Longitudinal Analysis of the ABCD® Study

Samuel W. Hawes
Andrew K. Littlefield
Daniel A. Lopez
Kenneth J. Sher
Erin L. Thompson
Raul Gonzalez
Laika Aguinaldo
Ashley R. Adams
Mohammadreza Bayat
Amy L. Byrd
Luis FS Castro-de-Araujo
Anthony Dick
Steven F. Heeringa
Christine M. Kaiver
Sarah M. Lehman
Lin Li
Janosch Linkersdörfer
Thomas J. Maullin-Sapey
Michael C. Neale
Thomas E. Nichols
Samantha Perlstein
Susan F. Tapert
Colin E. Vize
Margot Wagner
Rebecca Waller
Wesley K. Thompson

***Abstract*—The Adolescent Brain Cognitive Development® (ABCD) Study provides a unique opportunity to investigate developmental processes in a large, diverse cohort of youths, aged approximately 9-10 at baseline and assessed annually for 10 years. Given the size and complexity of the ABCD Study, researchers analyzing its data will encounter a myriad of methodological and analytical considerations. This review provides an examination of key concepts and techniques related to longitudinal analyses of the ABCD Study data, including: 1) characterization of the factors associated with variation in developmental trajectories; 2) assessment of how level and timing of exposures may impact subsequent development; 3) quantification of how variation in developmental domains may be associated with outcomes, including mediation models and reciprocal relationships. We emphasize the importance of selecting appropriate statistical models to address these research questions. By presenting the advantages and potential challenges of longitudinal 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 multidimensional longitudinal data currently available.**

***Index terms*—Longitudinal Analysis, ABCD Study, Development**

I. Introduction

The Adolescent Brain Cognitive Development (ABCD) Study® is the largest longitudinal 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 factors impacting neurodevelopment and their roles in behavioral and health outcomes across ten years of adolescence [1]. 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 commitment of the ABCD Study and its NIH funders 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. In this context, the ABCD Study is a collaboration with the broader 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, including: 1) characterization of the genetic and environmental factors associated with variation in developmental trajectories; 2) assessment of how the level and tim-

ing of exposures may impact subsequent neurodevelopment; 3) quantification of how variation in developmental domains may be associated with outcomes, including mediation models and reciprocal relationships. We instantiate these longitudinal analyses in worked examples using the ABCD Release 5.1 data with accompanying R scripts. Worked examples are available in Quarto files, accessible in the project's GitHub repository.

A. *The ABCD Study Data*

The ABCD Study enrolled a cohort of n=11,880 participants born between 2006-2008 and aged approximately 9-10 years at baseline, each with a parent/guardian. The study sample was recruited from households 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 organizations can be obtained at https://abcdstudy.org. The ABCD Study design is described in more detail in [2] and [3].

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 characterizing developmental trajectories and implementing longitudinal analyses that allow researchers to characterize normative development, to identify variables that presage deviations from normative development, and to assess a range of variables associated with biopsychosocial outcomes of interest. These data include: 1) a neurocognitive battery [4]; [5]; 2) mental and physical health assessments [6]; 3) measures of culture and environment [7]; [8]; 4) substance use [9]; 5) gender identity and sexual health [10]; 6) biospecimens [11]; 7) structural and functional brain imaging [12]; [13]; [14]; 8) geolocation-based environmental exposure data [15]; 9) wearables and mobile technology [16]; and 10) whole-genome genotyping [17]. 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 brief telephone or online assessments.

Data are publicly released approximately annually, currently through the NIMH Data Archive (NDA). The study's earliest data releases consisted primarily of one or two visits per participant. However, the most recent public release as of the writing of this paper (Release 5.1) contains data collected across five annual visits, including three brain imaging assessments (baseline, year 2 follow-up, and year 4 follow-up visits) for at least a subset of the cohort. Hence, starting with Release 5.0, it is feasible for researchers to begin focusing on the characterization of neurodevelopmental and other trajectories.

B.

i Organization and Aims

- Part I. Introduction
  - The ABCD Study®
- Part II. Developmental Research
  - Fundamental Concepts
- Part III. Longitudinal Analysis
  - Methods & Analysis
- Part IV. Supplemental materials
  - Linked open-source resources

II. Developmental Research

A. *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 the 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 represents the same or different phenomena.

Moreover, 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- from between-person 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 at different times enables estimation of individual rates of change (slopes) where more observations allow for more precise estimates of individual slopes (random slopes), as well as characterization of non-linear development. Rate of change or other trajectory characteristics may be more informative about individuals than the simple snapshots of level differences that cross-sectional data are limited to informing about. Cross-sectional age-related differences across individuals are poor substitutes for longitudinal trajectory estimates, except under highly restrictive assumptions, e.g., parallel trajectories and lack of age, cohort and experience effects [18]. 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.

1) *Vulnerable periods.:* Adolescent development progresses normatively 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 [19]; [20]; [21]; [22]. Similarly, rodent brains are especially sensitive to the neurotoxic effects of alcohol on brain structure and learning early in development, corresponding to early adolescence in humans [23]; [24]; [25]. In another example, longitudinal data from the National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) show that binge drinking is associated more strongly with decrements in gray matter volume early in adolescence compared to later [26]. 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 on the developing brain may be particularly pronounced– when trying to establish an accurate understanding of the association between exposures and outcomes.

2) *Developmental disturbances.:* Whereas vulnerable periods heighten neurobiological susceptibility to environmental influences, at other times, environmental exposures will tend to suppress stability and disrupt the orderly stochastic process of normative development (e.g., [27]). This situation reflects a developmental disturbance in that the normal course of development is "altered" for a 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 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.

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 deviant 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 [28]. 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., [29]; [30]). Although conceptually distinct from vulnerable periods, both types of developmental considerations highlight the importance of viewing behavior in the context of development and 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 substance use for the vast majority of participants).

4) *Mediational Processes.:* Questions regarding the biological mechanisms whereby exposures impact outcomes can often be framed in terms of mediation analyses [31]; [32]. Mediation analyses can be implemented using the causal steps approach [33] and structural equation models (SEM) [34]. More recently, mediation models have been adapted for longitudinal exposures, mediators, and/or outcomes [35]; [36]. All of these modeling approaches decompose the total effects of an exposure on an outcome into direct and indirect effects, where indirect effects of an exposure flow through its impact on a mediating process. [36] details conditions under which the direct and indirect causal effects can be in a longitudinal setting. An important example of mediational analyses in the ABCD Study is the impact of exposures on behavioral outcomes (e.g., neurocognition, mental health, substance use) via their impact on the brain, as quantified by imaging-derived phenotypes (IDPs). Methods for mediational analyses using multi-dimensional IDPs have been developed and applied to functional MRI data [37]; [38].

III. Longitudinal Data

A. Considerations and Challenges

The hallmark characteristic of longitudinal data analysis (LDA) 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 person-level factors that may impact these trajectories. Administering repeated measurements more frequently or over longer periods enables researchers to ask more nuanced questions and to make stronger inferences.

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 an exposure 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, analyses of data based on two assessments are inherently limited on multiple fronts. As [39] noted over forty years ago, "Two waves of data are better than one, but maybe not much better" (p. 744).

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 inferences. This is also consistent with recommendations 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 (given that linear change is the only estimable form for two assessment waves; (see [40]). 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 [41]– 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" [42]. 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 [LCM: SR]; [43]) require at least four assessments to parameterize fully and, more generally, increasingly accurate and nuanced parameter estimates are obtained as more assessment occasions are used [40].

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. Symbolically, developmental trajectories can be expressed as fi(t), a possibly multivariate function of time t, specific to the ith individual and typically taking values in the real numbers for continuous phenotypes and the integers for discrete phenotypes. Ideally, t is a biologically meaningful temporal index (e.g., calendar age) as opposed to an exogenous progression of events (e.g., study visit number). Properties of interest might include rate of change over time, degree of smoothness (e.g., continuously differentiable), shape (e.g., polynomial or asymptotic behavior), how and how much f(t) differs across individuals, and what factors predict either within-individual variation (at different times) or between-individual variation (either overall or at specific times).

There are a few different ways to think about patterns of stability and change (see Figure 1). Consider measuring school disengagement at the start of middle school and the end of middle school. A common first step may be to compare sixth graders' average disengagement values and eighth graders' disengagement values. This comparison of the average scores for the same group of individuals at multiple time points is referred to as "mean-level", as it provides information about change over time (or lack thereof) for an outcome of interest aggregated across members of a group. In contrast, "between-individual" stability could be assessed, e.g., by calculating the Spearman correlation between the values obtained at different time points (e.g., 'disengagement in sixth grade' with 'disengagement in eighth grade). This analysis focuses on the degree to which individuals retain their relative placement in a group across time. Consider someone who reported the lowest frequencies of disengagement in 6th grade and may report significantly higher disengagement over middle school (i.e., exhibit high levels of change), but report the lowest frequencies of disengagement in eighth grade. That is, the individual is manifesting rank-order stability, even in the context of high mean-level change.

Both types of stability and change are important. Mean-level change in certain traits might help to explain why, in general, populations of individuals tend to be particularly vulnerable to the effects of environmental factors in specific age ranges; rank-order stability might help to quantify the extent to which certain characteristics of the individual are more or less trait-like compared to others. For example, in some areas of development, considerable mean-level change occurs over time (e.g., changes in Big 5 personality traits [44], but exhibit relatively high rank-order stability, at least over shorter measurement intervals [44]; [45]; [46].

Despite the useful information afforded by examining mean-level and rank-order stability and change, these approaches are limited in that they provide little information about the overall patterns of within-individual change and, in turn, can result in fundamental misinterpretations about substantial or meaningful changes in an outcome of interest [47]. For example, questions related to the impact of early-onset substance use on brain development focus on changes within a given individual (i.e., intraindividual differences). The ABCD Study will provide researchers with over ten time points for certain constructs (e.g., substance use) across a ten-year pe-

riod, allowing for a detailed study of some within-person processes.

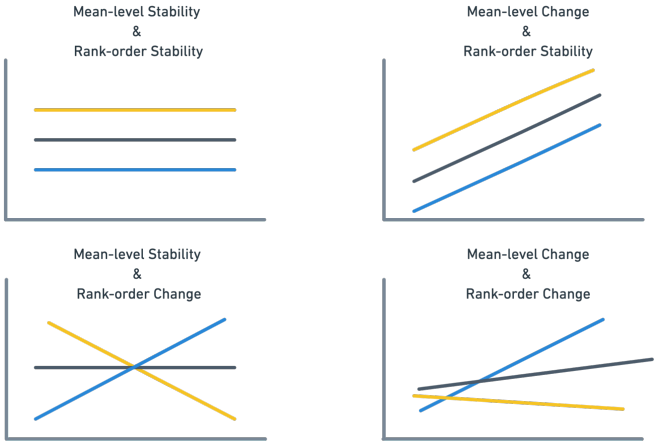


Figure 1: Types of Stability and Change

3) *Use of appropriate longitudinal models:* 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., [47]; [48]; [42]). First, developmental studies typically involve the use of prospective data to inform theories that are concerned with clear within-person processes (e.g., how phenotypes change or remain stable within individuals over time, (e.g., [47])). 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 [42]; [49]). 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., random-intercept cross-lagged panel model [RI-CLPM], [50]); latent curve models with structured residuals [LCM-SR], [43]). 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 [42], for further discussion).

Second, 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 and equality of variances. These assumptions should be carefully considered before finalizing analytical approaches, 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. Another pertinent modeling assumption is whether trajectories are linear or non-linear. With two or three assessments per individual, usually only a linear model of within-person change is feasible.

As the study progresses and more time points are assessed, the potentially nonlinear aspects of trajectories can be assessed, for example using quadratic functions of time. Methods that make even fewer assumptions about trajectory shapes, such as nonparametric curve estimation at the mean (e.g., Generalized Additive Mixed Models [GAMMs]; [51]) and at the individual level (e.g., Functional Data Analysis [FDA]; [52]) may also become useful. Note, baseline age in the ABCD Study ranges over two full years; for some outcomes it may be feasible to include a possibly nonlinear effect of baseline age along with a linear effect of within-person change in age even with only two or three assessment times [53].

4) *Continuous and Discrete Outcomes:* Repeated assessments within the ABCD Study 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., caregiver-reported items measuring emotional and behavioral concerns via the Child Behavior Checklist including the categories of "Not True", "Somewhat True", and "Very True"), and count variables (e.g., number of cigarettes smoked per day). 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 [54]). For example, the Mplus manual [55] 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 [56]. These models account for issues that may occur when working with discrete outcomes, including overdispersion, i.e., when the variance is higher than would be expected based on a given parametric distribution (see [54]). Given the sheer breadth of issues relevant to determining adequate models for discrete outcomes, it is not uncommon for texts on LDA to only cover models and approaches that assume continuous variables (e.g., [57]). 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: [54], Chapter 11 for matched pair/two-as-

sessment 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.

5) *Issues in attributing longitudinal change to development:* Systematic changes over time in a variable of interest are not always attributable to development: various pitfalls with longitudinal data can complicate or even invalidate this conclusion. For example, if data missingness or participant dropout are related to the values of the outcome, changing sample composition as the study progresses can bias mean trajectory estimates (we describe this in more detail in Section 3.1.7 below). Another prerequisite for valid developmental interpretations of longitudinal data is to establish whether a construct is measured consistently over time, i.e., longitudinal measurement invariance [58]; [59]; [60]. Establishing longitudinal measurement invariance ensures that change over time for a given construct is attributable to individual development rather than merely a measurement artifact. For instance, one study using data from the ABCD Study [61] found differential item functioning in two items from a brief delinquency measure, revealing significant bias in an arrest item across Black and White youth. More specifically, Black youth were more likely to report being arrested compared to White youth with similar levels of delinquency. Prevalence rates of delinquent behavior would have been severely biased if measurement invariance had not been tested. Alternatively, [62] showed partially strong to strong evidence of longitudinal measurement invariance across broad externalizing dimensions in youth taking in the ABCD Study, suggesting that changes observed over time in these constructs were not due to systematic measurement error, but likely reflect true developmental change.

Observed patterns of growth and decline often differ between cross-sectional vs. longitudinal effects [63] 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 [64] 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 [63]), 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., [65]; [66]) and this phenomenon has been well documented in research using structured diagnostic interviews [67]. 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 artifactually elevated (see [68]).

Some longitudinal studies, e.g., accelerated longitudinal designs (ALDs; [18]) are especially well suited for discovering these effects and modeling them. While ABCD is not an ALD, the variability in age (and grade in school) at the time of baseline recruitment (approximately 9-10 years old) allows some measures, collected every year, to be conceptualized as an ALD (e.g., substance use; prosocial behavior; family conflict; screen time). It is also 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 number of prior assessments and age are mostly confounded, and for, perhaps, most analyses, the possible influence of experience effects needs to be kept in mind.

6) *Modeling Covariance:* A central issue for repeated measurements on an individual is how to account for the correlated nature of the data. Lack of independence of residuals across time occurs for longitudinal data with repeated assessments on individuals and in other situations with nested data (e.g., visits nested within participants, children nested within schools; siblings nested within families). Note, the ABCD Study has multiple levels of nesting, depending on the analysis, including within-participant, within-family, within-school, within-MRI scanner, and within-site.

Statistical models for nested data include two main components, coupling a model for the mean response and its dependence on covariates with a model for the covariance among repeated outcomes on an individual. In contrast, traditional methods, such as multiple regression and ANOVAs, assume residuals are independent and thus are generally inappropriate for designs that incorporate some type of nesting. Specifically, given that residuals are no longer independent in a repeated measures design, standard errors from these models are biased and can produce misleading inferences. Therefore, an initial question to be addressed by a researcher analyzing prospective data is how to best model their covariance structure. A range of methods can be used 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 [69]).

The most common approach is to use random effects. Essentially, random effects allow for covariance estimates around fixed effects. A classic example (from [70]; [71]) involves math achievement measured among students nested within schools. In a basic, intercept-only model with no covariates (i.e., an unconditional growth model), there would be one fixed effect (the grand mean, or intercept, of math achievement),

one school random effect (representing variation in the intercept between schools) and the within-school student residuals (variation left over after accounting for fixed and random effects). In this framework, each student's score would be the sum of the fixed effect (the grand mean), the school random effect and the student's within-school residual. Assumptions about the variance and covariance components of this model dictate the form of the variance/covariance structure. For example, if we assume the random effects are independent and identically distributed, the implied structure would be compound symmetry, where it is assumed the covariance of any two students in a single school is captured by a school random intercept and the covariance of any two students in different schools is zero. The assumptions of this relatively simple covariance structure can be relaxed depending on the nesting structure of the data, resulting in different covariance structures with additional parameters (see [71]).

In longitudinal studies, visits are nested within individuals. Mixed-effect models can be fitted to longitudinal data that couple a model for growth (development) at the mean level with a model for capturing within-individual covariance of assessments. For example, a linear growth model would involve two fixed effects – one for the intercept (the average score when time is coded zero) and one for the linear slope (the change in scores for each unit increase in time). Random effects could include a random effect for intercept, capturing individual variation in scores at time zero, and a random effect for the linear slope, capturing individual variation in linear change across time. Within-individual residuals account for the remaining variation in assessments after accounting for the fixed and random intercepts and slopes. Assumptions regarding the covariation among the random effects also indicate different covariance structures. For example, it is typical to assume that the random intercept and slope components covary, i.e., an individual's score at time zero relates to the amount of change exhibited across time. Further, particularly in structural equation model forms of this model, it is sometimes assumed that the variance of the residuals varies across assessments [72].

An alternative to random effects is the autoregressive structure, which allows for correlations between repeated assessments to diminish across time. As the name suggests, the structure assumes the residual of a subsequent measurement occasion (e.g., visit 2) is regressed onto the residual of a prior measurement occasion (e.g., baseline visit). The most common type of autoregressive structure is the AR(1), where residuals at time t + 1 are regressed on residuals at time t. Identical to compound symmetry, this model assumes the variances are homogenous across time; however, it differs from compound symmetry in that the correlations between repeated assessments decline exponentially across visits 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 visit 1 and visit 2 data is thought to be .5, then the correlation between visit 1 and visit 3 data would be assumed to be .5 × .5 = .25, and the correlation between visit 1 and visit 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 residuals 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., CLPM). These designs still typically assume an AR(1) process. However, the magnitude of these relations is often allowed to differ across different AR(1) pairs of assessment (e.g., the relation between visit 1 and visit 2 can be different from the relation between visit 2 and visit 3). These 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 [50]. Note also, discrete-time autoregressive structures such as AR(1) implicitly assumes relatively constant time gaps between visits; this may not be true in many applications using the ABCD Study data.

7) *Missing Data/Attrition:* Attrition from a longitudinal study such as ABCD is inevitable and represents a potential threat to the external validity of analyses conducted at later visits, especially since attrition can only be expected to grow over time [73]. The ABCD Retention Workgroup 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, formal attrition in ABCD has been minimal [74]. Ideally, one tries to minimize attrition through good retention practices from the outset via strategies designed to maintain engagement in the project [75]; [76]; [77]. 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.

Three types of missingness are considered in the literature [78]; [57], namely: a) missing completely at random (MCAR), b) missing at random (MAR), and c) missing not at random (MNAR). Data that are MCAR are a simple random sample of all data in a given dataset. 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 associations even after conditioning on the observed covariates. [79] provides an excellent and easy-to-digest overview of further details involving missing data considerations.

Modern approaches for handling missing data, such as full-information maximum likelihood, propensity weighting, auxiliary variables and multiple imputation avoid the biases of older approaches (see [80]; [79]). [79] 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 listwise deletion in most circumstances, regardless of the model of missingness (i.e., MCAR, MAR, MNAR; see [79]; but also see [81]). The ABCD Biostatistics Workgroup is currently implementing several missing data approaches which are being implemented and compared to each other (and listwise deletion) in the 5.0 data release, including, propensity score weighting, and multiple (multilevel) imputation.

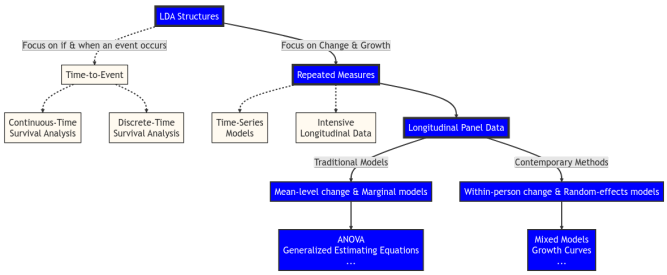
8) *Quantifying effect sizes longitudinally:* Given that longitudinal data involve multiple sources of variation, quantifying effect sizes longitudinally is more complex compared to deriving such estimates from cross-sectional data. An effect size can be defined as, "a population parameter (estimated in a sample) encapsulating the practical or clinical importance of a phenomenon under study." (Kraemer 2014). Common effect size metrics include the Pearson correlation r between two variables and the standardized difference between two means, Cohen's d [82]. An extensive discussion of cross-sectional effect sizes and their relevance for ABCD is given in [3].

Adjustments to common effect size calculations, such as Cohen's d, are required even when only two time points are considered (e.g., [83]). [84] note there are multiple approaches to obtaining standardized within-person effects, and that commonly suggested approaches (e.g., global standardization) can be problematic (see [84], 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 longitudinal data. [85] 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 cross-sectional analysis (see [85], for more details).

Given this issue, there have been various proposals for adjusting effect size measures in repeated assessments. [86] reviews the approach for effect size metrics for analyses based on growth modeling, including when considering linear and non-linear (e.g., quadratic) growth factors. [83] 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 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 relationships.

9) *Longitudinal Data Structures:* An ideal longitudinal analysis 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 [87]. To accommodate various research questions and contexts, different types of longitudinal data and data structures have emerged (see Figure 1). An understanding of these data structures is helpful, as they can warrant different types of LDA. Given that identifying a starting point for making comparisons is somewhat arbitrary, Curran and [88] 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 [88] 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 [89]; [90], including a recent analysis examining exclusionary discipline in schools using data from the ABCD Study [61], our emphasis will be given to the modeling of "repeated measures" data.



Link

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 ($\geqq$ 50-100 measurements) on a single or small number of variables of interest. Time-series analysis is often used to predict temporal trends and cyclic patterns and is geared toward making inferences about prospective outcomes within a population (with relatively 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 (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 to model individual difference outcomes.

The final type of repeated measures analysis that we will primarily focus on is the longitudinal panel study. These models follow a group of individuals— a panel (also referred to as a cohort) — across relatively fewer measurement occasions (~ 5-15) and are often focused on examining both change within- and between-individuals. The ABCD Study is primarily a longitudinal panel study, though some data streams (e.g., functional brain imaging, FitBit data) could be analyzed as ILP or even time series methods.

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.

IV. Longitudinal Analysis

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. This section aims to help researchers navigate these many options to identify the statistical approach most appropriate to their unique research question when deciding on how to measure change over time. 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 linear vs nonlinear models [87], the number of measurement occasions [41], and statistical equivalency (e.g., change scores vs. residualized change; see [91]). The organization we use below overlaps in several ways with these examples, and in particular with [88]. However, it is important to note that in each case, the chosen way of grouping is primarily intended to allow the reader to compare and contrast various analytical approaches. In the following sections, we briefly summarize the advantages/disadvantages of a series of longitudinal models organized into the following groupings: Traditional Models, Modern GLM Extensions, Structural Equation Models (SEM), and Advanced SEM (see Figure 2). We note that this is not an exhaustive review of each of these methods, and for more in-depth detail we do provide the reader with relevant resources. As aptly summarized by [88], "...there are many exceptions, alternatives, nuances, 'what ifs', and 'but couldn't you's that aren't addressed here."

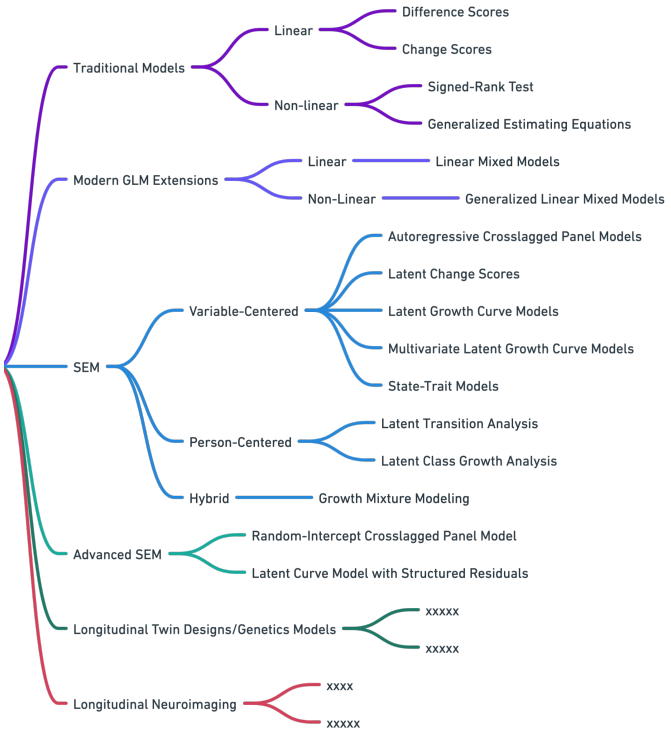


Figure 2: Longitudinal Models/Data Structures

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 internalizing symptoms 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, ANCOVA and Generalized Estimating Equations (GEEs). Mean-level change models are commonly used when data are only available from 2 measurement occasions. For example, computing a difference score (e.g., mean internalizing scores at visit 2 - mean internalizing scores at visit 1) 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 but can be used with ABCD Study data, is to use residualized change score analysis to assess the degree of change in a variable, while controlling for its initial level [91].

For example, to examine change in cortico-limbic connectivity among ABCD participants, [92] regressed cortico-limbic connectivity at the year 2 follow-up on baseline cortico-limbic connectivity, which allowed the authors to examine the associations between negative life events and the variance of cortico-limbic connectivity unexplained by baseline connectivity. Similarly, [93] used a residualized-change model to examine the bidirectional influences of executive functioning and a general psychopathology factor 'p' across the first two years of the ABCD Study. Both studies were able to conclude associations between their constructs of interest that could not be accounted for by prior frequencies at baseline.

Traditional longitudinal models, such as residualized change score models, can be useful in some contexts (e.g., two 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. 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, [47]). 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 (e.g., see [94], 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 three or more time points, as is currently true for many ABCD Study assessments.

Modern GLM Extensions :

Modern approaches to LDA 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 and is suitable for analyzing correlated longitudinal data and modeling population-averaged effects. For example, [95] used GEE to obtain relative risks for psychiatric diagnoses among children in the ABCD Study with a family history of depression and used the ABCD Study sampling weights to generalize prevalence rates among 9 and 10-year-olds across the US.

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. Within the ABCD Study, researchers may want to consider nesting by individual, family (i.e., siblings or twins), school or district, and/or site. GLMMs further extend the LMM framework to accommodate non-normal response variables, such as binary, count, or ordinal data, such as the use of ABCD data on substance use (e.g., [96]) screen media use [97], and microstructure of the brain [14].

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 [98].

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.

Structural Equation Modeling (SEM) :

Structural Equation Modeling (SEM) is a flexible modeling framework that integrates elements of path analysis and confirmatory factor analysis (CFA) to examine complex relationships between a set of observable variables and latent constructs [99]. The integration of structural (regression) and measurement (CFA) components within a unified framework supports a theory-driven approach that allows researchers to rigorously test hypothesized relationships among variables of interest and their underlying causes [99]; [100]. Over the years, the flexibility of the SEM framework has evolved to become particularly adept for modeling autoregressive processes (which often assume underlying stationarity) and growth processes which accommodate both, mean trajectories and individual differences in them [101]; [57].

Longitudinal SEM techniques share many similarities with mixed-effects methods and research demonstrates their mathematical equivalence in many situations [72]; [102]. However, these related approaches often cater to distinct theoretical and analytical needs. For instance, mixed-effects techniques are an extension of the regression framework and often excel when working with complex data structures such as multiple levels of nesting, small samples, and non-equidistant time points [103]; [104]. Alternatively, applying SEM methods to longitudinal analysis provides a flexible means for modeling the underlying process of change. It also addresses several challenges faced by competing approaches, including the ability to accommodate intricate error structures and deal effectively with missing data, as well as the implementation of numerous modeling extensions [103]; [98]. These models have grown increasingly popular for modeling longitudinal outcomes particularly due to their ability to build statistical models that match some particular underlying theory [105].

Considering the variety of available techniques, it can be helpful to classify longitudinal SEM, broadly (if not coarsely), into variable-centered, person-centered, and hybrid analyses, each with unique strengths and limitations. Variable-centered analyses (e.g., latent growth curves [72], latent change scores [106], latent state-trait models [107]) are primarily concerned with understanding covariation among variables at the group level and characterizing population-level patterns of change, while person-centered analyses (e.g., latent class and latent transition models) identify distinct subgroups or patterns within the data [108]; [109]; [110]. Hybrid models combine these perspectives to offer a comprehensive analysis of latent subgroups and growth parameter relationships [111]; [112]. The choice between these approaches is primarily driven by the research question, data structure, and relevant underlying assumptions.

Variable-centered models :

One key application of the SEM framework to the analysis of longitudinal data is the latent growth curve model (LGCM). This is a variable-centered approach that characterizes average group trajectories and individual variations (random effects) in an outcome over time [72]. These models are similar to their linear mixed effects counterpart in many ways, with the main conceptual difference being that LGCM includes a repeatedly measured outcome in the model as a function of time (closely resembling a standard CFA approach), rather than as an explanatory variable (as in a standard regression approach) [103]. Specifically, observed scores at each time point are treated as indicator variables with their factors loading scaled to reflect a hypothesized pattern of change (e.g., loadings of 0, 1, and 2 would assume equidistant, linear change). Latent intercepts (initial levels) and slopes (rates of change) are estimated, along with their variances and covariance to capture common trends and individual deviations over time. This method was used in a recent study by [113] to show a decreasing trajectory of parent-reported externalizing behaviors from ages 9-12 among youth taking part in the ABCD Study. This study also examined hypothesized predictors of the growth trajectory intercept and slope factors, highlighting a particular strength of these models–– their flexibility and extensibility. As an example, [114] used publicly available data from the ABCD Study and several other large-scale datasets to explore bivariate (parallel process) relationships between

white matter pathways and literacy over time. Beyond these examples, LGCMs can be extended in numerous ways, including to compare rates of growth across groups, investigate the consequences of change, and incorporate time-invariant or time-varying covariates, to highlight only a few (for a more detailed treatment of LGCM applications and methodologies, refer to [115]; [116]; [117]).

The latent change score model (LCSM) is a variable-centered approach uniquely tailored for analyzing temporal variations in how a construct changes over time [118]; [106]. These models share many features with growth curve analysis, but with a more explicit focus on how change occurs between measurement occasions [105]; [101]. Specifically, LCSM estimates a series of latent variables to model change in an outcome from one time point to the next, as a function of scores on that outcome at prior time points [106]; [119]. Some types of LCSM estimate two underlying latent factors: a constant change factor that remains fixed over time, and a proportional change factor that adjusts for previous scores. By disaggregating change into constant and proportional components, this approach facilitates a more nuanced understanding of whether prior changes in a given process are related to future changes in the same process [105]; [120]. Expanding upon the capabilities of this framework, LCSM also allows for comprehensive multivariate analyses that can facilitate investigations into how change in one construct is associated with change in another construct. The appeal of this approach is evidenced by several recent studies that have used data from the ABCD Study to explore bivariate associations between brain development and changes in several mental and physical health indicators [121]; [122]; [123]; [124]; [125].

Latent State-Trait Models (LSTM) offer another variable-centered approach to longitudinal analysis that also allows for the estimation of patterns of change over time. Unlike LGCM, which conceptualizes change as a function of time, and LCSM, which views change through sequential measurements, this approach disaggregates observed behaviors into distinct stable (trait) and occasion-specific (state) components [126]; [127]. Based on LST theory [128]; [129], these models hold that scores on a repeated measures outcome can be partitioned into an enduring latent trait variable that reflects between-individual differences, and a transient latent state residual that represents situational influences [130]; [107]. Beyond parsing out these key variance components, LSTM can be extended in many ways, such as by incorporating autoregressive effects to capture relative stability and the influence of past states on future responses (i.e. carry-over effects; [131]; [132]; [107]). The merits of this approach are highlighted in a recent review by [133] focused on strategies for modeling neurobehavioral development. These study authors encourage researchers to leverage data from the ABCD Study and other large-scale longitudinal and publicly available datasets and to apply state-trait methods to map neural and behavioral trajectories in youth (for a more detailed overview of these models, see [126]; [134]; [128]). In general, while many commonalities and important features are shared across different variable-centered approaches, selecting the most appropriate statistical model for assessing change hinges on the specific theoretical model of change and what is intended to be learned from the model (see [120]; [101]; [119], for discussion), which is critical for informing the interpretation and applicability of the research findings.

Person-centered models :

Despite the flexibility afforded by variable-centered analysis, these methods are not generally equipped to capture underlying developmental trajectories that are unique to distinct clusters of individuals. This limitation can be particularly notable for research that aims to characterize 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. For example, taking advantage of the large sample size of the ABCD Study, [135] found evidence of four subgroups of youth with unique longitudinal patterns of depressive symptoms over time and identified risk factors that were differentially associated with the various trajectories.

The use of such models allows for a more nuanced understanding of the associations between risk factors and change in symptomatology, as opposed to a snapshot of symptomatology at a single time point. Despite a range of potential model specifications for longitudinal mixture modeling, person-centered approaches tend to use parameterizations that default to settings found in popular software packages (e.g., Mplus). It has recently been demonstrated (see [136]) that the use of such specifications tends to identify the so-called "cat's cradle" solution (see [137]) that consists of "...(a) a consistently 'low' group, (b) an 'increase' group, (c) a 'decrease' group, and (d) a consistently 'high' group" [137], p. 322. Indeed, [135] describe their four-group solution as follows: "Of all participants, 536 (10.80%) were classified as increasing, 269 (5.42%) as persistently high, 433 (8.73%) as decreasing, and 3724 (75.05%) as persistently low" [135], p. 162. Although [137] cautioned that groups from these trajectory-based approaches should not be over-reified, this practice also remains common (e.g., [138]; [139]). Thus, though person-centered approaches can, in theory, help researchers understand the factors that contribute to differences in developmental trajectories, researchers should more thoughtfully consider alternative specifications (see

[140], as an example) and be especially skeptical when default specifications identify these four prototypic groups.

Hybrid approaches, such as growth mixture (jung2008; [108]) and factor-mixture [112]; [141] 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 LDA, 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 Structural Equation Models :

Advanced SEM approaches, such as the RI-CLPM and LCM-SR models, have emerged to provide a clearer understanding of important 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. Within-person variance in these models is 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. Using three waves of ABCD Study data, [142] found a prospective association between psychopathology and childhood obesity as well as between childhood obesity and later eating behavior. The authors also showed that reciprocal associations were overestimated when stable, interindividual trait differences were not included in the model (i.e., via the random intercept).

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 LDA 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. Given that these modeling approaches necessitate more waves of data, they are not yet commonly used with ABCD Study data. We anticipate that as more waves of ABCD data are publically released, these models can be used to address some of the pitfalls of the more traditional methods.

Longitudinal Analysis of Neuroimaging Data :

Neuroimaging data, characterized by its large scale, spatial structure and binary data formats, requires the use of specialized software for effective analysis. Fortunately, there are now several freely available software packages that provide options for statistical modeling of brain imaging data, thus facilitating analysis of the brain's function or structure at every voxel or vertex in an image (see Fig. 3). However, the most widely used packages have only rudimentary support for longitudinal data. Prominent software solutions such as SPM (https://www.fil.ion.ucl.ac.uk/spm) and FSL (https://fsl.fmrib.ox.ac.uk) packages offer limited support in analyzing longitudinal data due to their use of strong assumptions. SPM assumes a common longitudinal correlation structure over space, and FSL requires balanced designs and relies on the assumption of compound symmetry. Despite these limitations, there is a steady growth in the development of neuroimaging tools that provide for comprehensive longitudinal data analysis, typically via commonly adopted modeling approaches such as LMMs marginal models. These emerging tools are equipped to handle more

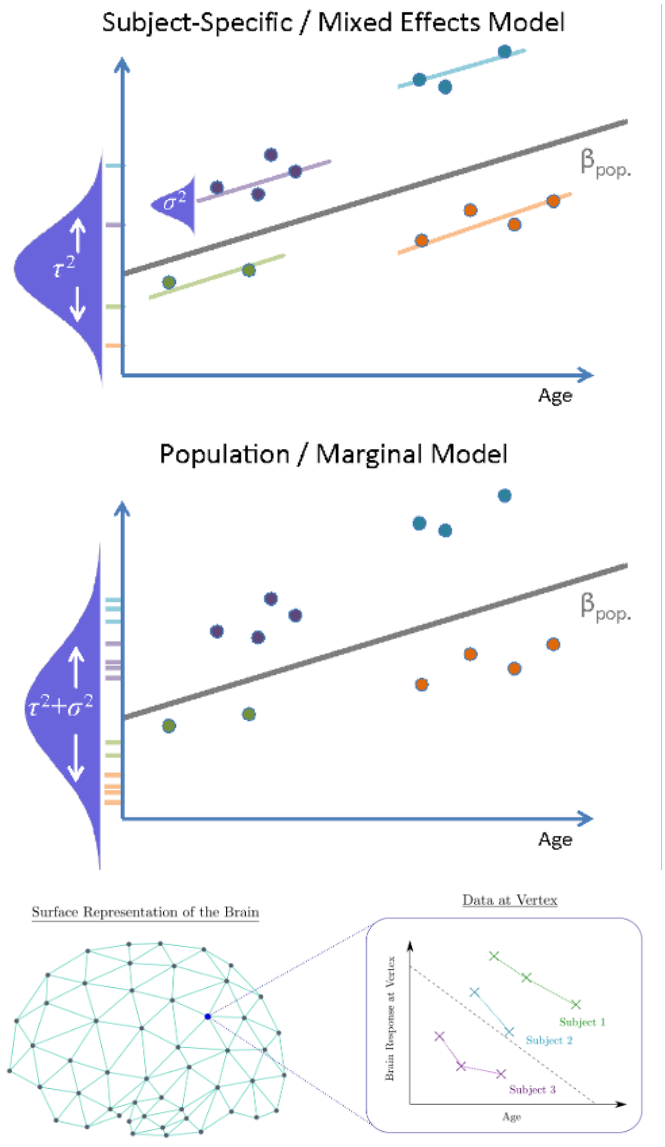
complex scenarios, including unbalanced designs and random covariate effects, among others.

AFNI (https://afni.nimh.nih.gov/), a well-established tool in neuroimaging, integrates 3dLMEr [143]; https://afni.nimh.nih.gov/pub/dist/doc/program_help/3dLMEr.html which adopts an LMM-based approach by providing access to the advanced capabilities of R's lme4's lmer function. For surface-based data, Freesurfer (https://surfer.nmr.mgh.harvard.edu/) provides the linear mixed effects (LME) package for modelling longitudinal data [144]; https://surfer.nmr.mgh.harvard.edu/fswiki/LinearMixedEffectsModels. This tool is able to apply spatial regularization of LMM parameters with surface-based ROIs to improve stability [145].

For imaging data, LMM's present a significant computational challenge, not only because they require iterative optimization but also because the computations cannot be vectorized as efficiently compared to ordinary least squares. To overcome this challenge, The Big Linear Mixed Models (BLMM, https://github.com/NISOx-BDI/BLMM) software addresses this by using Python's broadcasting operations to estimate LMM's as efficiently as possible [146]; [147]. BLMM further separates the computation of sufficient statistics and parameter estimation, allowing sensitive image data to remain private if needed.

A different yet efficient approach is used with Fast and Efficient Mixed-effects Analysis (FEMA, https://github.com/cmig-research-group/cmig_tools), which uses a non-iterative regression estimator of the LMM variance components plus variance parameter quantization. This allows vectorization within groups of voxels that share the same variance parameters [148]. While this method uses different approximations, the authors have shown it provides results that closely match a traditional LMM implementation.

An alternative method for modeling longitudinal data is the marginal model. This approach differs from others by modeling only the population-level factors and covariates rather than explicitly modeling individual intercepts and slopes. It employs a flexible intra-subject covariance model to account for residual dependence. Like the LMM, it allows for unbalanced designs and singleton subjects, and it implicitly measures any covariance that would otherwise be explained by random covariates. In a marginal model, a "working covariance" matrix is utilized. This matrix does not need to be accurately modelled and may even be constructed under the potentially incorrect assumption that the errors are independent. In this case, parameter estimation reduces to ordinary least squares, but remains consistent. A crucial component of this method is the use of a robust "sandwich estimator" for standard errors, which accounts for any dependence not captured by the working covariance. The Sandwich Estimator (SwE) is a toolbox for SPM (https://www.nisox.org/Software/SwE, including CIFTI support) and FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Swe) that provides marginal model inference using an independence working covariance matrix [149].



V. 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 an overview 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 LDA, 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. 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.

VI. Acknowledgements

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive DevelopmentSM (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

The ABCD data repository grows and changes over time. The ABCD data used in this report came from DOI: 10.15154/z563-zd24. DOIs can be found at the following link.

Bibliography

[1] N. D. Volkow *et al.*, "The conception of the ABCD study: From substance use to a broad NIH collaboration", *Developmental Cognitive Neuroscience*, vol. 32, pp. 4–7, Aug. 2018, doi: 10.1016/j.dcn.2017.10.002.

[2] H. Garavan *et al.*, "Recruiting the ABCD sample: Design considerations and procedures", *Developmental cognitive neuroscience*, vol. 32, pp. 16–22, 2018.

[3] A. S. Dick *et al.*, "Meaningful associations in the adolescent brain cognitive development study", *NeuroImage*, vol. 239, p. 118262, 2021.

[4] M. Luciana *et al.*, "Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery", *Developmental cognitive neuroscience*, vol. 32, pp. 67–79, 2018.

[5] W. K. Thompson *et al.*, "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*, vol. 36, p. 100606, 2019.

[6] D. M. Barch *et al.*, "Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description", *Developmental cognitive neuroscience*, vol. 32, pp. 55–66, 2018.

[7] R. Gonzalez *et al.*, "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*, vol. 52, p. 101021, 2021.

[8] R. A. Zucker *et al.*, "Assessment of culture and environment in the Adolescent Brain and Cognitive Development Study: Rationale, description of measures, and early data", *Developmental cognitive neuroscience*, vol. 32, pp. 107–120, 2018.

[9] K. M. Lisdahl *et al.*, "Substance use patterns in 9-10 year olds: Baseline findings from the adolescent brain cognitive development (ABCD) study", *Drug and alcohol dependence*, vol. 227, p. 108946, 2021.

[10] A. S. Potter *et al.*, "Measurement of gender and sexuality in the Adolescent Brain Cognitive Development (ABCD) study", *Developmental Cognitive Neuroscience*, vol. 53, p. 101057, 2022.

[11] K. A. Uban *et al.*, "Biospecimens and the ABCD study: Rationale, methods of collection, measurement and early data", *Developmental Cognitive Neuroscience*, vol. 32, pp. 97–106, Aug. 2018, doi: 10.1016/j.dcn.2018.03.005.

[12] B. J. Casey *et al.*, "The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites", *Developmental Cognitive Neuroscience*, vol. 32, pp. 43–54, Aug. 2018, doi: 10.1016/j.dcn.2018.03.001.

[13] D. J. Hagler *et al.*, "Image processing and analysis methods for the Adolescent Brain Cognitive Development Study", *NeuroImage*, vol. 202, p. 116091, Nov. 2019, doi: 10.1016/j.neuroimage.2019.116091.

[14] C. E. Palmer *et al.*, "Microstructural development from 9 to 14 years: Evidence from the ABCD Study", *Developmental cognitive neuroscience*, vol. 53, p. 101044, 2022.

[15] C. C. Fan *et al.*, "Adolescent Brain Cognitive Development (ABCD) study Linked External Data (LED): Protocol and practices for geocoding and assignment of environmental data", *Developmental cognitive neuroscience*, vol. 52, p. 101030, 2021.

[16] K. S. Bagot *et al.*, "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*, vol. 32, pp. 121–129, 2018.

[17] R. J. Loughnan, C. E. Palmer, W. K. Thompson, A. M. Dale, T. L. Jernigan, and C. C. Fan, "Polygenic score of intelligence is more predictive of crystallized than fluid performance among children", *bioRxiv*, p. 637512, 2020.

[18] W. K. Thompson, J. Hallmayer, R. O'Hara, and A. D. N. Initiative, "Design considerations for characterizing psychiatric trajectories across the lifespan: application to effects of APOE-\varepsilon4 on cerebral cortical thickness in Alzheimer's disease", *American Journal of Psychiatry*, vol. 168, no. 9, pp. 894–903, 2011.

[19] L. Arseneault, M. Cannon, R. Poulton, R. Murray, A. Caspi, and T. E. Moffitt, "Cannabis use in adolescence and risk for adult psychosis: longitudinal prospective study", *Bmj*, vol. 325, no. 7374, pp. 1212–1213, 2002.

[20] J. Bechtold, A. Hipwell, D. A. Lewis, R. Loeber, and D. Pardini, "Concurrent and sustained cumulative effects of adolescent marijuana use on subclinical psychotic symptoms", *American Journal of Psychiatry*, vol. 173, no. 8, pp. 781–789, 2016.

[21] A. Hasan *et al.*, "Cannabis use and psychosis: a review of reviews", *European archives of psychiatry and clinical neuroscience*, vol. 270, pp. 403–412, 2020.

[22] D. M. Semple, A. M. McIntosh, and S. M. Lawrie, "Cannabis as a risk factor for psychosis: systematic re-

view", *Journal of psychopharmacology*, vol. 19, no. 2, pp. 187–194, 2005.

[23] L. P. Spear, "Consequences of adolescent use of alcohol and other drugs: Studies using rodent models", *Neuroscience & Biobehavioral Reviews*, vol. 70, pp. 228–243, 2016.

[24] F. T. Crews, C. J. Braun, B. Hoplight, R. C. Switzer III, and D. J. Knapp, "Binge ethanol consumption causes differential brain damage in young adolescent rats compared with adult rats", *Alcoholism: Clinical and Experimental Research*, vol. 24, no. 11, pp. 1712–1723, 2000.

[25] Z. Ji, L. Yuan, X. Lu, H. Ding, J. Luo, and Z.-J. Ke, "Binge alcohol exposure causes neurobehavioral deficits and GSK3$b\eta$ activation in the hippocampus of adolescent rats", *Scientific Reports*, vol. 8, no. 1, p. 3088, 2018.

[26] M. Infante *et al.*, "Adolescent binge drinking is associated with accelerated decline of gray matter volume", *Cerebral Cortex*, vol. 32, no. 12, pp. 2611–2620, 2022.

[27] J. Schulenberg, J. Maslowsky, M. E. Patrick, and M. Martz, "Substance use in the context of adolescent development", *The Oxford handbook of adolescent substance abuse*. Oxford University Press, pp. 19–36, 2019.

[28] T. E. Moffitt, "Life-course-persistent versus adolescence-limited antisocial behavior", *Developmental Psychopathology: Volume Three: Risk, Disorder, and Adaptation*, pp. 570–598, 2015.

[29] A. S. Masten *et al.*, "Developmental cascades: linking academic achievement and externalizing and internalizing symptoms over 20 years.", *Developmental psychology*, vol. 41, no. 5, p. 733, 2005.

[30] F. A. Rogosch, A. Oshri, and D. Cicchetti, "From child maltreatment to adolescent cannabis abuse and dependence: A developmental cascade model", *Development and psychopathology*, vol. 22, no. 4, pp. 883–897, 2010.

[31] D. P. MacKinnon, A. J. Fairchild, and M. S. Fritz, "Mediation analysis", *Annu. Rev. Psychol.*, vol. 58, pp. 593–614, 2007.

[32] T. J. VanderWeele, "Mediation analysis: a practitioner's guide", *Annual review of public health*, vol. 37, pp. 17–32, 2016.

[33] R. M. Baron and D. A. Kenny, "The moderator--mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.", *Journal of personality and social psychology*, vol. 51, no. 6, p. 1173, 1986.

[34] K. J. Preacher, Z. Zhang, and M. J. Zyphur, "Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM", *Structural equation modeling*, vol. 18, no. 2, pp. 161–182, 2011.

[35] M.-A. Bind, T. Vanderweele, B. Coull, and J. Schwartz, "Causal mediation analysis for longitudinal data with exogenous exposure", *Biostatistics*, vol. 17, no. 1, pp. 122–134, 2016.

[36] T. J. VanderWeele and E. J. Tchetgen Tchetgen, "Mediation analysis with time varying exposures and mediators", *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 79, no. 3, pp. 917–938, 2017.

[37] M. A. Lindquist, "Functional causal mediation analysis with an application to brain connectivity", *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1297–1309, 2012.

[38] Y. Zhao, X. Luo, M. Lindquist, and B. Caffo, "Functional mediation analysis with an application to functional magnetic resonance imaging data", *arXiv preprint arXiv:1805.06923*, 2018.

[39] D. Rogosa, D. Brandt, and M. Zimowski, "A growth curve approach to the measurement of change.", *Psychological bulletin*, vol. 92, no. 3, p. 726, 1982.

[40] T. E. Duncan and S. C. Duncan, "The ABC's of LGM: An Introductory Guide to Latent Variable Growth Curve Modeling", *Social and Personality Psychology Compass*, vol. 3, no. 6, pp. 979–991, 2009, doi: 10.1111/j.1751-9004.2009.00224.x.

[41] K. M. King, A. K. Littlefield, C. J. McCabe, K. L. Mills, J. Flournoy, and L. Chassin, "Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology", *Developmental Cognitive Neuroscience*, vol. 33, pp. 54–72, Oct. 2018, doi: 10.1016/j.dcn.2017.11.009.

[42] A. K. Littlefield, K. M. King, S. F. Acuff, K. T. Foster, J. G. Murphy, and K. Witkiewitz, "Limitations of cross-lagged panel models in addiction research and alternative models: An empirical example using project MATCH.", *Psychology of Addictive Behaviors*, 2021, doi: 10.1037/adb0000750.

[43] P. J. Curran, A. L. Howard, S. Bainter, S. T. Lane, and J. S. McGinley, "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*, vol. 82, no. 5, pp. 879–894, Oct. 2014, doi: 10.1037/a0035297.

[44] W. Bleidorn *et al.*, "Personality stability and change: A meta-analysis of longitudinal studies.", *Psychological bulletin*, vol. 148, no. 7–8, p. 588, 2022.

[45] B. W. Roberts and W. F. DelVecchio, "The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies.", *Psychological bulletin*, vol. 126, no. 1, p. 3, 2000.

[46] B. W. Roberts, K. E. Walton, and W. Viechtbauer, "Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies.", *Psychological bulletin*, vol. 132, no. 1, p. 1, 2006.

[47] P. J. Curran and D. J. Bauer, "The disaggregation of within-person and between-person effects in longitudinal models of change", *Annual review of psychology*, vol. 62, pp. 583–619, 2011.

[48] L. Hoffman, *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge, 2015.

[49] U. Orth, D. A. Clark, M. B. Donnellan, and R. W. Robins, "Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models.", *Journal of Personality and Social Psychology*, vol. 120, no. 4, p. 1013, 2021.

[50] E. L. Hamaker, R. M. Kuiper, and R. P. P. P. Grasman, "A critique of the cross-lagged panel model", *Psychological Methods*, vol. 20, no. 1, pp. 102–116, 2015, doi: 10.1037/a0038889.

[51] S. N. Wood, *Generalized additive models: an introduction with R*. chapman, hall/CRC, 2017.

[52] J. O. Ramsay and B. W. Silverman, *Applied functional data analysis: methods and case studies*. Springer, 2002.

[53] W. K. Thompson *et al.*, "Characterizing trajectories of cognitive functioning in older adults with schizophrenia: does method matter?", *Schizophrenia research*, vol. 143, no. 1, pp. 90–96, 2013.

[54] S. T. Lenz, "Alan Agresti (2013): Categorical data analysis", *Statistical Papers*, vol. 57, no. 3, p. 849, 2016.

[55] L. K. Muthén, "Mplus user's guide. Los Angeles: Muthén & Muthén; 1998". 2017.

[56] J. Ren, S. Tapert, C. C. Fan, and W. K. Thompson, "A semi-parametric Bayesian model for semi-continuous longitudinal data", *Statistics in Medicine*, vol. 41, no. 13, pp. 2354–2374, 2022.

[57] T. D. Little, *The Oxford Handbook of Quantitative Methods, Vol. 2: Statistical Analysis*. Oxford University Press, 2013.

[58] Y. Liu, R. E. Millsap, S. G. West, J.-Y. Tein, R. Tanaka, and K. J. Grimm, "Testing measurement invariance in longitudinal data with ordered-categorical measures.", *Psychological methods*, vol. 22, no. 3, p. 486, 2017.

[59] R. Van De Schoot, P. Schmidt, A. De Beuckelaer, K. Lek, and M. Zondervan-Zwijnenburg, "Measurement invariance", vol. 6. Frontiers Media SA, p. 1064, 2015.

[60] M. T. Willoughby, R. Wirth, and C. B. Blair, "Executive function in early childhood: longitudinal measurement invariance and developmental change.", *Psychological assessment*, vol. 24, no. 2, p. 418, 2012.

[61] S. J. Brislin *et al.*, "Differential item functioning in reports of delinquent behavior between Black and White youth: evidence of measurement bias in self-reports of arrest in the adolescent brain cognitive development study", *Assessment*, p. 10731911231164627, 2023.

[62] C. Vize, W. R. Ringwald, E. R. Perkins, R. Waller, S. Hawes, and A. L. Byrd, "The Hierarchical Structure and Longitudinal Measurement Invariance of Externalizing Symptoms in the Adolescent Brain and Cognitive Development (ABCD) Study", 2023.

[63] T. A. Salthouse, "Why are there different age relations in cross-sectional and longitudinal comparisons of cognitive functioning?", *Current directions in psychological science*, vol. 23, no. 4, pp. 252–256, 2014.

[64] E. V. Sullivan *et al.*, "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*, vol. 24, pp. 72–83, 2017.

[65] A. T. Beck, C. H. Ward, M. Mendelson, J. Mock, and J. Erbaugh, "An inventory for measuring depression", *Archives of general psychiatry*, vol. 4, no. 6, pp. 561–571, 1961.

[66] D. P. French and S. Sutton, "Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it?", *British journal of health psychology*, vol. 15, no. 3, pp. 453–468, 2010.

[67] L. Robins, "Epidemiology: Reflections on Testing the Validity of Psychiatric Interviews | JAMA Psychiatry | JAMA Network". Accessed: Oct. 03, 2021. [Online]. Available: https://jamanetwork.com/journals/jamapsychiatry/article-abstract/493658

[68] P. E. Shrout *et al.*, "Initial elevation bias in subjective reports", *Proceedings of the National Academy of Sciences*, vol. 115, no. 1, p. E15--E23, 2018.

[69] C. Kincaid, "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*, 2005.

[70] A. S. Bryk and S. W. Raudenbush, *Hierarchical linear models: applications and data analysis methods*. Sage Publications, Inc, 1992.

[71] J. D. Singer, "Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models", *Journal of educational and behavioral statistics*, vol. 23, no. 4, pp. 323–355, 1998.

[72] P. J. Curran, "Have multilevel models been structural equation models all along?", *Multivariate behavioral research*, vol. 38, no. 4, pp. 529–569, 2003.

[73] A. K. Littlefield, K. M. King, S. F. Acuff, K. T. Foster, J. G. Murphy, and K. Witkiewitz, "Limitations of cross-lagged panel models in addiction research and alternative models: An empirical example using project MATCH.", *Psychology of Addictive Behaviors*, vol. 36, no. 3, p. 271, 2022.

[74] S. W. F. Ewing *et al.*, "Measuring retention within the adolescent brain cognitive development (ABCD) SM study", *Developmental Cognitive Neuroscience*, vol. 54, p. 101081, 2022.

[75] R. B. Cotter, J. D. Burke, M. Stouthamer-Loeber, and R. Loeber, "Contacting participants for follow-up: how much effort is required to retain participants in longitudinal studies?", *Evaluation and Program Planning*, vol. 28, no. 1, pp. 15–21, 2005.

[76] K. G. Hill, D. Woodward, T. Woelfel, J. D. Hawkins, and S. Green, "Planning for long-term follow-up: Strategies learned from longitudinal studies", *Prevention Science*, vol. 17, no. 7, pp. 806–818, 2016.

[77] N. Watson, E. Leissou, H. Guyer, and M. Wooden, "Best Practices for Panel Maintenance and Retention", *Advances in Comparative Survey Methods*. John Wiley & Sons, Ltd, pp. 597–622, 2018. doi: 10.1002/9781118884997.ch29.

[78] R. J. Little and D. B. Rubin, "The Analysis of Social Science Data with Missing Values", *Sociological Methods & Research*, vol. 18, no. 2–3, pp. 292–326, Nov. 1989, doi: 10.1177/0049124189018002004.

[79] J. W. Graham, "Missing Data Analysis: Making It Work in the Real World", *Annual Review of Psychology*, vol. 60, no. 1, pp. 549–576, 2009, doi: 10.1146/annurev.psych.58.110405.085530.

[80] C. K. Enders, *Applied Missing Data Analysis*. Guilford Press, 2010.

[81] J. Twisk, M. de Boer, W. de Vente, and M. Heymans, "Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis", *Journal of clinical epidemiology*, vol. 66, no. 9, pp. 1022–1028, 2013.

[82] J. Cohen, "Statistical power", *Analysis for the behavioral sciences*, pp. 273–406, 1988.

[83] S. B. Morris and R. P. DeShon, "Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs.", *Psychological methods*, vol. 7, no. 1, p. 105, 2002.

[84] L. Wang, Q. Zhang, S. E. Maxwell, and C. Bergeman, "On standardizing within-person effects: Potential problems of global standardization", *Multivariate Behavioral Research*, vol. 54, no. 3, pp. 382–403, 2019.

[85] A. Feingold, "Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis.", *Psychological methods*, vol. 14, no. 1, p. 43, 2009.

[86] A. Feingold, "Time-varying effect sizes for quadratic growth models in multilevel and latent growth modeling", *Structural equation modeling: a multidisciplinary journal*, vol. 26, no. 3, pp. 418–429, 2019.

[87] L. M. Collins, "Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model", *Annu. Rev. Psychol.*, vol. 57, pp. 505–528, 2006.

[88] D. J. Bauer and P. J. Curran, "Conducting Longitudinal Data Analysis: Knowing What to Do and Learning How to Do It.", 2019, [Online]. Available: https://centerstat.org/wp-content/uploads/2019/03/SRCD2019-2up.pdf

[89] D. W. Hosmer Jr, S. Lemeshow, and S. May, *Applied survival analysis: regression modeling of time-to-event data*, vol. 618. John Wiley & Sons, 2008.

[90] D. Rizopoulos, *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press, 2012.

[91] L. Castro-Schilo and K. J. Grimm, "Using residualized change versus difference scores for longitudinal research", *Journal of Social and Personal Relationships*, vol. 35, no. 1, pp. 32–58, Jan. 2018, doi: 10.1177/0265407517718387.

[92] A. E. Brieant, L. M. Sisk, and D. G. Gee, "Associations among negative life events, changes in corticolimbic connectivity, and psychopathology in the ABCD

Study", *Developmental Cognitive Neuroscience*, vol. 52, p. 101022, 2021.

[93] A. L. Romer and D. A. Pizzagalli, "Is executive dysfunction a risk marker or consequence of psychopathology? A test of executive function as a prospective predictor and outcome of general psychopathology in the adolescent brain cognitive development study®", *Developmental Cognitive Neuroscience*, vol. 51, p. 100994, 2021.

[94] A. K. Littlefield, "Longitudinal data analysis", *APA handbook of research methods in psychology: Data analysis and research publication, Vol. 3*. American Psychological Association, pp. 245–268, 2023.

[95] M. T. Van Dijk, E. Murphy, J. E. Posner, A. Talati, and M. M. Weissman, "Association of multigenerational family history of depression with lifetime depressive and other psychiatric disorders in children: Results from the Adolescent Brain Cognitive Development (ABCD) Study", *JAMA psychiatry*, vol. 78, no. 7, pp. 778–787, 2021.

[96] M. E. Martz *et al.*, "Individual-, peer-, and parent-level substance use-related factors among 9-and 10-year-olds from the ABCD Study: prevalence rates and sociodemographic differences", *Drug and alcohol dependence reports*, vol. 3, p. 100037, 2022.

[97] B. Lees, L. M. Squeglia, F. J. Breslin, W. K. Thompson, S. F. Tapert, and M. P. Paulus, "Screen media activity does not displace other recreational activities among 9--10 year-old youth: A cross-sectional ABCD study®", *BMC public health*, vol. 20, no. 1, pp. 1–11, 2020.

[98] P. J. Curran and G. R. Hancock, "The challenge of modeling co-developmental processes over time", *Child development perspectives*, vol. 15, no. 2, pp. 67–75, 2021.

[99] J. F. Hair Jr *et al.*, "An introduction to structural equation modeling", *Partial least squares structural equation modeling (PLS-SEM) using R: a workbook*, pp. 1–29, 2021.

[100] T. Raykov and G. A. Marcoulides, *A first course in structural equation modeling*. routledge, 2012.

[101] J. J. McArdle, "Latent variable modeling of differences and changes with longitudinal data", *Annual review of psychology*, vol. 60, pp. 577–605, 2009.

[102] P. D. Mehta and M. C. Neale, "People are variables too: multilevel structural equations modeling.", *Psychological methods*, vol. 10, no. 3, p. 259, 2005.

[103] D. McNeish and T. Matta, "Differentiating between mixed-effects and latent-curve approaches to growth modeling", *Behavior research methods*, vol. 50, pp. 1398–1414, 2018.

[104] D. Hedeker and R. D. Gibbons, *Longitudinal data analysis*. Wiley-Interscience, 2006.

[105] S. Serang, K. J. Grimm, and Z. Zhang, "On the correspondence between the latent growth curve and latent change score models", *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 26, no. 4, pp. 623–635, 2019.

[106] J. J. McArdle and F. Hamagami, "Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data.", 2001.

[107] C. Geiser and G. Lockhart, "A comparison of four approaches to account for method effects in latent state--trait analyses.", *Psychological methods*, vol. 17, no. 2, p. 255, 2012.

[108] B. Muthén and L. K. Muthén, "Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes", *Alcoholism: Clinical and experimental research*, vol. 24, no. 6, pp. 882–891, 2000.

[109] S. E. Woo, J. Hofmans, B. Wille, and L. Tay, "Person-Centered Modeling: Techniques for Studying Associations Between People Rather Than Variables", *Annual Review of Organizational Psychology and Organizational Behavior*, vol. 11, pp. 453–480, 2024.

[110] M. C. Howard and M. E. Hoffman, "Variable-centered, person-centered, and person-specific approaches: Where theory meets the method", *Organizational Research Methods*, vol. 21, no. 4, pp. 846–876, 2018.

[111] A. J. Morin, A. Bujacz, and M. Gagné, "Person-centered methodologies in the organizational sciences: Introduction to the feature topic", vol. 21, no. 4. Sage Publications Sage CA: Los Angeles, CA, pp. 803–813, 2018.

[112] G. H. Lubke and B. Muthén, "Investigating population heterogeneity with factor mixture models.", *Psychological methods*, vol. 10, no. 1, p. 21, 2005.

[113] A. D. Trevino, B. Jamil, J. Su, F. Aliev, K. K. Elam, and K. Lemery-Chalfant, "Alcohol Use Disorder Polygenic Risk Scores and Trajectories of Early Adolescent Externalizing Behaviors: Examining the Role of Parenting and Family Conflict in the Racially/Ethnically Diverse ABCD Sample", *Behavior Genetics*, pp. 1–18, 2023.

[114] E. Roy *et al.*, "White matter and literacy: a dynamic system in flux", *Developmental Cognitive Neuroscience*, vol. 65, p. 101341, 2024.

[115] K. J. Preacher, M. J. Zyphur, and Z. Zhang, "A general multilevel SEM framework for assessing multilevel me-

diation.", *Psychological methods*, vol. 15, no. 3, p. 209, 2010.

[116] K. J. Preacher, "Latent growth curve models", *The reviewer's guide to quantitative methods in the social sciences*. Routledge, pp. 178–192, 2018.

[117] P. J. Curran, K. Obeidat, and D. Losardo, "Twelve frequently asked questions about growth curve modeling", *Journal of cognition and development*, vol. 11, no. 2, pp. 121–136, 2010.

[118] J. McArdle and J. R. Nesselroade, "Using multivariate data to structure developmental change", *Life-span developmental psychology: Methodological contributions*, pp. 223–267, 1994.

[119] P. Ghisletta and J. J. McArdle, "Latent curve models and latent change score models estimated in R", *Structural equation modeling: a multidisciplinary journal*, vol. 19, no. 4, pp. 651–682, 2012.

[120] R. A. Kievit *et al.*, "Developmental cognitive neuroscience using latent change score models: A tutorial and applications", *Developmental cognitive neuroscience*, vol. 33, pp. 99–117, 2018.

[121] T. Wiker *et al.*, "Assessing the longitudinal associations between decision-making processes and attention problems in early adolescence", *Research on Child and Adolescent Psychopathology*, pp. 1–15, 2023.

[122] K. M. Rapuano *et al.*, "Longitudinal evidence of a vicious cycle between nucleus accumbens microstructure and childhood weight gain", *Journal of Adolescent Health*, vol. 70, no. 6, pp. 961–969, 2022.

[123] D. Beck *et al.*, "Puberty differentially predicts brain maturation in male and female youth: A longitudinal ABCD Study", *Developmental Cognitive Neuroscience*, p. 101261, 2023.

[124] T. Nweze *et al.*, "Trajectories of cortical structures associated with stress across adolescence: a bivariate latent change score approach", *Journal of Child Psychology and Psychiatry*, 2023.

[125] L. Mewton *et al.*, "Longitudinal relationships between lifestyle risk factors and neurodevelopment in early adolescence.", *Health Psychology*, 2023.

[126] D. A. Kenny and A. Zautra, "Trait–state models for longitudinal data", *New methods for the analysis of change*. American Psychological Association, Washington, DC, US, pp. 243–263, 2001. doi: 10.1037/10409-008.

[127] R. Steyer, A. Mayer, C. Geiser, and D. A. Cole, "A theory of states and traits—Revised", *Annual review of clinical psychology*, vol. 11, pp. 71–98, 2015.

[128] R. Steyer, M. Schmitt, and M. Eid, "Latent state--trait theory and research in personality and individual differences", *European Journal of Personality*, vol. 13, no. 5, pp. 389–408, 1999.

[129] R. Steyer, D. Ferring, and M. J. Schmitt, "States and traits in psychological assessment.", *European Journal of Psychological Assessment*, 1992.

[130] N. Stadtbaeumer, S. Kreissl, and A. Mayer, "Comparing revised latent state--trait models including autoregressive effects.", *Psychological Methods*, 2022.

[131] D. A. Cole, N. C. Martin, and J. H. Steiger, "Empirical and conceptual problems with longitudinal trait-state models: introducing a trait-state-occasion model.", *Psychological methods*, vol. 10, no. 1, p. 3, 2005.

[132] M. Eid, J. Holtmann, P. Santangelo, and U. Ebner-Priemer, "On the definition of latent-state-trait models with autoregressive effects", *European Journal of Psychological Assessment*, 2017.

[133] S. Sanchez-Alonso and R. N. Aslin, "Predictive modeling of neurobehavioral state and trait variation across development", *Developmental Cognitive Neuroscience*, vol. 45, p. 100855, 2020.

[134] R. Steyer, C. Geiser, and C. Loßnitzer, "Latent state--trait models.", 2023.

[135] Q. Xiang *et al.*, "Prediction of the trajectories of depressive symptoms among children in the adolescent brain cognitive development (ABCD) study using machine learning approach", *Journal of Affective Disorders*, vol. 310, pp. 162–171, 2022.

[136] D. McNeish and J. R. Harring, "Improving convergence in growth mixture models without covariance structure constraints", *Statistical Methods in Medical Research*, vol. 30, no. 4, pp. 994–1012, 2021.

[137] K. J. Sher, K. M. Jackson, and D. Steinley, "Alcohol use trajectories and the ubiquitous cat's cradle: cause for concern?", *Journal of abnormal psychology*, vol. 120, no. 2, p. 322, 2011.

[138] S. W. Hawes, S. B. Perlman, A. L. Byrd, A. Raine, R. Loeber, and D. A. Pardini, "Chronic anger as a precursor to adult antisocial personality features: The moderating influence of cognitive control.", *Journal of Abnormal Psychology*, vol. 125, no. 1, p. 64, 2016.

[139] S. W. Hawes *et al.*, "The developmental course of psychopathic features: Investigating stability, change, and long-term outcomes", *Journal of Research in Personality*, vol. 77, pp. 83–89, 2018.

[140] A. K. Littlefield, K. J. Sher, and D. Steinley, "Developmental trajectories of impulsivity and their association with alcohol use and related outcomes during emerging and young adulthood I", *Alcoholism: Clinical and Experimental Research*, vol. 34, no. 8, pp. 1409–1416, 2010.

[141] G. Lubke and B. O. Muthén, "Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters", *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 14, no. 1, pp. 26–47, 2007.

[142] L. K. Kulisch, K. Arumäe, D. A. Briley, and U. Vainik, "Triangulating causality between childhood obesity and neurobehavior: Behavioral genetic and longitudinal evidence", *Developmental Science*, p. e13392, 2023.

[143] G. Chen, Z. S. Saad, J. C. Britton, D. S. Pine, and R. W. Cox, "Linear mixed-effects modeling approach to FMRI group analysis", *Neuroimage*, vol. 73, pp. 176–190, 2013.

[144] J. L. Bernal-Rusiel *et al.*, "Statistical analysis of longitudinal neuroimage data with linear mixed effects models", *Neuroimage*, vol. 66, pp. 249–260, 2013.

[145] J. L. Bernal-Rusiel *et al.*, "Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data", *Neuroimage*, vol. 81, pp. 358–370, 2013.

[146] T. Maullin-Sapey and T. E. Nichols, "BLMM: Parallelised computing for big linear mixed models", *NeuroImage*, vol. 264, p. 119729, 2022.

[147] T. Maullin-Sapey and T. E. Nichols, "Fisher Scoring for crossed factor linear mixed models", *Statistics and computing*, vol. 31, no. 5, p. 53, 2021.

[148] P. Parekh *et al.*, "FEMA: Fast and efficient mixed-effects algorithm for large sample whole-brain imaging data", *BioRxiv*, pp. 2021–10, 2021.

[149] B. Guillaume *et al.*, "Fast and accurate modelling of longitudinal and repeated measures neuroimaging data", *Neuroimage*, vol. 94, pp. 287–302, 2014.