtain measurements at two occasions say, we will not be able to test linearity assumptions. If we want to test for higher order trends, we will need even more occasions. We should also keep in mind that although spacing measurement occasions unequally so that individuals are observed more often at the end can be beneficial in regards to needed sample size, this might result in problems, such as memory effects.

A treatment may increase or decrease the variance in the treatment group. A program aimed at stimulating cognitive abilities in children may stimulate already gifted children more than other children for example, and increase the variance in the treatment group. On the other hand, the program may be geared towards raising scores of disadvantaged children, and thus decrease the variance. If we suspect that treatment will have an effect on the variance, we should account for this in computing the needed sample sizes. In addition, a treatment may affect the covariance between the growth factors. A program that will keep on stimulating already gifted children may increase the covariance between the intercept and slope factors in the treatment group relative to the control group. Factor covariances may decrease or increase the required sample size and should therefore be considered in the design phase as well.

Correlations between errors of measurements also increase or decrease needed sample sizes. These correlations may arise because of memory effects, for example, if the same indicators are used repeatedly (e.g. Jöreskog & Sörbom, 1979). Correlated errors of measurements indicate that we measure our constructs with lower measurement precision. When the correlation between residuals at the first and last occasion is large, this makes up for lower measurement precision however. We will thus need larger sample sizes for small positive correlations when we have more than two occasions (compared to zero correlations), and smaller sample sizes for larger correlations. We examined the correlation between adjacent measurements, and an auto correlation of 0.5 for a two-occasion design means that the endpoint residuals are correlated 0.5, whereas a six occasion design with autocorrelation 0.5 means that the end point residuals are correlated $0.5^5 = 0.03$. The decrease in sample size is thus reached sooner for designs with few measurement occasions in our illustrations. It might sometimes be more appropriate to keep the correlation between the endpoints constant if one is to compare designs with different number of measurement occasions.

Although we did not specifically evaluate duration of the study, keeping the number of measurement occasions constant, this was specifically evaluated in Raudenbush and Liu (2001) and Wänström (2007). Our formulas include x_t, the elapsed time since the start of the study. If we want to evaluate effects of duration of the study or frequency of observation, keeping one or the other constant, we can compare models using appropriate values of x_t.

We used Cohen's (1988) standardized effect standards (e.g. d = .2 for a small effect size) in this paper. However, as discussed in Hancock (2001), Cohen's (1988) standards refer to observed variables, and these may be adjusted to apply to latent variables. Hancock (2001) showed how the standardized effect sizes could be adjusted for reliabilities of the indicators of latent constructs. We used observed variable standards in this paper because one of the aims was to explore reliability effects on sample size. However, researchers may use latent variable standards instead. For effect size standards of latent variables, see Hancock (2001).

As mentioned previously, several researchers have studied power and sample sizes of LGC models using either simulations or Satorra and Saris (1985) power approximation technique. The sample size formulas presented in Wänström (2007) were found to be practically equivalent to both simulations and Satorra-Saris approximations. Our formulas are a reasonable choice when choosing a technique for deciding required sample size. They hold approximately in cases of nonnormality and unknown covariance matrices as long as sample sizes are large. However, in cases of unknown covariance matrices, parameter values need to be guessed. Values from previous studies, pilot studies, or theory can be used as estimates in the formulas. Because estimates used in sample size formulas may not be very accurate, sample size boundaries may be computed from ranges of hypothetical parameter values.

7.2 Future work

Our formulas can be applied to slope differences between groups. Raudenbush and Liu (2001) presented formulas for trends in general, i.e. for linear slopes, quadratic slopes etc. Future work may look at extending our formulas to incorporate higher order polynomials. Another way of handling nonlinear growth in LGC models is to estimate some of the slope factor loadings. In other words, we may set x_t to be 0, 1, 2, X, where X is estimated (see e.g. Meredith & Tisak, 1990; Raykov, 2000). In this way, we can either constrain the estimated factor loading(s) to be equal across groups or we can estimate them in both groups. Effects of estimated factor loadings on sample sizes can be explored.

Our formulas can be extended in various other ways. For example, covariates may be included in the models, additional levels may be added, differences between more than two groups may be examined. In addition, other aspects than mean differences may be considered. As mentioned previously, a treatment may affect the variances and/or covariances in the treatment group. Muthén and Curran (1997) noted that larger sample sizes were generally needed to detect baseline-treatment interactions compared to detecting differences in slope means between groups. In addition, Hertzog et al. (2006) noted that power to detect differences in covariances between slopes in single groups were generally low. Future studies may investigate factors that will affect sample sizes needed when we are interested in variances or covariances in addition to (or instead of) means.

In conclusion, we extended work by Wänström (2007) and presented algebraic variance expressions that can be used to compute needed sample sizes for detection of differences in group slope means in second order LGC models. We found that observations at the beginning and end are more important than middle observations, that observations at the end are most important if we assume equal baseline means, and that attrition can substantially increase the needed starting sample size. We also suggested a correction to needed sample size that can be used when attrition occurs uniformly. In addition, we found that correlations between factors may increase or decrease needed sample size. Our formulas can be used for arbitrary variances, and in designs with generally spaced and individually varying measurement occasions when baseline levels may be assumed equal in groups.

8. References

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

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

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

Curran, P. J., & Muthén, B. O. (1999). The application of latent curve analysis to testing developmental theories in intervention research. *American Journal of Community Psychology, 27*, 567-595.

Desu, M. M., & Raghavarao, D. (1990). *Sample size methodology*. Boston: Academic Press.

Diggle, P. J., Heagerty, P. J., Liang, K., & Zeger, S. L. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford Statistical Science Series. New York: Oxford University Press Inc.

Duncan, S. C., & Duncan, T. E. (1996). A multivariate growth curve analysis of adolescent substance use. *Structural Equation Modeling, 3*, 323-347.

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. In G. A. Marcoulides (Ed.*). Quantitative Methodology Series. Methodology for Business and Management*. London: Lawrence Erlbaum.

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

Fan, X. & Fan, X. (2005). Power of latent growth modeling for detecting linear growth: number of measurements and comparison with other analytic approaches. *Journal of Experimental Education, 73*, 121-139.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1985). *Fundamentals of clinical trials* (2nd ed.). Littleton, MA: PSG Publishing.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). New York: Halsted.

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, *66*, 373-388.

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

Hox, J., & Stoel, R. D. (2005). Multilevel and SEM approaches to growth curve modeling. In B. S. Everett, & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioural science* (pp 1296-1305). Chisester: John Wiley and Sons, Ltd.

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Associates.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Liu, A., Boyett, J. M., & Xiong, X. (2000). Sample Size calculation for planning group sequential longitudinal trials. *Satistics in Medicine, 19*, 205-220.

Liu, G., & Liang, K. (1997). Sample size calculations for studies with correlated observations. *Biometrics*, *53*, 937-947.

Liu, A., Shih, W. J., & Gehan, E. (2002). Sample size and power determination for clustered repeated measurements. *Statistics in Medicine, 21*, 1787-1801.

Loehlin, J. C. (1992). *Latent variable models: An introduction to factor, path, and structural analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Longford, N. T. (1993). *Random coefficient models*. Oxford, UK: Clarendon.

Maxwell, S. E. (1998). Longitudinal designs in randomized group comparisons: When will intermediate observations increase statistical power? *Psychological Methods, 3*, 275-290.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In R. B. Cattell & J. Nesselroade (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 564-614). New York: Plenum Press.

McArdle, J. J., & Bell, R. Q. (2000). An introduction to latent growth models for developmental data analysis. In T. D. Little, K. U. Schnabel, J. and Baumert (Eds.). *Modeling longitudinal and multilevel data. Practical issues, applied approaches and specific examples* (pp.69-107). Mahwah, New Jersey: Lawrence Erlbaum Associates.

McDonald, R. P. (1999). *Test theory: a unified treatment*. Mahwah, New Jersey, Lawrence Erlbaum Associates.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55,* 107-122.

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, 4*, 599-620.

Overall, J. E., Shobaki, G., Shivakumar, C., & Steele, J. (1998). Adjusting sample sizes for anticipated dropouts in clinical trials. *Psychopharm Bulletin, 34*, 25-33.

Overall, J. E., Tonidandel, S., & Starbuck, R. R. (2006). Rule-of-thumb adjustment of sample sizes to accommodate dropouts in a two-stage analysis of repeated measurements. *International Journal of Methods in Psychiatric Research, 15*, 1-11.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics, 14*, 1-17.

Raudenbush, S. W. (2002). Toward a coherent framework for comparing trajectories of individual change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp.35-64). Washington, DC: American Psychological Association.

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, *5*, 199-213.

Raudenbush, S. W., & Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, *6*, 387-401.

Raykov, T. (2000). Modeling simultaneously individual and group patterns of ability growth or decline. In: T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling Longitudinal and Multilevel Data (pp. 127-146)*. Mahwaj, NJ: Lawrence Erlbaum.

Rochon, J. (1991). Sample size calculations for two-group repeated-measures experiments. *Biometrics, 47, 1383-1398*.
Rochon, J. (1998). Application of GEE procedures for sample size calculations in repeated measures experiments. *Statistics in Medicine, 17*, 1643-1658.
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. (2002). 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.
Scheaffer, R. L., Mendenhall, W., & Ott, R. L. (1996). *Elementary survey sampling* (5th ed.). New York: Wadsworth Publishing Company.
Stoel, R. D., van Der Wittenboer, G, & Hox, J. J. (2003). Analyzing longitudinal data using multilevel regression and latent growth curve analysis. *Metodologia de las Ciencas del Comportamiento, 5*, 21.42.
Tran, (1997). Estimating sample size in repeated-measures analysis of variance. *Measurement in Physical Education and Exercise Science, 1*, 89-102.
Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23*, 19-23.
Willet, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*, 363-381.
Wänström, L. (2007). Sample sizes for two-group second order latent growth curve models. (Submitted).
Yi, Q., & Panzarella, T (2002). Estimating sample size for tests on trends across repeated measurements with missing data based on the interaction term in a mixed model. *Controlled Clinical Trials, 23*, 481-496.
Zucker, D. M., & Denne, J. (2002). Sample-size redetermination for repeated measures studies. *Biometrics, 58*, 548-559.

Appendix

Assume that individual i, i = 1, …, n, is observed on each indicator k, k = 1, …, K, at each occasion t, t = 1, …, T_i where x_{ti} is the elapsed time since the start of the study for individual i at time t. Let Y_{kti} denote individual i's score on indicator k at occasion t. We assume, for simplicity, that all indicator factor loadings, $\lambda_{kt\varepsilon}$, are 1. Conditional on his or her intercept and slope parameters α_{1i} and α_{2i},

$$\bar{Y}_i \in N\left(\alpha_{1i} + \alpha_{2i}\bar{x}_i, V_{\bar{Y}_i}\right)$$

where

$$V_{\bar{Y}_i} = 1 \Big/ \sum_{t=1}^{T_i} \left(1 \Big/ \sigma^2_{\hat{Y}_{ti}}\right),$$

$$\sigma^2_{\hat{Y}_{ti}} = \sigma^2_{\eta_{ti}} + 1 \Big/ \sum_{k=1}^{K} \left(1 \Big/ \sigma^2_{\varepsilon_{kti}}\right),$$

and

$$r_i = \sum_{t=1}^{T_i} \hat{Y}_{ti}\left(x_{ti} - \bar{x}_i\right) \Big/ \sum_{t=1}^{T_i}\left(x_{ti} - \bar{x}_i\right)^2 \in N\left(\alpha_{2i}, V_{ri}\right)$$

where

$$V_{ri} = \sum_{t=1}^{T_i} \sigma^2_{\hat{Y}_{ti}} \Big/ \left(x_{ti} - \bar{x}_i\right)^2,$$

and

$$\bar{x}_i = \sum_{t=1}^{T_i} \left(x_{ti} \Big/ \sigma^2_{\hat{Y}_{ti}}\right) \Big/ \sum_{t=1}^{T_i} \left(1 \Big/ \sigma^2_{\hat{Y}_{ti}}\right),$$

and $\bar{Y}_i$ and r_i are independent.

Unconditionally,

$$\bar{Y}_i \in N\left(\alpha_1 + \alpha_2\bar{x}_i, \sigma^2_{\bar{Y}_i}\right)$$

where

$$\sigma^2_{\bar{Y}_i} = V_{\bar{Y}_i} + \sigma^2_{\pi_1} + \bar{x}_i^2\sigma^2_{\pi_2} + 2\bar{x}_i\sigma_{\pi_1\pi_2}$$

and

$$r_i \in N\left(\alpha_2, \sigma_{ri}^2\right)$$

where

$$\sigma_{ri}^2 = V_{ri} + \sigma_{\pi_2}^2 \qquad \text{(a)}$$

and

$$\operatorname{cov}\left(\overline{Y}_i, r_i\right) = \sigma_{\overline{Y}_i r_i}$$

where $\sigma_{\overline{Y}_i r_i} = \sigma_{\pi_1 \pi_2} + \sigma_{\pi_2}^2 \overline{x}_i$.

We can create a new variable, z_i, that is independent of r_i,

$$z_i = \overline{Y}_i - r_i \left(\sigma_{\overline{Y}_i r_i}\right) / \left(\sigma_{r_i}^2\right) \in N\left[\alpha_1 + \alpha_2 \overline{x}_i^* \quad ; \quad \sigma_{z_i}^2\right]$$

where

$$\overline{x}_i^* = \overline{x}_i - \sigma_{\overline{Y}_i r_i} / \sigma_{r_i}^2$$

and $\sigma_{z_i}^2 = \sigma_{\overline{Y}_i}^2 - \sigma_{\overline{Y}_i r_i}^2 / \sigma_{r_i}^2$.

We can estimate α_2, the mean slope, from all z_i :

$$\hat{b}_z = \sum_{i=1}^n \left(z_i \left(\overline{x}_i^* - \overline{\overline{x}}\right)^2 \Big/ \sigma_{z_i}^2\right) \Big/ \sum_{i=1}^n \left(\left(\overline{x}_i^* - \overline{\overline{x}}\right)^2 \Big/ \sigma_{z_i}^2\right) \text{ where}$$

$$\overline{\overline{x}} = \sum_{i=1}^n \left(\overline{x}_i^* \Big/ \sigma_{z_i}^2\right) \Big/ \sum_{i=1}^n \left(1 \Big/ \sigma_{z_i}^2\right)$$

and

$$\hat{b}_z \in N\left[\alpha_2 \quad ; \quad 1 \Big/ \sum_{i=1}^n \left(\left(\overline{x}_i^* - \overline{\overline{x}}\right)^2 \Big/ \sigma_{z_i}^2\right)\right]$$

We can estimate α_2 from all r_i:

$$\hat{b}_r = \sum_{i=1}^n \left(r_i \Big/ \sigma_{r_i}^2\right) \Big/ \sum_{i=1}^n \left(1 \Big/ \sigma_{r_i}^2\right),$$

and

$$\hat{b}_r \in N\left[\alpha_2 \quad ; \quad 1\Big/\sum_{i=1}^{n}\left(1/\sigma_{r_i}^2\right)\right]$$

These two estimates are combined:

$$\hat{b} = \frac{\hat{b}_z \sum_{i=1}^{n}\left(\left(\bar{x}_i^* - \bar{\bar{x}}\right)^2/\sigma_{z_i}^2\right) + \hat{b}_r \sum_{i=1}^{n}\left(1/\sigma_{r_i}^2\right)}{\sum_{i=1}^{n}\left(\left(\bar{x}_i^* - \bar{\bar{x}}\right)^2/\sigma_{z_i}^2\right) + \sum_{i=1}^{n}\left(1/\sigma_{r_i}^2\right)},$$

and

$$\hat{b} \in N\left[\alpha_2 \quad ; \quad 1\Big/\left(\sum_{i=1}^{n}\left(\left(\bar{x}_i^* - \bar{\bar{x}}\right)^2/\sigma_{z_i}^2\right) + \sum_{i=1}^{n}\left(1/\sigma_{r_i}^2\right)\right)\right] \qquad \text{(b)}$$

Assume now that we have two groups. We then have, from each group, an estimate of the mean that is independent of $\hat{b}_z$ and $\hat{b}_r$

$$\bar{\bar{z}}^{(g)} = \sum_{i=1}^{n}\left(z_i/\sigma_{z_i}^2\right)\Big/\sum_{i=1}^{n}\left(1/\sigma_{z_i}^2\right) \in N\left[\alpha_1^{(g)} + \alpha_2^{(g)}\bar{\bar{x}} \quad ; \quad 1\Big/\sum_{i=1}^{n}\left(1/\sigma_{z_i}^2\right)\right]$$

where the superscript (g) refers to group. In randomized designs, it is reasonable to assume that the intercepts are equal between groups, $\alpha_1^{(1)} = \alpha_1^{(2)}$. We can then estimate $\alpha_2^{(1)}\bar{\bar{x}}^{(1)} - \alpha_2^{(2)}\bar{\bar{x}}^{(2)}$ by

$$\bar{\bar{z}}^{(1)} - \bar{\bar{z}}^{(2)} \in N\left[\alpha_2^{(1)}\bar{\bar{x}}^{(1)} - \alpha_2^{(2)}\bar{\bar{x}}^{(2)} \quad ; \quad 1\Big/\sum_{i=1}^{n}\left(1/\sigma_{z_i}^{(1)2}\right) + 1\Big/\sum_{i=1}^{n}\left(1/\sigma_{z_i}^{(2)2}\right)\right]$$

If we, for simplicity, assume that $\bar{\bar{x}}^{(1)} = \bar{\bar{x}}^{(2)}$ and that we have equal variances between groups, then

$$\left(\bar{\bar{z}}^{(1)} - \bar{\bar{z}}^{(2)}\right)\Big/\bar{\bar{x}} \in N\left[\alpha_2^{(1)} - \alpha_2^{(2)} \quad ; \quad \left(1/\bar{\bar{x}}^2\right)\left(2\Big/\sum_{i=1}^{n} 1/\sigma_{z_i}^2\right)\right] \qquad \text{(c)}$$

Combining (b) and (c) gives

$$\hat{\pi}_3 = \frac{\left(\hat{b}^{(1)} - \hat{b}^{(2)}\right)\left(\sum_{i=1}^{n}\frac{\left(\bar{x}_i^* - \bar{\bar{x}}\right)^2}{\sigma_{z_i}^2} + \sum_{i=1}^{n}\frac{1}{\sigma_{r_i}^2}\right) + \left(\frac{\bar{\bar{z}}^{(1)} - \bar{\bar{z}}^{(2)}}{\bar{\bar{x}}}\right)\left(\bar{\bar{x}}^2\sum_{i=1}^{n}\frac{1}{\sigma_{z_i}^2}\right)}{\left(\sum_{i=1}^{n}\frac{\left(\bar{x}_i^* - \bar{\bar{x}}\right)^2}{\sigma_{z_i}^2} + \sum_{i=1}^{n}\frac{1}{\sigma_{r_i}^2}\right) + \left(\bar{\bar{x}}^2\sum_{i=1}^{n}\frac{1}{\sigma_{z_i}^2}\right)}$$

$$\in N\left[\alpha_2^{(1)} - \alpha_2^{(2)} \qquad 2\left(\sum_{i=1}^{n}\left(\frac{\left(\overline{x}_i^* - \overline{\overline{x}}\right)^2}{\sigma_{z_i}^2} + \frac{1}{\sigma_{r_r}^2} + \frac{\overline{\overline{x}}^2}{\sigma_{z_i}^2}\right)\right)^{-1}\right] =$$

$$\mathrm{N}\left[\alpha_2^{(1)} - \alpha_2^{(2)} \qquad ; \quad 2\left(\sum_{i=1}^{n}\left(\frac{\overline{x}_i^{*2}}{\sigma_{z_i}^2} + \frac{1}{\sigma_{r_i}^2}\right)\right)^{-1}\right].$$