Longitudinal Analysis Manuscript: Working Draft

Longitudinal Analysis using data from the ABCD© Study

Samuel Hawes

Andrew K. Littlefield

Daniel A. Lopez

Kenneth J. Sher

Erin L. Thompson

Raul Gonzalez

Additional Co-authors

Wesley K. Thompson

The Adolescent Brain Cognitive DevelopmentSM (ABCD) Study presents a unique opportunity for researchers to investigate developmental processes in a large, diverse cohort of children and adolescents. Given the breadth and complexity of the ABCD Study, researchers analyzing its data are likely to encounter a myriad of methodological and analytic considerations and concerns. This review provides an examination of key issues and techniques related to longitudinal data analyses of the ABCD data. We discuss the value of longitudinal data over cross-sectional data, focusing on the types of inferences that are possible when one assesses individuals across multiple time points, including: 1) characterization of normative variation in developmental trajectories and the genetic and environmental factors that may lead to deviations from normative development; 2) assessment of how variation in the development of one domain may or may not impact development in another; 3) identification of developmental disturbances, snares, and cascade effects; and 4) relative timing and reciprocal relationships of within-person changes from different developmental domains. We emphasize the importance of selecting appropriate methods to address these research questions, such as accounting for correlation in repeated measurements and using models for continuous or discrete outcomes as necessary. By addressing the advantages and potential challenges of developmental analyses in the ABCD Study, this review seeks to equip researchers with foundational knowledge and tools to make informed decisions as they navigate and effectively analyze and interpret the multi-dimensional longitudinal data currently available.

# 1. Introduction

[The Adolescent Brain Cognitive Development (ABCD) Study®](https://abcdstudy.org/) 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 outcomes. Central to achieving these goals is the ABCD Study’s® commitment to an open science framework, intended to facilitate sharing of data and analytical methods, by espousing practices that increase access, 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.

The size and scope of the ABCD Study data allow the research community to perform a large variety of developmental analyses of both substantive and methodological interest, presenting a unique opportunity to significantly advance our understanding of how a multitude of biopsychosocial processes unfold across critical periods of development. In this paper we describe models and methods for longitudinal analysis of ABCD Study data that can address these fundamental scientific aims, as well as some challenges inherent in a large, longitudinal observational study of adolescents. We instantiate many of the longitudinal analyses in worked examples with accompanying R scripts available at XXXX.

## 1.1 The ABCD Study Data

The ABCD Study enrolled a large cohort of youth (n=11,880) born between 2006-2008 and aged 9-10 years at baseline, as well as their parents/guardians. The study sample was recruited from household populations in defined catchment areas for each of the 21 (originally 22) study sites across the United States. Information regarding funding agencies, recruitment sites, investigators, and project organization can be obtained at https://abcdstudy.org/.

The ABCD Study is currently collecting longitudinal data on a rich variety of outcomes that will enable the construction of complex statistical models potentially incorporating factors from many domains. Each new wave of data collection provides another building block for estimating developmental trajectories and implementing probing longitudinal analyses that allow researchers to characterize normative development, to identify variables that presage deviations from prototypic development, and to assess a range of outcomes associated with biopsychosocial variables of interest. These data include: 1) a neurocognitive battery (Luciana et al. 2018; Thompson et al. 2019); 2) mental and physical health assessments [Barch et al. (2018)); 3) measures of culture and environment (Gonzalez et al. 2021; Zucker et al. 2018); 4) substance use (Lisdahl et al. 2021); 5) biospecimens (Uban et al. 2018); 6) structural and functional brain imaging (Casey et al. 2018; Hagler et al. 2019); 7) geolocation-based environmental exposure data (Fan et al. 2021); 8) wearables, and mobile technology (Bagot et al. 2018); and 9) 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, currently through the NIMH Data Archive.

By necessity, the study’s earliest data releases consisted primarily of cross-sectional (baseline) data; however, the most recent public data release (Release 5.0) contains data collected across four annual assessments, including three brain imaging assessments (baseline, year 2 follow-up, and year 4 follow-up visits). Thus, the time is ripe for implementing a more longitudinal perspective and a focus on estimation of trajectories (i.e., within-person change over time).

## 1.2

|  |
| --- |
| Organization and Aims |
| • Part I. Longitudinal Research   * Identify fundamental concepts   • Part II. Longitudinal Data   * Highlight key challenges   • Part III. Longitudinal Analysis   * Learn methods & analysis to apply   • Part IV. Supplemental materials   * Make use of linked open-source resources |

# 2. 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 but also brings with it greater methodological opportunities to make use of statistical models that distinguish within- and between-individual effects and that loosen constraints that need to be imposed on the furtherance of critical scientific questions. For example, collecting two or more within-person observations on the same construct but at different times enables estimation of within-person rate of change (slopes); more observations allow for more precise estimates of individual slopes, as well as characterization of non-linear development. Rate of change or other aspects of trajectory shapes may be more informative than the simple snapshots of level differences that cross-sectional data are limited to. 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.

### 2.1.1 Vulnerable periods.

Adolescent development progresses normaly 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 (Arseneault et al. 2002; Bechtold et al. 2016; Hasan et al. 2020; Semple, McIntosh, and Lawrie 2005); 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) (Spear 2016; Crews et al. 2000; Ji et al. 2018). For example, longitudinal data from the National Consortium on Alcohol in Adolescence (NCANDA) show that binge drinking is associated more strongly with decrements in gray matter volume early in adolescence compared to later (Infante et al. 2022). These examples highlight the importance of considering the role of vulnerable periods – e.g., temporal windows of rapid brain development or remodeling during which the effects of environmental stimuli (e.g. alcohol exposure) on the developing brain may be particularly pronounced– when trying to establish an accurate understanding of the association between exposures and outcomes.

### 2.1.2 Developmental disturbances.

Whereas vulnerable periods heighten neurobiological susceptibility to environmental influences, at other times, environmental presses will tend to suppress stability and disrupt the orderly stochastic process of normative development (e.g., Schulenberg et al. 2019). This situation reflects a developmental disturbance in that the normal course of development is “altered” 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 pressures are removed (or the individual is removed from the environment), patterns of individual differences (and autoregressive effects) recover to levels similar to those prior to entering the environment. For example, in Infante et al. (2022), recent binge drinking appears to be most predictive of gray matter volume trajectories, as opposed to prior binge drinking or cumulative number of binge drinks, suggesting the potential for recovery of gray matter trajectories to prior levels of growth if binge drinking subsides.

### 2.1.3 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 other substance-using or rule-breaking peers, precluding opportunities for learning various adaptive skills and prosocial behaviors, in effect creating a “snare” that delays psychosocial development, such as maturing out of adolescent antisocial behavior (Moffitt 2015). 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 (e.g., Masten et al. 2005; Rogosch, Oshri, and Cicchetti 2010). Although conceptually distinct from vulnerable periods, both 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. Longitudinal data are crucial in this context to assess individual levels of development prior to and following onset of experiences or other environmental factors; e.g., the ABCD Study collected data starting at ages 9-10 and hence before the onset of drug use for the vast majority of participants.

# 3. Longitudinal Data

## 3.1 Interpretation / Issues / Pitfalls & Assumption

The hallmark characteristic of longitudinal data analysis is the administration of repeated measurements of the same constructs on assessment targets (e.g., individuals, families) across time. The primary rationale for collecting longitudinal data is to assess within-person change over time, allowing researchers to estimate individual developmental trajectories and the genetic and/or environmental factors that may impact these. Generally speaking, administering repeated measurements more frequently or over longer time spans enables researchers to ask more nuanced questions and to make stronger inferences.

### 3.1.1 Two Time Points versus Three or More.

Although the clear leap from cross-sectional 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, Brandt, and Zimowski (1982) noted over forty years ago, “Two waves of data are better than one, but maybe not much better” (p XX).

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 trajectory estimation, 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 change is the only estimable form given two assessments (see Duncan and Duncan 2009). Research designs that include three (but preferably more) time points allow for non-linear trajectory estimation and 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 developmental research.

To illustrate, developmental theories are useful for 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” (Andrew K. 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; Patrick J. 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 and Duncan 2009).

### 3.1.2 Types of stability and change

If one were to try to sum up what developmental trajectories in a living organism are exactly, one could plausibly argue they are the patterns of stability and change in its phenotypes as the organism traverses the life course. Mathematically, this could be expressed as f(t), a possibly multivariate function of time t typically taking values in the real numbers for continuous phenotypes or the integers for discrete phenotypes. There are a few different ways to think about patterns of stability and change (see Figure 1). Consider measuring 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 over time for an outcome of interest, aggregated across members of a group.

In contrast, “between-individual” stability and change could be assessed, e.g., by calculating the Spearman correlation between the values obtained at different time points (e.g., ‘height in the fall’ with ‘height in the spring’). This analysis focuses 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 tend to be particularly vulnerable to the effects of environmental factors at some ages more than others; rank-order stability 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; Bleidorn et al. (2022)], but relatively high rank-order stability, at least over relatively shorter measurement intervals (Bleidorn et al. 2022; Roberts and DelVecchio 2000; Roberts, Walton, and Viechtbauer 2006). Despite the useful information afforded by examining mean-level and rank-order stability, these approaches are limited in that they provide little information about the patterns of within-individual change (i.e., trajectories) and, in turn, can result in fundamental misinterpretations about substantial or meaningful changes in an outcome of interest (Patrick J. Curran and Bauer 2011).

|  |
| --- |
| Figure 1: Types of Stability and Change |

There is growing recognition that statistical models commonly applied to longitudinal data often fail to align with the developmental theory they are being used to assess (e.g., Patrick J. Curran and Bauer 2011; Hoffman 2015; Andrew K. 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., Patrick J. Curran and 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. Fortunately, there exists a range of models that have been proposed to tease apart between- and within-person sources of variance across time (see Andrew K. 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 such as the random-intercept crosslagged panel model (e.g., RI-CLPM, Hamaker, Kuiper, and Grasman 2015) and the latent curve model with structured residuals (e.g., LCM-SR, Patrick J. 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), additional expertise to overcome issues with model convergence, and appreciation of modeling assumptions when attempting to adjudicate among potential models in each research context (see Andrew K. Littlefield et al. 2021, for further discussion).

### 3.1.3 Model assumptions

Many statistical models assume certain characteristics about the data to which they are being applied. Common assumptions of parametric statistical models (e.g., linear mixed-effects 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, as violation of a model’s assumptions can substantively invalidate the interpretation of results. For example, longitudinal data can exhibit heterogeneous variability (i.e., the variance of the response changes over the duration of the study) that may need to be accounted for within a model. Similarly, longitudinal measurement invariance can help to establish whether or not a construct is measured consistently over time (Liu et al. 2017; Van De Schoot et al. 2015; Willoughby, Wirth, and Blair 2012), providing researchers with greater confidence that any change over time identified for a construct is attributable to individual change rather than a measurement artifact.

### 3.1.4 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 during 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; French and 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 artifically 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…albeit to a modest extent. It is possible that in later waves, analyses will allow for disaggregating the confounded effects of age and the number of prior assessments. However, ABCD is fundamentally a single-cohort, longitudinal design, wherein the number of prior assessments and age may be confounded. As such, the possible influence of experience effects needs to be kept in mind.

### 3.1.5 Covariance Structures

A central issue for repeated measurements on an individual is how to account for the correlated nature of the data. Traditional techniques, such as a standard regression or ANOVA model, assume residuals are independent and thus are inappropriate for designs that assess the same individuals across time. That is, given that 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 and Watson 1950), independence is 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.

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 (involving the covariance among pairs of repeated outcomes on an individual), coupled with a model for the mean response and its dependence on covariates (e.g., sex at birth). There exists a range of methods to model covariance structures, each with its own set of tradeoffs between model fit and parsimony and which may be more or less appropriate for each specific application [e.g., see Kincaid (2005)).

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 measurement occasions rather than remaining constant. That is, we can think of the underlying process as a stochastic one that wears itself out over time. 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 × .5 = .25, and the correlation between Time 1 and Time 4 data would be assumed to be .5 × .5 × .5 = .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 a simple AR(1) process 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, Kuiper, and Grasman 2015). Finally, it is also worth noting that in many longitudinal contexts, the time intervals between assessments are not equidistant and researchers need to consider carefully how to appropriately model time in their model and what this model implies.

### 3.1.6 Continuous and Discrete Outcomes

Identification of valid and efficient statistical models requires an understanding of the type of data being analyzed. For example, repeated assessments can be based on 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, and d) linear growth models for a count outcome assuming a zero-inflated Poisson model. Beyond these highlighted examples, other distributions (e.g., negative binomial) can be assumed for the indicators when modeling longitudinal data (Ren et al. 2022). 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; see Lenz (2016)]. 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 variables (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 Missing Data/Attrition

As recently reviewed by Andrew K. Littlefield et al. (2022), investigators of prospective data are confronted by the certain prospect of study attrition (i.e., participants may not provide data at a given wave of assessment or may drop out of the study altogether) and thus approaches are needed to confront the issue of missing data. Three types of missingness are typically considered in the literature (see R. J. Little and Rubin 1989), namely: a) missing completely at random (MCAR), b) missing at random (MAR), and c) missing not at random (MNAR). Data that are MCAR means that the missing data are a simple random sample of all data in a given dataset. MAR means missing at random conditional on the observed covariates (see Graham 2009). That is, MAR implies missing data are a random sample (i.e., does not hinge on some unmeasured variables) within strata of the measured covariates in a dataset (e.g., biological sex). Data that are MNAR are missing as a function of unobserved variables and may bias assocaitions even after conditioning on the observed covariates. 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). Another common approach to imputing missing data is termed “last observation carried forward” (LOCF). LOCF replaces a participant’s missing values after dropout with the last available measurement (Molnar, Hutton, and Fergusson 2008). This approach assumes stability (i.e., a given participant’s score is not anticipated to systematically increase or decline after study dropout) and that the data are MCAR. However, as described by Molnar, Hutton, and Fergusson (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, Hutton, and Fergusson 2008, for more details).

More modern approaches, such as full information maximum likelihood, propensity weighting, auxiliary variables and multiple imputation 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 imputing 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 potential 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. The ABCD Retention Workgroup (RW) employs a data-driven approach to examine, track, and intervene in these issues and while preliminary findings show participant race and parent education level to be associated with late and missing visits, although to date, attrition in ABCD has been minimal (Ewing et al. 2022). 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. Such bias can attenuate generalizability, particularly if the pattern of missingness is not random (e.g., certain subsets of the population are more likely to drop out/not attend a visit). 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.

### 3.1.8 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 [i.e., the standardized difference between two means; 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 and 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 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 random-intercept cross-lagged panel model (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.

### 3.1.9 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 (Collins 2006). 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 longtiudinal data analysis. Given that identifying a starting point for making comparisons is somewhat arbitrary, Bauer and Curran (2019) provide a nice on-ramp in first 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 (Bauer and Curran 2019). 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 “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.

|  |
| --- |
| Longitudinal Data Analysis 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) 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 (~ 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 analaysis 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 occassions (~ 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.

# 4. Longitudinal Analysis

### 4.0.1 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 navigate these many options 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 statistical equivalency [e.g., change scores vs. residualized change; see Castro-Schilo and Grimm (2018)]. The organization we use below overlaps in a number of ways with these examples, and in particular with Bauer and Curran (2019). However, it is important to note that in each case, the ordering that is chosen is somewhat arbitrary and primarily intended to allow the reader to compare and contrast various analytic approaches. In the following we describe (briefly) and summarize the advantages/disadvantages of a series of longitudinal models organized into the following groupings: Traditional Models, Modern GLM Extensions, SEM, and Advanced SEM. 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 (e.g., Patrick J. Curran and Hancock 2021). As aptly summarized by Bauer and Curran (2019), “there are many exceptions, alternatives, nuances, ‘what ifs’, and ‘but couldn’t yous’ that aren’t addressed here.”

Longitudinal Models/Data Structures

#### Traditional Models

Traditional methods for longitudinal analysis primarily focus on modeling mean-level change, and how these changes may differ across groups or levels of some other variable. For example, is there a difference in average math scores obtained across multiple assessments between boys and girls? Longitudinal 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. Additionally, the longitudinal signed-rank test, a nonparametric alternative to the paired t-test, 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 analyses, such as for evaluating intervention effects in pre-test/post-test designs (Kisbu-Sakarya, MacKinnon, and Aiken 2013).

Traditional longitudinal methods can still be useful in some contexts (e.g., < 3 measurement occasions), but 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. This is a particularly important limitation since most psychological theories posit within-person processes (i.e., what will happen within a given individual). As such, traditional approaches often correspond poorly with most theoretical models of change and a failure to disaggregate between-person and within-person effects can result in consequential errors of inference (e.g., ecological fallacy) (Patrick J. Curran and Bauer 2011). 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 (i.e., when baseline scores are correlated with a covariate of interest), which can produce different conclusions across methods (see Andrew K. Littlefield 2023, for a review). Given these shortcomings, and the complexity of the issues surrounding some of these methods, 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, an extension of Generalized Linear Models, 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 [although, ARCL models are relatively less useful for teasing apart between-person and within-person sources of variances; see Patrick J. Curran and Hancock (2021)].

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 among variables over time. Within-person variance in these models are captured by a series of latent variables that reflect time specific variance (i.e., the residual variance from the random intercept). These time-specific variables are referred to as structured residuals. Distinguishing between-person variance subsumed by the random intercept from the structured residuals 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 RI-CLPM by including additional growth factors, such as a random linear slope. That is, the LCM-SR is a hybrid between a latent growth model and CLPM. 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, even more so than the RI-CLPM, LCM-SR 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 the necessity of multiple assessments (e.g., four or more for LCM-SR), 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.

|  |
| --- |
| LDA Table |

# 5. Discussion

As we enter the era of large-scale longitudinal investigations, it is essential to critically examine the various analytical methods that can be employed to glean insights from these rich datasets. The complex nature of longitudinal data demands sophisticated and well-suited methodologies to accurately address research questions and minimize biases. This paper aimed to provide a comprehensive review of diverse longitudinal analysis techniques, with a particular emphasis on their application to extensive longitudinal studies such as the ABCD Study. Beyond contributing to the ever-growing body of knowledge on longitudinal data analysis, we hope this manuscript also serves as a valuable resource for researchers seeking to optimize the use of large-scale longitudinal investigations in advancing our understanding of human development and behavior. In this discussion, we will focus on the key findings and recommendations of our review and discuss potential innovations that can further enhance the utility of these methods.

We began by addressing fundamental concepts and considerations in longitudinal research that are essential for generating accurate and meaningful insights into developmental processes. Concepts such as vulnerable periods, developmental disturbances and snares, or cascade and experience effects (among many others), are instrumental in shaping the design, analysis, and interpretation of longitudinal studies. For example, understanding and accounting for vulnerable periods and developmental disturbances in research designs allows researchers to investigate the timing and magnitude of the impact of specific factors on development, while snares and cascade effects demonstrate the importance of considering the interconnectedness of developmental processes. Together, these concepts provide a framework for understanding the mechanisms underlying the course of development, while also accounting for the complex interplay between individual development and the influence of environmental factors. By considering the intricate relationships among these factors, researchers can better identify the critical time periods, situations, and contexts that contribute to individual differences in developmental outcomes. This awareness enables more precise inferences regarding the causal relationships between exposures and outcomes, ultimately leading to more robust and meaningful findings that can help facilitate the translation of research findings into practical applications in clinical and public health settings.

We also discussed some of the opportunities, challenges, and pitfalls that arise when working with longitudinal data. Key issues include selecting appropriate methods to account for the intricacies of longitudinal data, addressing missing data in a way that minimizes biases, and determining suitable longitudinal data structures that align with research questions and context. To address these challenges, researchers should carefully consider issues such as study design, selection of methods that account for both within- and between-person sources of variance, and employing modern techniques, (e.g., FIML, multiple imputation) for handling missing data. By adhering to best practices in longitudinal research and remaining vigilant of potential pitfalls, researchers can effectively harness the power of longitudinal data to maximize the potential of their investigations and gain valuable insights into complex developmental processes, individual differences, and the underlying mechanisms that drive change over time.

The final 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. 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 research questions, researchers can ensure that their longitudinal analyses are both rigorous and informative. Notably, 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.

# 6. Part V: References

Arseneault, Louise, Mary Cannon, Richie Poulton, Robin Murray, Avshalom Caspi, and Terrie E Moffitt. 2002. “Cannabis Use in Adolescence and Risk for Adult Psychosis: Longitudinal Prospective Study.” *Bmj* 325 (7374): 1212–13.

Bagot, K. S., S. A. Matthews, M. Mason, Lindsay M. Squeglia, J. Fowler, K. Gray, M. Herting, A. May, Ian Colrain, and J. Godino. 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.” *Developmental Cognitive Neuroscience* 32: 121–29.

Barch, Deanna M., Matthew D. Albaugh, Shelli Avenevoli, Linda Chang, Duncan B. Clark, Meyer D. Glantz, James J. Hudziak, Terry L. Jernigan, Susan F. Tapert, and Debbie Yurgelun-Todd. 2018. “Demographic, Physical and Mental Health Assessments in the Adolescent Brain and Cognitive Development Study: Rationale and Description.” *Developmental Cognitive Neuroscience* 32: 55–66.

Bauer, D. J., and P. J. Curran. 2019. “Conducting Longitudinal Data Analysis: Knowing What to Do and Learning How to Do It.” <https://centerstat.org/wp-content/uploads/2019/03/SRCD2019-2up.pdf>.

Bechtold, Jordan, Alison Hipwell, David A Lewis, Rolf Loeber, and Dustin Pardini. 2016. “Concurrent and Sustained Cumulative Effects of Adolescent Marijuana Use on Subclinical Psychotic Symptoms.” *American Journal of Psychiatry* 173 (8): 781–89.

Beck, Aaron T., Calvin H. Ward, Mock Mendelson, Jeremiah Mock, and John Erbaugh. 1961. “An Inventory for Measuring Depression.” *Archives of General Psychiatry* 4 (6): 561–71.

Bleidorn, Wiebke, Ted Schwaba, Anqing Zheng, Christopher J Hopwood, Susana S Sosa, Brent W Roberts, and DA Briley. 2022. “Personality Stability and Change: A Meta-Analysis of Longitudinal Studies.” *Psychological Bulletin* 148 (7-8): 588.

Casey, B. J., Tariq Cannonier, May I. Conley, Alexandra O. Cohen, Deanna M. Barch, Mary M. Heitzeg, Mary E. Soules, et al. 2018. “The Adolescent Brain Cognitive Development (ABCD) Study: Imaging Acquisition Across 21 Sites.” *Developmental Cognitive Neuroscience*, The Adolescent Brain Cognitive Development (ABCD) Consortium: Rationale, Aims, and Assessment Strategy, 32 (August): 43–54. <https://doi.org/10.1016/j.dcn.2018.03.001>.

Castro-Schilo, Laura, and Kevin J. Grimm. 2018. “Using Residualized Change Versus Difference Scores for Longitudinal Research.” *Journal of Social and Personal Relationships* 35 (1): 32–58. <https://doi.org/10.1177/0265407517718387>.

Cohen, Jacob. 1988. “Statistical Power.” *Analysis for the Behavioral Sciences*, 273–406.

Collins, Linda M. 2006. “Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model.” *Annu. Rev. Psychol.* 57: 505–28.

Cotter, Robert B., Jeffrey D. Burke, Magda Stouthamer-Loeber, and Rolf Loeber. 2005. “Contacting Participants for Follow-up: How Much Effort Is Required to Retain Participants in Longitudinal Studies?” *Evaluation and Program Planning* 28 (1): 15–21.

Crews, Fulton T, Christopher J Braun, Blair Hoplight, Robert C Switzer III, and Darin J Knapp. 2000. “Binge Ethanol Consumption Causes Differential Brain Damage in Young Adolescent Rats Compared with Adult Rats.” *Alcoholism: Clinical and Experimental Research* 24 (11): 1712–23.

Curran, Patrick J., and Daniel J. Bauer. 2011. “The Disaggregation of Within-Person and Between-Person Effects in Longitudinal Models of Change.” *Annual Review of Psychology* 62: 583–619.

Curran, Patrick J, and Gregory R Hancock. 2021. “The Challenge of Modeling Co-Developmental Processes over Time.” *Child Development Perspectives* 15 (2): 67–75.

Curran, Patrick J., Andrea L. Howard, Sierra Bainter, Stephanie T. Lane, and James S. McGinley. 2014. “The Separation of Between-Person and Within-Person Components of Individual Change Over Time: A Latent Curve Model with Structured Residuals.” *J Consult Clin Psychol* 82 (5): 879–94. <https://doi.org/10.1037/a0035297>.

Duncan, Terry E., and Susan C. Duncan. 2009. “The ABC’s of LGM: An Introductory Guide to Latent Variable Growth Curve Modeling.” *Social and Personality Psychology Compass* 3 (6): 979–91. <https://doi.org/10.1111/j.1751-9004.2009.00224.x>.

Durbin, James, and Geoffrey S. Watson. 1950. “Testing for Serial Correlation in Least Squares Regression: I.” *Biometrika* 37 (3/4): 409–28.

Enders, Craig K. 2010. *Applied Missing Data Analysis*. Guilford Press.

Ewing, Sarah W Feldstein, Genevieve F Dash, Wesley K Thompson, Chase Reuter, Vanessa G Diaz, Andrey Anokhin, Linda Chang, et al. 2022. “Measuring Retention Within the Adolescent Brain Cognitive Development (ABCD) SM Study.” *Developmental Cognitive Neuroscience* 54: 101081.

Fan, Chun Chieh, Andrew Marshall, Harry Smolker, Marybel R Gonzalez, Susan F Tapert, Deanna M Barch, Elizabeth Sowell, et al. 2021. “Adolescent Brain Cognitive Development (ABCD) Study Linked External Data (LED): Protocol and Practices for Geocoding and Assignment of Environmental Data.” *Developmental Cognitive Neuroscience* 52: 101030.

Feingold, Alan. 2009. “Effect Sizes for Growth-Modeling Analysis for Controlled Clinical Trials in the Same Metric as for Classical Analysis.” *Psychological Methods* 14 (1): 43.

———. 2019. “Time-Varying Effect Sizes for Quadratic Growth Models in Multilevel and Latent Growth Modeling.” *Structural Equation Modeling: A Multidisciplinary Journal* 26 (3): 418–29.

French, David P., and Stephen Sutton. 2010. “Reactivity of Measurement in Health Psychology: How Much of a Problem Is It? What Can Be Done about It?” *British Journal of Health Psychology* 15 (3): 453–68.

Gonzalez, Raul, Erin L Thompson, Mariana Sanchez, Amanda Morris, Marybel R Gonzalez, Sarah W Feldstein Ewing, Michael J Mason, et al. 2021. “An Update on the Assessment of Culture and Environment in the ABCD Study: Emerging Literature and Protocol Updates over Three Measurement Waves.” *Developmental Cognitive Neuroscience* 52: 101021.

Graham, John W. 2009. “Missing Data Analysis: Making It Work in the Real World.” *Annual Review of Psychology* 60 (1): 549–76. <https://doi.org/10.1146/annurev.psych.58.110405.085530>.

Hagler, Donald J., SeanN. Hatton, M. Daniela Cornejo, Carolina Makowski, Damien A. Fair, Anthony Steven Dick, Matthew T. Sutherland, et al. 2019. “Image Processing and Analysis Methods for the Adolescent Brain Cognitive Development Study.” *NeuroImage* 202 (November): 116091. <https://doi.org/10.1016/j.neuroimage.2019.116091>.

Hamaker, Ellen L., Rebecca M. Kuiper, and Raoul P. P. P. Grasman. 2015. “A Critique of the Cross-Lagged Panel Model.” *Psychological Methods* 20 (1): 102–16. <https://doi.org/10.1037/a0038889>.

Hasan, Alkomiet, Rupert von Keller, Chris Maria Friemel, Wayne Hall, Miriam Schneider, Dagmar Koethe, F Markus Leweke, Wolfgang Strube, and Eva Hoch. 2020. “Cannabis Use and Psychosis: A Review of Reviews.” *European Archives of Psychiatry and Clinical Neuroscience* 270: 403–12.

Hill, Karl G., Danielle Woodward, Tiffany Woelfel, J. David Hawkins, and Sara Green. 2016. “Planning for Long-Term Follow-up: Strategies Learned from Longitudinal Studies.” *Prevention Science* 17 (7): 806–18.

Hoffman, Lesa. 2015. *Longitudinal Analysis: Modeling Within-Person Fluctuation and Change*. Routledge.

Infante, MA, SC Eberson, Y Zhang, T Brumback, SA Brown, IM Colrain, FC Baker, et al. 2022. “Adolescent Binge Drinking Is Associated with Accelerated Decline of Gray Matter Volume.” *Cerebral Cortex* 32 (12): 2611–20.

Ji, Zhe, Lin Yuan, Xiong Lu, Hanqing Ding, Jia Luo, and Zun-Ji Ke. 2018. “Binge Alcohol Exposure Causes Neurobehavioral Deficits and GSK3 Activation in the Hippocampus of Adolescent Rats.” *Scientific Reports* 8 (1): 3088.

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.” *Inc., Cary, North Carolina*.

King, Kevin M., Andrew K. Littlefield, Connor J. McCabe, Kathryn L. Mills, John Flournoy, and Laurie Chassin. 2018. “Longitudinal Modeling in Developmental Neuroimaging Research: Common Challenges, and Solutions from Developmental Psychology.” *Developmental Cognitive Neuroscience*, Methodological Challenges in Developmental Neuroimaging: Contemporary Approaches and Solutions, 33 (October): 54–72. <https://doi.org/10.1016/j.dcn.2017.11.009>.

Kisbu-Sakarya, Yasemin, David P MacKinnon, and Leona S Aiken. 2013. “A Monte Carlo Comparison Study of the Power of the Analysis of Covariance, Simple Difference, and Residual Change Scores in Testing Two-Wave Data.” *Educational and Psychological Measurement* 73 (1): 47–62.

Kraemer, Helena Chmura. 2014. “Effect Size.” *The Encyclopedia of Clinical Psychology*, 1–3.

Lenz, Sylvia Tamara. 2016. “Alan Agresti (2013): Categorical Data Analysis.” *Statistical Papers* 57 (3): 849.

Lisdahl, Krista M, Susan Tapert, Kenneth J Sher, Raul Gonzalez, Sara Jo Nixon, Sarah W Feldstein Ewing, Kevin P Conway, et al. 2021. “Substance Use Patterns in 9-10 Year Olds: Baseline Findings from the Adolescent Brain Cognitive Development (ABCD) Study.” *Drug and Alcohol Dependence* 227: 108946.

Little, Roderick J., and Donald B. Rubin. 1989. “The Analysis of Social Science Data with Missing Values.” *Sociological Methods & Research* 18 (2-3): 292–326. <https://doi.org/10.1177/0049124189018002004>.

Little, Todd D. 2013. *The Oxford Handbook of Quantitative Methods, Vol. 2: Statistical Analysis*. Oxford University Press.

Littlefield, Andrew K. 2023. “Longitudinal Data Analysis.” In *APA Handbook of Research Methods in Psychology: Data Analysis and Research Publication, Vol. 3*, edited by H. E. Cooper, M. N. Coutanche, L. M. McMullen, A. T. Panter, D. E. Rindskopf, and K. J. Sher, 245–68. American Psychological Association.

Littlefield, Andrew K., Kevin M. King, Samuel F. Acuff, Katherine T. Foster, James G. Murphy, and Katie Witkiewitz. 2021. “Limitations of Cross-Lagged Panel Models in Addiction Research and Alternative Models: An Empirical Example Using Project MATCH.” *Psychology of Addictive Behaviors*. <https://doi.org/10.1037/adb0000750>.

Littlefield, Andrew K, Kevin M King, Samuel F Acuff, Katherine T Foster, James G Murphy, and Katie Witkiewitz. 2022. “Limitations of Cross-Lagged Panel Models in Addiction Research and Alternative Models: An Empirical Example Using Project MATCH.” *Psychology of Addictive Behaviors* 36 (3): 271.

Liu, Yu, Roger E Millsap, Stephen G West, Jenn-Yun Tein, Rika Tanaka, and Kevin J Grimm. 2017. “Testing Measurement Invariance in Longitudinal Data with Ordered-Categorical Measures.” *Psychological Methods* 22 (3): 486.

Loughnan, Robert J., Clare E. Palmer, Wesley K. Thompson, Anders M. Dale, Terry L. Jernigan, and Chun Chieh Fan. 2020. “Polygenic Score of Intelligence Is More Predictive of Crystallized Than Fluid Performance Among Children.” *bioRxiv*, 637512.

Luciana, M., J. M. Bjork, B. J. Nagel, D. M. Barch, R. Gonzalez, S. J. Nixon, and M. T. Banich. 2018. “Adolescent Neurocognitive Development and Impacts of Substance Use: Overview of the Adolescent Brain Cognitive Development (ABCD) Baseline Neurocognition Battery.” *Developmental Cognitive Neuroscience* 32: 67–79.

Masten, Ann S, Glenn I Roisman, Jeffrey D Long, Keith B Burt, Jelena Obradović, Jennifer R Riley, Kristen Boelcke-Stennes, and Auke Tellegen. 2005. “Developmental Cascades: Linking Academic Achievement and Externalizing and Internalizing Symptoms over 20 Years.” *Developmental Psychology* 41 (5): 733.

Moffitt, Terrie E. 2015. “Life-Course-Persistent Versus Adolescence-Limited Antisocial Behavior.” *Developmental Psychopathology: Volume Three: Risk, Disorder, and Adaptation*, 570–98.

Molnar, Frank J., Brian Hutton, and Dean Fergusson. 2008. “Does Analysis Using ‘Last Observation Carried Forward’ Introduce Bias in Dementia Research?” *Cmaj* 179 (8): 751–53.

Morris, Scott B., and Richard P. DeShon. 2002. “Combining Effect Size Estimates in Meta-Analysis with Repeated Measures and Independent-Groups Designs.” *Psychological Methods* 7 (1): 105.

Muthén, L. K. 2017. “Mplus User’s Guide. Los Angeles: Muthén & Muthén; 1998.”

Orth, Ulrich, D. Angus Clark, M. Brent Donnellan, and Richard W. Robins. 2021. “Testing Prospective Effects in Longitudinal Research: Comparing Seven Competing Cross-Lagged Models.” *Journal of Personality and Social Psychology* 120 (4): 1013.

Ren, Junting, Susan Tapert, Chun Chieh Fan, and Wesley K Thompson. 2022. “A Semi-Parametric Bayesian Model for Semi-Continuous Longitudinal Data.” *Statistics in Medicine* 41 (13): 2354–74.

Roberts, Brent W, and Wendy F DelVecchio. 2000. “The Rank-Order Consistency of Personality Traits from Childhood to Old Age: A Quantitative Review of Longitudinal Studies.” *Psychological Bulletin* 126 (1): 3.

Roberts, Brent W, Kate E Walton, and Wolfgang Viechtbauer. 2006. “Patterns of Mean-Level Change in Personality Traits Across the Life Course: A Meta-Analysis of Longitudinal Studies.” *Psychological Bulletin* 132 (1): 1.

Robins, Lee. 1985. “Epidemiology: Reflections on Testing the Validity of Psychiatric Interviews JAMA Psychiatry JAMA Network.” <https://jamanetwork.com/journals/jamapsychiatry/article-abstract/493658>.

Rogosa, David, David Brandt, and Michele Zimowski. 1982. “A Growth Curve Approach to the Measurement of Change.” *Psychological Bulletin* 92 (3): 726.

Rogosch, Fred A, Assaf Oshri, and Dante Cicchetti. 2010. “From Child Maltreatment to Adolescent Cannabis Abuse and Dependence: A Developmental Cascade Model.” *Development and Psychopathology* 22 (4): 883–97.

Salthouse, Timothy A. 2014. “Why Are There Different Age Relations in Cross-Sectional and Longitudinal Comparisons of Cognitive Functioning?” *Current Directions in Psychological Science* 23 (4): 252–56.

Schulenberg, John, Julie Maslowsky, Megan E Patrick, and Meghan Martz. 2019. “Substance Use in the Context of Adolescent Development.” In *The Oxford Handbook of Adolescent Substance Abuse*, edited by Robert A Zucker and Sandra A Brown, 19–36. Oxford University Press.

Semple, David M, Andrew M McIntosh, and Stephen M Lawrie. 2005. “Cannabis as a Risk Factor for Psychosis: Systematic Review.” *Journal of Psychopharmacology* 19 (2): 187–94.

Shrout, Patrick E, Gertraud Stadler, Sean P Lane, M Joy McClure, Grace L Jackson, Frederick D Clavél, Masumi Iida, Marci EJ Gleason, Joy H Xu, and Niall Bolger. 2018. “Initial Elevation Bias in Subjective Reports.” *Proceedings of the National Academy of Sciences* 115 (1): E15–23.

Spear, Linda Patia. 2016. “Consequences of Adolescent Use of Alcohol and Other Drugs: Studies Using Rodent Models.” *Neuroscience & Biobehavioral Reviews* 70: 228–43.

Sullivan, Edith V., Ty Brumback, Susan F. Tapert, Devin Prouty, Rosemary Fama, Wesley K. Thompson, Sandra A. Brown, Kevin Cummins, Ian M. Colrain, and Fiona C. Baker. 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.” *Developmental Cognitive Neuroscience* 24: 72–83.

Thompson, Wesley K., Deanna M. Barch, James M. Bjork, Raul Gonzalez, Bonnie J. Nagel, Sara Jo Nixon, and Monica Luciana. 2019. “The Structure of Cognition in 9 and 10 Year-Old Children and Associations with Problem Behaviors: Findings from the ABCD Study’s Baseline Neurocognitive Battery.” *Developmental Cognitive Neuroscience* 36: 100606.

Uban, Kristina A., Megan K. Horton, Joanna Jacobus, Charles Heyser, Wesley K. Thompson, Susan F. Tapert, Pamela A. F. Madden, and Elizabeth R. Sowell. 2018. “Biospecimens and the ABCD Study: Rationale, Methods of Collection, Measurement and Early Data.” *Developmental Cognitive Neuroscience*, The Adolescent Brain Cognitive Development (ABCD) Consortium: Rationale, Aims, and Assessment Strategy, 32 (August): 97–106. <https://doi.org/10.1016/j.dcn.2018.03.005>.

Van De Schoot, Rens, Peter Schmidt, Alain De Beuckelaer, Kimberley Lek, and Marielle Zondervan-Zwijnenburg. 2015. “Measurement Invariance.” *Frontiers in Psychology*. Frontiers Media SA.

Wang, Lijuan, Qian Zhang, Scott E Maxwell, and CS Bergeman. 2019. “On Standardizing Within-Person Effects: Potential Problems of Global Standardization.” *Multivariate Behavioral Research* 54 (3): 382–403.

Watson, Nicole, Eva Leissou, Heidi Guyer, and Mark Wooden. 2018. “Best Practices for Panel Maintenance and Retention.” In *Advances in Comparative Survey Methods*, 597–622. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118884997.ch29>.

Willoughby, Michael T, RJ Wirth, and Clancy B Blair. 2012. “Executive Function in Early Childhood: Longitudinal Measurement Invariance and Developmental Change.” *Psychological Assessment* 24 (2): 418.

Zucker, Robert A., Raul Gonzalez, Sarah W. Feldstein Ewing, Martin P. Paulus, Judith Arroyo, Andrew Fuligni, Amanda Sheffield Morris, Mariana Sanchez, and Thomas Wills. 2018. “Assessment of Culture and Environment in the Adolescent Brain and Cognitive Development Study: Rationale, Description of Measures, and Early Data.” *Developmental Cognitive Neuroscience* 32: 107–20.