



# Multi-Level Multi-Growth Models: New opportunities for addressing developmental theory using advanced longitudinal designs with planned missingness

Ethan M. McCormick<sup>a,b,\*</sup>

<sup>a</sup> Department of Psychology and Neuroscience, University of North Carolina, Chapel Hill, NC, 27599, United States

<sup>b</sup> Cognitive Neuroscience Department, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, the Netherlands

## ARTICLE INFO

### Keywords:

Multi-level models  
Longitudinal methods  
Development and learning  
Quantitative methods  
Experience  
Puberty

## ABSTRACT

Longitudinal models have become increasingly popular in recent years because of their power to test theoretically derived hypotheses by modeling within-person processes with repeated measures. Growth models constitute a flexible framework for modeling a range of complex trajectories across time in outcomes of interest, including non-linearities and time-varying covariates. However, these models can be expanded to include the effects of multiple growth processes at once on a single outcome. Here, I outline such an extension, showing how multiple growth processes can be modeled as a specific case of the general ability to include time-varying covariates in growth models. I show that this extension of growth models cannot be accomplished by statistical models alone, and that study design plays a crucial role in allowing for proper parameter recovery. I demonstrate these principles through simulations to mimic important theoretical conditions where modeling the effects of multiple growth processes can address developmental theory including, disaggregating the effects of age and practice or treatment in repeated assessments and modeling age- and puberty-related effects during adolescence. I compare how these models behave in two common longitudinal designs, cohort and accelerated, and how planned missingness in observations is key to parameter recovery. I conclude with directions for future substantive research using the method outlined here.

## 1. Introduction

The aim of the developmental sciences is to understand the course, cause, and consequence of change across time (Curran et al., 2010). The earliest attempts to understand developmental processes utilized cross-sectional designs (Fig. 1A), and this approach remains prevalent. Cross-sectional designs can either involve comparing age groups or sampling ages from a distribution continuously. Known limitations of these designs, including the inability to test causal relationships or to understand individual differences in trajectories of change (Kraemer et al., 2000; Louis et al., 1986; Maxwell and Cole, 2007) stem from the between-person nature of estimated effects. Given these limitations, the developmental sciences have invested heavily in longitudinal (i.e., repeated-measures) designs (Card and Little, 2007) where the same individuals are observed across multiple occasions. Longitudinal designs have increased power to detect developmental effects, allow individuals to be compared to their own baseline, and explicitly model within- and

between-person variability (Kraemer et al., 2000; Louis et al., 1986). A common longitudinal design is the cohort design (Fig. 1B), where individuals are measured at the same initial age and followed roughly at equal spacing across time. However, despite the many advantages of these studies, one important limitation of this approach (aside from the practical challenges of cost and attrition), is a limited ability to model correlated developmental predictors (e.g., age, puberty, learning). Overcoming this limitation is a key challenge for testing complex developmental theory.

The fundamental limitation of cohort designs is the near or total confounding of age and number of observations (Cook and Ware, 1983; Telzer et al., 2018). Despite being known for over half a century (Bell, 1953; Palmore, 1978), relatively little attention has been paid to these issues in substantive work since. Failure to quantify the effect of repeated measures (e.g., practice, habituation, etc.) presents a threat to the internal validity of developmental inferences made from cohort longitudinal data. The effects of repeated exposure (e.g., practice

\* Correspondence to: 235 E. Cameron Avenue, Chapel Hill, NC, 27514, United States.

E-mail address: [emccormick@unc.edu](mailto:emccormick@unc.edu).

<https://doi.org/10.1016/j.dcn.2021.101001>

Received 8 January 2021; Received in revised form 9 May 2021; Accepted 4 August 2021

Available online 8 August 2021

1878-9293/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

effects) to observation could artificially enhance (or entirely account for) developmental effects estimated in longitudinal models (Ferrer et al., 2004; McArdle et al., 1997; McCormick et al., 2021; Rabbitt et al., 2001), or conversely interfere and mask the effects of maturation (e.g., habituation). Fortunately, an alternative, the accelerated longitudinal design (Fig. 1C), offers a potential solution. Accelerated longitudinal designs combined features of cross-sectional and cohort studies, measuring individuals on several occasions but not at the same initial ages (Bell, 1953). While motivations for adopting accelerated designs have mostly focused on the cost-efficiency and greater age coverage in shorter study durations (Galbraith et al., 2017), these designs also offer a solution to the age-experience confounding seen in cohort studies by disentangling the effects of repeated observations (i.e., experience) from maturational changes (i.e., development) (Cook and Ware, 1983; Van't Hof et al., 1977).

The availability of longitudinal data has prompted the development of many models for repeated measures data, including a variety of autoregressive (Kessler and Greenberg, 1981) latent curve models, and a class of linear regression models known as multi-level (or mixed-effect) models (MLMs; Bryk and Raudenbush, 1987; Raudenbush and Bryk, 2002). Under conditions common to many cohort designs, LCMs and MLMs are mathematically isomorphic (Bauer, 2003; Curran, 2003; MacCallum et al., 2010), however, MLMs are particularly well-suited to the conditions of accelerated longitudinal studies because of their ability to handle the planned-missingness (i.e., individuals are not observed at all possible ages) inherent in the design (Galbraith et al., 2017; Little and Rhemtulla, 2013; Rhemtulla and Little, 2012). Alternatively, an SEM growth modeling approach can accommodate continuous time in the SEM through definition variables (Mehta and Neale, 2005; Sterba, 2014) but comes with some drawbacks (e.g., lack of a chi-square value and associated fit statistics, not currently available in all software implementations). While MLMs are applicable to a range of questions, multilevel growth models specifically take the form where an observed outcome is modeled as a function of time (Curran and Bauer, 2011) in a level-one (i.e., person specific) equation

$$y_{it} = \beta_{0i} + \beta_{1i}Time_{it} + r_{it} \tag{1}$$

where  $y$  is the outcome at time  $t$  for person  $i$ ,  $\beta_0$  and  $\beta_1$  are the person-specific intercept and slope (i.e., effect of time) respectively, and  $r_{it}$  is the time- and person-specific residual. Individuals are allowed to vary by including level-two parameters which model individual variability in intercept and rate of change

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i} \end{aligned} \tag{2}$$

where  $\gamma_{00}$  and  $\gamma_{10}$  represent the group-level (i.e., fixed effects) intercept and slope, and  $u_{0i}$  and  $u_{1i}$  represent individual deviations in those terms (i.e., random effects). By substitution, we can see that the full equation describes a single growth process in the outcome (across  $Time_{it}$ ) that allows for estimating individual variability in the initial level and rate of change of  $y_{it}$ .

$$y_{it} = \underbrace{\gamma_{00} + \gamma_{10}Time_{it}}_{\text{Fixed Effects}} + \underbrace{u_{0i} + u_{1i}Time_{it} + r_{it}}_{\text{Random Effects}} \tag{3}$$

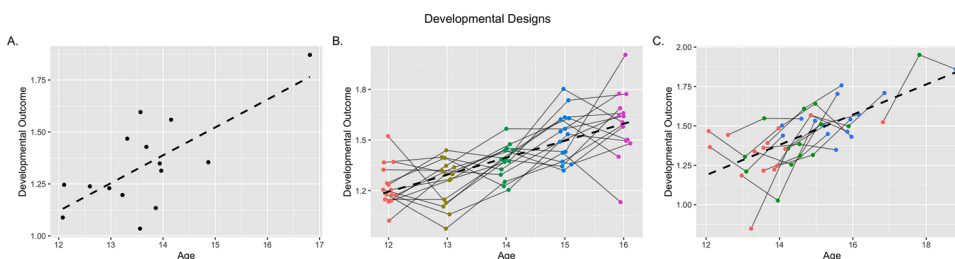
This model could be expanded to include more complex effects of time (e.g., polynomials or other nonlinear functions; (Cudeck and Haring, 2007; Cudeck and Klebe, 2002)), or additional covariates, but this general form constitutes the core of multilevel growth models.

While these models have primarily been used to characterize single growth process (i.e., to model only the effect of time), there are no inherent restrictions on including multiple growth processes in the same model (Ferrer et al., 2004; McArdle et al., 1997). Note that this is distinct from modeling the same growth predictor for multiple outcomes in a multivariate model (MacCallum et al., 1997). For instance, modeling two growth processes on a single outcome would involve a simple expansion of Eq. (3) to include additional predictors, taking the form:

$$y_{it} = \underbrace{\gamma_{00} + \gamma_{10}Growth\ 1_{it} + \gamma_{20}Growth\ 2_{it}}_{\text{Fixed Effects}} + \underbrace{u_{0i} + u_{1i}Growth\ 1_{it} + u_{2i}Growth\ 2_{it} + r_{it}}_{\text{Random Effects}} \tag{4}$$

where the growth predictors represent separable constructs of change (e.g., experience and maturation) that unfold across levels of  $t$  and  $s$  respectively. This model is not fundamentally different from including any other time-varying covariate (see (Curran and Bauer, 2011; Curran et al., 2012; McNeish and Matta, 2020) for more in-depth treatment of the general TVC model), but there are limitations for obtaining proper effect estimates under common conditions. That is, if the two growth processes are highly collinear, there will be instability in effect estimates for either process and standard errors will be inflated (Shieh and Fou-ladi, 1991), increasing Type II error rates. However, accelerated longitudinal designs offer a potential solution to this limitation since it is possible to attenuate the correlation between different growth processes through the use of planned missingness (Graham, 2009). Combining appropriate sampling designs with the extension to a multilevel, multi-growth modeling framework offers a number of opportunities that can be flexibly employed depending on the specific research question.

In this study, I will highlight the promise of multilevel, multi-growth models (MLMGs) for addressing developmental theory. To do so, I will develop a series of scenarios where multiple growth processes might be highly entangled. For each scenario, I will highlight how cohort designs fail to disaggregate between different influences (e.g., age and puberty) on developmental outcomes of interest, and how MLMGMs in combination with accelerated designs address these failures. I show that MLMGMs can be flexibly employed for a range of purposes, from controlling for repeated assessments while estimating developmental effects to probing interactions between chronological age and pubertal maturation. Finally, I outline how sampling strategies can be optimized to take advantage the promise of MLMGMs. While the primary target for this work is the developmental sciences, there are undoubtedly applications for these models in other fields where multiple growth processes influence outcomes of interest.



**Fig. 1.** Examples of A) Cross-Sectional, B) Cohort, and C) Accelerated Longitudinal developmental designs. While cross-sectional designs rely on one measurement per subject and rely on between-person change to estimate developmental effects (dashed line), cohort longitudinal designs rely on within-person changes across repeated measurements. Accelerated longitudinal designs combine these approaches by using both between- and within-person change to estimate developmental effects.

## 2. Methods

### 2.1. Simulation design

For each scenario, I simulated data consistent with growth along one or more separable processes (e.g., using variants of Eq. (4)). For a given scenario, I simulated data to have the structure of both cohort and accelerated longitudinal designs. I then fit both a properly-specified and mis-specified model (described in each scenario) and demonstrated both how parameters are distributed across different designs and how theoretical inferences would be impacted. To test these questions, I simulated 1000 replicants of each design for each scenario. Data were simulated in R, and all relevant code for replicating data and analyses are available online (<https://osf.io/65sdw/>). I plotted results for each scenario in terms of ages that are reasonable for each hypothetical for ease of interpretation and discussion, but they are neither intrinsically meaningful nor meant to directly reflect particular results from real data.

#### 2.1.1. Sample size and number of observations

The limitations of common developmental models highlighted in this investigation are not ones that reasonable increases in sample size or number of observations can easily solve. As such, I simulated 750 total observations for each simulated dataset to avoid issues of power and highlight the challenges inherent in confounding age and observations. To mimic real data features, I set 5 observations per person for the cohort design ( $n = 150$ ), and 3 observations per person for the accelerated data ( $n = 250$ ), with approximately one year between observations (jittered to prevent perfect collinearity  $N \sim [0, 0.1]$ ). Both of these sample sizes represent reasonably-powered studies for the respective developmental design (Fang et al., 2008).

#### 2.1.2. Missing data

Similar to issues of sample size, I simulated most scenarios to have complete data to focus the scope of the investigation towards issues of theory and inference. This is because planned missing data is accommodated in multilevel models generally and indeed it is what enables their use in accelerated longitudinal data in the first place (Card and Little, 2007). Furthermore, to the extent that missing data influences correlations among growth processes, the most likely contributor to unplanned missing data in longitudinal designs (i.e., attrition) actually attenuates correlations between number of observation and other growth processes. However, one feature of interest to the current investigation is the additional instability that missingness can introduce to parameter estimates in longitudinal data, especially with predictors in the model are highly correlated. To highlight this point, I simulated a separate iteration of one condition in Scenario 1 to have 10 % missing data at random ( $\sim 75$  of 750 total observations). The results did not substantively change with the inclusion of missing data, and results are reported in supplemental material (Table S1).

#### 2.1.3. Fixed effects parameters

Unless otherwise indicated, I simulated regression parameters for the fixed effects to show moderately strong effects ( $|\gamma_x| = 0.3$ ). This type of effect should be easily detected at the sample sizes shown and clearly highlights the effects of interest in plots. Parameters associated with interactions and polynomials were simulated with weaker absolute effects, and specifics can be found in the description for each scenario below.

#### 2.1.4. Random effects parameters

Finally, I generated random effects to allow for individual variability in initial level (i.e., intercept,  $u_{0i}$ ) and residual variability at level 1 (i.e.,  $r_{it}$ ). For simplicity, I excluded random effects of slope (i.e.,  $u_{0i}$ ) from the data generating model; however, MLMGMs, like multilevel, single-growth modes, can be easily expanded to include these effects. Unless

otherwise specified, I generated random intercepts ( $N \sim [0, 0.5]$ ), and residuals ( $N \sim [0, 1]$ ) from normal distributions. Note that the random effects are expressed as deviations from the fixed effects rather than absolute locations on the scale of the outcome. In later scenarios, I also simulated random slope effects and those effects are described there.

### 2.2. Parameter recovery

Observed variables from the data generation step were fit with a set of linear mixed effects models with the lme4 package in R (version 1.1–21; Bates et al., 2015) and significance levels were obtained using the lmerTest package (version 3.1–1) (Kuznetsova et al., 2017). I first fit a single growth process MLM (i.e., mis-specified model) as would conventionally be done for the scenario in question (e.g., including only age effects), and then fit an MLMG model with all relevant growth processes included. I extracted both point estimates for the relevant effects, as well as standard errors and significance decisions (i.e., adjusted and unadjusted  $p < \text{or} > 0.05$ ). Significance adjustment was done within each model using a Bonferroni correction, and accounted for the number of predictors in a given model (i.e., 1–3 predictors depending on the scenario). Based on the point estimate and standard error for each parameter, I also calculated a standardized measure of bias:

$$Bias_{std} = \frac{1}{SE(\hat{\theta}_j)} * \left( \frac{\sum_{j=1}^N \hat{\theta}_j}{N} - \theta \right) \quad (3)$$

Bias was calculated as the difference between the mean parameter estimate ( $\hat{\theta}_j$ ) across replications ( $j$ ) and the true generating parameter ( $\theta$ ), scaled by the standard error (Collins et al., 2001; Gottfredson et al., 2014). Values represented the distance between the estimated and true parameter as the percentage of the standard deviation of the sampling distribution. For example, a value of .500 would represent a distance of half a standard deviation away from the true parameter. This metric has advantages over significance testing, since differences are very likely to be statistically significant given the large number of iterations, or simple differences that do not take into account the spread of parameter estimates. Previous work has suggested that absolute values above .400 begin to introduce challenges for model performance (Collins et al., 2001). To take a conservative approach, I considered absolute bias values greater than .250 to reflect poor performance. However, no substantive conclusions were altered by the choice of a .250 versus .400 threshold.

## 3. Results

### 3.1. Scenario set 1: age and practice retest effects

Perhaps the most common candidate for a confounding growth process in developmental research is the effect of repeated across multiple occasions on task measures. At each observation, individuals in the sample are not only older (roughly indicating maturation), but they have been further exposed to the specific task or instrument being used to assess outcomes. Retest effects can take many forms (e.g., practice, habituation, sensitization) that can confound developmental effects (Salthouse, 2014). Here, I use the example of a difficult cognitive control task. Previous research has shown performance gains across adolescence (Luna et al., 2010; Somerville and Casey, 2010), but these total effects may confound potentially separable impacts of developmental maturation and repeated exposure to the task.

#### 3.1.1. Practice effects only

Using this framework, I first simulated longitudinal data with no effect of development, but a positive effect of practice ( $\gamma_{prac} = 0.3$ ). In

other words, individuals improve their ability to inhibit automatic responses solely through practicing the task, and this practice effect is consistent across age. An exemplar replicant of cohort and accelerated longitudinal data are plotted below (Fig. 2A & B respectively). As expected, the correlation between age and practice was consistently near perfect ( $M_p = .998$ ,  $SE < .001$ ) for the cohort design, leading to high levels of variance inflation ( $M_{VIF} = 201.58$ ,  $SE = 10.22$ ) and standard errors ~14-times larger than expected. In contrast, these effects were attenuated for the accelerated data with moderate correlations ( $M_p = .379$ ,  $SE = .015$ ) and little inflation of standard errors ( $M_{VIF} = 1.17$ ,  $SE = .016$ ; standard error inflation = 1.08 times).

I then fit a traditional, single-growth process model (using age as a predictor) to each dataset and extracted parameter estimates and compared them with their population value (Fig. 2C & D). For the cohort data, the practice effect completely aliased as an age effect in the model, whereas there was much less bias in the accelerated data (2.98 vs 11.46; see Table 1 for full parameter details) in the estimate of the age effect due to the attenuated correlation between the two growth predictors. When fitting the correctly-specified multi-growth model, the challenges of disentangling these effects in traditional longitudinal data became clear. While the effects are unbiased (i.e., the distribution center on their respective population values; all bias < .050) even in the presence of high multi-collinearity (Shieh and Fouladi, 1991), the estimates across samples were highly unstable for the cohort data and the distributions of effects for age and practice overlap considerably (Fig. 2E). This instability did not lead to inflated false positive rates for the effect of age (age effects were only significant for 4.8 % of replicants; 2 % when correcting for multiple predictors in the model), but there was a significant elevation of false negatives for the effects of practice, only being significant for 12.6 % of replicants without correction (8.2 % with correction). This was due to both the increased instability in effect point estimates, as well as inflated standard errors for hypothesis testing.

Neither of these issues were present in the accelerated longitudinal data, as there was virtually no bias in either effect estimates and false positive (3.8 % for effects of age without correction; 1.5 % with correction) and false negative (0% for effects of practice) rates were appropriately low. All models correctly captured the variance of the residual ( $r_{it}$ ) and person-level ( $\tau_{00}$ ) random effects, with some downward bias (Shieh and Fouladi, 1991). When I simulated only an effect of age instead of practice, age-only models were unbiased, but results were otherwise comparable to those here (see Supplemental Material; Fig. S1).

### 3.1.2. Additive effects of age and practice

Combining the previous two simulations, I then simulated data to contain effects of both age and practice ( $\gamma$ 's = 0.3). This is a more plausible scenario than the two previous conditions as both maturity and exposure likely contribute to increases in task performance. However, based on results thus far, we would expect that effects in cohort models with only an age predictor would be inflated relative to the true effect due to the high correlation ( $M_p = .998$ ,  $SE < .001$ ) and unstable due to the variance inflation ( $M_{VIF} = 202.18$ ,  $SE = 10.67$ ) if both predictors were included. Indeed, in this additive-effects model, a single growth model with the effect of age substantially inflates the predicted effect ( $Bias_{std} = 11.32$ ,  $SE = .026$ ) in the cohort data, while inflating the effect much less in the accelerated design ( $Bias_{std} = .291$ ,  $SE = .022$ ; Fig. 3C & D). The MLMG models show unbiased effect estimates for both age and practice for both design types, but false negatives are controlled appropriately only in the accelerated data (Fig. 3; Table 2). In a real data application (McCormick et al., 2021), we observed that an unmodeled effect of repeated practice artificially inflated the effects of age on learning performance (compare estimates in Tables 1 and 3) exactly as these results would suggest.

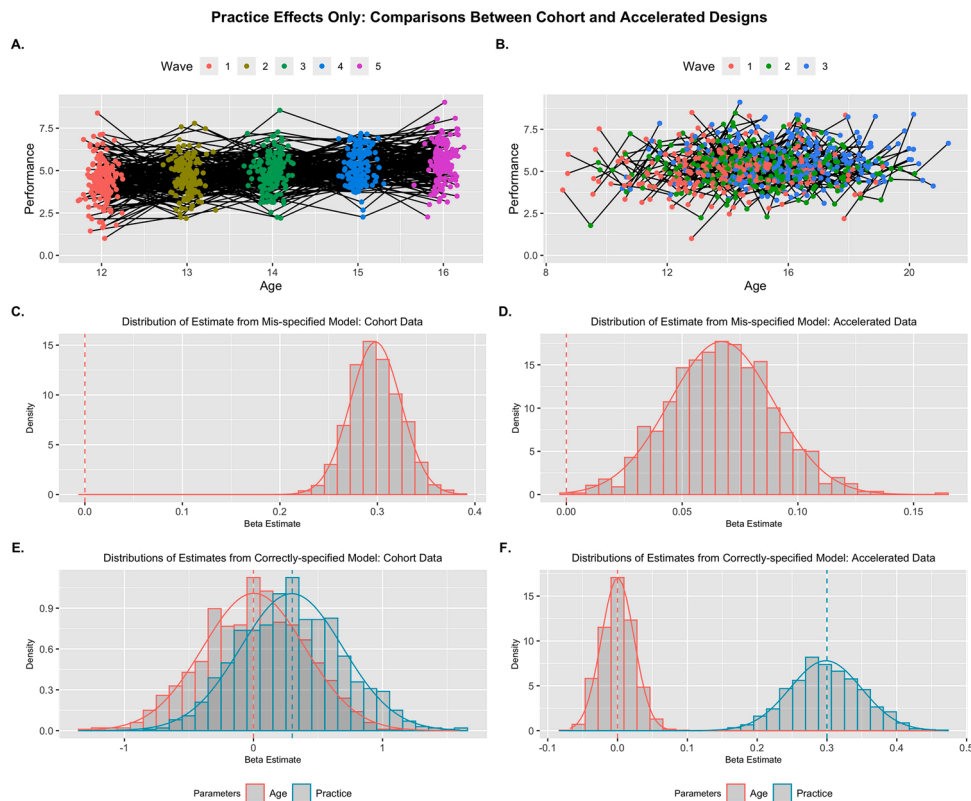


Fig. 2. Practice Effects Only: Example replicants for Cohort (A) and Accelerated (B) designs. Distribution of parameter estimates across replicants are presented for mis-specified models where age is the sole predictor (C & D), and properly-specified models where age and practice are included (E & F). Only a MLMG model with accelerated data showed unbiased and precise beta estimates. Generating parameters for each predictor are denoted by vertical dashed lines.



**Table 1**  
Estimated Effects for Growth Models from Scenario Set 1: Single Effects.

	Cohort Design						Accelerated Design					
	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig. (adj.)	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig. (adj.)
<i>Practice Effects Only</i>												
$\rho_{growth}$	.998	.000	.997	.998			.379	.015	.335	.426		
VIF	201.6	10.2	171.7	233.8			1.17	.015	1.13	1.22		
$\gamma_{age}$ (mis)	.298	.026	.223	.385	<b>11.46</b>	1	.067	.023	.002	.162	<b>2.98</b>	.865
$\sigma$ (mis)	1.00	.029	.903	1.09	.011		1.03	.033	.920	1.15	<b>.784</b>	
$\sqrt{\tau_{00}}$ (mis)	.496	.054	.303	.669	-.082		.497	.062	.245	.660	-.042	
$\gamma_{age}$	.003	.395	-1.31	1.12	.008	.048 (.020)	.000	.023	-.068	.089	.014	.038 (.015)
$\gamma_{prac}$	.296	.396	-.777	1.60	-.009	.126 (.082)	.298	.051	.147	.470	-.034	1 (1)
$\sigma$	1.00	.029	.903	1.08	-.005		.999	.032	.888	1.12	-.021	
$\sqrt{\tau_{00}}$	.496	.054	.302	.663	-.082		.498	.057	.290	.648	-.035	
<i>Age Effects Only</i>												
$\rho_{growth}$	.998	.000	.997	.998			.378	.014	.335	.417		
VIF	202.1	10.7	171.7	243.4			1.17	.015	1.13	1.21		
$\gamma_{age}$ (mis)	.299	.025	.218	.373	-.024	1	.301	.021	.233	.361	.040	1
$\sigma$ (mis)	1.00	.029	.893	1.10	.021		1.00	.031	.924	1.09	.008	
$\sqrt{\tau_{00}}$ (mis)	.498	.053	.314	.646	-.031		.497	.058	.257	.658	-.057	
$\gamma_{age}$	.298	.387	-.848	1.64	-.005	.114 (.071)	.301	.025	.223	.379	.041	1 (1)
$\gamma_{prac}$	.001	.388	-1.41	1.15	.004	.045 (.023)	-.001	.051	-.150	.166	-.016	.044 (.020)
$\sigma$	1.00	.029	.893	1.10	.022		1.00	.031	.924	1.09	.008	
$\sqrt{\tau_{00}}$	.498	.054	.314	.646	-.030		.497	.058	.257	.658	-.057	

Note: Mean Est. is the expected parameter value across replicants and the Std. Err is the standard deviation of the sampling distribution. The proportion significant (Prop. Sig) represents the number of inferential tests that return a significant result for each parameter (unadjusted first, adjusted in parentheses). Std. Bias is the standardized bias estimate; values > .25 are bolded and represent unacceptable bias. The proportion significant only includes significant results where the inference is correct (e.g., significant negative effects are not included of the population parameter is positive). Rhos ( $\rho_{growth}$ ) represent the correlation between growth predictors, VIF represents variance inflation factor, gammas ( $\gamma$ ) represent the fixed effects, sigma ( $\sigma$ ) is the square-root of the residual random variance, and tau ( $\tau_{00}$ ) is the person-level random variance (square-root shown for consistency). The (mis) denotes parameters from the mis-specified (i.e., age-only) model.



**Fig. 3.** Additive Practice and Age Effects: Example replicants for Cohort (A) and Accelerated (B) designs. Distribution of parameter estimates across replicants are presented for mis-specified models where age is the sole predictor (C & D), and properly-specified models where age and practice are included (E & F). Only a MLMG model with accelerated data showed unbiased and precise beta estimates. Generating parameters for each predictor are denoted by vertical dashed lines.

**Table 2**  
Estimated Effects for Growth Models from Scenario Set 1: Additive Effects.

	Cohort Design						Accelerated Design					
	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig.(adj.)	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig.(adj.)
<b>Additive Effects</b>												
$\rho_{growth}$	.998	.000	.997	.998			.379	.015	.338	.434		
VIF	202.2	10.7	169.2	238.8			1.17	.016	1.13	1.23		
$\gamma_{age}$ (mis)	.598	.026	.502	.686	<b>11.32</b>	1	.365	.022	.293	.450	<b>2.91</b>	1
$\sigma$ (mis)	1.00	.028	.921	1.09	-.016		1.03	.033	.922	1.14	<b>-.811</b>	
$\sqrt{\tau_{00}}$ (mis)	.497	.051	.305	.660	-.049		.492	.052	.215	.657	-.124	
$\gamma_{age}$	.316	.364	-.968	1.42	.045	.123 (.071)	.299	.024	.222	.383	-.049	1 (1)
$\gamma_{prac}$	.283	.365	0.825	1.56	-.045	.091 (.053)	.300	.051	.147	.478	.002	1 (1)
$\sigma$	.999	.028	.921	1.09	-.026		1.00	.032	.904	1.09	-.003	
$\sqrt{\tau_{00}}$	.497	.051	.307	.660	-.051		.493	.058	.244	.651	-.116	
<b>Quadratic Effects of Age</b>												
$\rho_{growth}$	.998	.000	.997	.998			.379	.015	.334	.427		
VIF	202.7	10.2	171.62	244.29			1.17	.016	1.13	1.22		
$\gamma_{age}$ (mis)	.599	.026	.506	.684	<b>11.67</b>	1 (1)	.367	.026	.277	.453	<b>2.57</b>	1 (1)
$\gamma_{age^2}$ (mis)	-.101	.022	-.160	-.030	-.047	.999 (.996)	-.100	.007	-.125	-.080	-.051	1 (1)
<i>vertex</i> (mis)	3.12	.790	1.73	9.67	<b>2.06</b>		1.83	.129	1.49	2.45	<b>2.58</b>	
$\sigma$ (mis)	1.00	.029	.919	1.10	.039		1.03	.033	.912	1.13	<b>.783</b>	
$\sqrt{\tau_{00}}$ (mis)	.499	.053	.325	.673	-.018		.497	.062	.270	.667	-.052	
$\gamma_{age}$	.312	.382	-.887	1.50	.031	.112 (.063)	.300	.027	.217	.398	-.015	1 (1)
$\gamma_{age^2}$	-.101	.022	-.160	-.031	-.048	.999 (.992)	-.100	.007	-.129	-.080	-.042	1 (1)
$\gamma_{prac}$	.289	.382	-.909	1.49	-.029	.109 (.050)	.299	.053	.112	.475	-.013	1 (.999)
<i>vertex</i>	1.63	2.10	-5.57	16.40	.060		1.50	.128	1.08	1.95	-.029	
$\sigma$	1.00	.029	.917	1.10	.026		.999	.031	.885	1.10	-.028	
$\sqrt{\tau_{00}}$	.499	.053	.322	.674	.029		.498	.057	.296	.650	-.043	

Note: Mean Est. is the expected parameter value across replicants and the Std. Err is the standard deviation of the sampling distribution. The proportion significant (Prop. Sig) represents the number of inferential tests that return a significant result for each parameter (unadjusted first, adjusted in parentheses). Std. Bias is the standardized bias estimate; values  $> .25$  are bolded and represent unacceptable bias. The proportion significant only includes significant results where the inference is correct (e.g., significant negative effects are not included of the population parameter is positive). Rhos ( $\rho_{growth}$ ) represent the correlation between growth predictors, VIF represents variance inflation factor, gammas ( $\gamma$ ) represent the fixed effects, "vertex" represents the implied inflection point in the trajectory, sigma ( $\sigma$ ) is the square-root of the residual random variance, and tau ( $\tau_{00}$ ) is the person-level random variance (square-root shown for consistency). The (mis) denotes parameters from the mis-specified (i.e., age-only) model.

### 3.1.3. Shifting quadratic effects of age

For more-complex trends across time (e.g., quadratic) the impacts of correlated growth predictors can seep into additional model-based inferences. Of particular concern is the increased instability in model-implied inflection points when correlated growth processes are included in the same model, making inferences about developmental transitions highly unreliable (for complete details, see Table 2; Fig. S3).

### 3.2. Scenario set 2: interference retest effects

Of course, the effects of repeated assessments can go beyond practice, which can generally be assumed to boost task performance. Other confounders may operate against the axis of the primary growth process of interest. In the next simulations, I considered cases when multiple growth processes interfere with one another, using examples common in developmental research.

#### 3.2.1. Habituation effects

A prime example of this kind of interferences is the process of habituation in brain reactivity to a set of stimuli. This could take the form of a conflict (e.g., dorsal anterior cingulate) or threat (e.g., amygdala) response that decreases once stimuli are less unexpected (Breiter et al., 1996; Denny et al., 2014; Ellwanger et al., 2003). In contrast, we might expect general developmental increases in negative emotionality (e.g., depression) across some periods of development (e.g., adolescence; Cyranowski et al., 2000; Garber et al., 2002), which habituation effects could work to attenuate. To demonstrate this kind of masking by the confounder, I simulated data to have a positive effect of age ( $\gamma_{age} = 0.3$ ) but a negative effect of exposure ( $\gamma_{exp} = -0.3$ ) indicating habituation of a hypothetical brain response across repeated

measures (Fig. 4A & B). As seen with the additive models, the age-only growth model in the cohort data showed substantial bias ( $Bias_{std} = -12.17$ ,  $SE = .025$ ), and results would infer that there is no change in reactivity across age. In the age-only model with the accelerated data, the correlation between age and exposure does attenuate the true effect of age somewhat, but much less ( $Bias_{std} = -3.02$ ,  $SE = .022$ ) and all replicants showed a significant positive effect of age despite this attenuation (see Table 3 for full details). With the correctly-specified MLMGM model, the effect estimates showed the same instability as previously shown, while in the accelerated design, the two were well-separated. Additionally, the MLMGM with accelerated data provided appropriate false negative control. For the inverse condition (i.e., a negative effect of age and positive effect of practice), see Table 3 for similar performance of the single-predictor and MLMG models (see Supplemental Material and Fig. S4 for more-complete information).

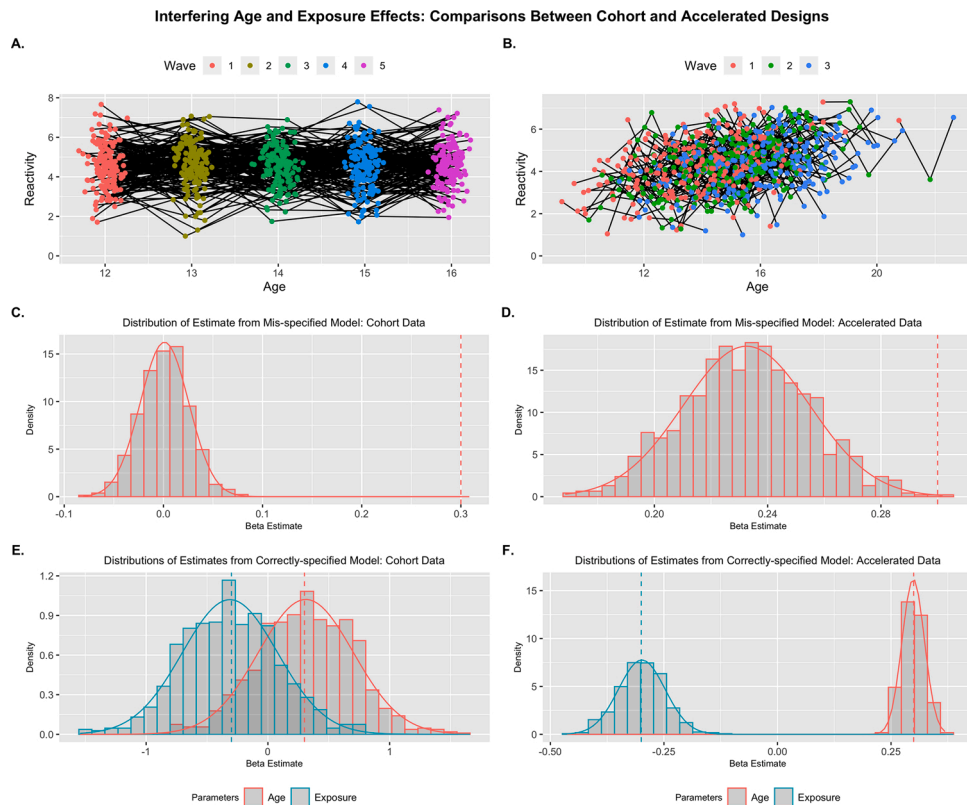
#### 3.3. Scenario set 3: when cohort data is appropriate

Thus far, I have highlighted the limitations of cohort for estimating multiple growth processes (e.g., age and practice) that are highly colinear, and how accelerated designs can be leveraged to address these limitations. At the risk of undermining this relatively straightforward point, I then considered scenarios where a cohort-type design would be capable of disentangling colinear growth processes of interest. However, these scenarios specifically build on the principles of an accelerated design (i.e., planned missingness), which enables proper estimation. Consider the previous scenarios. Here, experience (i.e., repeated assessments) is consistent across individuals such that each person is measured at every level of the predictor (unplanned missingness aside). In contrast, age is distributed across individuals such that each person is

**Table 3**  
Estimated Effects for Growth Models from Scenario Set 2: Interference Effects.

	Cohort Design						Accelerated Design					
	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig. (adj.)	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig. (adj.)
<b>Habituation Effects</b>												
$\rho_{growth}$	.998	.000	.997	.998			.379	.015	.333	.426		
VIF	201.6	10.6	169.7	234.9			1.17	.015	1.12	1.22		
$\gamma_{age}$ (mis)	.001	.025	-.081	.084	<b>-12.17</b>	.021	.233	.022	.1169	.302	<b>-3.02</b>	1
$\sigma$ (mis)	1.00	.028	.905	1.08	.049		1.03	.033	.913	1.12	<b>.795</b>	
$\sqrt{\tau_{00}}$ (mis)	.493	.053	.281	.660	-.124		.494	.062	.265	.665	-.101	
$\gamma_{age}$	.312	.291	-.809	1.56	.030	.109 (.065)	.299	.025	.233	.369	-.040	1 (1)
$\gamma_{prac}$	-.312	.391	-1.54	.791	-.030	.112 (.061)	-.299	.051	-.464	-.115	.024	1 (.999)
$\sigma$	1.00	.028	.905	1.08	.031		1.00	.032	.887	1.09	-.009	
$\sqrt{\tau_{00}}$	.493	.053	.279	.664	-.124		.495	.057	.285	.652	-.095	
<b>SNR Effects</b>												
$\rho_{growth}$	.998	.000	.997	.998			.379	.014	.338	.423		
VIF	202.4	10.3	169.4	237.6			1.17	.015	1.13	1.22		
$\gamma_{age}$ (mis)	-.001	.205	-.080	.090	<b>11.51</b>	.023	-.233	.023	-.296	-.149	<b>2.92</b>	1
$\sigma$ (mis)	.999	.029	.886	.108	-.028		1.03	.035	.929	1.13	<b>.734</b>	
$\sqrt{\tau_{00}}$ (mis)	.498	.055	.294	.709	-.031		.498	.061	.292	.657	-.039	
$\gamma_{age}$	-.300	.377	-1.53	.913	.000	.107 (.064)	-.300	.024	-.378	-.222	-.016	1 (1)
$\gamma_{prac}$	.300	.377	-.917	.155	.001	.104 (.062)	.301	.052	.150	.492	.022	1 (1)
$\sigma$	.999	.029	.887	1.08	-.041		.998	.033	.907	1.09	-.047	
$\sqrt{\tau_{00}}$	.498	.055	.294	.708	-.032		.498	.057	.317	.646	-.027	

Note: Mean Est. is the expected parameter value across replicants and the Std. Err is the standard deviation of the sampling distribution. The proportion significant (Prop. Sig) represents the number of inferential tests that return a significant result for each parameter (unadjusted first, adjusted in parentheses). Std. Bias is the standardized bias estimate; values > 0.25 are bolded and represent unacceptable bias. The proportion significant only includes significant results where the inference is correct (e.g., significant negative effects are not included of the population parameter is positive). Rhos ( $\rho_{growth}$ ) represent the correlation between growth predictors, VIF represents variance inflation factor, gammas ( $\gamma$ ) represent the fixed effects, sigma ( $\sigma$ ) is the square-root of the residual random variance, and tau ( $\tau_{00}$ ) is the person-level random variance (square-root shown for consistency). The (mis) denotes parameters from the mis-specified (i.e., age-only) model.



**Fig. 4.** Interfering Age and Exposure Effects: Example replicants for Cohort (A) and Accelerated (B) designs. Distribution of parameter estimates across replicants are presented for mis-specified models where age is the sole predictor (C & D), and properly-specified models where age and exposure are included (E & F). Only a MLMG model with accelerated data showed unbiased and precise beta estimates. Generating parameters for each predictor are denoted by vertical dashed lines.

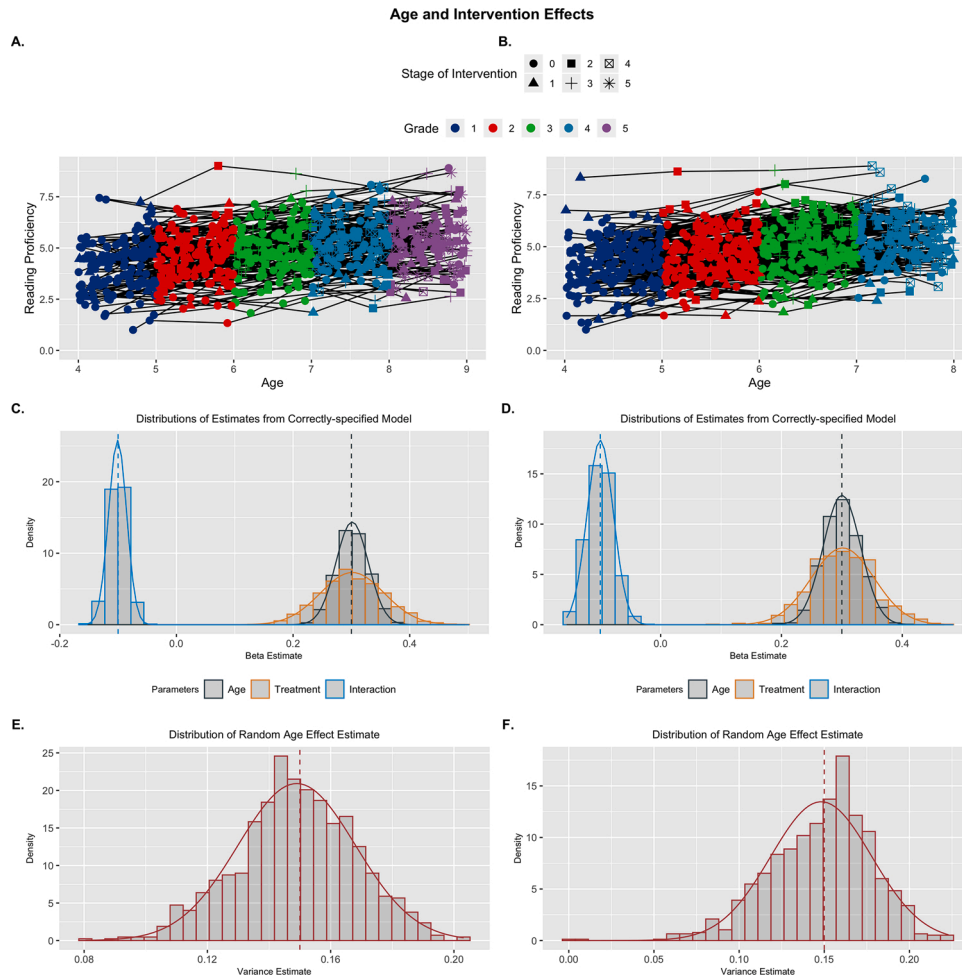
only measured on a subset of the levels of the predictor. While this requires an assumption that some degree of exchangeability exists across individuals with no age overlap (i.e., age convergence; [Sliwinski et al., 2010](#)), it allows for a decoupling of the two predictors in the sample. For a cohort design to achieve this decoupling, these features of sampling approach need to be reversed; that is, individuals are measured at all levels of the age predictor but not the experience predictor. In this way, these designs are accelerated, but with respect to the experience variable, not age.

### 3.3.1. Staggered school-based interventions

Here I simulated an application of this idea, where individuals are measured in a cohort fashion based on age, but some intervention is applied differentially across subjects ([Cook & Ware, 1983](#)). For example, perhaps we are interested in piloting how an intervention might boost reading proficiency across early education. Due to resource limitations, it might not be possible to apply treatment to all students across the study period, and therefore, we might start the intervention randomly across measurement occasions. Here I simulated a positive linear effect of both age and the intervention on reading proficiency ( $\gamma$ 's = 0.3), but a decelerating interaction effect ( $\gamma$ 's = -0.1), suggesting that intervening later in school is less effective. Previously I simulated cohort data to follow a lab-based assessment design where individuals are measured in tight clusters around discrete ages. However, in a school-based setting, we would expect individuals to be uniformly distributed in terms of age within grade, and I adopted this design for these

simulations. I simulated a 4- and 5-wave observation condition to probe the effects of study duration on the challenge of multicollinearity (more observations should induce higher correlations between the growth processes). While thus far I have only focused on a relatively simple random variance structure (only a level 1 residual and level 2 random intercept), MLMGMs can accommodate further random effects just as well as other MLMs. To highlight this fact, I also included a random effect of age ( $\sigma_{age} = .15$ ).

The MLMGM in this scenario recovered all parameters appropriately (i.e., unbiased and precise estimates), demonstrating the ability to have designs which are accelerated (i.e., planned missing) with respect to alternative growth processes ([Fig. 5](#)). There were minimal differences between the 4- and 5-wave simulations with respect to appropriate parameter estimation (see [Table 4](#) for full details). The correlations between age and intervention stage were slightly inflated for the 5-wave simulations ( $M_\rho = .554, SE = .025$ ) compared with the 4-wave design ( $M_\rho = .532, SE = .024$ ), and the false negative rate was larger but still well-controlled (5 wave: age = 0%, intervention = .2%, interaction = 1.7 %; 4 wave: 0% for all predictors) due to greater variance inflation (5 wave:  $M_{VIF} = 1.45, SE = .059$ ; 4 wave:  $M_{VIF} = 1.40, SE = .050$ ). The key novelty here is that the data are ostensibly structured as cohort data (which failed in previous examples), but because we have induced missingness in the other growth process (i.e., intervention), the MLMGM model is able to obtain unbiased and efficient estimates.



**Fig. 5.** Treatment and Age Effects: Example replicants for 5-wave (A) and 4-wave (B) designs. Distributions of parameter estimates across replicants (C & D), and random effects of age (E & F) are presented for properly-specified models. Both models recovered parameters well. Generating parameters for each predictor are denoted by vertical dashed lines.



**Table 4**  
Estimated Effects for Growth Models from Scenario Set 3: Cohort Models.

	5-Wave Design						4-Wave Design					
	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig. (adj.)	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig. (adj.)
<i>Intervention Model</i>												
$\rho_{growth}$	.554	.025	.474	.634			.532	.024	.450	.602		
VIF	1.45	.059	1.29	1.67			1.40	.050	1.25	1.57		
$\gamma_{age}$	.301	.028	.222	.389	.043	1 (1)	.300	.031	.186	.406	.003	1 (1)
$\gamma_{cx}$	.301	.055	.145	.495	.027	1 (1)	.301	.052	.094	.469	.026	.999 (.998)
$\gamma_{int}$	-.101	.015	-.152	-.048	-.053	1 (1)	-.100	.022	-.156	-.026	-.001	.994 (.983)
$\sigma$	.501	.017	.444	.556	.034		.499	.018	.446	.565	-.057	
$\sqrt{\tau_{00}}$	.999	.060	.825	1.17	-.012		.997	.055	.831	1.19	-.046	
$\sqrt{\tau_{11}}$	.149	.019	.080	.203	-.053		.148	.030	.001	.223	-.060	
<i>TVC Model</i>												
$\rho_{growth}$							.410	.030	.297	.502		
VIF							1.20	.036	1.10	1.34		
$\gamma_{age}$							.429	.068	.185	.742	1.89	1 (1)
$\gamma_{stress}$							.300	.042	.168	.426	-.004	1 (1)
$\gamma_{int}$							-.101	.035	-.236	.018	-.026	.829 (.700)
$\sigma$							.500	.019	.444	.557	-.001	
$\sqrt{\tau_{00}}$							.999	.055	.810	1.18	-.026	
$\sqrt{\tau_{11}}$							.148	.029	.021	.221	-.054	

Note: Mean Est. is the expected parameter value across replicants and the Std. Err is the standard deviation of the sampling distribution. The proportion significant (Prop. Sig) represents the number of inferential tests that return a significant result for each parameter (unadjusted first, adjusted in parentheses). Std. Bias is the standardized bias estimate; values > 0.25 are bolded and represent unacceptable bias. The proportion significant only includes significant results where the inference is correct (e.g., significant negative effects are not included of the population parameter is positive). Rhos ( $\rho_{growth}$ ) represent the correlation between growth predictors, VIF represents variance inflation factor, gammas ( $\gamma$ ) represent the fixed effects, sigma ( $\sigma$ ) is the square-root of the residual random variance, and tau ( $\tau_{00}$ ) is the person-level random variance (square-root shown for consistency). The (mis) denotes parameters from the mis-specified (i.e., age-only) model.

### 3.3.2. MLMGMs as a subset of TVC models

These results highlight the link between MLMG models and a general class of longitudinal models with time-varying covariates (TVCs). Indeed, replacing intervention in the previous design with a different covariate (e.g., stress or anxiety) would be representative of a traditional single-growth process multi-level model (Curran and Bauer, 2011). For instance, if we were interested in how depressive symptoms change across age during adolescence and included stress as a TVC of interest, this would follow a similar form to the MLMGMs I have demonstrated previously. However, there are some key differences between these predictors. First, the non-age growth process predictors (e.g., practice, intervention) are monotonic and intrinsically related to age (i.e., accumulating exposure does not decrease across age). While a general TVC can be correlated with age (i.e., stress also increases during adolescence), stress is not constrained to monotonic growth (e.g., stress may increase between wave 1 and 2, but decrease between 2 and 3). Additionally, the stress predictor will be subject to measurement error, whereas the MLMGM predictors I have considered thus far are truly fixed and known (e.g., number of exposure events), although this may not always be the case.

I simulated data to follow this form, where adolescent depressive symptoms are modeled as a function of age, stress, and their interaction. I also included a random effect of age, mimicking the intervention scenario above. To create a better comparison with the MLMGM simulations, I simulated stress to show age-related increases, but unlike the MLMGM, I included a random error term in the model for the stress predictor ( $\sigma_{error} = .5$ ; Fig. 6B). Results showed that this error introduced bias into the estimate of the age predictor ( $Bias_{std} = 1.89$ ,  $SE = .068$ ), but not the estimates for the stress or interaction predictors (compare Fig. 6C with Fig. 5C & D; see Table 4). Aside from this bias introduced by the measurement error in the TVC (see Curran and Bauer, 2011 for more on this issue), this model performs like an MLMGM, highlighting the firm grounding of the current approach in previous methods.

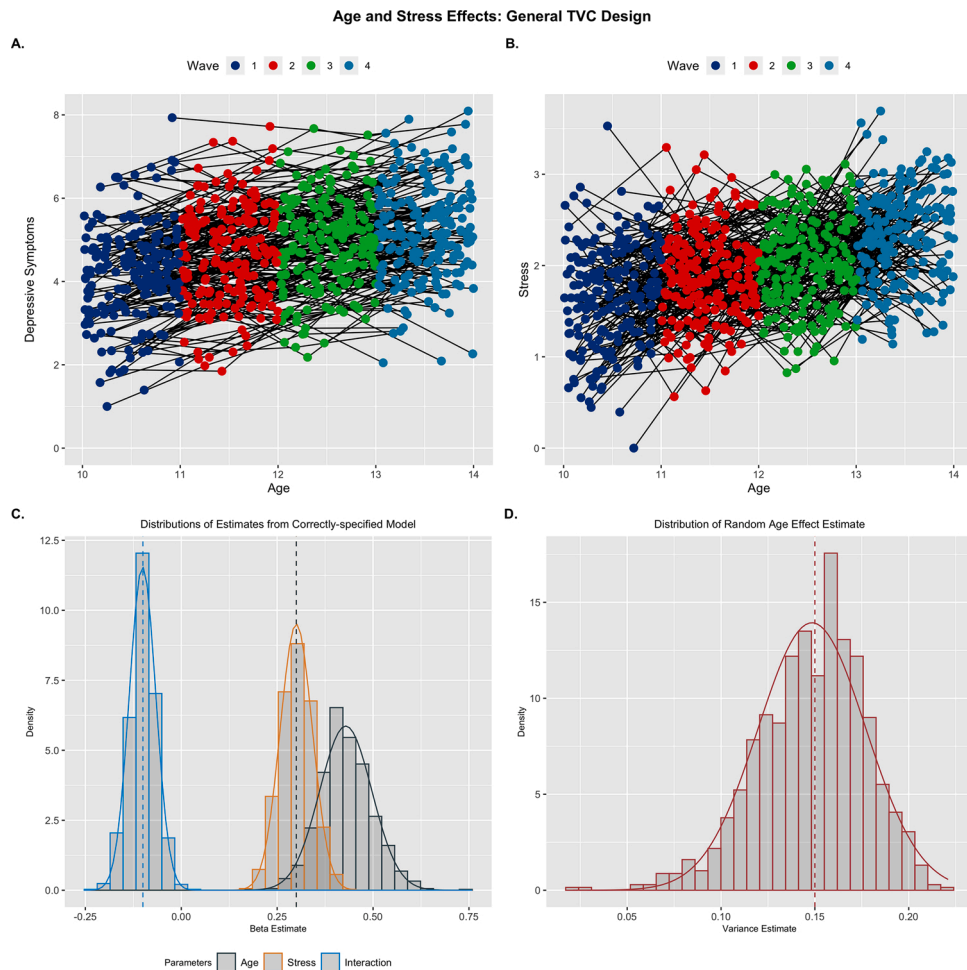
### 3.4. Scenario set 4: disentangling puberty and age effects

The inability to jointly model the effects of pubertal hormones and chronological age has been a stumbling block for developmental science in understanding a range of phenomenon of interest (e.g., risk-taking and reward-seeking behavior). Previous work (Blakemore et al., 2010; van Duijvenvoorde et al., 2019; Wierenga et al., 2018) has highlighted these challenges and called for implementing research designs which can address these limitations. One potential solution posed by Blakemore and colleagues (2010) was to recruit individuals at the same age, but different pubertal development status, and then follow them longitudinally. This idea fits perfectly into the framework of MLMGMs that I have presented thus far, if we think of this design as having planned missingness with respect to puberty instead of age. That is, individuals are measured across all levels of age, but not all levels of puberty. To demonstrate the ability of multi-level multi-growth models to capture these unique effects, I constructed a model to disentangle the impact of pubertal development and age on reward sensitivity across adolescence.

I considered a number of factors when constructing the simulations of age and puberty effects to better reflect the real-world nature of pubertal trends. I built ground-truth pubertal trajectories for each set of simulates from a sigmoid function (van Duijvenvoorde et al., 2019; Wierenga et al., 2018):

$$p = \frac{1}{1 + e^{-k*(x-z)}} \quad (6)$$

Here,  $p$  is the level of puberty development where  $0 < p < 1$  (lower bound = pre-puberty, upper bound = post-puberty),  $z$  is the mid-point of the trajectory ( $p = .5$ ),  $k$  is the shape parameter which governs the steepness of the pubertal curve (steeper curves related to a shorter window of pubertal development), and  $x$  ranges from -4 to 4 around the mean age of the sample ( $M_{age} = 12$  for all simulations). I drew individual values for the location parameter from a truncated normal distribution with mean 0 and bounds at -3 and 3 ( $N_T[0, \sigma, -3, +3]$ ), and values for the shape parameter from a gamma distribution ( $\Gamma[10, 3]$ ) to produce more relatively slow (> 2 years for complete transition) pubertal



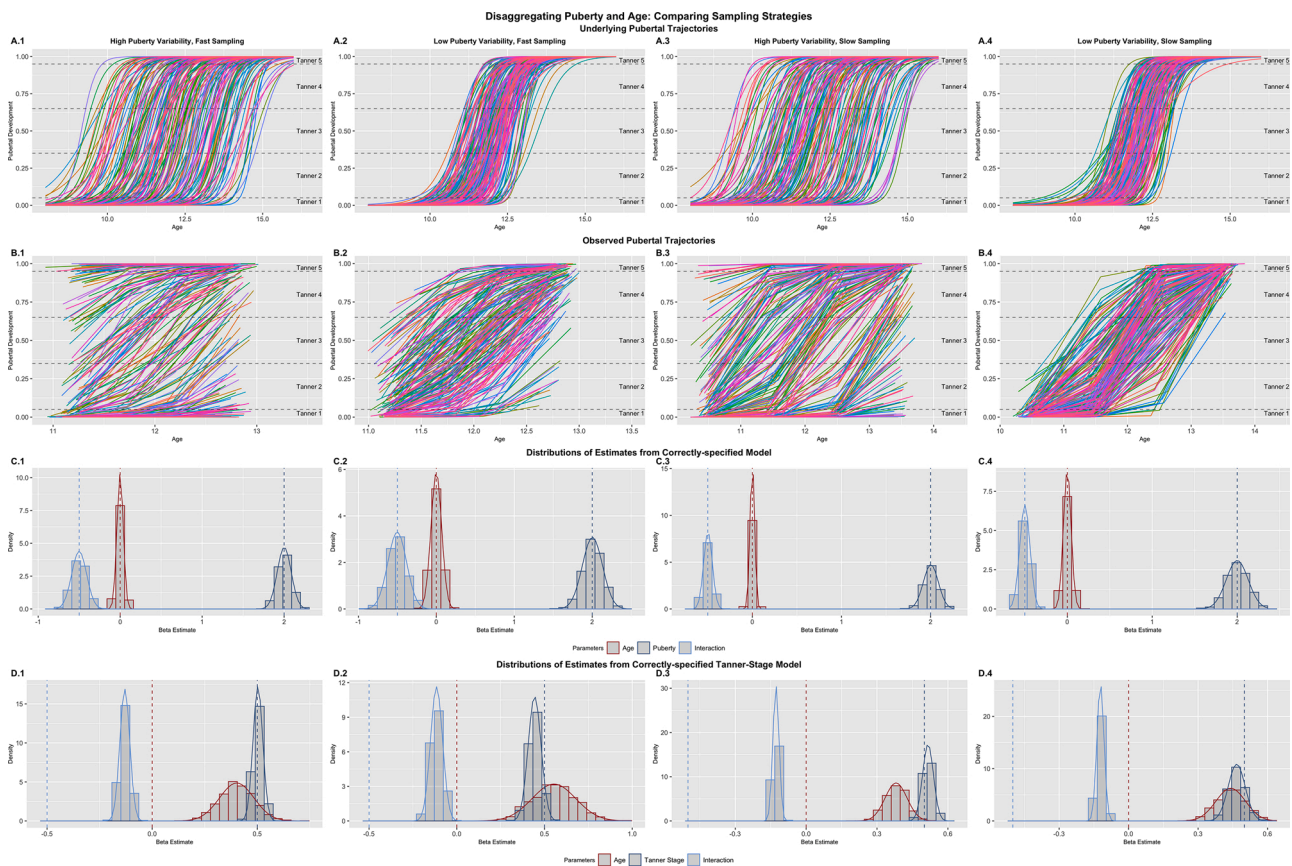
**Fig. 6.** Effects from a TVC Model: Example replicant for trajectories of depression (A) and stress (B) across early adolescence. Distributions of fixed effects (C) and random age effect (D) parameter estimates are plotted. Generating parameters for each predictor are denoted by vertical dashed lines. The random measurement error in the stress predictor caused inflation in the parameter estimate of the age predictor in the model (black histogram in C).

trajectories. I simulated a correlated multivariate distribution for the location and shape parameters such that simulates who started puberty earlier were more likely to have protracted pubertal time-courses and those who initiated puberty relatively late had faster transitions through puberty (Marti-Henneberg and Vizmanos, 1997). Examples of underlying pubertal trajectories for each condition are shown below (Fig. 7A.1–A.4) for descriptive purposes.

I explored a number of variations in order to assess how MLMGMs might be used to disentangle the effects of puberty. In addition to the differences between cohort and accelerated (with respect to age) designs, I also varied the degree of variability in the midpoints of pubertal trajectories and the rate of sampling. For conditions with high pubertal variability, I set  $\sigma = 1.5$  when drawing location parameters, and  $\sigma = .5$  for low variability conditions. I defined high-sampling as an observation each 6 months, and low sampling as the traditional annual observation design. By centering both sampling designs around a mean of 12, this condition highlights the influence of the tails of puberty trajectories (where little change is occurring) on the ability to accurately estimate parameters of interest. Finally, while sampling from the underlying pubertal trajectories (e.g., hormonal assays) would be ideal from a measurement perspective, pubertal research in practice often relies on much coarser measurement categories (e.g., Tanner Stages; Marshall and Tanner, 1969, 1970). To reflect this reality, I defined cut-off values to transform the true pubertal measures into 5 stages. Values less than .05 reflected Tanner Stage 1 (pre-puberty), values  $.05 \leq p < .35$  reflected Tanner Stage 2 (early puberty), values  $.35 \leq p < .65$  reflected Tanner

Stage 3 (middle puberty), values  $.65 \leq p < .95$  reflected Tanner Stage 4 (late puberty), and values greater than .95 reflected Tanner Stage 5 (post-puberty). I fit model with both the true and transformed Tanner scores to assess the impact of this transformation on parameter recovery. Based on prior work, this coarsening should introduce bias into the model based on attenuated correlations between the variables (Bollen and Barb, 1981; Taylor et al., 2006). For all models, I simulated reward sensitivity (e.g., ventral striatum response) to have no main effect of age ( $\gamma_{age} = 0$ ), a strong effect of puberty ( $\gamma_{puberty} = 2$ ), and a negative interaction effect ( $\gamma_{int} = -0.5$ ) such that the impact of puberty on reward sensitivity is blunted at older ages. It is important to note that while the relationship between age and puberty is clearly non-linear (indeed I generated values from a truly non-linear model), the relationship between puberty and the outcome is linear, allowing for these effects to be modeled as usual in a linear model framework.

Results confirmed that the MLMGM worked to disaggregate the effects of puberty and age with high fidelity when the design was cohort with respect to age and accelerated with respect to puberty. A high sampling rate (i.e., biannual observations), coupled with a high degree of variability in pubertal timing (Fig. 7B.1) showed the greatest ability to de-couple the age and puberty predictors, with between-predictor correlations near those seen in the age-practice scenarios ( $M_p = .345$ ,  $SE = .024$ ). A slower sampling approach (i.e., annual observations) inflated these correlations (Fig. 7B.3;  $M_p = .599$ ,  $SE = .025$ ), however, with sufficient power, these effects can still be estimated reliably (i.e., false negative rates were low at the simulated effect sizes). Low pubertal



**Fig. 7.** Effects from the Puberty Models: Example replicant for underlying (A) and observed (B) trajectories of pubertal development. Distributions of fixed effects are plotted for models where the puberty predictor is sampled from the continuous puberty curves (C) and models where puberty is assessed by the transformed Tanner Stage values (D). This transformation introduced substantial bias into the age predictor estimates, whereas no bias was introduced when using the underlying values. Only data from the cohort model are shown.

variability (Fig. 8B.2 & B.4) presented more significant challenges, although the high ( $M_p = .750, SE = .019$ ) versus low ( $M_p = .898, SE = .007$ ) sampling strategy was still relatively more able to attenuate the inter-predictor correlation. As in previous simulations, the high correlations do not introduce bias into the parameter estimates, but rather inflate standard errors leading to increases in false positives (for full details, see Tables 5–8). Across most conditions, an accelerated design (with respect to age) increased the correlations between age and puberty predictors relative to the respective cohort design, with the exception of the case where there is low pubertal variability and slow sampling. For conciseness, I displayed pubertal curves and parameter recovery for the cohort design results only (Fig. 7); however, full details for both model types are provided elsewhere (see Tables 5–8; Figs. S5–S8).

The models with pubertal values sampled from the true underlying trajectories (Fig. 8A.1–A.4) showed unbiased parameter estimates for both the main effects and the interaction term (Fig. 7C.1–C.4). In contrast, coarsening the continuous trajectories into an ordinal Tanner Stage predictor introduced some bias into pubertal parameter estimates as expected (Fig. 7D.1–D.4), except in the high variability, high sampling condition ( $Bias_{std} = -.087, SE = .023$ ). Interestingly this bias was not uniformly in one direction across all conditions. However, the bias introduced to the estimate of the Tanner Stage effect was minor in comparison to the bias introduced into the age parameter estimates ( $Bias_{std} \text{ range} = 4.45\text{--}8.20$ ). The biases for the interaction estimates ( $Bias_{std} \text{ range} = -.210 \text{ to } .283$ ) were marginally problematic as well. Overall, the same principles that lead to successful model performance in previous scenarios held here by considering data to be accelerated with respect to puberty. However, differences due to sampling time and measurement error highlight additional challenges that will need to be

overcome in real-world applications of these models to disentangle effects of correlated predictors across time. While this challenge is not unique to longitudinal designs, they are uniquely problematic since predictors like age, experience, and puberty monotonically increase across time. Even if the true effects of these predictors are nonlinear (e.g., quadratic), standard modeling approaches necessitate that linear terms also be included. When these linear terms are too highly colinear, they can introduce significant problems for proper estimation of model parameters and impact the substantive conclusions about inflection points and sensitive periods in developmental trajectories. However, simply excluding one of the correlated predictors threatens the internal validity of the model and introduces significant bias into the retained parameters except under very restrictive assumptions (i.e., the true effect of the excluded predictor is zero). Fortunately, leveraging a natural extension of linear mixed-effects growth models in combination with appropriate sampling designs offers a solution to these challenges. Here I outlined a multi-level, multi-growth model (MLMGM) form and highlighted how it can be flexibly used across a number of scenarios to address questions from substantive developmental theory.

#### 4. Discussion

A key challenge for the study of change over time is disentangling effects of correlated predictors across time. While this challenge is not unique to longitudinal designs, they are uniquely problematic since predictors like age, experience, and puberty monotonically increase across time. Even if the true effects of these predictors are nonlinear (e.g., quadratic), standard modeling approaches necessitate that linear terms also be included. When these linear terms are too highly colinear, they can introduce significant problems for proper estimation of model parameters and impact the substantive conclusions about inflection points and sensitive periods in developmental trajectories. However, simply excluding one of the correlated predictors threatens the internal validity of the model and introduces significant bias into the retained parameters except under very restrictive assumptions (i.e., the true effect of the excluded predictor is zero). Fortunately, leveraging a natural extension of linear mixed-effects growth models in combination with appropriate sampling designs offers a solution to these challenges. Here I outlined a multi-level, multi-growth model (MLMGM) form and highlighted how it can be flexibly used across a number of scenarios to address questions from substantive developmental theory.

**Table 5**  
Estimated Effects for Puberty Models: High Variability, Fast Sampling.

	Cohort Design						Accelerated Design					
	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig.(adj.)	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig.(adj.)
<i>Puberty Model</i>												
$\rho_{growth}$	.345	.024	.272	.423			.548	.037	.419	.653		
VIF	1.14	.022	1.08	1.22			1.44	.084	1.21	1.74		
$\gamma_{age}$ (mis)	.500	.045	.364	.644	<b>11.13</b>	1 (1)	.432	.035	.308	.548	<b>12.41</b>	1 (1)
$\gamma_{age^2}$ (mis)	-.116	.075	-.382	.152	<b>5.14</b>	.287 (.196)	-.088	.022	-.155	-.024	<b>18.74</b>	.980 (.970)
$\sigma$ (mis)	.592	.017	.543	.660	<b>5.30</b>		.564	.018	.507	.619	<b>3.48</b>	
$\sqrt{\tau_{00}}$ (mis)	.869	.042	.727	1.00	<b>8.80</b>		.836	.043	.684	.980	<b>7.87</b>	
$\gamma_{age}$	-.001	.039	-.127	.121	-.021	.051 (.013)	-.001	.035	-.121	.108	-.039	.060 (.024)
$\gamma_{puberty}$	2.00	.085	1.71	2.30	.017	1 (1)	2.00	.098	1.71	2.31	-.003	1 (1)
$\gamma_{int}$	-.500	.091	-.814	-.235	-.005	1 (1)	-.498	.067	-.710	-.249	.036	1 (1)
$\sigma$	.500	.015	.458	.545	-.026		.500	.016	.451	.551	.012	
$\sqrt{\tau_{00}}$	.497	.033	.396	.602	-.087		.497	.031	.387	.601	-.084	
<i>Tanner-Stage Model</i>												
$\rho_{growth}$	.350	.023	.284	.427			.556	.036	.438	.665		
$\gamma_{age}$	.402	.082	.148	.736	<b>4.90</b>	.999 (.996)	.387	.064	.177	.573	<b>6.04</b>	1 (1)
$\gamma_{tanner}$	.503	.023	.418	.577	.117	1 (1)	.498	.026	.420	.572	-.087	1 (1)
$\gamma_{int}$	-.130	.024	-.215	-.054	-.225	1 (.999)	-.125	.018	-.181	-.054	-.023	1 (1)
$\sigma$	.500	.015	.458	.545	-.026		.500	.016	.451	.551	.012	
$\sqrt{\tau_{00}}$	.497	.033	.396	.602	-.087		.497	.031	.387	.601	-.084	

Note: Mean Est. is the expected parameter value across replicants and the Std. Err is the standard deviation of the sampling distribution. The proportion significant (Prop. Sig) represents the number of inferential tests that return a significant result for each parameter (unadjusted first, adjusted in parentheses). Std. Bias is the standardized bias estimate; values > .25 are bolded and represent unacceptable bias. The proportion significant only includes significant results where the inference is correct (e.g., significant negative effects are not included of the population parameter is positive). Rhos ( $\rho_{growth}$ ) represent the correlation between growth predictors, VIF represents variance inflation factor, gammas ( $\gamma$ ) represent the fixed effects, sigma ( $\sigma$ ) is the square-root of the residual random variance, and tau ( $\tau_{00}$ ) is the person-level random variance (square-root shown for consistency). The (mis) denotes parameters from the mis-specified (i.e., age-only) model.

**Table 6**  
Estimated Effects for Puberty Models: Low Variability, Fast Sampling.

	Cohort Design						Accelerated Design					
	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig.(adj.)	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig.(adj.)
<i>Puberty Model</i>												
$\rho_{growth}$	.750	.019	.691	.808			.832	.013	.787	.871		
VIF	2.29	.150	1.92	2.87			3.26	.224	2.63	4.13		
$\gamma_{age}$ (mis)	.915	.040	.794	1.05	<b>23.02</b>	1 (1)	.606	.034	.483	.721	<b>17.59</b>	1 (1)
$\gamma_{age^2}$ (mis)	-.178	.079	-.426	.090	<b>4.09</b>	.655 (.548)	-.096	.024	-.185	-.021	<b>16.84</b>	.988 (.978)
$\sigma$ (mis)	.556	.017	.502	.608	<b>3.33</b>		.564	.018	.506	.628	<b>3.50</b>	
$\sqrt{\tau_{00}}$ (mis)	.633	.038	.507	.754	<b>3.52</b>		.611	.036	.497	.749	<b>3.10</b>	
$\gamma_{age}$	.000	.068	-.227	.226	.003	.048 (.015)	.002	.049	-.139	.175	.049	.063 (.027)
$\gamma_{puberty}$	2.00	.130	1.52	2.45	.026	1 (1)	2.00	.130	.149	2.34	-.022	1 (1)
$\gamma_{int}$	-.501	.121	-.924	-.160	-.010	.992 (.954)	-.499	.074	-.794	-.269	.016	1 (1)
$\sigma$	.500	.015	.451	.548	.028		.500	.016	.445	.550	-.012	
$\sqrt{\tau_{00}}$	.498	.033	.392	.595	-.058		.500	.031	.405	.596	-.014	
<i>Tanner-Stage Model</i>												
$\rho_{growth}$	.728	.018	.664	.788			.838	.013	.794	.882		
$\gamma_{age}$	.554	.124	.196	.981	<b>4.45</b>	.998 (.987)	.434	.078	.168	.733	<b>5.56</b>	1 (.999)
$\gamma_{tanner}$	.444	.037	.323	.563	-.152	1 (1)	.468	.037	.348	.566	-.878	1 (1)
$\gamma_{int}$	-.115	.034	-.219	-.007	<b>.283</b>	.928 (.833)	-.121	.019	-.189	-.065	.231	1 (1)
$\sigma$	.500	.015	.451	.548	.028		.500	.016	.445	.550	-.012	
$\sqrt{\tau_{00}}$	.498	.033	.392	.595	-.058		.500	.031	.405	.596	-.014	

Note: Mean Est. is the expected parameter value across replicants and the Std. Err is the standard deviation of the sampling distribution. The proportion significant (Prop. Sig) represents the number of inferential tests that return a significant result for each parameter (unadjusted first, adjusted in parentheses). Std. Bias is the standardized bias estimate; values > .25 are bolded and represent unacceptable bias. The proportion significant only includes significant results where the inference is correct (e.g., significant negative effects are not included of the population parameter is positive). Rhos ( $\rho_{growth}$ ) represent the correlation between growth predictors, VIF represents variance inflation factor, gammas ( $\gamma$ ) represent the fixed effects, sigma ( $\sigma$ ) is the square-root of the residual random variance, and tau ( $\tau_{00}$ ) is the person-level random variance (square-root shown for consistency). The (mis) denotes parameters from the mis-specified (i.e., age-only) model.

4.1. Quantifying the effects of repeated exposure

The strength of repeated-measures designs (i.e., the ability to assess within-person change), can simultaneously be a threat to internal val-

idity. That is, individuals may change in their response to an instrument (e.g., survey, fMRI task) due to being assessed multiple times independent of changes that are due to developmental effects (Bell, 1953; Palmore, 1978; Telzer et al., 2018). Whether because of familiarity or



**Table 7**  
Estimated Effects for Puberty Models: High Variability, Slow Sampling.

	Cohort Design						Accelerated Design					
	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig.(adj.)	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig.(adj.)
<i>Puberty Model</i>												
$\rho_{growth}$	.599	.025	.521	.679			.615	.025	.502	.701		
VIF	1.57	.073	1.38	1.86			1.61	.079	1.34	1.97		
$\gamma_{age}$ (mis)	.456	.025	.382	.537	<b>18.50</b>	1 (1)	.415	.026	.329	.486	<b>16.01</b>	1 (1)
$\gamma_{age^2}$ (mis)	-.106	.024	-.176	-.037	<b>16.72</b>	.995 (.987)	-.084	.014	-.129	-.041	<b>29.96</b>	1 (1)
$\sigma$ (mis)	.666	.020	.606	.730	<b>8.12</b>		.658	.020	.590	.702	<b>7.56</b>	
$\sqrt{\tau_{00}}$ (mis)	.736	.042	.609	.873	<b>5.64</b>		.744	.041	.604	.889	<b>5.98</b>	
$\gamma_{age}$	.000	.025	-.083	.072	-.009	.056 (.018)	.000	.026	-.084	.084	.001	.065 (.023)
$\gamma_{puberty}$	2.00	.084	1.71	2.26	.027	1 (1)	2.00	.085	1.73	2.28	-.031	1 (1)
$\gamma_{int}$	-.498	.048	-.660	-.337	.042	1 (1)	-.501	.045	-.640	-.346	-.028	1 (1)
$\sigma$	.499	.015	.456	.557	-.039		.499	.016	.447	.546	-.076	
$\sqrt{\tau_{00}}$	.497	.032	.387	.589	-.082		.501	.031	.402	.613	.046	
<i>Tanner-Stage Model</i>												
$\rho_{growth}$	.606	.024	.534	.677			.623	.025	.510	.707		
$\gamma_{age}$	.381	.046	.207	.546	<b>8.20</b>	1 (1)	.382	.044	.233	.521	<b>8.76</b>	1 (1)
$\gamma_{tanner}$	.516	.023	.439	.600	<b>6.89</b>	1 (1)	.512	.023	.447	.586	<b>5.05</b>	1 (1)
$\gamma_{int}$	-.128	.013	-.169	-.080	-.205	1 (1)	-.127	.012	-.167	-.092	-.211	1 (1)
$\sigma$	.499	.015	.456	.557	-.039		.499	.016	.447	.546	-.076	
$\sqrt{\tau_{00}}$	.497	.032	.387	.589	-.082		.501	.031	.402	.613	.046	

Note: Mean Est. is the expected parameter value across replicants and the Std. Err is the standard deviation of the sampling distribution. The proportion significant (Prop. Sig) represents the number of inferential tests that return a significant result for each parameter (unadjusted first, adjusted in parentheses). Std. Bias is the standardized bias estimate; values > .25 are bolded and represent unacceptable bias. The proportion significant only includes significant results where the inference is correct (e.g., significant negative effects are not included of the population parameter is positive). Rhos ( $\rho_{growth}$ ) represent the correlation between growth predictors, VIF represents variance inflation factor, gammas ( $\gamma$ ) represent the fixed effects, sigma ( $\sigma$ ) is the square-root of the residual random variance, and tau ( $\tau_{00}$ ) is the person-level random variance (square-root shown for consistency). The (mis) denotes parameters from the mis-specified (i.e., age-only) model.

**Table 8**  
Estimated Effects for Puberty Models: Low Variability, Slow Sampling.

	Cohort Design						Accelerated Design					
	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig.(adj.)	Mean Est.	Std. Err.	Min	Max	Std. Bias	Prop. Sig.(adj.)
<i>Puberty Model</i>												
$\rho_{growth}$	.898	.007	.875	.917			.853	.007	.829	.876		
VIF	5.21	.334	4.28	6.30			3.67	.160	3.19	4.29		
$\gamma_{age}$ (mis)	.656	.018	.598	.710	<b>37.49</b>	1 (1)	.552	.025	.453	.625	<b>21.72</b>	1 (1)
$\gamma_{age^2}$ (mis)	-.132	.022	-.201	-.048	<b>16.52</b>	1 (1)	-.087	.017	-.146	-.041	<b>24.59</b>	1 (1)
$\sigma$ (mis)	.586	.018	.526	.638	<b>4.89</b>		.633	.019	.570	.704	<b>7.09</b>	
$\sqrt{\tau_{00}}$ (mis)	.535	.037	.414	.649	<b>9.43</b>		.533	.037	.430	.656	<b>9.00</b>	
$\gamma_{age}$	.000	.045	-.122	.176	-.004	.046 (.018)	-.001	.036	-.101	.123	-.023	.053 (.018)
$\gamma_{puberty}$	2.00	.129	1.56	2.36	.002	1 (1)	2.00	.107	1.62	2.33	.024	1 (1)
$\gamma_{int}$	-.499	.060	-.680	-.284	.014	1 (1)	-.500	.050	-.663	-.342	.005	1 (1)
$\sigma$	.499	.015	.457	.555	-.039		.500	.016	.442	.549	-.005	
$\sqrt{\tau_{00}}$	.497	.033	.398	.606	-.089		.498	.032	.399	.609	-.060	
<i>Tanner-Stage Model</i>												
$\rho_{growth}$	.897	.008	.870	.920			.861	.008	.836	.884		
$\gamma_{age}$	.441	.064	.239	.622	<b>6.91</b>	1 (1)	.383	.053	.228	.554	<b>7.28</b>	1 (1)
$\gamma_{tanner}$	.467	.037	.353	.591	-.904	1 (1)	.510	.030	.406	.591	<b>3.23</b>	1 (1)
$\gamma_{int}$	-.122	.015	-.171	-.064	.218	1 (1)	-.123	.013	-.166	-.075	.172	1 (1)
$\sigma$	.499	.015	.457	.555	-.039		.500	.016	.442	.549	-.005	
$\sqrt{\tau_{00}}$	.497	.033	.398	.606	-.089		.498	.032	.399	.609	-.060	

Note: Mean Est. is the expected parameter value across replicants and the Std. Err is the standard deviation of the sampling distribution. The proportion significant (Prop. Sig) represents the number of inferential tests that return a significant result for each parameter (unadjusted first, adjusted in parentheses). Std. Bias is the standardized bias estimate; values > .25 are bolded and represent unacceptable bias. The proportion significant only includes significant results where the inference is correct (e.g., significant negative effects are not included of the population parameter is positive). Rhos ( $\rho_{growth}$ ) represent the correlation between growth predictors, VIF represents variance inflation factor, gammas ( $\gamma$ ) represent the fixed effects, sigma ( $\sigma$ ) is the square-root of the residual random variance, and tau ( $\tau_{00}$ ) is the person-level random variance (square-root shown for consistency). The (mis) denotes parameters from the mis-specified (i.e., age-only) model.

practice with a task, changing sensitivity of certain questionnaire items (e.g., drug use), or increased comfort with the testing environment, the effect of repeated exposure can introduce bias into estimates of developmental trajectories (Salthouse, 2014). I began by outlining a number

of scenarios where this bias is highlighted. Except in scenarios where the true of effect of repeated exposure was zero, the presence of these effects posed challenges at multiple stages of model estimation and interpretation. However, merely including relevant predictors does not

necessary ameliorate all of these challenges. Traditional cohort-longitudinal designs, where all individuals are assessed at every age levels (Fig. 1B), limit the utility of including predictors of repeated exposure because the high correlation between age and assessment ( $M_p \approx .998$ ) prevents obtaining unique effects with any precision, or detecting significant effects when they do exist (i.e., standard errors are highly inflated). I showed that only MLMGMs combined with accelerated-longitudinal designs could reliably recover the underlying model generating parameters under commonly-encountered retest conditions. I highlight how this approach can address several specific confounds in assessing developmental effects.

#### 4.1.1. Bias from omitted predictors

Across all of the presented scenarios, I first fit models that follow conventional practices (Curran et al., 2010; Duncan et al., 1996; Herting et al., 2018; Sliwinski et al., 2010; Soden et al., 2015; Telzer et al., 2018), where only the effects of age are assessed as a growth process in the model (i.e., single-growth models). These models highlighted the various challenges that can arise from omitting relevant predictors. In cohort data, the effect of repeated measures (e.g., practice, habituation) was almost completely absorbed into the parameter estimates for age due to the high correlations between predictors (Frees, 2001; Kim and Frees, 2006). Depending on the nature of the two effects, this led to serious over- or under-estimation of the true age effect. However, the bias in the accelerated design is substantially lower than in the cohort data (standardized values of  $\sim 3$  versus  $\sim 11.5$ ) due to the attenuation of the correlations between predictors ( $M_p \approx .380$ ).

However, the key advantage of this attenuation is that it allows for even greater validity in modeling developmental trajectories, since both age and the repeated measure predictor can be included in the same model. When including these multiple growth predictors, the model appropriately estimates the independent effects of age and repeated exposure from accelerated data. While effects are unbiased in the aggregate with multiple predictors in the cohort data, the parameters from any one replicant (i.e., study) may be quite biased due to instability in the estimates, and the likelihood for false-negatives is high ( $> 85\%$ ) due to inflated standard errors ( $> 14$  times as large as if predictors were uncorrelated). As demonstrated by the simulation results, failure to account for these omitted effects could lead to biased estimates and contribute to a lack of reproducibility and generalization of effects across studies if the effects of repeated exposure vary between measures, populations, or other assessment parameters.

#### 4.1.2. Bias in models with higher-order effects

The issue of high correlations between predictors impacts linearly related independent variables, however, there are any number of situations where the relationship between growth predictors and outcomes of interest might be expected to be non-linear, or some combination of linear and non-linear effects. Because of the centering method for creating product terms (i.e., interactions or polynomials; Aiken et al., 1991), the predictors for these higher order effects will be uncorrelated with the linear terms in the model. The simulations with a quadratic effect of age demonstrate this (see Supplemental Material for complete details), as this term is estimated without bias and with high precision in both cohort and accelerated models (Fig. S4). Unfortunately, including correlated linear predictors are still necessary to properly model any higher-order effects, nor does it solve issues for model-implied features of the developmental trajectories. These implied features can include the rate of the linear trend, or where the quadratic or interaction inflection (i.e., vertex) occurs within the overall developmental trajectory. Because of the instability in the linear term, the properly specified quadratic effect can still lead to implied inflections that vary widely. For instance, the implied vertex in the mis-specified (i.e., age-only) model using a cohort design ranged across 5 years (15–20), and in the properly-specified model, across more than a full decade. Because real studies do not have the benefit of obtaining a sampling distribution,

results from any one study could lead to radically different substantive conclusions based on parameter instability. This complication could further decrease replicability of effects across samples, even though the point estimates of product term effects could be nearly identical.

#### 4.2. Generalizing the idea of accelerated designs

While distinguishing between experience- (e.g., practice, habituation, etc.) and age-related changes in outcomes of interest is a persistent issue, the utility of the MLMGM generalizes to many modeling contexts. Experience effects are often treated as a nuisance variable ((Ferrer et al., 2004; Rabbitt et al., 2001) but see (McCormick et al., 2021)) that contaminates the developmental effect. As such, MLMGMs can be used to partial out variance associated with repeated assessment. In other contexts, both growth processes may be of interest and the focus can be on characterizing trajectories across both levels of time. In many of these scenarios, accelerated designs (as typically conceived) may be impossible or undesirable. Nevertheless, the MLMGM framework can be leveraged. The key insight to understand how MLMGMs separate different growth effects; by measuring all individuals across the full range of one growth predictor, but on only a subset of the levels of the other predictor. In the age/experience simulations, every individual is measured at three occasions (i.e., across all levels of the experience predictor) but on only a subset of ages. This fits perfectly within a generalized framework of planned missingness in study design (Little and Rhemtulla, 2013; Rhemtulla and Little, 2012). Under this broad framework, researchers have flexibility in defining planned missing designs for many sets of correlated growth processes. To highlight this flexibility, I presented one such example, where individuals are measured at all levels of age (within the scope of the study), which results in a cohort data frame, but where the levels of the intervention vary across individuals at any given timepoint. As such, researchers can utilize both accelerated and cohort longitudinal designs in implementing this modeling framework, depending on the theoretical question at hand.

I also showed how the MLMGM is a special case of a general growth model with a time-varying covariate (TVC) with a model distinguishing effects of age and stress on depression (Curran and Bauer, 2011). However, the multi-growth model introduces challenges that may not impact more general TVC models. TVCs, like stress, can fluctuate non-monotonically across measurement occasions, even when there is a general trend across age (e.g., stress generally increases across the transition to adolescence). In contrast, the predictors in growth-processes (including age, experience, and puberty) cannot decrease from one measurement occasion to the other, which drives the need to leverage accelerated designs to attenuate otherwise near-perfect multi-collinearity. The advantage of these multiple-growth predictors, relative to other TVCs, is that there can potentially be very low – or non-existent – measurement error (although this may not always be the case) since the number of assessments or interventions for each individual is truly fixed and known. This set of scenarios, where planned missingness occurs in a growth predictor other than age is a rich target for novel and innovative future work and where existing cohort datasets can be used to empirically test the types of effects I have presented here.

#### 4.3. Using multi-level, multi-growth models to address developmental theory

Finally, I applied the principles derived from the first three sets of scenarios to demonstrate an approach for disentangling the effects of pubertal and age-related development. This challenge is a long-standing one in the developmental literature (Blakemore et al., 2010; van Duijvenvoorde et al., 2019), and interestingly, results suggest that data needed to address this gap likely already exist. In particular, I showed that cohort designs could be leveraged to estimate both age and puberty effects with reasonably low correlations, especially when sampling

occurs more often than typical annual visits ( $M_p = .345$ ). However, even though annual observations increased this correlation ( $M_p = .599$ ), a sufficiently powered design can overcome these challenges (the current simulations had 188 individuals measured four times) with standard errors being only  $\sim 1.25$  times larger than expected. In the universe of longitudinal studies, similar data likely exists and will with increasing frequency in the coming years (e.g., the ABCD study) for a range of theoretical questions. The limiting factor may be the variability of pubertal trajectories within the observation period, which is difficult to account for at the onset of data collection. In simulations with low pubertal variability, the correlations between age and puberty were highly inflated, especially with slower (i.e., annual) observations ( $M_p = .898$ ). One strategy for addressing this limitation could be to sub-sample individuals from a larger study to maximize the amount of pubertal variability, although this would have to be done carefully to ensure that the properties of the sub-sample match the larger study characteristics, and avoid selecting on the outcome (Elwert and Winship, 2014). Of course, this could be further complicated by the likely need to separate out effects by groups (e.g., gender), but in principle the same approach could be done for each group if ad hoc sampling techniques did not (by chance or design) result in high levels of pubertal variability. Alternatively, high-density sampling attenuated the correlation between predictors even with low pubertal variability, offering researchers multiple avenues for building these models.

While the MLMGM framework was able to appropriately capture pubertal and age-related effects when using measurements of the underlying trajectory, coarse categorization of those values to an ordinal scale (representing Tanner Stages) introduced bias into both the estimated effect of Tanner Stage and uniformly inflated the estimate of age based on attenuating correlations among predictors (Bollen and Barb, 1981; Taylor et al., 2006). Importantly, this bias was not due to meaningful differences in the correlations between the growth predictors (see Tables 5–8). This suggests that age effects in models using Tanner Stage as a predictor may need to be interpreted with caution, as they may be biased upwards. The bias in the Tanner Stage estimate itself was smaller and directionally inconsistent across simulation conditions (Fig. 7D), but also suggests that unlike the effect of stress in the TVC scenario, the coarse categorization also impacts the estimate of the pubertal effect. These results suggest that continuous measures of puberty (e.g., hormone concentrations) are likely to be more reliable in their effect estimates, despite the near perfect correlation between the underlying puberty measure and the Tanner Stage ( $M_p = .965\text{--}.985$ ). While continuous measures of puberty are often more expensive and can be challenging to collect, these results suggest that they are likely necessary for maintaining internal validity.

Finally, while not the main focus of the pubertal analyses, I did assess how the use of an accelerated design (with respect to age) impacted results (see Tables 5–8 for details; Figs. S5–S8). In general, these models performed similarly to those fit to cohort data and were able to deliver unbiased estimates for the models using underlying pubertal trajectories, despite an inflated correlation between the puberty and age predictors, as well as increasing standard errors for all estimates. However, it is important to note that while the accelerated design simulated reflects common practice, it could be modified to minimize this correlation. For example, instead of planning the missingness around puberty, researchers could use an accelerated (i.e., missing age) design and measure individuals across the complete range of puberty instead, or sub-sample from a large accelerated study to approximate this design. This is obviously a more challenging design to realize than one where researchers can simply track age, given the variability in timing and rate of pubertal development and potentially complicated relationships between underlying maturation and easily-identifiable physical markers. Nevertheless, some theoretical questions may require individuals sampled across the full range of puberty, making a focused study of this nature extremely valuable.

#### 4.4. Future work

While the current study suggests that MLMGMs offer a promising framework for simultaneously modeling growth processes that co-occur across time, there remain two avenues for future work to explore. As I have hopefully demonstrated with a small subset of potential scenarios, these models can be applied to understand a wide array of substantive theoretical questions. In addition to being useful for testing novel questions, MLMGMs could be used to re-visit previous data where simultaneous growth processes were not considered (i.e., age-only models; see (McCormick et al., 2021) for an example). Additionally, the findings here highlight the clear importance of sampling design for fitting models appropriately in longitudinal data. While this is hardly a new revelation (Bell, 1953; Kraemer et al., 2000; Louis et al., 1986; Palmore, 1978; Van't Hof et al., 1977), it bears repeating due to the crucial role design plays in reducing correlations between growth predictors and keeping standard errors appropriately narrow. With these considerations in mind, I hope that the current work can help guide analysis of current data, as well as motivate new data collection with planned missing designs to disentangle multiple growth processes that occur across time.

From a methodological perspective, there are additional steps that could broaden our knowledge about how MLMGMs perform across a variety of conditions that are less ideal, but nevertheless common. For instance, while I attempted to choose sample sizes that would reasonably reflect reality and avoid potential issues of power, the sample sizes encountered in practice likely vary widely (e.g.,  $> 1000$  might be reasonable for a school-based survey study, but  $< 100$  might be more typical of neuroimaging studies). In particular, future work should seek to characterize the boundary conditions where correlations between growth predictors are too inflated for a given sample size to yield satisfactory inferential performance. While this work can build off principles with work done with correlated predictors in general (Shieh and Fouladi, 1991), the unique characteristics of growth predictors (e.g., monotonic across time) may require innovative designs and tailored solutions to decouple them. Additionally, while the accelerated design takes advantage of planned missingness, longitudinal studies are often subject to more pernicious missing data mechanisms, including attrition and other non-random processes (Enders, 2011; Schafer and Graham, 2002). Finally, growth factors in the scenarios that I explored all varied at each measurement occasion at level 1. For instance, age or puberty was necessarily different at each observation (although the degree to which they changed might differ). However, it seems possible that growth factors might change differentially across time and potentially enter as predictors at different levels of the model. An example could arise if researchers were interested in modeling changes within-session in addition to between-session growth, where age would be consistent across all levels of the within-session observations. One potential issue is that nesting under a predictor like age or wave might cause estimation or inferential problems that are not entirely clear. Additional work will be needed to clarify this point.

#### 4.5. Conclusions

I proposed a natural extension of the multi-level growth model to accommodate estimating multiple growth process simultaneously: the multi-level, multi-growth model (MLMGM). To deal with the correlations between growth predictors (e.g., age and repeated exposure), I demonstrated how accelerated-longitudinal designs could attenuate these predictor relationships through planned missingness. I further generalized the idea of accelerated designs to show that you could plan missingness in growth predictors other than age (e.g., treatment) and how these models behaved similarly to general growth models with time-varying covariates. Finally, I demonstrated how MLMGMs could be used to address the thorny substantive question of simultaneously estimating the disparate impacts of puberty and age on behavior and

cognition. Taken together, this work motivates innovation in developmental designs and modeling of trajectories across time across a broad domain of empirical research.

## Data and code availability

All relevant code for replicating simulated data and analyses are available online (<https://osf.io/65sdw/>).

## Declaration of Competing Interest

The authors report no declarations of interest.

## Acknowledgements

Thanks go to P.J. Curran for providing thoughtful conversations on the topic of longitudinal models in accelerated designs and to P.J. Curran and J.A. Andrews for feedback on the initial draft of the manuscript.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.dcn.2021.101001>.

## References

- Aiken, L.S., West, S.G., Reno, R.R., 1991. *Multiple Regression: Testing and Interpreting Interactions*. Sage Publications.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67 (1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Bauer, D.J., 2003. Estimating multilevel linear models as structural equation models. *J. Educ. Behav. Stat.* 28 (2), 135–167.
- Bell, R.Q., 1953. Convergence: an accelerated longitudinal approach. *Child Dev.* 24 (2), 145. <https://doi.org/10.2307/1126345>.
- Blakemore, S.J., Burnett, S., Dahl, R.E., 2010. The role of puberty in the developing adolescent brain. *Hum. Brain Mapp.* 31 (6), 926–933. <https://doi.org/10.1002/hbm.21052>.
- Bollen, K.A., Barb, K.H., 1981. Pearson's  $r$  and coarsely categorized measures. *Am. Sociol. Rev.* 232–239.
- Breiter, H.C., Etcoff, N.L., Whalen, P.J., Kennedy, W.A., Rauch, S.L., Buckner, R.L., et al., 1996. Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* 17 (5), 875–887. [https://doi.org/10.1016/S0896-6273\(00\)80219-6](https://doi.org/10.1016/S0896-6273(00)80219-6).
- Bryk, A.S., Raudenbush, S.W., 1987. Application of hierarchical linear models to assessing change. *Psychol. Bull.* 101.
- Card, N.A., Little, T.D., 2007. Longitudinal modeling of developmental processes. *Int. J. Behav. Dev.* 31 (4), 297–302. <https://doi.org/10.1177/0165025407077750>.
- Collins, L.M., Schafer, J.L., Kam, C.-M., 2001. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol. Methods* 6 (4), 330–351. <https://doi.org/10.1037/1082-989x.6.4.330>.
- Cook, N.R., Ware, J.H., 1983. Design and analysis methods for longitudinal research. *Ann. Rev. Public Health* 4, 1–23. <https://doi.org/10.1146/annurev.pu.04.050183.000245>.
- Cudeck, R., Harring, J.R., 2007. Analysis of nonlinear patterns of change with random coefficient models. *Ann. Rev. Psychol.* 58, 615–637. <https://doi.org/10.1146/annurev.psych.58.110405.085520>.
- Cudeck, R., Klebe, K.J., 2002. Multiphase mixed-effects models for repeated measures data. *Psychol. Methods* 7 (1), 41–63. <https://doi.org/10.1037/1082-989X.7.1.41>.
- Curran, P.J., 2003. Have Multilevel Models Been Structural Equation Models All Along? *Multivariate Behav. Res.* 38 (4), 529–569. <https://doi.org/10.1207/s15327906mbr3804.5>.
- Curran, P.J., Bauer, D.J., 2011. The disaggregation of within-person and between-person effects in longitudinal models of change. *Ann. Rev. Psychol.* 62 (1), 583–619. <https://doi.org/10.1146/annurev.psych.093008.100356>.
- Curran, P.J., Obeidat, K., Losardo, D., 2010. Twelve frequently asked questions about growth curve modeling. *J. Cogn. Dev.* 11 (2), 121–136. <https://doi.org/10.1080/15248371003699969>.
- Curran, P.J., Lee, T., Howard, A.L., Lane, S., MacCallum, R., 2012. Disaggregating within-person and between-person effects in multilevel and structural equation growth models. - *PsycNET*. In: Hancock, G.R., Harring, J.R. (Eds.), *CILVR Series on Latent Variable Methodology. Advances in Longitudinal Methods in the Social and Behavioral Sciences*. IAP Information Age Publishing, pp. 217–253. Retrieved from <https://psycnet.apa.org/record/2012-31006-009>.
- Cyranowski, J.M., Frank, E., Young, E., Shear, M.K., 2000. Adolescent onset of the gender difference in lifetime rates of major depression. A theoretical model. *Arch. Gen. Psychiatry* 57 (1), 21–27. <https://doi.org/10.1001/archpsyc.57.1.21>.
- Denny, B.T., Fan, J., Liu, X., Guerrerri, S., Mayson, S.J., Rinsky, L., et al., 2014. Insula-amygdala functional connectivity is correlated with habituation to repeated negative images. *Soc. Cogn. Affect. Neurosci.* 9, 1660–1667. <https://doi.org/10.1093/scan/nst160>.
- Duncan, S.C., Duncan, T.E., Hops, H., 1996. Analysis of longitudinal data within accelerated longitudinal designs. *Psychol. Methods* 1 (3), 236–248. <https://doi.org/10.1037/1082-989X.1.3.236>.
- Ellwanger, J., Geyer, M.A., Braff, D.L., 2003. The relationship of age to prepulse inhibition and habituation of the acoustic startle response. *Biol. Psychol.* 62 (3), 175–195. [https://doi.org/10.1016/S0301-0511\(02\)00126-6](https://doi.org/10.1016/S0301-0511(02)00126-6).
- Elwert, F., Winship, C., 2014. Endogenous selection bias: the problem of conditioning on a collider variable. *Annu. Rev. Sociol.* 40 (1), 31–53. <https://doi.org/10.1146/annurev-soc-071913-043455>.
- Enders, C.K., 2011. Missing not at random models for latent growth curve analyses. *Psychol. Methods* 16 (1), 1–16. <https://doi.org/10.1037/a0022640>.
- Fang, H., Brooks, G.P., Rizzo, M.L., Espy, K.A., Barcikowski, R.S., 2008. A Monte Carlo power analysis of traditional repeated measures and hierarchical multivariate linear models in longitudinal data analysis. *J. Mod. Appl. Stat. Methods* 7 (1). <https://doi.org/10.22237/jmasm/1209614880>.
- Ferrer, E., Salthouse, T.A., Stewart, W.F., Schwartz, B.S., 2004. Modeling age and retest processes in longitudinal studies of cognitive abilities. *Psychol. Aging* 19 (2), 243–259. <https://doi.org/10.1037/0882-7974.19.2.243>.
- Frees, E.W., 2001. Omitted variables in longitudinal data models. *Can. J. Stat. / La Rev. Can. Stat.* 29 (4), 573–595. <https://doi.org/10.2307/3316008>.
- Galbraith, S., Bowden, J., Mander, A., 2017. Accelerated longitudinal designs: an overview of modelling, power, costs and handling missing data. *Stat. Methods Med. Res.* 26 (1), 374–398. <https://doi.org/10.1177/0962280214547150>.
- Garber, J., Keiley, M.K., Martin, N.C., 2002. Developmental trajectories of adolescents' depressive symptoms: predictors of change. *J. Consult. Clin. Psychol.* 70 (1), 79–95. <https://doi.org/10.1037/0022-006X.70.1.79>.
- Gottfredson, N.C., Bauer, D.J., Baldwin, S.A., Okiishi, J.C., 2014. Using a shared parameter mixture model to estimate change during treatment when termination is related to recovery speed. *J. Consult. Clin. Psychol.* 82 (5), 813–827. <https://doi.org/10.1037/a0034831>.
- Graham, J.W., 2009. Missing data analysis: making it work in the real world. *Annu. Rev. Psychol.* 60 (1), 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>.
- Herting, M.M., Johnson, C., Mills, K.L., Vijayakumar, N., Dennison, M., Liu, C., et al., 2018. Development of subcortical volumes across adolescence in males and females: a multisample study of longitudinal changes. *NeuroImage* 172, 194–205. <https://doi.org/10.1016/j.neuroimage.2018.01.020>.
- Kessler, R.C., Greenberg, D.F., 1981. *Linear Panel Analysis: Models of Quantitative Change*. Academic Press, New York, NY. <https://doi.org/10.1016/C2009-0-21992-1>.
- Kim, J.-S., Frees, E.W., 2006. Omitted variables in multilevel models. *Psychometrika* 71 (4), 659–690. <https://doi.org/10.1007/s11336-005-1283-0>.
- Kraemer, H.C., Yesavage, J.A., Taylor, J.L., Kupfer, D., 2000. How can we learn about developmental processes from cross-sectional studies, or can we? *Am. J. Psychiatry* 157 (2), 163–171. <https://doi.org/10.1176/appi.ajp.157.2.163>.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82 (13) <https://doi.org/10.18637/jss.v082.i13>.
- Little, T.D., Rhemtulla, M., 2013. Planned missing data designs for developmental researchers. *Child Dev. Perspect.* 7 (4), 199–204. <https://doi.org/10.1111/cdep.12043>.
- Louis, T.A., Robins, J., Dockery, D.W., Spiro, A., Ware, J.H., 1986. Explaining discrepancies between longitudinal and cross-sectional models. *J. Chronic Dis.* 39, 831–839.
- Luna, B., Padmanabhan, A., O'Hearn, K., 2010. What has fMRI told us about the Development of Cognitive Control through Adolescence? *Brain Cogn.* 72 (1), 101–113. <https://doi.org/10.1016/j.bandc.2009.08.005>.
- MacCallum, R.C., Kim, C., Malarkey, W.B., Kiecolt-Glaser, J.K., 1997. Studying multivariate change using multilevel models and latent curve models. *Multivariate Behav. Res.* 32 (3), 215–253. <https://doi.org/10.1207/s15327906mbr3203.1>.
- MacCallum, R.C., Kim, C., Malarkey, W.B., Kiecolt-Glaser, J.K., 2010. Studying multivariate change using multilevel models and latent curve models. *Multivariate Behav. Res.* 32 (3), 215–253. <https://doi.org/10.1207/s15327906mbr3203.1>.
- Marshall, W.A., Tanner, J.M., 1969. Variations in pattern of pubertal changes in girls. *Arch. Dis. Child.* 44, 291–303.
- Marshall, W.A., Tanner, J.M., 1970. Variations in the pattern of pubertal changes in boys. *Arch. Dis. Child.* 45 (239), 13–23. <https://doi.org/10.1136/adc.45.239.13>.
- Marti-Henneberg, C., Vizmanos, B., 1997. The duration of puberty in girls is related to the timing of its onset. *J. Pediatr.* 131 (4), 618–621. [https://doi.org/10.1016/S0022-3476\(97\)70073-8](https://doi.org/10.1016/S0022-3476(97)70073-8).
- Maxwell, S.E., Cole, D.A., 2007. Bias in cross-sectional analyses of longitudinal mediation. *Psychol. Methods* 12 (1), 23–44. <https://doi.org/10.1037/1082-989X.12.1.23>.
- McArdle, J.J., Woodcock, R.W., Aggen, S., Boker, S., Hamagami, A., Horn, J., et al., 1997. Expanding test-retest designs to include developmental time-lag components. *Psychol. Methods* 2 (4), 403–435. <https://doi.org/10.1037/1082-989X.2.4.403>.
- McCormick, E.M., Peters, S., Crone, E.A., Telzer, E.H., 2021. Longitudinal network re-organization across learning and development. *NeuroImage* 229, 117784. <https://doi.org/10.1016/j.neuroimage.2021.117784>.
- McNeish, D., Matta, T.H., 2020. Flexible treatment of time-varying covariates with time unstructured data. *Struct. Equ. Model.* 27 (2), 298–317. <https://doi.org/10.1080/10705511.2019.1627213>.



- Mehta, P.D., Neale, M.C., 2005. People are variables too: Multilevel structural equations modeling. *Psychol. Methods* 10 (3), 259–284. <https://doi.org/10.1037/1082-989X.10.3.259>.
- Palmore, E., 1978. When can age, period, and cohort be separated? *Soc. Forces* 57 (1), 282–295. <https://doi.org/10.2307/2577639>.
- Rabbitt, P., Diggle, P., Smith, D., Holland, F., Mc Innes, L., 2001. Identifying and separating the effects of practice and of cognitive ageing during a large longitudinal study of elderly community residents. *Neuropsychologia* 39 (5), 532–543. [https://doi.org/10.1016/S0028-3932\(00\)00099-3](https://doi.org/10.1016/S0028-3932(00)00099-3).
- Raudenbush, S.W., Bryk, A.S., 2002. *Hierarchical Linear Models. Applications and Data Analysis Methods*, 2nd ed. Sage Publications, Thousand Oaks, CA.
- Rhemtulla, M., Little, T.D., 2012. Planned missing data designs for research in cognitive development. *J. Cogn. Dev.* 13 (4), 425–438. <https://doi.org/10.1080/15248372.2012.717340>.
- Salthouse, T.A., 2014. Why are there different age relations in cross-sectional and longitudinal comparisons of cognitive functioning? *Curr. Dir. Psychol. Sci.* 23 (4), 252–256. <https://doi.org/10.1177/0963721414535212>.
- Schafer, J.L., Graham, J.W., 2002. Missing data: our view of the state of the art. *Psychol. Methods* 7 (2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>.
- Shieh, Y.-Y., Fouladi, R.T., 1991. The effect of multicollinearity on multilevel modeling parameter estimates and standard errors. *Educ. Psychol. Meas.* 63 (6), 951–985. <https://doi.org/10.1177/0013164403258402>.
- Sliwinski, M., Hoffman, L., Hofer, S.M., 2010. Evaluating convergence of within-person change and between-person age differences in age-heterogeneous longitudinal studies. *Res. Hum. Dev.* 7 (1), 45–60. <https://doi.org/10.1080/15427600903578169>.
- Soden, B., Christopher, M.E., Hulslander, J., Olson, R.K., Cutting, L., Keenan, J.M., et al., 2015. Longitudinal stability in reading comprehension is largely heritable from grades 1 to 6. *PLoS One* 10 (1). <https://doi.org/10.1371/journal.pone.0113807>.
- Somerville, L.H., Casey, B.J., 2010. Developmental neurobiology of cognitive control and motivational systems. *Curr. Opin. Neurobiol.* 20 (2), 236–241. <https://doi.org/10.1016/j.conb.2010.01.006>.
- Sterba, S.K., 2014. Fitting nonlinear latent growth curve models with individually varying time points. *Struct. Equ. Model. A Multidiscip. J.* 21, 630–647. <https://doi.org/10.1080/10705511.2014.919828>.
- Taylor, A.B., West, S.G., Aiken, L.S., 2006. Loss of power in logistic, ordinal logistic, and probit regression when an outcome variable is coarsely categorized. *Educ. Psychol. Meas.* 66 (2), 228–239. <https://doi.org/10.1177/0013164405278580>.
- Telzer, E.H., McCormick, E.M., Peters, S., Cosme, D., Pfeifer, J.H., van Duijvenvoorde, A. C.K., 2018. Methodological considerations for developmental longitudinal fMRI research. *Dev. Cogn. Neurosci.* 33, 149–160. <https://doi.org/10.1016/J.DCN.2018.02.004>.
- van Duijvenvoorde, A.C.K., Westhoff, B., de Vos, F., Wierenga, L.M., Crone, E.A., 2019. A three-wave longitudinal study of subcortical–cortical resting-state connectivity in adolescence: testing age- and puberty-related changes. *Hum. Brain Mapp.* 40 (13), 3769–3783. <https://doi.org/10.1002/hbm.24630>.
- Van't Hof, M.A., Roede, M.J., Kowalski, C.J., 1977. A mixed longitudinal data analysis model. *Hum. Biol.* 49 (2), 165–179. Retrieved from. <https://www.jstor.org/stable/41465091>.
- Wierenga, L.M., Bos, M.G.N., Schreuders, E., vd Kamp, F., Peper, J.S., Tamnes, C.K., Crone, E.A., 2018. Unraveling age, puberty and testosterone effects on subcortical brain development across adolescence. *Psychoneuroendocrinology* 91, 105–114. <https://doi.org/10.1016/j.psyneuen.2018.02.034>.