frontiers in PSYCHOLOGY

IMPACT 2.8 FACTOR

Quantitative Psychology and Measurement

LIFESPAN: A Tool for the Computer-Aided Design of Longitudinal Studies

Andreas M Brandmaier, Timo von_Oertzen, Paolo Ghisletta, Christopher Hertzog and Ulman Lindenberger

Journal Name:	Frontiers in Psychology
ISSN:	1664-1078
Article type:	Methods Article
Received on:	19 Jan 2015
Accepted on:	24 Feb 2015
Provisional PDF published on:	24 Feb 2015
Frontiers website link:	www.frontiersin.org
Citation:	Brandmaier AM, Von_oertzen T, Ghisletta P, Hertzog C and Lindenberger U(2015) LIFESPAN: A Tool for the Computer-Aided Design of Longitudinal Studies. *Front. Psychol.* 6:272. doi:10.3389/fpsyg.2015.00272
Copyright statement:	

LIFESPAN: A Tool for the Computer-Aided Design of Longitudinal Studies

Andreas M. Brandmaier[1]*, Timo von Oertzen[2], Paolo Ghisletta[3,4], Christopher Hertzog[5], and Ulman Lindenberger[1, 6]

[1] Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany
[2] Department of Psychology, University of Virginia, Charlottesville, USA
[3] Faculty of Psychology and Educational Sciences (FPSE), University of Geneva, Geneva, Switzerland
[4] Distance Learning University Switzerland, Switzerland
[5] Adult Cognition Lab, School of Psychology Georgia Institute of Technology, Atlanta, USA
[6] Max Planck University College London Centre for Computational Psychiatry and Ageing Research, London, UK

*** Correspondence:** Andreas M. Brandmaier (brandmaier@mpib-berlin.mpg.de), Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee 94, Berlin 14195, Germany

Keywords: statistical power, structural equation modeling, latent growth curve modeling, optimal design, power equivalence theory, effective error.

Abstract

Researchers planning a longitudinal study typically search, more or less informally, a multivariate space of possible study designs that include dimensions such as the hypothesized true variance in change, indicator reliability, the number and spacing of measurement occasions, total study time, and sample size. The main search goal is to select a research design that best addresses the guiding questions and hypotheses of the planned study while heeding applicable external conditions and constraints, including time, money, feasibility, and ethical considerations. Because longitudinal study selection ultimately requires optimization under constraints, it is amenable to the general operating principles of optimization in computer-aided design. Based on power equivalence theory (MacCallum et al., 2010;von Oertzen, 2010), we propose a computational framework to promote more systematic searches within the study design space. Starting with an initial design, the proposed framework generates a set of alternative models with equal statistical power for detecting hypothesized effects, and delineates tradeoff relations among relevant parameters, such as total study time and the number of measurement occasions. We present LIFESPAN (Longitudinal Interactive Front End Study Planner) that implements this framework. LIFESPAN boosts the efficiency, breadth, and precision of the search for optimal longitudinal designs. Its initial version, which is freely available at http://www.brandmaier.de/lifespan, is geared towards the power to detect variance in change as specified in a linear latent growth curve model.

1. Introduction

Describing, explaining, and modifying between-persons differences in change are central goals in research on lifespan development (Baltes and Nesselroade, 1979;Hertzog, 1996;Baltes et al., 2006;Ferrer and McArdle, 2010;Lindenberger et al., 2011). Numerous studies have shown that people differ in rates of change in many functional domains, both at neural and behavioral levels of

analysis (e.g., Lindenberger, 2014). To delineate the antecedents, correlates, and consequents of these differences, differences in change in variables of interest must be measured with sufficient reliability. Hence, researchers have begun to examine the relative importance of factors that contribute to the statistical power to detect between-person differences in change (represented by the variance in change), such as the true variance in change, the number and precision of indicators, the number and distribution of measurement occasions, and the total time elapsing from the beginning to the end of the study (henceforth referred to as total study time; Hertzog et al., 2006;von Oertzen et al., 2010;von Oertzen and Brandmaier, 2013;Rast and Hofer, 2014). The search for optimally powerful longitudinal research designs requires close and simultaneous attention to the relative contributions of each of these factors to statistical power.

There is a dire need for a coherent and unified approach to the a priori estimation of statistical power that can efficiently assist researchers in identifying longitudinal research designs with optimal statistical power to detect key effects under a given set of assumptions and design constraints (Maxwell et al., 2008;Moerbeek, 2011). Current statistical power analysis is often based on Monte Carlo simulations (e.g., Hertzog et al., 2008;Ke and Wang, 2014;Rast and Hofer, 2014), which can be carried out by help of statistical software packages such as M*plus* (Muthén and Muthén, 2007). However, the Monte Carlo simulation approach can be cumbersome, and requires scientists to choose how and when to simulate possible design configurations. What is currently needed is a method for an efficient yet comprehensive overview of the ways in which different parameter values or design configurations contribute to statistical power. Currently available dedicated software can be used for the a priori power analysis of hypotheses about repeated measures means and interactions in a general linear model context (G*power; Faul et al., 2007) and for group differences in mean growth curve parameters, as in intervention contexts (Hedeker et al., 1999;Kelley and Rausch, 2011) or observational studies with time-varying exposure (Barrera-Gomez et al., 2013). However, power tools with a focus on individual differences in change as specified by latent variable models are still lacking. Given recent advances in the formal understanding of statistical power in longitudinal structural equation modeling (e.g., von Oertzen, 2010), the time is ripe to introduce a software tool for the computer-aided design of longitudinal studies. Hence, we propose LIFESPAN, a freely available computer tool for designing linear latent growth curve model (LGCM) designs and for deriving approximate estimates of their statistical power. The currently available version of LIFESPAN allows researchers to explore alternative study designs with equivalent power to detect individual differences in linear change.

In the remainder of this article, we introduce the design principles and specific features of our computational approach, discuss limitations of its current implementation, and lay out a research agenda for the computer-aided design of longitudinal studies.

2. Computer-aided design of longitudinal studies: A Structural Equation Modeling approach

Human designers typically envision a design problem in terms of one or more goals they wish to attain, and then consider dimensions that put constraints on the space of admissible solutions, such as cost, time, feasibility, elegance (aesthetics), and ethics. In engineering and in the natural sciences, computers often assist humans in finding solutions to design problems of this sort. Computer-aided design (CAD) is devoted to reduce the elapsed time and resources spent during the design task supported by computational facilities (Coons and Mann, 1960). When the goal of a design task is not only feasibility but has further design objectives, the task at hand may be formalized in terms of

optimization under constraints (see Rao, 2009). The auspicious role assigned to the computer is to find a solution (e.g., a product) that optimizes one or more criteria under a given set of constraints. In mechanical design, typical goals are the reduction of stress, wear, or weight, e.g., minimizing the overall weight in aerospace design or minimizing manufacturing cost in civil engineering design.

Likewise, the planning of a longitudinal study, which involves repeated measurements of one or more variables over time, can be regarded as an engineering task. Generally, researchers have a good sense of their phenomena of interest, and select their measurement instruments on that basis. They then consider various longitudinal study designs based on a collection of reasons that include assumptions about the nature of the change process as well as practical considerations such as available resources (e.g., time and money). This selection process comes with many degrees of freedom, and decisions are often made without full knowledge of their implications. For instance, longitudinal design decisions entail choosing an observational time span, and, within that time span, the frequency and distribution of measurement occasions. Given the complexity and size of the longitudinal design search space, it is surprising to notice that computer-aided approaches to optimal longitudinal design have been largely neglected thus far, despite the longstanding availability of appropriate statistical approaches (e.g., Schlesselman, 1973).

Structural Equation Modeling (SEM; e.g., Bollen, 1989) is a statistical framework that formalizes the relationship between observed and latent variables. SEM notation includes diagrams that represent the entire set of equations underlying a given model (see McArdle and Nesselroade, 2014). This feature greatly facilitates the creation, modification, and communication of models, and is particularly useful for comparing different research designs (von Oertzen et al., 2015). Within SEM, latent growth curve models (LGCM) are widely used to capture change in longitudinal data on human behavioral development (e.g., Meredith and Tisak, 1990;Muthén and Curran, 1997;Ferrer and McArdle, 2003;Ferrer and McArdle, 2010;Duncan et al., 2013). In LGCM, factor loadings represent hypothesized trends over time, such as initial level and linear change. The mean vector, $\boldsymbol{\mu}$, and the covariance matrix, $\boldsymbol{\Sigma}$, of the observed variables are a function of factor loadings, $\boldsymbol{\Lambda}$, variables' intercepts, $\boldsymbol{\nu}$, a latent covariance matrix, $\boldsymbol{\Psi}$, and a residual covariance matrix, $\boldsymbol{\Theta}$ (e.g., Bollen, 1989):

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$
$$\boldsymbol{\mu} = \boldsymbol{\Lambda}\boldsymbol{\nu}$$

Under the assumption of homoscedastic and uncorrelated residual errors, and the intercept anchored at the first measurement occasion, the matrices for a linear LGCM are

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{1} & \boldsymbol{t_1} \\ \mathbf{1} & \boldsymbol{t_2} \\ \mathbf{1} & \vdots \\ \mathbf{1} & \boldsymbol{t_M} \end{bmatrix}$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\sigma_I^2} & \boldsymbol{\sigma_{IS}} \\ \boldsymbol{\sigma_{IS}} & \boldsymbol{\sigma_S^2} \end{bmatrix}$$

$$\boldsymbol{\nu} = \begin{bmatrix} \boldsymbol{\mu_I} \\ \boldsymbol{\mu_S} \end{bmatrix}$$

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\sigma_\epsilon^2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma_\epsilon^2} \end{bmatrix}$$

The parameters in the model are the number of measurement occasions, M, at times t_1 to t_M, the residual error, σ_ϵ^2, the mean, μ_I, and variance, σ_I^2, of the latent intercept, and the mean, μ_S, and the variance, σ_S^2, of the latent slope, and the latent intercept-slope-covariance, σ_{IS}.

When planning a longitudinal study, statistical consultants are typically approached with questions such as of how big does a sample need to be to approach a level of statistical power that is deemed adequate (e.g., 80%). Questions of this kind have been the target of a large number of simulation studies (e.g., Muthén and Muthén, 2002;Maxwell et al., 2008), which in turn have informed researchers on reasonable ranges for selected designs and effect sizes. However, the curse of dimensionality so far has rendered an exhaustive simulation-based treatment of statistical power for all potential combination of design parameters intractable. This impasse can be overcome by statistical theories that formalize parameter tradeoff relations in SEM (MacCallum et al., 2010;von Oertzen, 2010).

Specifically, von Oertzen and Brandmaier (2013) have proposed a formal approach, based on power equivalence theory (von Oertzen, 2010), that allows researchers to examine tradeoff relations among design parameters of a LGCM while holding statistical power constant. In the context of SEM, power equivalence theory allows the generation of alternative models with different designs parameters but equal power according to likelihood-ratio tests. von Oertzen and Brandmaier (2013) show how power-equivalent operations can be used to transform a given LGCM into alternative models. For tests of interindividual differences of change, σ_S^2, they present an empirical example of tradeoff relations between total study time and the number of measurement occasions. Of course, many more such tradeoffs are possible. If multiple measurement instruments are available, combinations of them might be used in multiple-indicator LGCM to increase power (von Oertzen et al., 2010), or the number of participants can be traded off for additional bursts or waves of measurement (Schlesselman, 1973;see Raudenbush and Liu, 2001;von Oertzen and Brandmaier, 2013). As is true for any engineering task, the optimal choice among models will depend on external criteria, such as the study time elapsing until targeted effects are reliable, the strain put onto research participants, and resource expenditures, such as laboratory space or money.

3. Power equivalence theory and effective error

Comparing alternative study designs under equal power allows the optimization of a study design with respect to a given design objective, for example, the minimization of the total study time or the number of measurement occasions or waves. To permit the manipulation of design parameters of a given study design without changing statistical power, we rely on power equivalence theory as introduced by von Oertzen (2010). Two study designs measuring the same effect of interest are power-equivalent if they exhibit the same statistical power to detect the effect. Translating this definition to study designs targeting interindividual differences in change, two study designs are power-equivalent if they have the same power to detect non-zero slope variance in a likelihood-ratio test. Two such study design may differ in any aspect that does not change the variables involved in the statistical hypothesis. In the context of a 1 degree of freedom (1-*df*) test, any parameter other than the linear slope may be changed. For example, two alternative study designs may have the same power while differing in a combination of parameters, such as the number of measurement occasions

(and thus in the number of observed variables), in the total study time, distribution of measurement occasions in time, precision of the measurement instrument, or the number of participants.

von Oertzen (2010) noted that a given, potentially complex, SEM together with a given statistical hypothesis can be reduced to a minimal power-equivalent model. For hypotheses about a single latent variable, as in a 1-*df* test of slope variance, power equivalence theory allows the reduction of a structurally complex study design to a simple model with a single *effective error*. This effective error may be interpreted as the hypothetical measurement error encountered had it been possible to directly measure the latent construct of interest. It follows that two alternative study designs with the same effective error are power-equivalent. Thus, the effective error acts as pivotal point allowing the derivation of power-equivalent models from an initial design. von Oertzen and Brandmaier (2013) have elaborated this approach for LGCM and hypotheses about the intercept and slope variance. In the following, we reiterate how the effective error in a linear LGCM can be used to arrive at alternative designs given an initial study design.

The effective error of measuring slope variance in a linear LGCM can be written as follows (adapted from von Oertzen and Brandmaier (2013), Equation 2):

$$\boldsymbol{\sigma^2_{eff}} = \frac{\boldsymbol{\sigma^2_\epsilon}}{\sum_{j=1}^{\boldsymbol{M}} \boldsymbol{t_j^2} - \frac{\mathbf{1}}{\boldsymbol{M} + \boldsymbol{\sigma^2_\epsilon}/\boldsymbol{\sigma^2_I}} \left(\sum_{j=1}^{\boldsymbol{M}} \boldsymbol{t_j}\right)^2}$$

where M is the number of measurement occasions at time points t_j, $\boldsymbol{\sigma^2_\epsilon}$ the residual error, and $\boldsymbol{\sigma^2_I}$ the intercept variance. Assuming equally spaced measurements and linear growth over time with T being the total study time, $T = t_M$, we can substitute the sums with the following terms:

$$\sum_{j=1}^{M} t_j = \sum_{j=1}^{M} \left(\frac{j}{M} T\right) = \frac{1}{2}(M+1)T$$

$$\sum_{j=1}^{M} t_j^2 = \sum_{j=1}^{M} \left(\frac{j}{M} T\right)^2 = \frac{(M+1)(2M+1)T^2}{6M}$$

It follows that the effective error is a function of a given set of parameters $\theta = \left(\boldsymbol{\sigma^2_\epsilon}, \boldsymbol{\sigma^2_I}, T, M\right)$ including residual variance, intercept variance, total study time and number of measurement occasions. Let θ represent the specification of the initial study design. Then, we can define an alternative study design by a second parameter vector $\theta' = \left(\boldsymbol{\sigma'^2_\epsilon}, \boldsymbol{\sigma'^2_I}, T', M'\right)$. Both study designs are power equivalent if their effective errors are equal, that is, $\boldsymbol{\sigma^2_{eff}}(\theta) = \boldsymbol{\sigma^2_{eff}}(\theta')$. To guarantee power equivalence during manipulation of an alternative design, we allow all but a single parameter in θ' to be freely varied. Henceforth, we refer to this excluded parameter as *computer-adjusted*. Whenever any value on one of the dimensions of θ' was changed during the design process, an optimization algorithm is used to adapt the computer-adjusted dimension of θ' such that $\boldsymbol{\sigma^2_{eff}}(\theta) = \boldsymbol{\sigma^2_{eff}}(\theta')$. To accomplish this end we employ a gradient descent algorithm (e.g., Luenberger, 1973) to find the root of $\boldsymbol{\sigma^2_{eff}}(\theta) - \boldsymbol{\sigma^2_{eff}}(\theta')$. This general-purpose optimization technique allows us to arrive at alternative models under equivalent statistical power without the need to run computationally expensive Monte Carlo simulations at each optimization step. Sample size can also be a modifiable parameter under power-equivalence when the optimization scheme is augmented by numerical approximations of

statistical power (see Satorra and Saris, 1985). The automatic layout of a path diagram for an automatically created, alternative study design given by the parameter vector θ' can either be implemented for a particular design or, generally, left to an automatic layout algorithm (e.g., Boker et al., 2002).

Based on the design parameters of a LGCM, various indices of design quality other than statistical power itself can be derived. When normalizing the absolute effect size, $\boldsymbol{\sigma_S^2}$, with the effective error, $\boldsymbol{\sigma_{eff}^2}$, we obtain an index of reliability of the specific likelihood-ratio test of slope variance, effective curve reliability (ECR), which can be interpreted as an effect size estimate of slope variance:

$$\boldsymbol{ECR} = \frac{\boldsymbol{\sigma_S^2}}{\boldsymbol{\sigma_S^2} + \boldsymbol{\sigma_{eff}^2}}$$

Similarly, growth rate reliability (GRR), introduced by Willett (1989), was used by Rast and Hofer (2014) as an index of statistical power in LGCM. GRR can be regarded as a special case of ECR, in the sense that the two indices yield identical results when the effect of intercept variance on effective error is asymptotically large so that the denominator of the effective error simplifies to a term proportional to the variance of the occasions of measurements. GRR may more appropriate than ECR if the statistical test used to detect slope variance does not account for the effect of intercept variance, e.g., a one-dimensional Wald test.

In contrast to both ECR and GRR, growth curve reliability (GCR; e.g., McArdle and Epstein, 1987) is a measure of variance explained in the observed variables, and reduces to a scaling of intercept variance and residual variance at the point in time where the regression of the observed variable on the latent slope is zero (i.e., at occasion j for which t_j=0):

$$\boldsymbol{GCR_0} = \frac{\boldsymbol{\sigma_I^2}}{\boldsymbol{\sigma_I^2} + \boldsymbol{\sigma_\epsilon^2}}$$

The different indices serve complementary functions in planning, selecting, and communicating a study. Effective error is particularly useful as a proxy for statistical power when researchers have no clear expectations about the corresponding true score, such the true variance of change. ECR relates a true score to its effective error, and serves as proxy to statistical power when sample size and alpha level are left undetermined.

4. The LIFESPAN tool

Based on power equivalence theory (von Oertzen, 2010;von Oertzen and Brandmaier, 2013), we have designed LIFESPAN (Longitudinal, Interactive Front End, Study Planner; LIFESPAN) to aid researchers in the design phase of longitudinal studies. The program is freely available at http://www.brandmaier.de/lifespan. With LIFESPAN, our primary goal is to help researchers in designing, manipulating, and optimizing a longitudinal study design. To this end, researchers using LIFESPAN can: (a) generate a graphical rendition of the model implied by an initial study design using; (b) freely and systematically explore the space of alternative power-equivalent study designs; (c) compute and display relevant design indices, such as effective error, GCR, GRR or ECR; (d) run

a Monte-Carlo simulation engine to estimate statistical power for a given sample size; and (e) convert the final model from a planning into a data analysis tool. To facilitate this transition, LIFESPAN is based on Ωnyx (von Oertzen et al., 2015), a SEM software environment that is also freely available (http://onyx.brandmaier.de), but distributed as a standalone program. LIFESPAN is written in JAVA and runs on all major operating systems, including Linux/Unix, OSX, and Windows. To streamline researchers' workflow and increase accessibility, we consider integrating LIFESPAN as a module directly with Ωnyx such that users need not switch between programs when planning a study, running Monte Carlo simulations, and conducting data analyses.

Currently, LIFESPAN is limited to linear LGCM, and is geared towards evaluating the power to detect variance in linear change. Further specification modes, design indicators, and simulation tools related to other design parameters will be added to future releases of the program (see below).

The main screen of LIFESPAN features three elements (see Figure 1). The top half of the screen displays the path diagram of the initial or target study design. The center shows a summary with a set of study design indices, such as the effective error, GCR, GRR, and ECR. The bottom half of the screen features a control panel. LIFESPAN offers four modes of operation, each corresponding to one of the tabs in the control panel: (1) model specification; (2) alternative models; (3) iso-power plots; and (4) Monte Carlo simulation. In the following, each of these modes is described in detail.

4.1. Model specification

In model specification mode, researchers can specify an initial study design in form of a linear LGCM. Specification does not require knowledge of syntax or algebra, as researchers are asked to directly manipulate the design parameters of the LGCM. These parameters include the number of measurement occasions, the total time span of the study, and the residual variance of the indicator. In addition, parameters referring to population values at the latent level need to be specified, that is, intercept variance, slope variance, and intercept-slope covariance. Once model specification is completed, clicking *Done* generates a path diagram that corresponds to the specified study design, and delivers the design indices GCR, GRR, ECR, and effective error.

4.2. Alternative models

Proceeding from model specification, LIFESPAN allows researchers to generate alternative models that have equal statistical power to detect variance in linear change. To this end, the parameters specified during model specification are represented as sliders. In this manner, researchers can observe how different parameter combinations result in identical statistical power.

Specifically, choice buttons allow selecting one design parameter to be computer-adjusted while the remaining parameters remain user-modifiable. Whenever any of the user-modifiable parameters is changed, the optimization algorithm described above adapts the computer-adjusted parameter such that the resulting alternative study design is power-equivalent to the initial design. As the researcher is exploring alternative designs, the corresponding path diagram and associated design indices are being updated.

4.3. Plots of iso-power contours

To attain a more complete understanding of parameter trade-offs relations, the next mode of operation allows researchers to plot iso-power curves (MacCallum et al., 2010;von Oertzen and Brandmaier, 2013). Iso-power curves display power-equivalent alternative models in two-

dimensional parameter space; they display bivariate associations between parameters while statistical power to detect linear variance of change is held constant. This feature allows researchers to identify parameter constellations that optimize one or more external criteria, such as total study time and indicator reliability.

4.4. Monte Carlo simulation

Finally, LIFESPAN estimates the statistical power to detect variance in linear change for a given sample size. To this end, researchers can choose between two tests: (i) a 1 degree-of-freedom test of the slope variance; (ii) a generalized variance-covariance likelihood ratio (LR) test with two degrees of freedom (for a discussion, see Hertzog et al., 2008). The current version of LIFESPAN uses a Monte Carlo simulation approach (e.g., Muthén and Muthén, 2002) for estimating actual statistical power. Researchers can specify the sample size and the number of Monte Carlo replications. In each replication, data are simulated from the currently specified study design and are fitted to the same model under no restriction and once under the restrictions imposed by the selected variance test. Parameter estimation is performed by the estimation engine of Ωnyx (for details, see von Oertzen et al., 2015). By counting the resulting significant LR tests, one obtains an unbiased approximation to the statistical power of the study design.

4.5. Workflow

In its current form, LIFESPAN allows the specification of a longitudinal study design with repeated measures over time in the form of a LGCM. Researchers can enter their initial model parameters and a best guess of the true variance in linear change (e.g. effect size) to obtain approximations to the statistical power to detect between-person difference in linear change. By using sliders that represent various design parameters, researchers can intuitively explore alternative models. Plotting associations between pairs of selected parameters under equal power allows visualizing critical design aspects based on power equivalence theory.

The final model specification can be exported to and directly used in Ωnyx. Ωnyx allows using the selected design option for Maximum Likelihood estimation of parameters once empirical data have been collected. Also, the graphical interface of Ωnyx allows researchers to expand the model beyond the limitations of LIFESPAN, for instance, by imposing constraints or expanding the model beyond the unconditional LGCM. Further capabilities of Ωnyx include the generation of publication-ready figures, further simulation, and export of the syntax of the final model to three freely available R packages, OpenMx (Boker et al., 2011), lavaan (Rosseel, 2012), and sem (Fox, 2006), and to the commercially available software package M*plus* (Muthén and Muthén, 2007).

5. A sample application of LIFESPAN

For illustration, we have recreated a study design taken from the study, "Origins of Variance in the Oldest-Old: Octogenarian Twins" (see Johansson et al., 1999;Johansson et al., 2004). Following the values reported by Rast and Hofer (2014; Table 5, line 1, p.11) for the measure, *Memory-in-Reality Free Recall*, we specified an initial study design with slope variance $\sigma_S^2 = .53$**,** intercept variance $\sigma_I^2 = 39.63$, residual error $\sigma_\epsilon^2 = 9.20$, intercept-slope covariance $\sigma_{IS} = -0.69$ (corresponding to an intercept-slope correlation of -0.15), and three measurement occasions spanning a total of 4 years, that is, $T = 4$, and $M = 3$. As reliability and effect size indicators, we obtain GCR of .81, GRR of .32, ECR of .36, and an effective error of .96. Using the Monte Carlo estimation functionality, we

estimate the power of the design with a sample size of $N = 250$ to be close to 80%.

Based on this empirically realized study design and its observed statistical parameters, we ask four questions concerning possible modifications of the initial design (see Figure 2): (1) In case we add (or subtract) measurement occasions, in how far can we afford to use a less reliable (or do we need a more reliable) measurement instrument? (2) Again, in case we add (or subtract) measurement occasions, by how much can we reduce (or do we need to increase) total study time? (3) If the true variance in linear change was larger (or smaller) than observed, in how far can we afford a less reliable (or do we need a more reliable) measurement instrument? (4) If individual differences at baseline were higher (or lower) than observed, by how much more time would we need to extend (or were we allowed to reduce) total study time to achieve the same power to detect between-person differences in linear change? The four panels of Figure 2 show iso-power curves that provide answers to each of these questions. Residual variance trades off almost linearly against the number of occasions and the variance of slope (left panels). The number of measurement occasions and total study time span trade off against each other in a quadratic relationship, in the sense that the effect of adding occasions on power is reduced with each additional measurement occasion (upper right panel; cf. von Oertzen and Brandmaier, 2013). Finally, the effect of intercept variance on power quickly reaches an asymptote such that increasing intercept variance needs to be compensated for by only small increments of total study time span to achieve equal statistical power (lower right panel).

6. Discussion

6.1. Current limitations of LIFESPAN

We see LIFESPAN as a computational tool that helps researchers to gain insights into tradeoff relations among design parameters, and hence enables them to make better decisions about the design of a planned longitudinal study. At the same time, we acknowledge that the current version has at least three important limitations.

First, LIFESPAN is currently limited to linear LGCM. We decided to formalize longitudinal study design in terms of a LGCM because models of this type are widely used for longitudinal data analysis, particularly in life span research (Hertzog, 1996;McArdle and Nesselroade, 2003;Lindenberger et al., 2011). We emphasize that the assumption of homogeneous linear change is strong, and quite likely to be incorrect in many empirical settings. For instance, in studies of cognitive aging, changes often accelerate with advancing age (cf. Ghisletta et al., submitted). Hence, we recommend some caution when searching for alternative models, as the linearity assumption may entail substantial misspecification at higher ages, especially when the model covers a large age range.

Second, the current version of LIFESPAN has an exclusive focus on the statistical power to detect between-person differences in linear change. In our judgment, this focus is well justified because the description, explanation, and modification of individual differences in change is central to lifespan theory (Baltes et al., 1977), and arguably the most important reason for conducting longitudinal work in the first place. Accordingly, the indices currently provided by LIFESPAN reflect our substantive research interest in between-person differences in change (Hertzog, 2008;Lindenberger, 2014) and complement our earlier work on statistical power (Hertzog et al., 2006;Hertzog et al., 2008;von Oertzen et al., 2010;von Oertzen and Brandmaier, 2013).

Third, in the present version of LIFESPAN, power equivalence is based on the 1 degree-of-

freedom test, which refers to the specific test of zero variance. Note that the hypothesis tested by this test is that there is no *unique* variance in linear slope. If the intercept-slope covariance is different from zero, then testing this hypothesis is different from testing the hypothesis of *total* zero variance. When confusing these two hypotheses, manipulating the covariance may yield unintuitive results. To reject the hypothesis of no slope variance in the presence of a non-zero intercept-slope covariance, the 2 degree-of-freedom test, or the generalized variance test, is the correct test. It draws power from both the intercept-slope covariance and the slope variance, which also makes it more powerful than the specific variance test (Hertzog et al., 2008;Ke and Wang, 2014). We are currently working on a derivation of the effective error corresponding to this two-dimensional null hypothesis (Brandmaier et al., in prep.) and will implement this derivation in a future version of LIFESPAN. Facilities for the Monte Carlo simulation of statistical power are already available for both the specific and the generalized test of slope variance.

6.2. LIFESPAN as a vehicle for progress in longitudinal study design

The LGCM is just one class of models for evaluating change. It does not directly address the issue of capturing various forms of causality (see Pearl, 2012). Future developments can consider the power to detect fixed and random regression coefficients in alternative structural regression models such as the bivariate dual-change score model (McArdle and Hamagami, 2001;e.g., Ferrer and McArdle, 2003; see Prindle and McArdle, 2012) as well as continuous time models (Voelkle et al., 2012).

LIFESPAN can be augmented in a number of ways that will enhance its usefulness as a tool for selecting and evaluating longitudinal study designs. The hope is to make LIFESPAN sufficiently flexible to serve as an instrument for promoting progress in longitudinal study design. From this perspective, the current emphasis on linear change as specified in a LGCM is a conservative design limitation that future versions of the program need to overcome. Power equivalence theory, in general, and the notion of effective error, in particular, will play a central role in this endeavor, as the concept of effective error is not limited to testing hypotheses about true variance in change, but can be extended to other effects of a given statistical model. von Oertzen and Brandmaier (2013) derived an effective error term for detecting intercept variance in the context of a LGCM.

In particular, we envision that future versions of LIFESPAN will ultimately include options to specify *variable spacing of measurement occasions* (Willett, 1989;Sliwinski et al., 2010), *selective attrition* (Lindenberger et al., 2002), *cohort-sequential designs* (Schaie, 1965;Baltes, 1968), *nonlinear change* (Ghisletta et al., submitted), and *planned missingness* (e.g., McArdle, 1994;Graham et al., 2001;Little et al., 2013;Rhemtulla et al., 2013). Some of these options are discussed in more detail below.

Alternative approaches to sampling time (i.e., occasions of measurement) are important because the density and distribution of measurement occasions influences the statistical power to detect variance in change (Willett, 1989;von Oertzen and Brandmaier, 2013;Rast and Hofer, 2014). More work is needed to find out which time-sampling schemes are well suited separating long-term change from forms of within-person variability that operate on shorter timescales (Nesselroade, 1991;Lindenberger and von Oertzen, 2006;Sliwinski et al., 2010). Following the original work by Willett (1989), increasing the variance of measurement intervals by giving up the longstanding habit of equally spaced measurement intervals seems highly commendable. Taken to the extreme, the sampling of time can be regarded as a variable that varies randomly across participants (e.g., Voelkle

and Oud, 2013)

Regarding selective attrition, future versions of LIFESPAN or related programs would allow researchers to specify drop-out rates, including selection equations to capture possible effects of non-random attrition (cf. Lawley, 1943;Lindenberger et al., 2002). As a first step in this direction, von Oertzen and Brandmaier (2013) derived power-equivalence relations based on score-independent drop-out to examine how power contributions shift from total study time to observation density depending on dropout rate.

It also would be useful to evaluate power in sequential sampling designs that incorporate convergence assumptions (Bell, 1953; 1954;McArdle and Hamagami, 2001;e.g., Moerbeek, 2011), and to formally explore the potential consequences of misspecification on the statistical power to detect variance in change (e.g., Sliwinski et al., 2010).

We remind readers that the generation of power-equivalent models requires the specification of population parameters. To the extent that these parameters are biased, unreliable, or simply wrong, the set of power-equivalent models derived on the basis of these parameters will be less useful than desired. Of course, this limitation also applies to Monte Carlo simulations, and to any other method for selecting and evaluating study designs. von Oertzen and Brandmaier (2013) advised to rely on conservative population values to obtain lower bounds on expected statistical power. Alternatively, it might be useful to formally treat the uncertainty in population values (see Kelley and Rausch, 2011;Lai and Kelley, 2011;Gribbin et al., 2013).

6.3. Outlook

Power evaluation programs such as LIFESPAN serve the purpose of helping researchers in crafting and selecting longitudinal designs that have optimal power to detect random effects of change, based on what is currently known about the change process under investigation. The goal of a fully flexible program that enhances longitudinal study design and obeys the principles of computer-aided design is more attainable than ever before, though a number of difficult problems still need to be resolved.

7. References

Baltes, P.B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Hum Dev* 11**,** 145-171.

Baltes, P.B., and Nesselroade, J.R. (1979). "History and rationale of longitudinal research," in *Longitudinal research in the study of behavior and development,* eds. J.R. Nesselroade & P.B. Baltes. (New York: Academic Press), 1-39.

Baltes, P.B., Reese, H.W., and Nesselroade, J.R. (1977). Life-span devlopmental psychology: Introduction to research methods. *Montcrey, California: Brooks/Colc*.

Baltes, P.B., Reuter-Lorenz, P.A., and Rösler, F. (2006). *Setting the Stage across the Ages of the Lifespan: Lifespan Development and the Brain.* Cambridge University Press.

Barrera-Gomez, J., Spiegelman, D., and Basagana, X. (2013). Optimal combination of number of participants and number of repeated measurements in longitudinal studies with time-varying exposure. *Statistics in Medicine* 32**,** 4748-4762. doi: Doi 10.1002/Sim.5870.

Bell, R.Q. (1953). Convergence: An accelerated longitudinal approach. *Child development***,** 145--152.

Bell, R.Q. (1954). An experimental test of the accelerated longitudinal approach. *Child Development***,** 281--286.

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T.,

Mehta, P., and Fox, J. (2011). OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika* 76**,** 306-317.

Boker, S.M., Mcardle, J., and Neale, M. (2002). An algorithm for the hierarchical organization of path diagrams and calculation of components of expected covariance. *Structural Equation Modeling: A Multidisciplinary Journal* 9**,** 174--194.

Bollen, K.A. (1989). *Structural Equations with Latent Variables.* Oxford, UK: John Wiley.

Brandmaier, A.M., Von Oertzen, T., Ghisletta, P., Lindenberger, U., and Hertzog, C. (in prep.). Effect Sizes of Random Slopes in Latent Growth Curve Models: Implications for Statistical Power Analysis. .

Coons, S.A., and Mann, R.W. (1960). "Computer-Aided Design Related To The Engineering Design Process". (Cambridge, Massachusetts: Electronic Systems Laboratory. Department of Electrical Engineering. Massachusetts Institute of Technology.).

Duncan, T.E., Duncan, S.C., and Strycker, L.A. (2013). *An introduction to latent variable growth curve modeling: Concepts, issues, and application.* Routledge Academic.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39**,** 175--191.

Ferrer, E., and Mcardle, J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling* 10**,** 493--524.

Ferrer, E., and Mcardle, J.J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science* 19**,** 149-154.

Fox, J. (2006). Teacher's Corner: structural equation modeling with the sem package in R. *Structural equation modeling* 13**,** 465--486.

Ghisletta, P., Von Oertzen, T., Hertzog, C., Nilsson, L., and Lindenberger, U. (submitted). Nonlinear Growth Models When Studying Long-Term Change: A Simulation Study of Exponential Decline in Adulthood.

Graham, J.W., Taylor, B.J., and Cumsille, P.E. (2001). Planned missing-data designs in analysis of change. *New Methods for the Analysis of Change***,** 335-353. doi: Doi 10.1037/10409-011.

Gribbin, M.J., Chi, Y.Y., Stewart, P.W., and Muller, K.E. (2013). Confidence regions for repeated measures ANOVA power curves based on estimated covariance. *BMC Med Res Methodol* 13**,** 57. doi: 10.1186/1471-2288-13-57.

Hedeker, D., Gibbons, R.D., and Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: Comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics* 24**,** 70-93. doi: Doi 10.3102/10769986024001070.

Hertzog, C. (1996). "Research design in studies of aging and cognition," in *Handbook of the Psychology of Aging* eds. J.E. Birren & K.W. Schaie. 4th ed: NY: Academic Press), 24--37.

Hertzog, C. (2008). "Theoretical approaches to the study of cognitive aging: An individual-differences perspective," in *Handbook of cognitive aging: Interdisciplinary perspectives,* ed. S.M.H.D.F. Alwin. (Thousand Oaks, CA, US: Sage Publications, Inc), 34-49.

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

Hertzog, C., Von Oertzen, T., Ghisletta, P., and Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling: A Multidisciplinary Journal* 15**,** 541--563.

Johansson, B., Hofer, S.M., Allaire, J.C., Maldonado-Molina, M.M., Piccinin, A.M., Berg, S., Pedersen, N.L., and Mcclearn, G.E. (2004). Change in cognitive capabilities in the oldest old: The effects of proximity to death in genetically related individuals over a 6-year period. *Psychology and Aging* 19**,** 145-156. doi: Doi 10.1037/0882-7974.19.1.145.

Johansson, B., Whitfield, K., Pedersen, N.L., Hofer, S.M., Ahern, F., and Mcclearn, G.E. (1999). Origins of individual differences in episodic memory in the oldest-old: a population-based study of identical and same-sex fraternal twins aged 80 and older. *J Gerontol B Psychol Sci Soc Sci* 54**,** P173-179.

Ke, Z., and Wang, L.P. (2014). Detecting Individual Differences in Change: Methods and Comparisons. *Structural Equation Modeling: A Multidisciplinary Journal*. doi: 10.1080/10705511.2014.936096.

Kelley, K., and Rausch, J.R. (2011). Sample Size Planning for Longitudinal Models: Accuracy in Parameter Estimation for Polynomial Change Parameters. *Psychological Methods* 16**,** 391-405.

Lai, K.K., and Kelley, K. (2011). Accuracy in Parameter Estimation for Targeted Effects in Structural Equation Modeling: Sample Size Planning for Narrow Confidence Intervals. *Psychological Methods* 16**,** 127-148. doi: Doi 10.1037/A0021764.

Lawley, D.N. (1943). XXIII.—On Problems connected with Item Selection and Test Construction. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences* 61**,** 273-287.

Lindenberger, U. (2014). Human cognitive aging: Corriger la fortune? *Science* 346**,** 572-578.

Lindenberger, U., Singer, T., and Baltes, P.B. (2002). Longitudinal selectivity in aging populations: separating mortality-associated versus experimental components in the Berlin Aging Study (BASE). *J Gerontol B Psychol Sci Soc Sci* 57**,** P474-482.

Lindenberger, U., and Von Oertzen, T. (2006). Variability in cognitive aging: From taxonomy to theory. *Lifespan cognition: Mechanisms of change***,** 297-314.

Lindenberger, U., Von Oertzen, T., Ghisletta, P., and Hertzog, C. (2011). Cross-sectional age variance extraction: what's change got to do with it? *Psychology and aging* 26**,** 34.

Little, T.D., Jorgensen, T.D., Lang, K.M., and Moore, E.W.G. (2013). On the joys of missing data. *Journal of pediatric psychology***,** jst048.

Luenberger, D.G. (1973). *Introduction to linear and nonlinear programming.* Reading, Mass. u.a.: Addison-Wesley.

Maccallum, R., Lee, T., and Browne, M.W. (2010). The issue of isopower in power analysis for tests of structural equation models. *Structural Equation Modeling* 17**,** 23--41.

Maxwell, S.E., Kelley, K., and Rausch, J.R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annu. Rev. Psychol.* 59**,** 537--563.

Mcardle, J.J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research* 29**,** 409--454.

Mcardle, J.J., and Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development* 58**,** 110--133.

Mcardle, J.J., and Hamagami, F. (2001). "Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data: New methods for the analysis of change.," in *New methods for the analysis of change,* eds. L.M. Collins & A.G. Sayer. (Washington, DC: American Psychological Association).

Mcardle, J.J., and Nesselroade, J.R. (2003). Growth curve analysis in contemporary psychological research. *Comprehensive handbook of psychology* 2**,** 447--480.

Mcardle, J.J., and Nesselroade, J.R. (2014). "Using the simplified reticular action model notation.," in *Longitudinal data analysis using structural equation models.* (Washington, DC: American Psychological Association), 59--66.

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

Moerbeek, M. (2011). The effects of the number of cohorts, degree of overlap among cohorts, and frequency of observation on power in accelerated longitudinal designs. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 7**,** 11.

Muthén, B.O., and 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., and Muthén, B.O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling* 9**,** 599-620. doi: Doi 10.1207/S15328007sem0904_8.

Muthén, L.K., and Muthén, B.O. (2007). "Mplus: Statistical analysis with latent variables: User's guide". (Los Angeles, CA).

Nesselroade, J.R. (1991). "The warp and woof of the developmental fabric". Hillsdale, NJ: Lawrence Erlbaum).

Pearl, J. (2012). "The Causal Foundations of Structural Equation Modeling," in *Handbook of Structural Equation Modeling,* ed. R.H. Hoyle. (New York: Guilford Press), 68--91.

Prindle, J.J., and Mcardle, J.J. (2012). An Examination of Statistical Power in Multigroup Dynamic Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal* 19**,** 351--371.

Rao, S.S. (2009). *Engineering Optimization: Theory and Practice.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Rast, P., and Hofer, S.M. (2014). Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: Simulation results based on actual longitudinal studies. *Psychological methods* 19**,** 133.

Raudenbush, S.W., and Liu, X.F. (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. doi: Doi 10.1037//1082-989x.6.4.387.

Rhemtulla, M., Savalei, V., and Little, T.D. (2013). On the asymptotic relative efficiency of planned missingness designs. *Psychometrika***,** 1-30.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48**,** 1--36.

Satorra, A., and Saris, W.E. (1985). Power of the Likelihood Ratio Test in Covariance Structure-Analysis. *Psychometrika* 50**,** 83-90. doi: Doi 10.1007/Bf02294150.

Schaie, K.W. (1965). A General Model for the Study of Developmental Problems. *Psychol Bull* 64**,** 92-107.

Schlesselman, J.J. (1973). Planning a longitudinal study: II. Frequency of measurement and study duration. *Journal of Chronic Diseases* 26**,** 561-570. doi: http://dx.doi.org/10.1016/0021-9681(73)90061-1.

Sliwinski, M., Hoffman, L., and Hofer, S.M. (2010). Evaluating convergence of within-person change and between-person age differences in age-heterogeneous longitudinal studies. *Research in Human Development* 7**,** 45--60.

Voelkle, M.C., Oud, J.H., Davidov, E., and Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological methods* 17**,** 176.

Voelkle, M.C., and Oud, J.H.L. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *British Journal of Mathematical & Statistical Psychology* 66**,** 103-126. doi: Doi 10.1111/J.2044-8317.2012.02043.X.

Von Oertzen, T. (2010). Power equivalence in structural equation modelling. *British Journal of Mathematical and Statistical Psychology* 63**,** 257--272.

Von Oertzen, T., and Brandmaier, A.M. (2013). Optimal study design with identical power: An application of power equivalence to latent growth curve models. *Psychology and aging* 28**,** 414.

Von Oertzen, T., Brandmaier, A.M., and Tsang, S. (2015). Structural Equation Modeling with Ωnyx. *Structural Equation Modeling* 22**,** 148-161. doi: doi:10.1080/10705511.2014.935842.

Von Oertzen, T., Hertzog, C., Lindenberger, U., and Ghisletta, P. (2010). The effect of multiple indicators on the power to detect inter-individual differences in change. *British Journal of Mathematical and Statistical Psychology* 63**,** 627--646.

Willett, J.B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement* 49**,** 587--602.

8. Figure legends

Figure 1 **Main Screen of LIFESPAN.** This screenshot shows the specification mode of LIFESPAN. Text fields allow researchers to type in study design parameters, for instance, time span or the number of measurement occasions, and best guesses about true variances in intercept and linear change. At the top, the current study design is displayed as a path diagram.

Figure 2 **Iso-power plots for bivariate trade-offs between parameters in a LGCM based on the OCTO-Twin Study**: number of occasions and residual variance (top left), number of occasions and time span (top right), variance of slope and residual variance (bottom left), and variance of intercept and time span (bottom right). The original study design is marked with a cross in each panel.

Figure 1.TIF

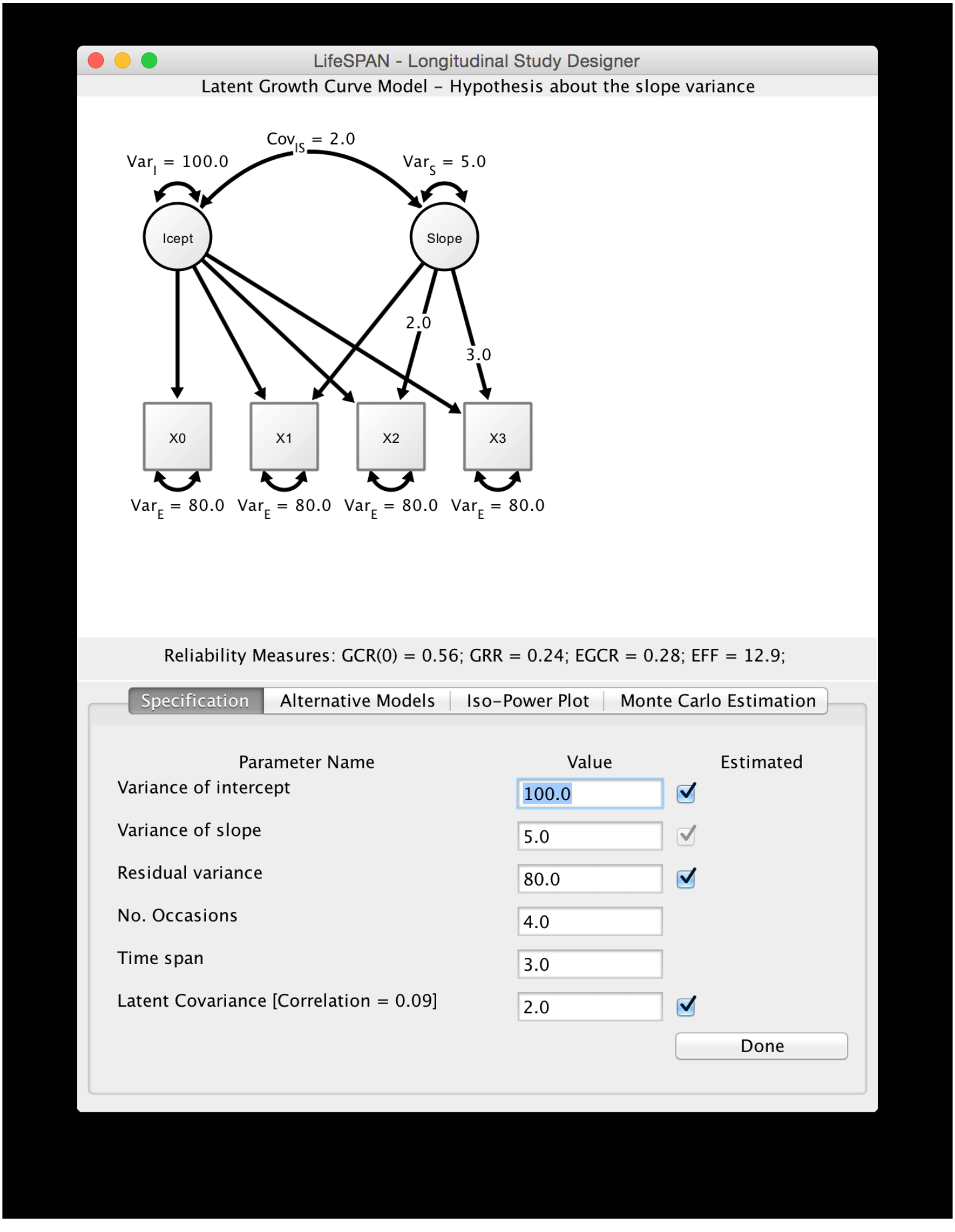
LifeSPAN - Longitudinal Study Designer
Latent Growth Curve Model – Hypothesis about the slope variance
Cov_IS = 2.0
Var_I = 100.0
Var_S = 5.0
Icept
Slope
2.0
3.0
X0
X1
X2
X3
Var_E = 80.0
Var_E = 80.0
Var_E = 80.0
Var_E = 80.0
Reliability Measures: GCR(0) = 0.56; GRR = 0.24; EGCR = 0.28; EFF = 12.9;
Specification
Alternative Models
Iso-Power Plot
Monte Carlo Estimation
Parameter Name
Value
Estimated
Variance of intercept
100.0
Variance of slope
5.0
Residual variance
80.0
No. Occasions
4.0
Time span
3.0
Latent Covariance [Correlation = 0.09]
2.0
Done

Figure 2.TIFF