

On the use of growth models to study normal cognitive aging

Paolo Ghisletta,^{1,2,3} Fabio Mason,¹ Timo von Oertzen,⁴ Christopher Hertzog,⁵ Lars-Göran Nilsson,⁶ and Ulman Lindenberger^{7,8,9}

International Journal of Behavioral Development
2020, Vol. 44(1) 88–96
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0165025419851576
journals.sagepub.com/home/jbd



Abstract

Growth models (GM) of the mixed-effects and latent curve varieties have become popular methodological tools in lifespan research. One of the major advantages of GM is their flexibility in studying individual differences in change. We scrutinized the change functions of GM used in five years of publications on cognitive aging. Of the 162 publications that we identified, 88% test linear or quadratic polynomials, and fewer than 5% apply functions that are nonlinear in their parameters, such as exponential decline. This apparent bias in favor of polynomial decomposition calls for exploring what conclusions about individual differences in change are likely to be drawn if one applies linear or quadratic GMs to data simulated under a conceptually and empirically plausible model of exponential cognitive decline from adulthood to old age. Hence, we set up a simulation that manipulated the rate of exponential decline, measurement reliability, number of occasions, interval width, and sample size. True rate of decline and interval width influenced results strongly, number of occasions and measurement reliability exerted a moderate effect, and the effects of sample size appeared relatively minor. Critically, our results show that fit statistics generally do not differentiate misspecified linear or quadratic models from the true exponential model. Moreover, power to detect variance in change for the linear and quadratic GMs is low, and estimates of individual differences in level and change can be highly biased by model misspecification. We encourage researchers to also consider plausible nonlinear change functions when studying behavioral development across the lifespan.

Keywords

Growth model, long-term change, normal cognitive aging, nonlinear mixed-effects models, longitudinal research designs

The last 25 years have seen major advances in statistical models for characterizing developmental change (e.g., Hertzog & Nesselrode, 2003; Little, 2013; Singer & Willett, 2003). One popular method is the growth model (GM), implemented as a linear mixed-effects model (LME); Bryk & Raudenbush, 1987; Laird & Ware, 1982) or a latent curve model (McArdle, 1988; Meredith & Tisak, 1990). The critical feature of the GM is that it models change by specifying an intercept and a slope component describing the overall level and change across time, respectively. For both components, one can estimate the mean parameter and variance parameter reflecting individual deviations around the mean. Given individual differences in level and/or in change, the GM can be readily expanded to test antecedents and/or correlates of such individual differences.

The Growth Model

The linear GM specifies a variable Y_{ij} for an individual j at time i ($i=0, 1, \dots, T-1$) in terms of a level π_{0j} and a linear rate of growth π_{1j} multiplied by age a_{ij} (often centered around its mean) at time i for each individual j , plus a residual E_{ij} :

$$\begin{aligned} Y_{ij} &= \pi_{0j} + \pi_{1j}a_{ij} + E_{ij} \\ \pi_{0j} &= \beta_0 + U_{0j} \\ \pi_{1j} &= \beta_1 + U_{1j} \end{aligned} \quad (1)$$

The linear GM predicts all individual growth trajectories to be straight lines. The level corresponds to the intercept (prediction when $a_{ij}=0$) and the rate of growth to the linear slope (predicted

change in Y for one unit change in a). Both level and change have a mean value of β_0 and β_1 and a subject-specific deviation from the mean of U_{0j} and U_{1j} , respectively (Bryk & Raudenbush, 1987; Laird & Ware, 1982). β_0 and β_1 are fixed effects and U_{0j} and U_{1j} are random effects, assumed to be normally distributed around zero (with variances σ^2_L and σ^2_C) and to possibly covary (σ_{LC}). The

¹ Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland

² Swiss Distance Learning University, Switzerland

³ Swiss National Centre of Competence in Research LIVES – Overcoming Vulnerability: Life Course Perspectives, Universities of Lausanne and of Geneva, Switzerland

⁴ Department of Psychology, University of the Federal Defense Forces, Germany

⁵ School of Psychology, Georgia Institute of Technology, USA

⁶ Department of Psychology, Stockholm University, Sweden

⁷ Center for Lifespan Psychology, Max Planck Institute for Human Development, Germany

⁸ Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany, and London, UK

⁹ European University Institute, San Domenico di Fiesole, Italy

Corresponding author:

Paolo Ghisletta, Faculty of Psychology and Educational Sciences, University of Geneva, Boulevard du Pont d'Arve 40, 1211 Genève, Switzerland.
Email: paolo.ghisletta@unige.ch

errors E_{ij} are typically assumed independently and normally distributed with time-invariant error variance σ^2_E .

Another popular specification of the GM is the *quadratic GM* of Equation (2) that allows the predicted trajectories to have a single change in direction. This model augments the linear GM by adding a second (quadratic) change component that linearly relates age squared to Y_{ij} :

$$\begin{aligned} Y_{ij} &= \pi_{0j} + \pi_{1j}a_{ij} + \beta_2 a_{ij}^2 + E_{ij} \\ \pi_{0j} &= \beta_0 + U_{0j} \\ \pi_{1j} &= \beta_1 + U_{1j} \end{aligned} \quad (2)$$

Compared to the linear model, the quadratic additionally estimates one additional parameter, β_2 , which allows the predicted trajectories to follow a quadratic polynomial. Although the quadratic change component could also be specified with random effects, their estimation is typically not included in extant applications (in our literature review, of the 40 records that tested a quadratic model, only nine declared having tested random effects of the quadratic slope, and of those only three found positive evidence).

Linear and quadratic GMs are often contrasted with a simpler *no growth or level model* that postulates no change whatsoever, with individuals maintaining the same level score over time (Equation (3); Widaman & Thompson, 2003):

$$\begin{aligned} Y_{ij} &= \pi_{0j} + E_{ij} \\ \pi_{0j} &= \beta_0 + U_{0j}, \end{aligned} \quad (3)$$

to test the hypothesis that Y_{ij} changes over time.

The Growth Model as a nonlinear Mixed-effects Model

So far, the specified GMs represent change as a linear combination of level, change(s), and errors, and are thus linear in their parameters (Davidian & Giltinan, 1995). The linearity of the parameters is not to be confused with the linearity of the shape of specified change (Cudeck & Harring, 2007). In particular, the quadratic GM specifies a nonlinear trajectory of change, despite being a LMEM.

LMEMs can be generalized to allow for nonlinear relations between Y_{ij} and the parameters of the growth components (Davidian & Giltinan, 1995), generating nonlinear mixed-effects models (NLMEMs). These models have been successfully applied in various disciplines (e.g., pharmacokinetics, biology, demography, finance), to capture underlying change processes that are inherently nonlinear (e.g., drug effects and population growth often follow a logistic function, accumulation of financial interests is often exponential). One can argue that many psychological constructs change nonlinearly over time. For instance, effects of accumulating age-related insults to an adult's central nervous system could be modeled by exponential decline. Indeed, nonlinear GM have occasionally been applied to study adult cognitive change, both normal (e.g., Grimm, McArdle, & Hamagami, 2007; McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002;) and pathological (e.g., Driscoll et al., 2006).

One variant of the (nonlinear) exponential model, as employed by several authors (e.g., McArdle et al., 2002; Grimm et al., 2007) is:

$$\begin{aligned} Y_{ij} &= \pi_{0j} + \pi_{1j}e^{\gamma a_{ij}} + E_{ij} \\ \pi_{0j} &= \beta_0 + U_{0j} \\ \pi_{1j} &= \beta_1 + U_{1j} \end{aligned} \quad (4)$$

This nonlinear mixed-effects GM specifies an exponential functional form of change, so that as age increases successive gains or losses are not constant but proportional to the current value. Compared to the linear GM, this model estimates one additional parameter, the exponential rate of change γ . Typically, random effects for γ are not estimated in extant applications. However, because of random effects of π_{1j} , which are estimated as the variance of change σ^2_C , individuals are allowed to differ in the amount of exponential decline, rather than in its rate.

The Present Study

We start by reviewing five years of scientific literature on applications of the GM to study normal cognitive aging. This review established that the linear and quadratic GMs are very often applied by substantive researchers in this field, whereas nonlinear mixed-effects GMs, on the other hand, are very rarely used.

Justified by this pattern, we evaluated what happens if longitudinal data, created by an underlying exponential model, are analyzed with a linear or quadratic GM. We investigate how well the linear and quadratic GMs preserve or distort individual differences in level and change in such instances. If the individual differences specified by the underlying generating model are largely concordant with those estimated by the misspecified GMs, one would be justified to claim that LMEM GMs are practically useful despite their misspecification. That is, the models would be of the "wrong but useful" kind (cf. Box & Draper, 1987). If, however, the two sets of results lead to discordant inferences about individual differences in change, we would conclude that the misspecified GMs are misleading.

Method

Literature Review

We computed a detailed literature search to study how longitudinal data are commonly analyzed within GMs to study normal cognitive aging. By covering five years of literature we initially obtained 1253 records, which reduced to 162 relevant records for full analysis. The detailed search and screening procedure are described in supplementary material (Appendices 1–2, Table 1, and Figure 1).

Simulation Design

Population model. The exponential population model of Equation (4) was used to generate the data for the Monte Carlo simulation (McArdle et al., 2002). We specified the existence of individual differences in both level and change, but not in decline rate, γ . This choice was guided by the facts that (a) in our empirical illustration (see below) estimating random effects for the exponential rate failed, (b) extant applications of the exponential model usually exclude random effects for γ , and, more importantly, (c) estimating multiple random effects in both linear and especially NLMEMs requires data denser than those of standard longitudinal panels.

Simulation factors. We based our simulation study on parameter estimates from an exponential model fitted to data from the Betula Project (Nilsson et al., 1997), a well-known study of aging, memory, and dementia. We analyzed the episodic recall scores of 1000 Betula participants assessed four times, at five-year intervals, as a function of age (final range: 35–95 years). We tested the level,

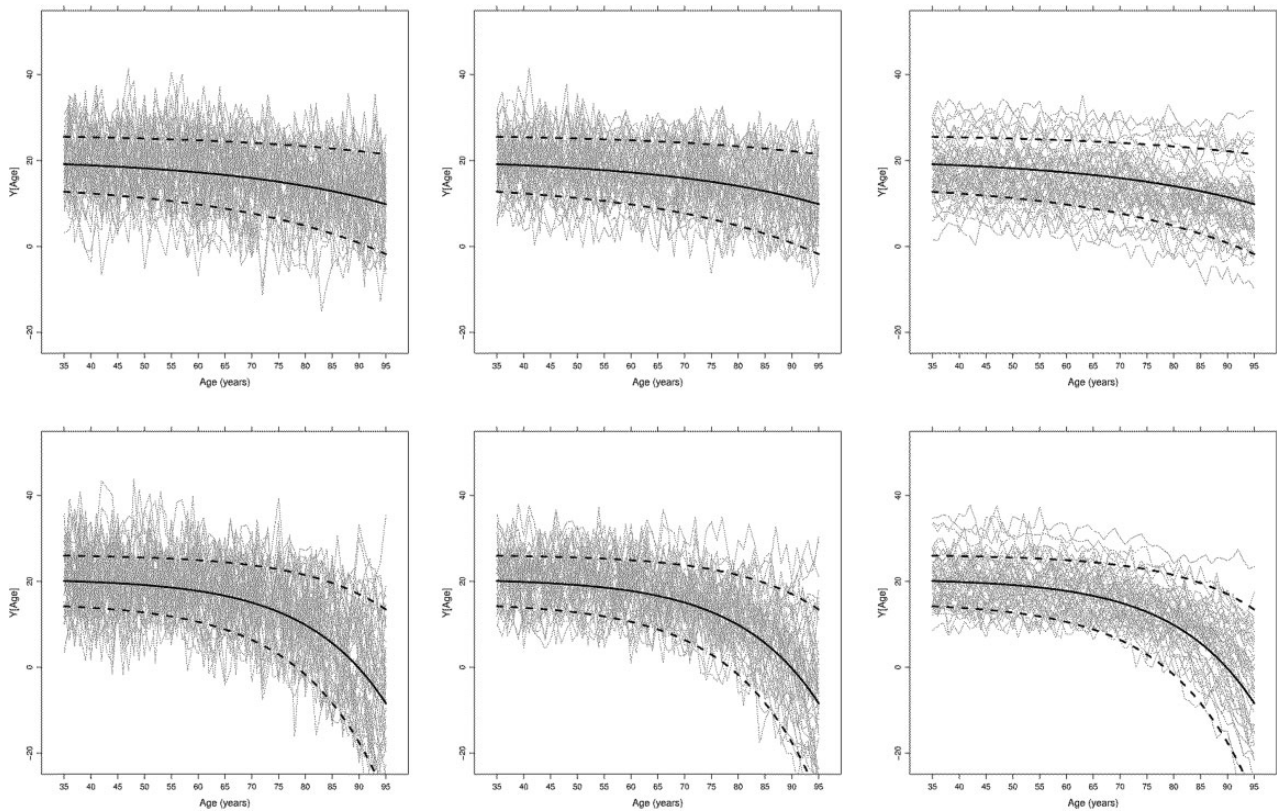


Figure 1. Simulated trajectories generated under the exponential decline model ($N = 50$).

Note. Row 1: $\gamma = 0.033$, row 2: $\gamma = 0.066$; Column 1: $GCR = .500$, column 2: $GCR = .681$, column 3: $GCR = .900$; Thick continuous line is sample average without residuals, thick dashed lines are 1 SD above/below the mean; Thin continuous lines are 50 simulated individual trajectories.

linear, quadratic, and exponential GMs and found that the exponential best described the data. We used the estimated population parameters as the basis for the simulation experiment (for full results, see supplementary Table 2 and supplementary Figure 2).

The constant factors were the level mean ($\beta_0 = 20.603$), change mean ($\beta_1 = -4.007$), level variance ($\sigma^2_L = 30.980$), change variance ($\sigma^2_C = 5.041$), and level-change covariance ($\sigma_{LC} = -3.163$, corresponding to a correlation of $-.253$). The varying factors were the rate of exponential decline ($\gamma = 0.066$, corresponding to the illustration, and $\gamma = 0.033$, for a gentler decline) and the Growth Curve Reliability ($GCR = .500, .681$ (cf. illustration), or $.900$, corresponding to $\sigma^2_E = 30.980, 14.541$ (cf. illustration), and 3.442 , respectively), defined as the proportion of level variance over total variance at time=0 (age 65; $GCR = \sigma^2_L / (\sigma^2_L + \sigma^2_E)$; McArdle, 1988).

We generated data under six conditions, 2 (γ) by 3 (GCR), to obtain continuous trajectories spanning over a maximum of 71 time points (i.e., from 35 to 105 years). Figure 1 presents 50 random individual trajectories simulated under the six different conditions.

Design factors. In practice, longitudinal studies of long-term change processes often adopt a longitudinal sequence design, which begins with a cross-sectional age-heterogeneous sample of persons and then follows them longitudinally over time. This sampling design generates multiple parallel subsamples differing in initial age, each providing longitudinal data on different segments of the targeted overall age range, with coverage increasing as the number of measurement occasions increases or as the measurement interval

widens (Baltes, Reese, & Nesselroade, 1988). We included 10 longitudinal subsamples with 50, 100, or 200 units each, yielding a total sample size of $N = 500, 1000$, or 2000 . Each subsample contained observations representing individuals of initial age 35, 40, 45, etc. to 80 years, observed repeatedly either $T = 3$ or $T = 6$ times. Furthermore, we varied Δ , the interval width between adjacent occasions of measurement, going from $\Delta = 1$ to $\Delta = 5$ years in steps of 1 year.

This design mimics a study where the youngest subsample observed the least ($T = 3$) at the shortest interval ($\Delta = 1$) would be assessed at ages 35, 36, and 37 years, the next subsample at ages 40, 41, and 42, etc. and the oldest subsample at ages 80, 81, and 82 years. The youngest subsample observed the most ($T = 6$) at the shortest interval ($\Delta = 1$) would be assessed at ages 35, 36, 37, 38, 39, and 40 years, thus overlapping with the next subsample (ages 40, 41, 42, 43, 44, and 45). The oldest subsample observed the most ($T = 6$) at the shortest interval ($\Delta = 1$) would have ages 80, 81, 82, 83, 84, and 85, while if Δ increases to 5, the ages would be 80, 85, 90, 95, 100, and 105 years, ultimately covering the entire adult life span. Crossing low numbers of repeated observations ($T = 3$ or 6) with very short to relatively long intervals within repeated observations ($\Delta = 1$ to 5) more closely mimics existing longitudinal studies than assuming that all units have been observed during the full length of the study and at all occasions (e.g., Schaie & Hofer, 2001). Our illustration from the Betula study counts 10 subsamples, each of 100 individuals (hence $N = 1000$), of initial age 35, 40, 45, 50, 55, 60, 65, 70, 75, or 80 years, assessed on four occasions ($T = 4$), with an interval between adjacent occasions of $\Delta = 5$ years.

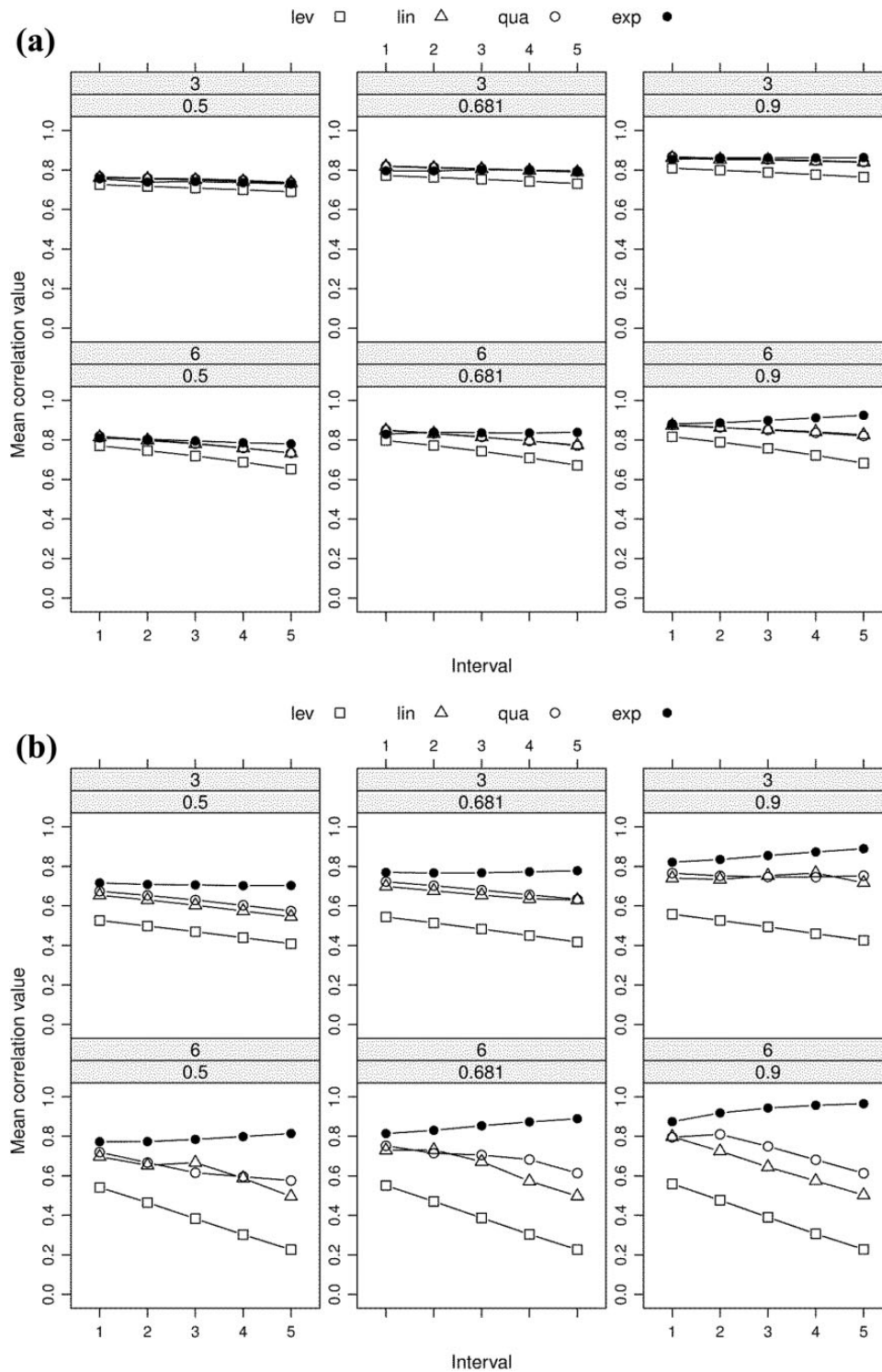


Figure 2. (a) Concordance in levels by interval width and model, for $\gamma = 0.033$ (averaged over $N = 500, 1000, 2000$). (b) Concordance in levels by interval width and model, for $\gamma = 0.066$ (averaged over $N = 500, 1000, 2000$).
 Note. Rows 1 and 3: $T = 3$, rows 2 and 4: $T = 6$; Column 1: GCR = .500, column 2: GCR = .681, column 3: GCR = .900.

Summary. In total, we fully crossed the varying simulation factors and design features to obtain 180 conditions: 2 (γ) by 3 (GCR) by 2 (T) by 5 (I) by 3 (N) (see Supplementary Table 3 for a full description of the simulation design). For each condition we generated 100 data sets, for a total of 18,000 data sets.

Procedure

Growth model analyses. For each generated data set we estimated a series of GMs. We first applied the exponential decline GM. By doing so we tested whether the estimation procedure recovered the

values of the generating function despite sampling limitations on a number of occasions and varying interval lengths, and also provided a basis for comparing results obtained from the alternative GMs. We subsequently implemented the level, linear, and quadratic GMs. For the level GM we estimated three parameters (β_0 , σ^2_L , and σ^2_E ; cf. Equation (3)); for the linear GM, 6 parameters (β_0 , β_1 , σ^2_L , σ^2_C , σ_{LC} , and σ^2_E ; cf. Equation (1)); and for the quadratic GM, 7 parameters (β_0 , β_1 , β_2 , σ^2_L , σ^2_C , σ_{LC} , and σ^2_E ; cf. Equation (2)); finally, for the exponential decline GM, we estimated seven parameters (β_0 , β_1 , γ , σ^2_L , σ^2_C , σ_{LC} , and σ^2_E ; cf. Equation (4)).

In all GM analyses we centered the time predictor a_{ij} (age) around its grand (sample) mean, to reduce estimation bias in the level-change covariance (σ_{LC}), and in the level variance (σ^2_L ; Mehta & West, 2000; Rovine & Molenaar, 1998; Wainer, 2000).

The level, linear, and quadratic GMs are LMEMs, for which the marginal likelihood can be expressed and estimated without approximation. However, because the exponential model is a NLMEM, we had to apply an approximation method (Davidian & Giltinan, 1995). To rule out that the estimation procedure acted as a generator of divergent results we applied the same estimation procedure to all models. We used PROC NLMIXED in SAS (v. 9.3) to estimate all GMs with the adaptive Gaussian quadrature approximation, which produces the same estimates as maximum likelihood in linear models (Littell, Milliken, Stroup, & Wolfinger, 1996).

Acceptable solutions. We deemed a solution acceptable only if (a) the estimation converged to a solution and (b) the solution contained parameter estimates that were statistically viable. In particular, the estimated variance of change (σ^2_C) had to be non-negative.

Goodness of fit. From each analysis we saved the Schwarz Bayesian Information Criterion (BIC), defined as

$$BIC = 2 \cdot f(\hat{\theta}) + p \cdot \ln(N),$$

where f is the negative of the marginal log-likelihood function, $\hat{\theta}$ is the vector of parameter estimates, p is the number of estimated parameters, \ln is the natural logarithm, and N is the number of subjects (Schwarz, 1978). This index is not normed, but allows comparing alternative models tested on the same data. A difference in BIC values of more than 10 between competing models can be interpreted as strong evidence in favor of the model with the lower BIC value (Kass & Raftery, 1995).

Power to detect individual differences in change. We estimated the power to detect interindividual differences in change by comparing each GM to a nested GM where the variance in change (σ^2_C) and the level-change covariance (σ_{LC}) were fixed at zero. We then compared each fully specified GM to its nested counterpart, and computed a likelihood ratio test ($\alpha = 5\%$).

Concordance in individual differences. We evaluated the validity of individual difference estimates from each model by inspecting the concordance of individual differences in growth components between the generated data and the results obtained from the GM analyses. To generate the data according to the exponential decline model of Equation (4) we first generated random effects in level and change scores (i.e., U_{0j} and U_{1j} of Equation (4)) in accordance to the constant simulation factors (cf. Supplementary Table 3). We then combined these random growth scores according to Equation

(4) and added the randomly generated time-specific error component (E_{ij}), to obtain the individual repeated observations (Y_{ij}).

We analyzed each simulated data set of individual repeated observations with the four GMs and each time saved the empirical Bayes estimates of the random effects of the level and the change component. To inspect the concordance of the individual differences in growth components, we correlated the random level and change scores used to generate the data with the empirical Bayes estimates of the random effects in level and change obtained from the GM analyses.

Results

Literature Review

The literature search produced 1253 records, and after exclusion of duplicate and irrelevant records, we fully analyzed 162 records (see Supplementary Table 1). Of the 162, 102 records (63%) tested only a linear function of change, whereas 40 (25%) tested also a quadratic function, and one a cubic function. Thirteen (8%) records tested a broken-stick function (aka single-node spline; ten of which compared it to a linear or quadratic function). Of the remaining records, five (3%) tested an exponential function, two (1%) estimated the change function from the data (within a latent curve model), one used a multiple-node spline approach, and one did not specify the change function (see Supplementary Appendix 3 for detailed results).

Overall, then, 142 (88%) of the 162 records either used the linear or the quadratic function, whereas eight (less than 5%) used a function that is nonlinear in its parameters.

Simulation Results

Number of acceptable solutions. Generally, the greater the sample, the number of repeated measurements, the interval of measurements, the variable's reliability, and the rate of exponential decline, the higher the number of acceptable solutions (AS). The percentages of AS across models were: 100, 93.53, 85.57, and 97.74 for the level (which does not estimate σ^2_C), linear, quadratic and exponential model, respectively. Supplementary Figure 3 displays the average number of AS as a function of the design features.

To aid understanding of these effects, we used logistic regression to predict the probability of obtaining an AS as a function of the simulation factors. Model type most strongly influenced the probability of an AS ($\sim R^2 = 13\%$ -25%), GCR, T , and Δ have low effects ($\sim R^2 = 3\%$ -8%), and sample size was the least influential factor ($\sim R^2 < 1\%$) (see supplementary Appendix 4 for detailed results).

Goodness of fit. We differenced BIC values of the level, linear, and quadratic GMs against the exponential GM's BIC and compared this index of evidence favoring the exponential as a function of simulation condition and design features. The rejection rates based on BIC differences for the level, linear, and quadratic GMs were, respectively, 100%, 99.955%, and 96.295% for $\gamma = 0.066$ and 99.884%, 69.625%, and 47.191% for $\gamma = 0.033$. Thus, for a shallow declining process, the quadratic model would be judged to fit the data as well as the exponential more than half the time. Supplementary Figure 4 displays this information in detail.

We computed a logistic regression predicting the odds that the linear and quadratic GMs be rejected in favor of the exponential model as a function of GCR, T , Δ , and N , for $\gamma = 0.033$ (and not for

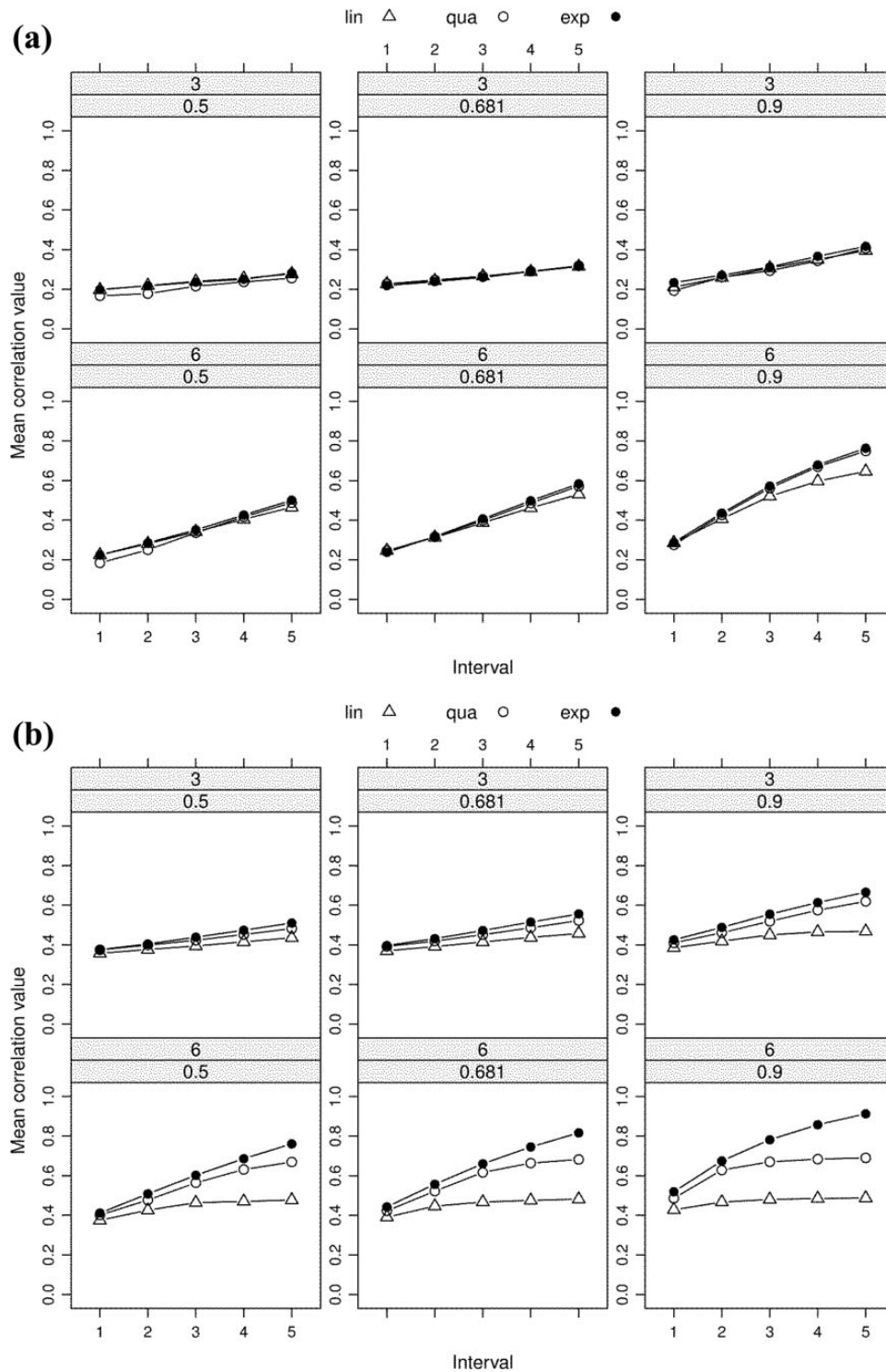


Figure 3. (a) Concordance in changes by interval width and model, for $\gamma = 0.033$ (averaged over $N = 500, 1000, 2000$). (b) Concordance in changes by interval width and model, for $\gamma = 0.066$ (averaged over $N = 500, 1000, 2000$).
 Note. Rows 1 and 3: $T = 3$, rows 2 and 4: $T = 6$; Column 1: $GCR = .500$, column 2: $GCR = .681$, column 3: $GCR = .900$.

$\gamma = 0.066$ or the level GM, because of the extremely high rejection rates). The number of occasions (T), and interval width (Δ) were the most influential factors ($\sim R^2 = 13\%–25\%$), GCR had a low effect ($\sim R^2 = 8\%–16\%$), and N was rather unimportant ($\sim R^2 = 3\%–12\%$) (see supplementary Appendix 5 for detailed results).

Power to detect individual differences in change. Supplementary Figure 5 shows the average proportions of times the nested analogous model with $\sigma^2_C = \sigma_{LC} = 0$ was rejected compared to the fully specified linear, quadratic, and exponential model, only for $\gamma = 0.033$. For $\gamma = 0.066$ these rejection curves are virtually at 1 across

all other conditions. Generally, the rejection curve of each model (hence estimated power) increased monotonically as Δ increases, with greater values at $GCR = .900$, $T = 6$, and Δ greater than 3. As GCR decreased, the rejection curves needed wider Δ to achieve high values, especially when $T = 3$. With the lowest GCR , the quadratic model had the least power to detect variance in change, especially when $T = 3$ and with small Δ . The most flagrant disparity between the quadratic model and the others was visible when $T = 3$, $N = 2000$, $GCR = .500$, and $\Delta = 1$: the power estimated for the quadratic GM in that cell is .105, while for the linear and exponential models it was .625 and .612, respectively.

To understand further the effects influencing detection of change variance we computed a logistic regression for each GM predicting the probability of a significant likelihood ratio test as a function of GCR , T , Δ , and N , for $\gamma = 0.033$. For all models the strongest factor was again Δ ($\sim R^2 = 10\%–20\%$), followed by T for the linear and quadratic models ($\sim R^2 = 7\%–14\%$), and by N for the exponential ($\sim R^2 = 8\%–13\%$). In the linear and exponential models, GCR influenced power the least ($\sim R^2 = 2\%–8\%$), while in the quadratic it was N ($\sim R^2 = 3\%–5\%$) (see supplementary Appendix 6 for detailed results).

Concordance in individual differences in level. We computed analyses of variance (ANOVAs) predicting correlations between the level scores used to generate the data and the Bayes estimates of levels from each GM (in all ANOVAs the residuals were normally distributed, centered on zero, and homoscedastic). A first ANOVA testing only main effects obtained that they were significant, although their sizes were moderate to strong for type of model (partial eta squared $\eta_p^2 = .536$), γ ($\eta_p^2 = .464$), Δ ($\eta_p^2 = .112$), GCR ($\eta_p^2 = .159$), but virtually nil for T and N (both $\eta_p^2 < .001$). All two-way interactions about N were irrelevant (all $\eta_p^2 < .0001$). However, some two-way interactions about T were associated to small, yet nonzero, effect sizes. In the end, we display the results graphically with respect to all effects but N .

Figure 2a displays the average level correlations for $\gamma = 0.033$, by Δ (abscissa), T (rows), and GCR (columns). In general, all indices are around .8. There appeared to be little effect of the model and of Δ when $GCR = .500$ and $T = 3$, but when both T and GCR increased, the level model produces lower concordance rates, especially as Δ increased. The correlations of the linear and quadratic models were hardly distinguishable, and they both diminished as Δ increased (especially when $T = 6$). For the exponential model the concordance rates remain stable or even rise as Δ increased. An ANOVA for the $\gamma = 0.033$ condition obtained the following effect size estimates (η_p^2): .308, .097, and .333 for type of model, Δ , and GCR , and less than .002 for T and N .

The disparities in concordance rates across the models increased dramatically when $\gamma = 0.066$ (cf. Figure 2b). Already when $T = 3$ and $GCR = .500$, the exponential model produced the highest level correlations. As Δ increases, the correlations diminished for all models except for the exponential, for which they remained stable or grew even further, approaching 1.0, whereas for the level, linear, and quadratic they dropped at, respectively, .2, .6, and .6–.7. An ANOVA for the $\gamma = 0.066$ condition obtained the following effect size estimates (η_p^2): .821, .273, and .244 for type of GM, Δ , and GCR , and less than .001 for T and N .

Concordance in individual differences in change. We repeated the same ANOVAs with respect to individual differences in change

information. An ANOVA with main effects only obtained moderate to strong effects for all factors but sample size (for type of GM, γ , Δ , GCR , and T , the η_p^2 estimates were .195, .568, .527, .281, and .472, respectively; for N , $\eta_p^2 = .002$). N was also irrelevant with respect to all two-way interactions (all $\eta_p^2 < .001$). We again graphically display all effects but sample size, in Figure 3a for $\gamma = 0.033$ and 3b for $\gamma = 0.066$, respectively.

With $\gamma = 0.033$, concordance indices started very low (around .2) also for the exponential model, and only increased considerably as a function of wider Δ when GCR increased. The three models obtain similar correlations, except with $GCR = .900$ and $T = 6$, where the linear model performed slight worse than the quadratic and exponential models. An ANOVA for the condition $\gamma = 0.033$ obtained the following effect size estimates (η_p^2): .014, .599, .389, and .569 for type of GM, Δ , GCR , and T and less than .005 for N .

When $\gamma = 0.066$, in general the lowest correlations started around .4 and increased with wider Δ . When $T = 3$, the quadratic and exponential models are again hardly discernable, while the linear model obtains lower correlations. With $T = 6$, however, starting at $\Delta = 5, 4$, and 3, for $GCR = .500, .681$, and .900, respectively, the exponential model clearly obtained stronger concordance indices about change than the quadratic model (reaching .9 vs. .6). An ANOVA for the condition $\gamma = 0.066$ obtained the following effect size estimates (η_p^2): .509, .570, .275, and .487 for type of GM, Δ , GCR , and T and less than .002 for N .

Prediction Quality in the Betula Project

To evaluate one possible implication of the application of the various GMs considered here, we assessed the quality of the linear, quadratic, and exponential GMs applied to analyze the Betula Project data used in the illustration and that informed our simulation study. For each participant we combined, according to Equations (1), (2), and (4) the estimated level and change scores and their age of assessment to predict their episodic recall scores. We then calculated the mean squared errors (MSE) between the observed and predicted scores under each GM for each participant.

We obtained MSE of 11.40, 10.25, and 9.94, respectively, for the linear, quadratic, and exponential GM. Thus, the exponential GM, which adjusted best to the data in terms of BIC (cf. supplementary Table 2), also produced predictions that were closest to the data, thereby showing the highest internal validity.

Survival Prediction in the Betula Project

In terms of predictive validity, we compared the three GMs in terms of the terminal decline hypothesis in the Betula Project data. This hypothesis states that many individuals during the final phase of life manifest accelerated decrements in functional capacities (Gerstorf & Ram, 2013). We used Bayesian factor score estimates of each person's level and change scores on the basis of the linear, quadratic, and exponential GMs to predict survival in the 1000 Betula participants considered here, 529 of whom were deceased by the study's most recent mortality update.

The Cox proportional hazard model predicted survival and the age of either death (for the deceased) or the last assessment (for the right-censored participants). The exponential level and change estimated scores predicted best survival, with a BIC of 5681, whereas the quadratic and linear obtained worse adjustments (BIC

values of 5685 and 5691, respectively). The estimated level and change scores from the exponential GM lowered the fit of the survival the most (drop in χ^2 statistic of 32.33 for two degrees of freedom, compared to 28.69 and 22.29 for the quadratic and linear, respectively – all p 's < .01).

Discussion

As expected, our literature review found that the vast majority (86%) of GMs applied to study normal cognitive aging implemented a linear or a quadratic function of change, whereas very few (less than 5%) considered a truly nonlinear function of change.

Hypothesized Consequences for Data Analysts

We explored what conclusions are likely to be reached when data generated under an exponential GM are analyzed with a linear or quadratic GM.

First, the exponential model can be implemented without great difficulty, despite its complexity. Exponential models using adaptive Gaussian quadrature failed to converge in less than 3% of the 18,000 data sets we generated. In contrast, the quadratic model failed to converge 14.43% of the cases. Thus, within the constraints present in our simulation study, the exponential model is a practicable alternative model to the linear and quadratic GMs.

Second, we found that in many conditions of our simulation study, it is not possible to tease apart the quadratic and the exponential model based on the commonly used BIC goodness of fit index. Thus, the chances of retaining the quadratic over the exponential model are high. Most analysts typically testing linear and quadratic models will not consider other functions; nevertheless, those scientists evaluating an exponential model using comparative fit may end up discarding it for the more familiar and equally well-fitting quadratic model.

Third, we evaluated possible consequences of retaining the misspecified quadratic model in our simulation study. Would doing so necessarily lead to inaccurate substantive conclusions about the change process analyzed? We addressed this question by considering two aspects: the power to detect individual differences in change, and the concordance between the level and change scores used to generate the data and the analogous scores predicted from the growth models. With respect to power, the quadratic GM clearly performed worst (across all conditions at 72.22%), whereas the linear and exponential fared similarly and in a range deemed by many to be satisfactory (82.21% and 81.26%, respectively). This gap increases in many conditions commonly encountered in empirical research. Thus, based on the quadratic model, we would often conclude that the individuals in the sample cannot be discriminated with respect to amount of change, whereas the exponential model may detect that important source of heterogeneity.

In contrast, the characterization of individual differences in growth components is often seriously compromised in quality if we rely on too simple a GM, especially when the decline rate is substantial. Concordance rates were generally higher for the level than for the change component. This was expected, as level information is more easily estimated than change information. Whereas the level correlations either remain stable or increase as testing intervals widen for the exponential model, for the level, linear, and quadratic models the concordance rates decrease monotonically. This effect is influenced by measurement

reliability (the higher the GCR, the stronger the discordance; this effect would become even more pernicious with multiple indicator latent curve models), but not by number of occasions or by sample size. For change, concordance rates increase with more occasions of measurement and wider intervals, but not considerably with greater sample sizes. The exponential model clearly outperforms both the linear and the quadratic model, and is the only model obtaining high concordance rates.

In sum then, under the empirically based simulation conditions studied here, if a change process is exponential in nature but is analyzed with a quadratic model, we may not pick up that individuals differ in amount of change. If we do detect variance in change, then we are very likely to obtain biased estimates of individuals' change scores. Hence, not only is the detection of change compromised; even when change is detected, it is not well characterized by derived individual indices of change.

Finally, in terms of the empirical illustration on the Betula Project data, we found that the exponential model (a) fit the data best, (b), implied the expected individual cognitive scores that were closest to the observed scores, and (c) predicted mortality better than the linear and quadratic models. We deduce that in this sample the exponential GM likely provides the representation that best captures cognitive aging processes relative to its two competitors. We further hypothesize that the advantage of the exponential over both the linear and the quadratic models would increase with greater intervals of testing ($\Delta > 5$ years), more reliable variables (GCR >.681), more occasions ($T > 4$), and a greater sample size ($N > 1000$).

Recommendations for Data Analysts

In light of these results, what should one do when analyzing real long-term change data? First, rather than continuing the common practice of solely testing the linear and the quadratic GMs, one might consider alternative, possibly nonlinear functions, which allow for substantive interpretations of their parameters. The exponential model is one such function that has proven useful in many disciplines, but other functions deserve to be considered as well (e.g., Grimm et al., 2007). Omitting nonlinear functions may cloud our understanding of the phenomena under investigation.

Second, when planning a longitudinal study, one should try to use reliable instruments (e.g., Little, Lindenberger, & Nesselrode, 1999) and assess the instruments repeatedly at carefully planned intervals (Hertzog & Nesselrode, 2003; Willett, 1989). What our results show is that even under ideal conditions, such as those of our simulation experiment (e.g., no unwanted retest effects, no longitudinal dropout, and group homogeneity), the discovery of change-related information often requires many widely spaced assessments. One practical implication is the need for patience in longitudinal sampling. A lack of change-related evidence at the second or third wave of a longitudinal study of long-term change does not preclude the sought-after effect manifesting itself after additional data collection (especially when waves are separated by wide time intervals).

Third, we found that overall the effect of sample size was rather minor, compared to the other effects of our simulation experiment. Although psychologists have become extremely conscious of the overall importance of sample size, the design features related to density and span of change assessments appear to be more important in enabling accurate characterization of individual differences

in change (Brandmaier, von Oertzen, Ghisletta, Lindenberger, & Hertzog, 2018).

In conclusion, much progress has been made in estimating nonlinear growth models and in showing their relevance in lifespan developmental psychology. We argue that the time is right for ending exclusive reliance on linear GMs for describing change. Embracing conceptual, statistical, and computational advances embedded in nonlinear GMs may deepen our understanding of psychological change phenomena.

Acknowledgment

We thank Mikael Stiernstedt and Lars Nyberg for providing us with the Betula data used in the empirical illustration.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work benefitted from support of the Lifebrain grant, financed by the European Commission in the Horizon2020 Programme, Grant agreement no: SCI-PM-04-2016.

Supplemental material

Supplemental material for this article is available online.

References

- Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1988). *Life-span developmental psychology: An introduction to research methods (Vol. reprint of 1977)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York, NY: John Wiley and Sons.
- Brandmaier, A. M., von Oertzen, T., Ghisletta, P., Lindenberger, U., & Hertzog, C. (2018). Precision, reliability, and effect size of slope variance in latent growth curve models: Implications for statistical power analysis. *Frontiers in Psychology, 9*, 294. doi:https://doi.org/10.3389/fpsyg.2018.00294.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147–158.
- Cudeck, R., & Harring, J. R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology, 58*, 615–637. doi:https://doi.org/10.1146/annurev.psych.58.110405.085520.
- Davidian, W., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. London, UK: Chapman & Hall.
- Driscoll, I., Resnick, S. M., Troncoso, J. C., An, Y., O'Brien, R., & Zonderman, A. B. (2006). Impact of Alzheimer's pathology on cognitive trajectories in nondemented elderly. *Annals of Neurology, 60*, 688–695. doi:https://doi.org/10.1002/ana.21031.
- Gerstorf, D., & Ram, N. (2013). Inquiry into terminal decline: five objectives for future study. *The Gerontologist, 53*, 727–737. doi:https://doi.org/10.1093/geront/gnt046.
- Grimm, K. J., McArdle, J. J., & Hamagami, F. (2007). Nonlinear growth mixture models in research on cognitive aging. In K. v. Monfort, H. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioural and related sciences* (pp. 267–294). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging, 18*, 639–657. doi:https://doi.org/10.1037/0882-7974.18.4.639.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795. doi:https://doi.org/10.1080/01621459.1995.10476572.
- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*, 963–974.
- Littell, R., Milliken, F., Stroup, W., & Wolfinger, R. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute.
- Little, T. D., & Card, N. A. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.
- Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When “good” indicators are bad and “bad” indicators are good. *Psychological Methods, 4*, 192–211. doi:https://doi.org/10.1037/1082-989X.4.2.192.
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). New York, NY: Plenum Press.
- McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology, 38*, 115–142.
- Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5*, 23–43.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107–122.
- Nilsson, L.-G., Adolfsson, R., Bäckman, L., Frias, C. M. de, Molander, B., & Nyberg, L. (2004). Betula: A prospective cohort study on memory, health and aging. *Aging, Neuropsychology, and Cognition, 11*, 134–148. doi:https://doi.org/10.1080/13825580490511026.
- Rovine, M. J., & Molenaar, P. C. M. (1998). The covariance between level and shape in the latent growth curve model with estimated basis vector coefficients. *Methods of Psychological Research Online, 3*, 95–107.
- Schaie, K. W., & Hofer, S. M. (2001). Longitudinal studies in aging research. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 53–77). San Diego, CA: Academic Press.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford, UK: Oxford University Press.
- Wainer, H. (2000). The centercept: An estimable and meaningful regression parameter. *Psychological Science, 11*, 434–436. doi:https://doi.org/10.1111/1467-9280.00284.
- Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8*, 16–37.
- Willett, J. B. (1989). Some results on reliability for the measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement, 49*, 587–602.