

**Longitudinal Analysis Manuscript: Working Draft:  
Longitudinal Analysis using data from the ABCD Study**

Samuel Hawes<sup>1</sup>, Andrew K. Littlefield<sup>2</sup>, Daniel A. Lopez<sup>3</sup>, Kenneth J. Sher<sup>4</sup>, Erin L. Thompson<sup>5</sup>, Raul Gonzalez<sup>5</sup>, Additional Co-authors<sup>6</sup>, and Wesley K. Thompson<sup>7</sup>

<sup>1</sup>Florida International University

<sup>2</sup>Texas Tech University

<sup>3</sup>University of Rochester Medical Center

<sup>4</sup>University of Missouri

<sup>5</sup>Florida International University

<sup>6</sup>xxxxxx University

<sup>7</sup>Laureate Institute for Brain Research

**Author note**

Samuel Hawes, Associate Research Professor, Center for Children & Families,  
Florida International University

Andrew K. Littlefield

Daniel A. Lopez

Kenneth J. Sher

Erin L. Thompson

Raul Gonzalez

Additional Co-authors

Wesley K. Thompson

**Correspondence**

Correspondence concerning this article should be addressed to Samuel Hawes,  
Center for Children & Families, Florida International University, 11200 SW 8th St,  
AHC-IV, Miami, FL 33199, USA

### Abstract

The Adolescent Brain Cognitive Development (ABCD) Study presents a unique opportunity for researchers to investigate developmental processes in a large, diverse cohort of children and adolescents. Given the complex nature of the longitudinal data collected in the ABCD Study, researchers are likely to encounter a myriad of methodological and analytic considerations and concerns. This review provides a comprehensive examination of key issues and techniques related to longitudinal data analysis, specifically focusing on the ABCD Study. The text discusses model assumptions, common violations (e.g., independent and identically distributed residuals, *heterogeneous variability*) and their implications for valid inference. The importance of appropriately modeling covariance structures, understanding trade-offs between model fit and parsimony, and challenges related to sample size, attrition, missing data are highlighted. Consideration is given to the importance of selecting appropriate statistical models to account for correlations in repeated measurements and the assumptions underlying these models. The review also differentiates between linear and non-linear models in the context of continuous and discrete data, emphasizing various distributional assumptions and the necessity of choosing appropriate models and statistical methods. By addressing these complexities, the review seeks to equip researchers with the necessary knowledge and tools to make informed decisions as they navigate effectively analyzing and interpreting data available in the ABCD Study.

**Keywords**—ABCD Study, longitudinal analyses, development

**Longitudinal Analysis Manuscript: Working Draft:  
Longitudinal Analysis using data from the ABCD Study**

## 1 Introduction

The Adolescent Brain Cognitive Development (ABCD) Study® is the largest long-term investigation of neurodevelopment and child health in the United States. Conceived and initiated by the National Institutes of Health (NIH), this landmark prospective longitudinal study aims to transform our understanding of the genetic and environmental influences on brain development and their roles in behavioral and health outcomes in adolescents (Volkow et al., 2018). At its heart, the study is designed to chart the course of human development across multiple, interacting domains from late childhood to early adulthood and to identify factors that lead to both positive and negative developmental outcomes. Central to achieving these goals is the ABCD Study’s® commitment to an open science framework designed to facilitate access to and sharing of scientific knowledge by espousing practices that increase openness, integrity, and reproducibility of scientific research (e.g., public data releases). In this sense, the ABCD Study® is a collaboration with the larger research community, with the rich longitudinal nature of the ABCD Study dataset allowing researchers to perform a variety of analyses of both methodological and substantive interest. Together, this presents a unique opportunity to significantly advance our understanding of how a multitude of biopsychosocial processes emerge and unfold across critical periods of development. In the following text, we describe methods for longitudinal analyses of ABCD Study data that can address its fundamental scientific aims, as well as challenges inherent in a large population-based long-

term study of adolescents. We begin with a brief orientation to the ABCD Study and its accompanying [publicly available dataset](#).

### 1.1 The ABCD Study® Data

Participants enrolled in the ABCD Study include a large cohort of youth (n=11880) aged 9-10 years at baseline and their parents/guardians. The study sample was recruited from household populations in defined catchment areas for each of the 21 study sites across the United States (information regarding funding agencies, recruitment sites, investigators, and project organization can be obtained at the ABCD Study website). The ABCD Study is collecting longitudinal data on a rich variety of outcomes that will enable the construction of realistically-complex etiological models by incorporating factors from many domains simultaneously. Each new wave of data collection provides the building blocks for conducting probing longitudinal analyses that allow us to characterize normative development, identify variables that presage deviations from prototypic development, and assess a range of outcomes associated with variables of interest. This data includes a neurocognitive battery (Luciana et al., 2018; Thompson et al., 2019), mental and physical health assessments (Barch et al., 2018), measures of culture and environment (Zucker et al., 2018), substance use [add citation], biospecimens (Uban et al., 2018), structural and functional brain imaging (Casey et al., 2018; Hagler et al., 2019), geolocation-based environmental exposure data, wearables, and mobile technology (Bagot et al., 2018), and whole genome genotyping (Loughnan et al., 2020). Many of these measures are collected at in-person annual visits, with brain imaging collected at baseline and every other year going forward. A limited number of assessments are collected in semi-annual telephone interviews between in-person visits. Data are publicly

released on an annual basis through the NIMH Data Archive. By necessity, the study's earliest data releases were cross-sectional (i.e., the baseline data), however, the most recent public data release (NDA Release 4.0) contains data collected across three annual assessments, including two imaging assessments (baseline and year 2 follow-up visits).

## 1.2

### Note

This manuscript will go through the following sections :xxxxxx, xxxx, xxxxx. # Part I: Longitudinal Research # Part II: Longitudinal Data # Part III: Longitudinal Analysis Overview of **concepts and considerations specific to developmentally-focused longitudinal research**

## 1.3 Organization and Aims

Aims/Goals: Use this xxxxx during any xxxxx and you're virtually guaranteed to xxxxx

- Concepts and Considerations when working with large and multimodal longitudinal data
- Concepts and considerations of developmental research (longitudinal research via developmental perspective)
- Learn specific methods
- Use our open-source resources to get started more quickly
- Learn How to Apply xxxxx to large longitudinal data such as the ABCD Study

identify common xxxxx of longitudinal research

1.3 Objectives and scope The focus of this xxxx is on the analysis of event history data using xxxx model with the objectives • to demonstrate the use of event history analysis in the analysis of xxxxx, xxxx and to identify the methodological problems in the analysis, • to propose procedures to choose a basic time scale, • to discuss the connections between the methods for modeling dual time scales and to perform quantitative comparisons between them, • to compare existing methods to deal with longitudinal measurements in the xxxx model with two proposed methods.

## 2 Part I: Longitudinal Research

### 2.1 Basic Concepts and Considerations

There are several important concepts to consider when conducting longitudinal analyses in a developmental context. These include different ways of thinking about developmental course, whether certain periods of development are relatively sensitive or insensitive to various types of insults or stressors, whether some time periods or situations inhibit the expression of individual differences due to extreme environmental pressures, and whether the same behavior manifested at different times represent the same phenomenon or different ones. Further, in the case of developmentally focused longitudinal research, each new measurement occasion not only provides a more extended portrait of the child's life course (and not just characterize growth during this period but also assesses the durability/chronicity of prior effects/consequences) but also brings with it greater methodological opportunities to exploit the statistical properties of longitudinal data in the furtherance of critical scientific questions. That is, we can ask more nuanced questions and make stronger inferences

as our number of time-ordered observations grows, assuming we have assessed the “right” variables and the timings of our observations comport with the temporal dynamics of the mechanisms of interest. Appreciation of these and other issues can help to guide the analysis and interpretation of data and aid translation to clinical and public health applications.

**Vulnerable periods.** Development normatively progresses from less mature to more mature levels of functioning. However, unique epochs and experiences can alter the course of this idealized form of development. Consider research that shows cannabis use during adolescence is associated with later psychosis to a greater degree than cannabis use initiated later in development [add citation]; or, similarly, experimental research on rodents that shows rodent brains to be especially sensitive to the neurotoxic effects of alcohol on brain structure and learning early in development (corresponding to early adolescence in humans)[add citation]. These examples highlight the importance of considering the role of vulnerable periods – temporal windows of rapid brain development or remodeling during which the effects of environmental stimuli (e.g. cannabis exposure) on the developing brain may be particularly pronounced– when trying to establish an accurate understanding of the association between exposures and outcomes.

**Developmental disturbances.** Whereas vulnerable periods heighten neurobiological susceptibility to environmental influences, at other times environmental pressures will tend to suppress stability and disrupt the orderly stochastic process of normative development (e.g., xxx-xxx). This situation reflects a developmental disturbance in that the normal course of development is “disturbed” for some time by some time-limited process. In such cases, we might find that prediction of behavior in the period of the disturbance is reduced and/or, similarly, the behavior exhibited during the disturbance might have less predictive

power with respect to distal outcomes compared to the behavior exhibited before and following the disrupted period. That is, once the environmental stimuli are removed (or the individual is removed from the environment), individual differences are again more freely expressed and the autoregressive effects increase to levels similar to those before entering the environment.

**Developmental snares and cascade effects.** Normative development can also be upended by experiences (e.g., drug use) that, through various mechanisms, disrupt the normal flow of development wherein each stage establishes a platform for the next. For instance, substance use could lead to association with deviant peers, precluding opportunities for learning various adaptive skills and prosocial behaviors, in effect, creating a “snare” that retards psychosocial development. Relatedly, the consequences of these types of events can cascade (e.g., school dropout, involvement in the criminal justice system) so that the effects of the snare are amplified. Although conceptually distinct from vulnerable periods, both of these types of developmental considerations highlight the importance of viewing behavior in the context of development and the importance of attempting to determine how various developmental pathways unfold.

**Distinguishing developmental change from experience effects.** One can often observe systematic changes over time in a variable of interest and assume this change is attributable to development. To this point, cognitive abilities (e.g, verbal ability, problem-solving) normatively grow earlier in development and often decline in late life (e.g., memory, speed of processing). However, the observed patterns of growth and decline often differ between cross-sectional vs. longitudinal effects (Salthouse, [2014](#)) where subjects gain increasing experience with the assessment with each successive measurement occasion. Such experience



effects on cognitive functioning have been demonstrated in adolescent longitudinal samples similar to ABCD (Sullivan et al., 2017) and highlight the need to consider these effects and address them analytically. In the case of performance-based measures [e.g., matrix reasoning related to neurocognitive functioning; see Salthouse (2014)], this can be due to “learning” the task from previous test administrations (e.g., someone taking the test a second time performs better than they did the first time simply as a function of having taken it before). Even in the case of non-performance-based measures (e.g., levels of depression), where one cannot easily make the argument that one has acquired some task-specific skill through learning, it has been observed that respondents tend to endorse lower levels on subsequent assessments (e.g., Beck et al., 1961; see French & Sutton, 2010) and this phenomenon has been well documented in research on structured diagnostic interviews (Robins, 1985). While it is typically assumed that individuals are rescinding or telling us less information on follow-up interviews, there is reason to suspect that in some cases the initial assessment may be artefactually elevated (see Shrout et al., 2018). Some designs (specifically, accelerated longitudinal designs) are especially well suited for discovering these effects and modeling them. While ABCD was not designed as an accelerated longitudinal design, the variability in age at the time of baseline recruitment (9 years, 0 months to 10 years, 11 months) allows some measures, collected every year, to be conceptualized as an accelerated longitudinal design. Moreover, it is possible that in later waves, patterns of longitudinal missing data will allow some analyses to assess the confounded effects of age and the number of prior assessments. However, ABCD is fundamentally a single-cohort, longitudinal design, where a number of prior assessments and age are highly confounded, and for, perhaps, most analyses, the possible influence of experience effects needs to be kept in mind.

### 3 Part II: Longitudinal Data

#### 3.1 Interpretation / Issues / Pitfalls & Assumption

**Defining Features of Longitudinal Data Analysis.** The hallmark characteristic of longitudinal data analysis is its application to repeated assessments of the same assessment targets (e.g., individuals, families) across time. While the primary reason for collecting longitudinal data is in pursuit of addressing scientific questions, from a methodological perspective, having multiple observations over time allows researchers to identify potentially problematic observations when highly improbable longitudinal patterns are observed. That is, we can ask more nuanced questions and make stronger inferences as our number of time-ordered observations grows assuming we have assessed the “right” variables and the timings of our observations comport with the temporal dynamics of the mechanisms of interest .

[“traditional approaches”: marginal models] [random effects: Growth Curves- SEM (lgcm) vs ME (lmm)] [(between- and within-person levels?)]

##### ***3.1.1 Modeling Data Across Two Time Points versus Three or More Time Points.***

Although the clear leap to the realm of longitudinal data involves going from one assessment to two or more assessments, there are also notable distinctions in designs based on two-assessment points versus three or more measurement occasions. Just as cross-sectional data can be informative in some situations, two waves of data can be beneficial in contexts such as when experimental manipulation is involved (e.g., pre/post tests), or if the central goal is prediction (e.g., trying to predict scores on Variable A at time T as a function of prior scores on Variable A and Variable B at time T-1). At the same time, data based

on two assessments are inherently limited on multiple fronts. As (Rogosa et al., 1982) noted approximately forty years ago, “Two waves of data are better than one, but maybe not much better”. These sentiments are reflected in more contemporary recommendations regarding best-practice guidelines for prospective data, which increasingly emphasize the benefits of additional measurement occasions for model identification and accurate parameter estimation. It is also consistent with research recommending that developmental studies include three or more assessment points, given it is impossible for data based on two-time points to determine the shape of development (since linear, straight line change is the only possible form, given two assessments; see (Duncan & Duncan, 2009)). Research designs that include three or more time points allow for increasingly nuanced analyses that more adequately tease apart sources of variation and covariation among the repeated assessments (King et al., 2018)—a key aspect of inferential research. To illustrate, developmental theories are typically interested in understanding patterns of within-individual change over time (discussed in further detail, below); however, two data points provide meager information on change at the person level. This point is further underscored in a recent review of statistical models commonly touted as distinguishing within-individual vs between-individual sources of variance in which the study authors concluded “... researchers are limited when attempting to differentiate these sources of variation in psychological phenomenon when using two waves of data” and perhaps more concerning, “...the models discussed here do not offer a feasible way to overcome these inherent limitations” Littlefield et al. (2021). It is important to note, however, that despite the current focus on two-wave designs versus three or more assessment waves, garnering three assessment points is not a panacea for longitudinal modeling. Indeed, several contemporary longitudinal models designed to isolate within-individual variability

[e.g., the Latent Curve Model with Structured Residuals; Curran et al. (2014)] require at least four assessments to parameterize fully and, more generally, increasingly accurate parameter estimates are obtained as more assessment occasions are used (Duncan & Duncan, 2009).


### ***3.1.2 Types of stability and change***

If one were to try to sum up what development in a living organism is exactly, one could plausibly argue it's the characterization of stability and change as the organism traverses the life course. There are a few different ways to think of stability (and change). Consider we measure the height of all youth in a 6th-grade class, once in the fall at the beginning of the school year and once again in the spring at the end of the school year. A common first step may be to compare the class's average height values obtained at these two different measurement occasions. This comparison of the average scores for the same group of individuals at multiple time points is referred to as "mean-level" stability as it provides information about continuity and change in the group level of an outcome of interest (e.g., height) over time. Another type of stability involves calculating the correlation between the values obtained at different time points (e.g., 'height in the fall' with 'height in the spring'). This type of "rank-order" stability evaluates between-individual change by focusing on the degree to which individuals retain their relative placement in a group across time. Consider, someone who is the shortest person in their class in 6th grade may grow considerably over the school year (i.e., exhibit mean level change), but remain the shortest person among their classmates. That is, the individual is manifesting a type of rank-order stability. Both types of stability and change are important. Mean-level change in certain traits might help to explain why, in general, individuals are particularly vulnerable to social influences at some ages more than others; rank order change might help to quantify the extent to which certain

characteristics of the individual are more trait-like. For example, in some areas of development, there is considerable mean-level change that occurs over time (e.g., changes in Big 5 personality traits), but relatively high rank-order stability. Despite the useful information afforded by examining mean-level and rank-order change, these approaches are limited in that they provide little information about patterns of “within-individual” change and, in turn, can result in fundamental misinterpretations about substantial or meaningful changes in an outcome of interest.

There is growing recognition that statistical models commonly applied to longitudinal data often fail to comport with the developmental theory they are being used to assess (e.g., Curran, Lee, Howard, Lane, & MacCallum, 2012; Hoffman, 2015; Littlefield et al., 2021). Specifically, developmental studies typically involve the use of prospective data to inform theories that are concerned with clear within-person (i.e., intraindividual) processes (e.g., how phenotypes change or remain stable within individuals over time) (e.g., see Curran & Bauer, 2011). Despite this, methods generally unsuited for disaggregating between- and within-person effects (e.g., cross-lagged panel models [CLPM]) remain common within various extant literatures. As a result, experts increasingly caution about the need to xxxxxxxx [add citation]. Fortunately, there exists a range of models that have been proposed to tease apart between- and within-person sources of variance across time (see Littlefield et al., 2021; Orth et al., 2021). Most of these contemporary alternatives incorporate time-specific latent variables to capture between-person sources of variance and model within-person deviations around an individual’s mean (or trait) level across time (e.g., RI-CLPM, Hamaker et al., 2015; LCM-SR, Curran et al., 2014). It is important to note however that these models require multiple assessments waves (e.g., four or more to fully specify the LCM-SR), addi-

tional expertise to overcome issues with model convergence, and appreciation of modeling assumptions when attempting to adjudicate among potential models in each research context (see Littlefield et al., 2021, for further discussion).

This is illustrated well by .

### 3.1.3 *Model Assumptions*

Many statistical models assume certain characteristics about the data to which they are being applied. As an example, common assumptions of parametric statistical models include normality, linearity, and equality of variances. These assumptions must be carefully considered before conducting analysis so that valid inferences can be made from the data; that is, violation of a model's assumptions can substantively alter the interpretation of results. Similarly, statistical models employed in the analyses of longitudinal data often entail a range of assumptions that must be closely inspected. One central issue for repeated measurements on an individual is how to account for the correlated nature of the data; another common feature of longitudinal data is heterogeneous variability; that is, the variance of the response changes over the duration of the study. Traditional techniques, such as a standard regression or ANOVA model, assume residuals are independent and thus are inappropriate for designs that assess (for example) the same individuals across time. That is, given the residuals are no longer independent, the standard errors from the models are biased and can produce misleading inferential results. Although there are formal tests of independence for time series data (e.g., the Durbin-Watson statistic (Durbin & Watson, 1950)), more commonly independence is assumed to be violated in study designs with repeated assessments. Therefore, an initial question to be addressed by a researcher analyzing prospective data is how to best model the covariance structure of said data.

### ***3.1.4 Covariance Structures***

Statistical models for longitudinal data include two main components to account for assumptions that are commonly violated when working with repeated measures data: a model for the covariance among repeated measures (both the correlations among pairs of repeated measures on an individual and the variability of the responses on different occasions), coupled with a model for the mean response and its dependence on covariates (eg, treatment group in the context of clinical trials). This allows for the specification of a range of so-called covariance structures, each with its own set of tradeoffs between model fit and parsimony (e.g., see Kincaid, 2005).

### ***3.1.5 Accounting for Correlated Data***

As an example, one alternative structure that attempts to handle the reality that correlations between repeated assessments tend to diminish across time is the autoregressive design. As the name implies, the structure assumes a subsequent measurement occasion (e.g., assessment at Wave 2) is regressed onto (that is, is predicted by) a prior measurement occasion (e.g., assessment at Wave 1). The most common type of autoregressive design is the AR(1), where assessments at time  $T + 1$  are regressed on assessments at Time  $T$ . Identical to compound symmetry, this model assumes the variances are homogenous across time. Diverting from compound symmetry, this model assumes the correlations between repeated assessments decline exponentially across time rather than remaining constant. For example, per the AR(1) structure, if the correlation between Time 1 and Time 2 data is thought to be .5, then the correlation between Time 1 and Time 3 data would be assumed to be  $.5 \cdot .5 = .25$ , and the correlation between Time 1 and Time 4 data would be assumed to be  $.5 \cdot .5 \cdot .5 = .125$ .

.125. As with compound symmetry, the basic AR(1) model is parsimonious in that it only requires two parameters (the variance of the assessments and the autoregressive coefficient). Notably, the assumption of constant autoregressive relations between assessments is often relaxed in commonly employed designs that use autoregressive modeling (e.g., cross-lagged panel models [CLPM]). These designs still typically assume an AR(1) process (e.g., it is sufficient to regress the Time 3 assessment onto the Time 2 assessment and is not necessary to also regress the Time 3 assessment onto the Time 1 assessment, which would result in an AR(2) process). However, the magnitude of these relations is often allowed to differ across different AR(1) pairs of assessment (e.g., the relation between Time 1 and Time 2 can be different from the relation between Time 2 and Time 3). These more commonly employed models also often relax the assumption of equal variances of the repeated assessments. Although the AR(1) structure may involve a more realistic set of assumptions compared to compound symmetry, in that the AR(1) model allows for diminishing correlations across time, the basic AR(1) model, as well as autoregressive models more generally, can also suffer from several limitations in contexts that are common in prospective designs. In particular, recent work demonstrates that if a construct being assessed prospectively across time is trait-like in nature, then autoregressive relations fail to adequately account for this trait-like structure, with the downstream consequence that estimates derived from models based on AR structures (such as the CLPM) can be misleading and fail to adequately demarcate between- vs. within-person sources of variance (Hamaker et al., 2015).

### ***3.1.6 Linear vs non-linear models***

Identification of optimal statistical models and appropriate mathematical functions requires an understanding of the type of data being used. Repeated assessments can be



based on either continuous or discrete measures. Examples of discrete measures include repeated assessments of binary variables (e.g., past 12-month alcohol use disorder status measured across ten years), ordinal variables (e.g., a single item measuring the level of agreement to a statement on a three-point scale including the categories of “disagree”, “neutral”, and “agree” in an ecological momentary assessment study that involves multiple daily assessments), and count variables (e.g., number of cigarettes smoked per day across a daily diary study). In many ways, the distributional assumptions of indicators used in longitudinal designs mirror the decision points and considerations when delineating across different types of discrete outcome variables, a topic that spans entire textbooks (e.g., see Lenz, 2016). For example, the Mplus manual (Muthén, 2017) includes examples of a) censored and censored-inflated models, b) linear growth models for binary or ordinal variables, c) linear growth models for a count outcome assuming a Poisson model, d) linear growth models for a count outcome assuming a zero-inflated Poisson model and e) discrete- and continuous-time survival analysis for a binary outcome. Beyond these highlighted examples, other distributions (e.g., negative binomial) can be assumed for the indicators when modeling longitudinal data. These models can account for issues that can occur when working with discrete outcomes, including overdispersion (when the variance is higher than would be expected based on a given distribution) and zero-inflation [when more zeros occur than is expected based on a given distribution; see Lenz (2016)]. Models involving zero-inflation parameters are referred to as two-part models, given one part of the model predicts the zero-inflation whereas the other part of the model predicts outcomes consistent with a given distribution [e.g., Poisson distribution; see Farewell et al. (2017), for a review of two-part models for longitudinal data]. Although there exist several alternative models for discrete indicators, some more

recent models that have been proposed for prospective data are only feasible in cases where indicators are assumed to be continuous rather than discrete [e.g., LCM-SR; Curran et al. (2014)]. Given the sheer breadth of issues relevant to determining better models for discrete outcomes, it is not uncommon for texts on longitudinal data analysis to only cover models and approaches that assume continuous indicators (e.g., T. D. Little, 2013). However, some textbooks on categorical data analysis provide more detailed coverage of the myriad issues and modeling choices to consider when working with discrete outcomes [e.g., Lenz (2016), Chapter 11 for matched pair/two-assessment designs; Chapter 12 for marginal and transitional models for repeated designs, such as generalized estimating equations, and Chapter 13 for random effects models for discrete outcomes].

### ***3.1.7 Marginal vs Conditional Models (?)***

[section still under development...]

### ***3.1.8 Missing Data/Attrition***

As recently reviewed by Littlefield (in press), investigators of prospective data are confronted with study attrition (i.e., participants may not provide data at a given wave of assessment) and thus approaches are needed to confront the issue of missing data. Three models of missingness are typically considered in the literature (see R. J. Little & Rubin, 1989). These three models are data: a) missing completely at random (MCAR), b) missing at random (MAR), and c) missing not at random (MNAR). Data that are MCAR means missing data is a random sample of all the types of participants (e.g., males) in a given dataset. MAR suggests conditionally missing at random (see Graham, 2009). That is, MAR implies missingness is completely random (i.e., does not hinge on some unmeasured variables) once missingness has been adjusted by all available variables in a dataset (e.g., biological

sex). Data that are MNAR are missing as a function of unobserved variables. Graham (2009) provides an excellent and easy-to-digest overview of further details involving missing data considerations.

Multiple approaches have been posited to handle missing data. Before the advent of more contemporary approaches, common methods included several ad hoc procedures. These include eliminating the data of participants with missing data (e.g., listwise or pairwise deletion) or using mean imputation (i.e., replacing the missing value with the mean score of the sample that did participate). However, these methods are not recommended because they can contribute to biased parameter estimates and research conclusions (see Graham, 2009). More specifically, the last observation carried forward (LOCF) is a common approach to imputing missing data. LOCF replaces a participant's missing values after dropout with the last available measurement (Molnar et al., 2008). This approach assumes stability (i.e., a given participant's score is not anticipated to increase or decline after study dropout) and that the data are MCA R. However, as described by Molnar et al. (2008), it is common for treatment groups to show higher attrition compared to control groups in studies of dementia drugs. Given that dementia worsens over time, using LOCF biases the results in favor of the treatment group (see Molnar et al., 2008, for more details).

More modern approaches, such as using maximum likelihood or multiple imputation to estimate missing data, are thought to avoid some of the biases of older approaches (see Enders, 2010; Graham, 2009). Graham (2009) noted several “myths” regarding missing data. For example, Graham notes many assume the data must be minimally MAR to permit estimating procedures (such as maximum likelihood or multiple imputation) compared to other, more traditional approaches (e.g., using only complete case data). Violations of MAR

impact both traditional and more modern data estimation procedures, though as noted by Graham, violations of MAR tend to have a greater effect on older methods. Graham thus suggests that estimating missing data is a better approach compared to the older procedures in most circumstances, regardless of the model of missingness [i.e., MCAR, MAR, MNAR; see Graham (2009)].

Attrition from a longitudinal panel study such as ABCD is inevitable and represents a threat to the validity of longitudinal analyses and cross-sectional analyses conducted at later time points, especially since attrition can only be expected to grow over time. While, to date, attrition in ABCD has been minimal (some cite here), it remains an important focus for longitudinal analysis and its significance is likely to only grow as the cohort ages. Ideally, one tries to minimize attrition through good retention practices from the outset via strategies designed to maintain engagement in the project (Cotter et al., 2005; Hill et al., 2016; Watson et al., 2018). However, even the best-executed studies need to anticipate growing attrition over the length of the study and implement analytic strategies designed to provide the most valid inferences. Perhaps the most key concern when dealing with data that is missing due to attrition is determining the degree of bias in retained variables that is a consequence of attrition. Assuming that the data are not missing completely at random, attention to the nature of the missingness and employing techniques designed to mitigate attrition-related biases need to be considered in all longitudinal analyses. Several different approaches can be considered and employed depending upon the nature of the intended analyses, the degree of missingness, and data available to help estimate missing and unobserved values.

### ***3.1.9 Quantifying effect sizes longitudinally***

Given longitudinal data involve different sources of variance, quantifying effect sizes longitudinally is a more difficult task compared to deriving such estimates from cross-sectional data. Effect size can be defined as, “a population parameter (estimated in a sample) encapsulating the practical or clinical importance of a phenomenon under study.” (Kraemer, 2014). Common effect size metrics include  $r$  (i.e., the standardized covariance, or correlation, between two variables) and Cohen’s  $d$  (Cohen, 1988). Adjustments to common effect size calculations, such as Cohen’s  $d$ , are required even when only two time points are considered (e.g., see Morris & DeShon, 2002). Wang et al. (2019) note there are multiple approaches to obtaining standardized within-person effects, and that commonly suggested approaches (e.g., global standardization) can be problematic (see Wang et al., 2019, for more details). Thus, obtaining effect size metrics based on standardized estimates that are relatively simple in cross-sectional data (such as  $r$ ) becomes more complex in the context of prospective data. Feingold (2009) noted that equations for effects sizes used in studies involving growth modeling analysis (e.g., latent growth curve modeling) were not mathematically equivalent, and the effect sizes were not in the same metric as effect sizes from traditional analysis (see Feingold, 2009, for more details). Given this issue, there have been various proposals for adjusting effect size measures in repeated assessments. Feingold (2019) reviews the approach for effect size metrics for analyses based on growth modeling, including when considering linear and non-linear (i.e., quadratic) growth factors. Morris and DeShon (2002) review various equations for effect size calculations relevant to when combining estimates in meta-analysis with repeated measures and independent-groups designs. Other approaches to quantifying

effect sizes longitudinally may be based on standardized estimates from models that more optimally disentangle between- and within-person sources of variance (as reviewed above). As an example, within a RI-CLPM framework, standardized estimates between random intercepts (i.e., the correlation between two random intercepts for two different constructs assessed repeatedly) could be used to index the between-person relation, whereas standardized estimates among the structured residuals could be used as informing the effect sizes of within-person relations.

## 4 Part III: Longitudinal Analysis

### 4.0.1 *Longitudinal Data Structures*

An ideal longitudinal study integrates (a) a well-articulated theoretical model, (b) an appropriate longitudinal data structure, and (c) a statistical model that is an operationalization of the theoretical model (collins2006). To accommodate various research questions and contexts, different types of longitudinal data and data structures have emerged (see Figure x). An understanding of these data structures is helpful, as they can warrant different types of longitudinal data analysis. Given that identifying a starting point for making comparisons is somewhat arbitrary, Curran and Bauer (xxxx) provide a nice on-ramp in distinguishing between the use of “time-to-event” and “repeated measures” data. Although both model time, the former is concerned with *whether* and *when* an event occurs, whereas the later is focused on *growth* and *change* [Curran and Bauer, xxxx]. Time-to-event structures measure time from a well-defined origin point up to the occurrence of an event of interest. This data structure is most often analyzed using survival analysis methods (e.g., hazard rate models, event history analysis, failure-time models) and the time-to-event data can be based on a

single assessment or include multiple recurrent or competing events. While much has been written about about “time-to-event” data (see xxx; xxx; xxxx), including several recent xxxx (see), our emphasis will be given to the modeling of “repeated measures” data.

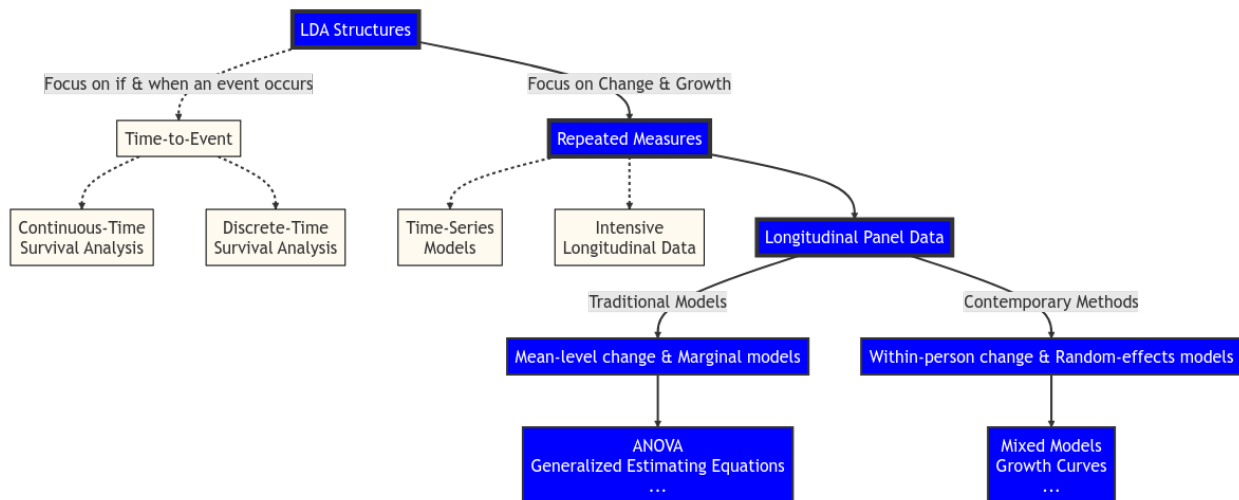


Figure 1: LDA Structural Diagram

When discussing longitudinal analysis, we are most often talking about data collected on the same unit (e.g., individuals) across multiple measurement occasions. However, repeated-measures analysis is not a monolith and it will serve us well to distinguish between a few of the most common types. One such approach to repeated measures analysis is the use of **time-series models**. These models generally consist of a long sequence of repeated measurements ( 50-100 measurements) on a single (or small number of) some variable of interest. Time-series analysis is often used to predict/forecast temporal trends and cyclic patterns and is geared toward making inferences about prospective outcomes within a population (with relatively with less focus on inferring individual-level mechanisms and risk factors). A related type of repeated measures analysis is **Intensive Longitudinal Data (ILD)**. Similar to time-series analysis, ILD models involve frequent measurements (~ 30-40 measurements)

of the same individuals in a relatively circumspect period of time (e.g., experience sampling to obtain time series on many individuals). Although ILD models may include slightly fewer measurement occasions than time-series data, ILD models tend to have more subjects than time series models ( $\sim 50$ -100 subjects). This allows ILD models to examine short-term patterns by incorporating a time series model that can sometimes fit parameter estimates to each individual's data in order to model individual difference outcomes. The final type of repeated measures analysis that we will discuss is the longitudinal panel model. These models follow a group of individuals— a panel (also referred to as a cohort) — across relatively fewer measurement occasions ( $\sim 5$ -15), and are often interested in examining change across both, within-individuals and between-individuals.

While other longitudinal designs have their own unique strengths and applications, the longitudinal panel design is particularly well-suited for investigating developmental processes in the context of the ABCD Study. In the following sections, we will discuss various analytic methods commonly used to analyze longitudinal panel data, including growth models, mixed models, and a number of additional trajectory models. These methods provide valuable insights into within- and between-individual differences and are highly relevant for researchers working with the ABCD Study dataset. By focusing on these methods, we aim to equip readers with the knowledge necessary to conduct longitudinal research and perform analyses using the rich, longitudinal, and publicly available data from the ABCD Study.

The following text includes collated information about each of the categories of longitudinal analyses that we have discussed during previous questions and responses that has been written up for publication in a scientific journal:



**Types of longitudinal panel models.** With the large and continually expanding body of research on statistical methods for longitudinal analyses, determining which longitudinal model to implement can be challenging. The aim of this section is to help researchers to navigate these many options in order to identify the statistical approach most appropriate to their unique research question. Notably, there are a myriad of viable ways one can go about grouping various types of longitudinal models for presentation. Common examples include grouping by linearity (linear vs nonlinear models; see Collins, 2006(?)), the number of measurement occasions (see King et al., 2018), and xxxxxx (see xxxx, xxx). The ordering/organization we use below overlaps in a number of ways with these examples, and with Curran & Bauer (2019); however, it is important to note that in each case, the ordering/organization that is chosen is somewhat arbitrary and primarily intended to allow the reader to compare and contrast of various analytic approaches. In the following we will describe (briefly) and summarize the advantages/disadvantages for a series of 16(?) distinct longitudinal models organized into the following x# **fuzzy** groupings: Traditional Models, Modern GLM Extensions, SEM, Advanced SEM, xxx, xxxx. We note that this is not an exhaustive review of each of these methods, but for more in-depth detail we do provide the reader relevant resources. As aptly summarized by Curran & Bauer (2019), “there are many exceptions, alternatives, nuances, “what ifs”, and “but couldn’t you” that aren’t addressed here.”

***Traditional (linear and nonlinear).*** Traditional methods for longitudinal analysis primarily focus on modeling mean-level change. For example, is there a difference in average math scores obtained across multiple assessments between boys and girls? Lon-

gitudinal models that focus on mean-level change are also referred to as **marginal models** and examples of specific methods include **repeated measures ANOVA** (and **MANOVA**), **ANCOVA**, and **Generalized Estimating Equations (GEE)**. These methods are commonly used when data is only available from 2 measurement occasions. For example, computing a **difference score** (e.g., math scores at t2 - math scores at t1) that can be used as an outcome in a subsequent GLM analysis (e.g., paired-samples t-test, repeated measures ANOVA) to test for differences in patterns of change over time and between groups. We also note the longitudinal **signed-rank test** which is a nonparametric alternative to the paired t-test that can be a useful tool for analyzing non-normal paired data. Another common approach, often used in pre-/post-design studies, is to use **residualized change score analysis** to assess the degree of change in a variable, while controlling for its initial level. For example, regressing post-treatment scores on pre-treatment scores and then using the resulting residuals as a measure of change that is adjusted for baseline scores (ignoring any prior group assignments/differences). Similar to difference scores, the residualized change score is often included in subsequent analysis, such as for evaluating intervention effects in pre-test/post-test designs (**Kisbu-Sakarya2013**).

Traditional longitudinal methods can still be useful in some contexts (e.g.,  $< 3$  measurement occasions), but overall their practical utility for answering questions about developmental processes is limited. Perhaps most notably, these models do not allow for characterizing patterns of within-person change, which is often an emphasis of longitudinal research. In turn, these models often correspond poorly with most theoretical models of change.

Moreover, even determining which of these procedures to use for comparing change over two time points across groups can be surprisingly complicated. A particularly vexing

example is that of imbalanced baseline scores which can produce biased effects when using these methods. Indeed, it is well-documented that this can result in different conclusions between these approaches, which has led to many articles debating the merits of which procedures is most appropriate to use and when [see xxxxx for a review]. Beyond these concerns, additional limitations include xxxxxx. Given these shortcomings and the complexity of the issues surrounding some of these methods, when possible, it is typically recommended that researchers make use of more modern approaches for analyzing longitudinal data (and preferably make use of data collected across 3 or more time points).

***Modern GLM Extensions.*** Modern approaches to longitudinal data analysis have advanced beyond traditional methods by offering greater flexibility and a more in-depth understanding of within-person and between-person variability. **Generalized Estimating Equations (GEE)**, **Linear Mixed Models (LMM)**, **Generalized Linear Mixed Models (GLMM)**, and **Autoregressive Cross-Lagged Panel Models (ARCL)** are examples of such contemporary techniques. GEE is an extension of Generalized Linear Models. This model combines the generalized linear model for non-normal outcomes with repeated measures (marginal) model and is suitable for analyzing correlated longitudinal data and modeling population-averaged effects. LMMs, also known as multilevel or hierarchical linear models, facilitate the simultaneous analysis of within-person and between-person variability, making them ideal for nested data structures or repeated measures. GLMMs further extend the LMM framework to accommodate non-normal response variables, such as binary, count, or ordinal data. Finally, ARCL models are used to investigate reciprocal relationships between variables over time, as they estimate both autoregressive and cross-lagged effects.

The strengths of these modern methods lie in their ability to account for individual differences, within-person change, and time-varying predictors, thereby providing a more comprehensive understanding of complex relationships in longitudinal data. Despite these advantages, modern approaches may require more complex modeling assumptions and higher computational demands compared to traditional methods. Additionally, proper model specification and the interpretation of results can be more challenging, especially in cases of high multicollinearity or missing data. However, modern longitudinal analysis methods have generally surpassed traditional methods in addressing a wider range of research questions, accommodating diverse data structures, and elucidating the intricate dynamics of developmental processes.

***SEM.*** Structural equation modeling (SEM) approaches have gained prominence in longitudinal data analysis due to their ability to estimate complex relationships among observed and latent variables while accounting for measurement error. SEM techniques can be categorized into variable-centered, person-centered, and hybrid approaches, each with unique strengths and limitations. The choice of method depends on the research question, data structure, and underlying assumptions.

Variable-centered approaches, such as Latent Change Scores, Latent Growth Curve Models, Multivariate (Parallel Process) Latent Growth Curve Models, and Latent State-Trait Models, focus on examining relationships among variables and population-level patterns. These models offer a powerful means to estimate individual change trajectories and the relationships between growth parameters of different variables, while also decomposing observed measurements into latent state and trait components. However, these approaches

may not adequately capture distinct subgroups of individuals who share similar patterns of change over time, which can be crucial for understanding heterogeneous developmental processes.

Person-centered approaches, including Latent Transition Analysis and Latent Class Growth Analysis, address this limitation by identifying subgroups of individuals who share similar patterns of change. These models can reveal meaningful subpopulations and help researchers understand the factors that contribute to differences in developmental trajectories. Nevertheless, person-centered methods may overlook important relationships among variables, which can be essential for understanding the dynamics of change over time.

Hybrid approaches, such as Growth Mixture Modeling, combine aspects of both variable-centered and person-centered models, allowing for the identification of latent subgroups while also modeling relationships among growth parameters. This combination provides a more comprehensive understanding of longitudinal data by capturing both within- and between-person variability. However, hybrid models can be more complex, necessitating careful model specification, selection, and interpretation. Additionally, these methods may require larger sample sizes to ensure the stability and accuracy of results.

In summary, SEM approaches offer powerful tools for longitudinal data analysis, enabling researchers to investigate complex relationships, individual differences, and change dynamics over time. The choice between variable-centered, person-centered, and hybrid approaches depends on the research objectives and the nature of the data. Despite their limitations, these models have greatly advanced our understanding of developmental processes and the factors that contribute to individual differences in change trajectories.

***Advanced SEM.*** Advanced structural equation modeling (SEM) approaches, such as the Random-Intercept Cross-Lagged Panel Model (RI-CLPM) and Latent Curve Models with Structured Residuals (LCM-SR), have emerged to address more complex research questions and data structures in longitudinal analysis. These advanced models extend traditional SEM techniques, enabling researchers to disentangle within-person and between-person effects, as well as capture additional time-specific dependencies and associations that may not be accounted for by the latent growth factors.

The RI-CLPM enhances the traditional cross-lagged panel model by incorporating random intercepts, which allow for the separation of stable individual differences from the dynamic within-person associations between variables over time. This distinction is particularly valuable for understanding the time-specific effects of one variable on another, while accounting for the influence of individual differences. However, RI-CLPM may require larger sample sizes to ensure stability and accuracy of the estimates and can be computationally demanding.

LCM-SR, on the other hand, extends the Latent Growth Curve Model by modeling structured residuals, such as autoregressive or cross-lagged effects among the residuals. This approach allows for a more comprehensive understanding of within-person change dynamics and factors influencing change over time. By including structured residuals, LCM-SR can capture additional time-specific relationships that are not explained by the latent growth factors. However, similar to RI-CLPM, LCM-SR also comes with increased model complexity and requires careful specification and interpretation.

In conclusion, advanced SEM approaches for longitudinal data analysis provide valuable tools for addressing complex research questions and data structures. While they offer

more nuanced insights into within-person change dynamics and the influence of individual differences, these models also come with certain limitations, such as increased complexity, computational demands, and the need for careful model specification and interpretation. As with any statistical method, researchers should carefully consider their research objectives, data characteristics, and the assumptions of each model when selecting the most appropriate advanced SEM approach for longitudinal analysis.

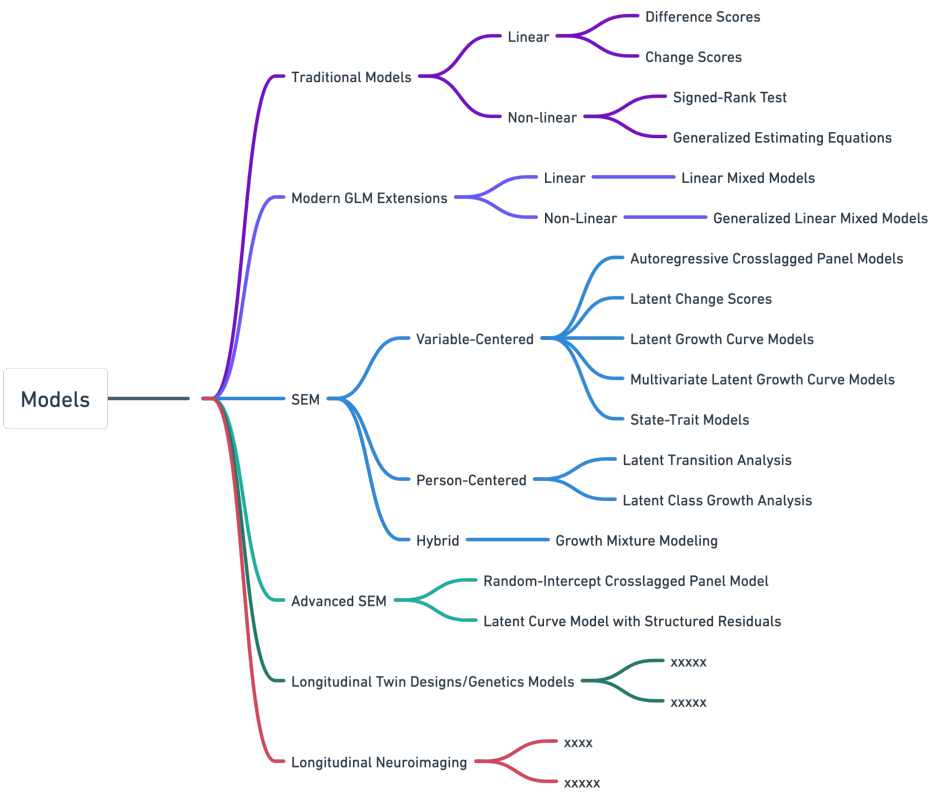


Figure 2: Longitudinal Models/Data Structures

*(possibly others xxxx?).* (longitudinal neuroimaging?) (longitudinal twins/genetics?)

#### ***4.0.2 Longitudinal Analysis Section Wrap-up***

This section, along with associated code and additional resources made available as online supplements ([\(\)](#)), aims to serve as a resource for researchers seeking to understand and implement various longitudinal panel models. By providing an overview of different approaches, their strengths and limitations, and key considerations for their use, we hope to facilitate the selection of appropriate models tailored to specific research questions and data structures. It is essential for researchers to consider their research objectives, the characteristics of their data, and the assumptions underlying each model when choosing the most suitable approach for longitudinal analysis. In this vast and continually evolving field, with numerous models and approaches available to address a wide range of research questions, no single model is universally applicable or without limitations. The diversity of methods ensures that researchers can find an appropriate tool for their specific needs. By familiarizing themselves with the various types of longitudinal models, researchers can more effectively navigate the complexities of longitudinal data and contribute valuable insights into the developmental processes and individual differences that shape human experience.

We encourage researchers to consult the cited literature and online supplements for further guidance in selecting and implementing longitudinal models when using the ABCD Study dataset. As the field continues to advance, we anticipate the emergence of new methods and refinements to existing approaches, further expanding the toolkit available to researchers for the analysis of longitudinal data. By staying informed about developments in this area and critically evaluating the appropriateness of different models for their re-



search questions, researchers can ensure that their longitudinal analyses are both rigorous and informative.

---

## 5 Part IV: Discussion

### *5.0.1 Subsection TBD.*

## References

- Bagot, K. S., Matthews, S. A., Mason, M., Squeglia, L. M., Fowler, J., Gray, K., Herting, M., May, A., Colrain, I., & Godino, J. (2018). Current, future and potential use of mobile and wearable technologies and social media data in the ABCD study to increase understanding of contributors to child health [00046 Publisher: Elsevier]. *Developmental cognitive neuroscience*, 32, 121–129.
- Barch, D. M., Albaugh, M. D., Avenevoli, S., Chang, L., Clark, D. B., Glantz, M. D., Hudziak, J. J., Jernigan, T. L., Tapert, S. F., & Yurgelun-Todd, D. (2018). Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description [00192 Publisher: Elsevier]. *Developmental cognitive neuroscience*, 32, 55–66.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression [45810 Publisher: American Medical Association]. *Archives of general psychiatry*, 4(6), 561–571.
- Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., ... Dale, A. M. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites [00232]. *Developmental Cognitive Neuroscience*, 32, 43–54. <https://doi.org/10.1016/j.dcn.2018.03.001>
- Cohen, J. (1988). Statistical power [03369]. *Analysis for the behavioral sciences*, 273–406.

- Cotter, R. B., Burke, J. D., Stouthamer-Loeber, M., & Loeber, R. (2005). Contacting participants for follow-up: How much effort is required to retain participants in longitudinal studies? [00105 Publisher: Elsevier]. *Evaluation and Program Planning*, 28(1), 15–21.
- Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change [01308 Publisher: Annual Reviews]. *Annual review of psychology*, 62, 583–619.
- Curran, P. J., Howard, A. L., Bainter, S., Lane, S. T., & McGinley, J. S. (2014). The Separation of Between-person and Within-person Components of Individual Change Over Time: A Latent Curve Model with Structured Residuals [00194]. *J Consult Clin Psychol*, 82(5), 879–894. <https://doi.org/10.1037/a0035297>
- Duncan, T. E., & Duncan, S. C. (2009). The ABC's of LGM: An Introductory Guide to Latent Variable Growth Curve Modeling [00000 \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.9004.2009.00224.x>]. *Social and Personality Psychology Compass*, 3(6), 979–991. <https://doi.org/10.1111/j.1751-9004.2009.00224.x>
- Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I [03276 Publisher: JSTOR]. *Biometrika*, 37(3/4), 409–428.
- Enders, C. K. (2010, April). *Applied Missing Data Analysis* [06821 Google-Books-ID: MN8ruJd2tvvC]. Guilford Press.
- Farewell, V., Long, D., Tom, B., Yiu, S., & Su, L. (2017). Two-Part and Related Regression Models for Longitudinal Data [\_eprint: <https://doi.org/10.1146/annurev-statistics-060116-054131>]. *Annual Review of Statistics and Its Application*, 4(1), 283–315. <https://doi.org/10.1146/annurev-statistics-060116-054131>

- Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. [00633 Publisher: American Psychological Association]. *Psychological methods*, 14(1), 43.
- Feingold, A. (2019). Time-varying effect sizes for quadratic growth models in multilevel and latent growth modeling [00008 Publisher: Taylor & Francis]. *Structural equation modeling: a multidisciplinary journal*, 26(3), 418–429.
- French, D. P., & Sutton, S. (2010). Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? [00185 Publisher: Wiley Online Library]. *British journal of health psychology*, 15(3), 453–468.
- Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World [05760 \_\_eprint: <https://doi.org/10.1146/annurev.psych.58.110405.085530>]. *Annual Review of Psychology*, 60(1), 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Hagler, D. J., Hatton, S., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., Sutherland, M. T., Casey, B. J., Barch, D. M., Harms, M. P., Watts, R., Bjork, J. M., Garavan, H. P., Hilmer, L., Pung, C. J., Sicat, C. S., Kuperman, J., Bartsch, H., Xue, F., ... Dale, A. M. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study [000000]. *NeuroImage*, 202, 116091. <https://doi.org/10.1016/j.neuroimage.2019.116091>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model [00848 Place: US Publisher: American Psychological Association]. *Psychological Methods*, 20(1), 102–116. <https://doi.org/10.1037/a0038889>

- Hill, K. G., Woodward, D., Woelfel, T., Hawkins, J. D., & Green, S. (2016). Planning for long-term follow-up: Strategies learned from longitudinal studies [00036 Publisher: Springer]. *Prevention Science*, 17(7), 806–818.
- Kincaid, C. (2005). Guidelines for selecting the covariance structure in mixed model analysis, paper 198-30 in Proceedings of the Thirtieth Annual SAS Users Group Conference [00013]. *Inc., Cary, North Carolina*.
- King, K. M., Littlefield, A. K., McCabe, C. J., Mills, K. L., Flournoy, J., & Chassin, L. (2018). Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology [00044]. *Developmental Cognitive Neuroscience*, 33, 54–72. <https://doi.org/10.1016/j.dcn.2017.11.009>
- Kraemer, H. C. (2014). Effect size [00178 Publisher: Wiley Online Library]. *The encyclopedia of clinical psychology*, 1–3.
- Lenz, S. T. (2016). Alan Agresti (2013): Categorical data analysis [00002 Publisher: Springer Nature BV]. *Statistical Papers*, 57(3), 849.
- Little, R. J., & Rubin, D. B. (1989). The Analysis of Social Science Data with Missing Values [01328 Publisher: SAGE Publications Inc]. *Sociological Methods & Research*, 18(2-3), 292–326. <https://doi.org/10.1177/0049124189018002004>
- Little, T. D. (2013, February). *The Oxford Handbook of Quantitative Methods, Vol. 2: Statistical Analysis* [00006 Google-Books-ID: bCFwAgAAQBAJ]. Oxford University Press.
- Littlefield, A. K., King, K. M., Acuff, S. F., Foster, K. T., Murphy, J. G., & Witkiewitz, K. (2021). Limitations of cross-lagged panel models in addiction research and alternative models: An empirical example using project MATCH. [00001 Publisher: US: American

- Psychological Association]. *Psychology of Addictive Behaviors*. <https://doi.org/10.1037/adb0000750>
- Loughnan, R. J., Palmer, C. E., Thompson, W. K., Dale, A. M., Jernigan, T. L., & Fan, C. C. (2020). Polygenic score of intelligence is more predictive of crystallized than fluid performance among children [00002 Publisher: Cold Spring Harbor Laboratory]. *bioRxiv*, 637512.
- Luciana, M., Bjork, J. M., Nagel, B. J., Barch, D. M., Gonzalez, R., Nixon, S. J., & Banich, M. T. (2018). Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery [00000 Publisher: Elsevier]. *Developmental cognitive neuroscience*, 32, 67–79.
- Molnar, F. J., Hutton, B., & Fergusson, D. (2008). Does analysis using “last observation carried forward” introduce bias in dementia research? [00139 Publisher: Can Med Assoc]. *Cmaj*, 179(8), 751–753.
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. [02272 Publisher: American Psychological Association]. *Psychological methods*, 7(1), 105.
- Muthén, L. K. (2017). Mplus user’s guide. Los Angeles: Muthén & Muthén; 1998 [00000].
- Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. [00047 Publisher: American Psychological Association]. *Journal of Personality and Social Psychology*, 120(4), 1013.

- Robins, L. (1985). Epidemiology: Reflections on Testing the Validity of Psychiatric Interviews | JAMA Psychiatry | JAMA Network [00479]. Retrieved October 3, 2021, from <https://jamanetwork.com/journals/jamapsychiatry/article-abstract/493658>
- Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. [01525 Publisher: American Psychological Association]. *Psychological bulletin*, 92(3), 726.
- Salthouse, T. A. (2014). Why are there different age relations in cross-sectional and longitudinal comparisons of cognitive functioning? [00069 Publisher: Sage Publications Sage CA: Los Angeles, CA]. *Current directions in psychological science*, 23(4), 252–256.
- Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavél, F. D., Iida, M., Gleason, M. E. J., Xu, J. H., & Bolger, N. (2018). Initial elevation bias in subjective reports [00089 Publisher: National Academy of Sciences Section: PNAS Plus]. *PNAS*, 115(1), E15–E23. <https://doi.org/10.1073/pnas.1712277115>
- Sullivan, E. V., Brumback, T., Tapert, S. F., Prouty, D., Fama, R., Thompson, W. K., Brown, S. A., Cummins, K., Colrain, I. M., & Baker, F. C. (2017). Effects of prior testing lasting a full year in NCANDA adolescents: Contributions from age, sex, socioeconomic status, ethnicity, site, family history of alcohol or drug abuse, and baseline performance [00008 Publisher: Elsevier]. *Developmental cognitive neuroscience*, 24, 72–83.
- Thompson, W. K., Barch, D. M., Bjork, J. M., Gonzalez, R., Nagel, B. J., Nixon, S. J., & Luciana, M. (2019). The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neu-

- rocognitive battery [00000 Publisher: Elsevier]. *Developmental cognitive neuroscience*, 36, 100606.
- Uban, K. A., Horton, M. K., Jacobus, J., Heyser, C., Thompson, W. K., Tapert, S. F., Mad-den, P. A. F., & Sowell, E. R. (2018). Biospecimens and the ABCD study: Rationale, methods of collection, measurement and early data [00034]. *Developmental Cognitive Neuroscience*, 32, 97–106. <https://doi.org/10.1016/j.dcn.2018.03.005>
- Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., Pérez-Stable, E. J., Riley, W. T., Bloch, M. H., Conway, K., Deeds, B. G., Dowling, G. J., Grant, S., Howlett, K. D., Matochik, J. A., Morgan, G. D., Murray, M. M., Noronha, A., Spong, C. Y., ... Weiss, S. R. B. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration [00147 tex.ids= volkow2018a ISBN: 1878-9293 publisher: Elsevier]. *Developmental Cognitive Neuroscience*, 32, 4–7. <https://doi.org/10.1016/j.dcn.2017.10.002>
- Wang, L., Zhang, Q., Maxwell, S. E., & Bergeman, C. S. (2019). On standardizing within-person effects: Potential problems of global standardization [00014 Publisher: Taylor & Francis]. *Multivariate behavioral research*, 54(3), 382–403.
- Watson, N., Leissou, E., Guyer, H., & Wooden, M. (2018). Best Practices for Panel Maintenance and Retention [00006 Section: 29 \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118884997.ch29> In *Advances in Comparative Survey Methods* (pp. 597–622). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118884997.ch29>
- Zucker, R. A., Gonzalez, R., Ewing, S. W. F., Paulus, M. P., Arroyo, J., Fuligni, A., Morris, A. S., Sanchez, M., & Wills, T. (2018). Assessment of culture and environment in the Adolescent Brain and Cognitive Development Study: Rationale, description of



measures, and early data [00064 Publisher: Elsevier]. *Developmental cognitive neuroscience*, 32, 107–120.