Longitudinal Analysis Manuscript: Working Draft

Longitudinal Analysis using data from the ABCD Study

Samuel W. Hawes Andrew K. Littlefield Daniel A. Lopez
Kenneth J. Sher Erin L. Thompson Raul Gonzalez
Additional Co-authors Wesley K. Thompson

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

1 Introduction

The Adolescent Brain Cognitive Development (ABCD) Study® is the largest long-term investigation of neurodevelopment and child health in the United States. Conceived and initiated by the National Institutes of Health (NIH), this landmark prospective longitudinal study aims to transform our understanding of the genetic and environmental influences on brain development and their roles in behavioral and health outcomes in adolescents (Volkow et al. 2018). At its heart, the study is designed to chart the course of human development across multiple, interacting domains from late childhood to early adulthood and to identify factors that lead to both positive and negative developmental outcomes. Central to achieving these goals is the ABCD Study's®

commitment to an open science framework designed to facilitate access to and sharing of scientific knowledge by espousing practices that increase openness, integrity, and reproducibility of scientific research (e.g., public data releases). In this sense, the ABCD Study® is a collaboration with the larger research community, with the rich longitudinal nature of the ABCD Study dataset allowing researchers to perform a variety of analyses of both methodological and substantive interest. Together, this presents a unique opportunity to significantly advance our understanding of how a multitude of biopsychosocial processes emerge and unfold across critical periods of development.

[section still be developed…]

1.1 The ABCD Study® Data

Participants enrolled in the ABCD Study include a large cohort of youth (n=11880) aged 9-10 years at baseline and their parents/guardians. The study sample was recruited from household populations in defined catchment areas for each of the 21 study sites across the United States (information regarding funding agencies, recruitment sites, investigators, and project organization can be obtained at the ABCD Study website). The ABCD Study is collecting longitudinal data on a rich variety of outcomes that will enable the construction of realistically-complex etiological models by incorporating factors from many domains simultaneously. Each new wave of data collection provides the building blocks for conducting probing longitudinal analyses that allow us to characterize normative development, identify variables that presage deviations from prototypic development, and assess a range of outcomes associated with variables of interest. This data includes a neurocognitive battery (Luciana et al. 2018; Thompson et al. 2019), mental and physical health assessments (Barch et al. 2018), measures of culture and environment (Zucker et al. 2018), substance use [add citation], biospecimens (Uban et al. 2018), structural and functional brain imaging (Casey et al. 2018; Hagler et al. 2019), geolocation-based environmental exposure data, wearables, and mobile technology (Bagot et al. 2018), and whole genome genotyping (Loughnan et al. 2020). Many of these measures are collected at in-person annual visits, with brain imaging collected at baseline and every other year going forward. A limited number of assessments are collected in semi-annual telephone interviews between in-person visits. Data are publicly released on an annual basis through the NIMH Data Archive. By necessity, the study's earliest data releases were cross-sectional (i.e., the baseline data), however, the most recent public data release (NDA Release 4.0) contains data collected across three annual assessments, including two imaging assessments (baseline and year 2 follow-up visits).

1.2 Organization of current manuscript

The rich longitudinal nature of the ABCD Study dataset will allow researchers to perform analyses of both methodological and substantive interest. This report describes methods for longitudinal analyses of ABCD Study data that can address its fundamental scientific aims, as well as challenges inherent in a large population-based long-term study of adolescents. The manuscript is organized as follows:

[section still be developed…]

2 Part I: Longitudinal Research

2.1 Basic Concepts and Considerations

There are several important concepts to consider when conducting longitudinal analyses in a developmental context. These include different ways of thinking about developmental course, whether certain periods of development are relatively sensitive or insensitive to various types of insults or stressors, whether some time periods or situations inhibit the expression of individual differences due to extreme environmental pressures, and whether the same behavior manifested at different times represent the same phenomenon or different ones. Further, in the case of developmentally focused longitudinal research, each new measurement occasion not only provides a more extended portrait of the child's life course (and not just characterize growth during this period but also assesses the durability/chronicity of prior effects/consequences) but also brings with it greater methodological opportunities to exploit the statistical properties of longitudinal data in the furtherance of critical scientific questions. That is, we can ask more nuanced questions and make stronger inferences as our number of time-ordered observations grows, assuming we have assessed the "right" variables and the timings of our observations comport with the temporal dynamics of the mechanisms of interest. Appreciation of these and other issues can help to guide the analysis and interpretation of data and aid translation to clinical and public health applications.

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

Developmental disturbances. Whereas vulnerable periods heighten neurobiological susceptibility to environmental influences, at other times environmental pressures will tend to suppress stability and disrupt the orderly stochastic process of normative development (e.g., xxx-xxx). This situation reflects a developmental disturbance in that the normal course of development is "disturbed" for some time by some time-limited process. In such cases, we might find that prediction of behavior in the period of the disturbance is reduced and/or, similarly, the behavior exhibited during the disturbance might have less predictive power with respect to distal outcomes compared to the behavior exhibited before and following the disrupted period. That is, once the environmental stimuli are removed (or the individual is removed from the environment), individual differences are again more freely expressed and the autoregressive effects increase to levels similar to those before entering the environment.

Developmental snares and cascade effects. Normative development can also be upended by experiences (e.g., drug use) that, through various mechanisms, disrupt the normal flow of

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

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

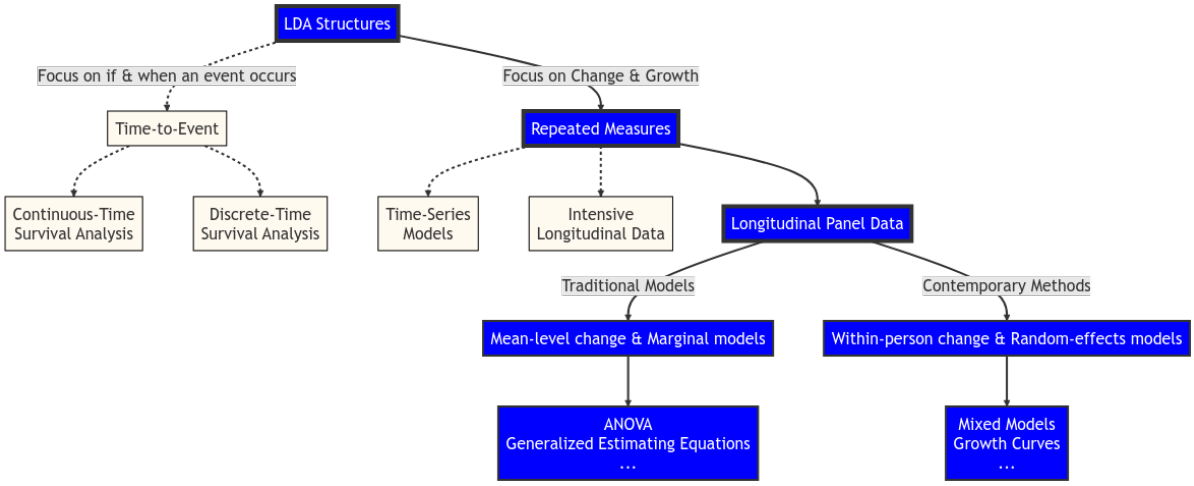

Figure 1: LDA Structural Diagram

3 Part II: Longitudinal Data

3.1 Interpretation / Issues / Pitfalls & Assumption

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

"traditional approaches": marginal models

random effects: Growth Curves- SEM (lgcm) vs ME (lmm)

(between- and within-person levels?)

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

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

fronts. As (Rogosa, Brandt, and Zimowski 1982) noted approximately forty years ago, "Two waves of data are better than one, but maybe not much better". These sentiments are reflected in more contemporary recommendations regarding best-practice guidelines for prospective data, which increasingly emphasize the benefits of additional measurement occasions for model identification and accurate parameter estimation. It is also consistent with research recommending that developmental studies include three or more assessment points, given it is impossible for data based on two-time points to determine the shape of development (since linear, straight line change is the only possible form, given two assessments; see (Duncan and Duncan 2009)). Research designs that include three or more time points allow for increasingly nuanced analyses that more adequately tease apart sources of variation and covariation among the repeated assessments (King et al. 2018)– a key aspect of inferential research. To illustrate, developmental theories are typically interested in understanding patterns of within-individual change over time (discussed in further detail, below); however, two data points provide meager information on change at the person level. This point is further underscored in a recent review of statistical models commonly touted as distinguishing within-individual vs between-individual sources of variance in which the study authors concluded "… researchers are limited when attempting to differentiate these sources of variation in psychological phenomenon when using two waves of data" and perhaps more concerning, "…the models discussed here do not offer a feasible way to overcome these inherent limitations" Littlefield et al. (2021). It is important to note, however, that despite the current focus on two-wave designs versus three or more assessment waves, garnering three assessment points is not a panacea for longitudinal modeling. Indeed, several contemporary longitudinal models designed to isolate within-individual variability [e.g., the Latent Curve Model with Structured Residuals; Curran et al. (2014)] require at least four assessments to parameterize fully and, more generally, increasingly accurate parameter estimates are obtained as more assessment occasions are used (Duncan and Duncan 2009).

3.1.2 Types of stability and change

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

why, in general, individuals are particularly vulnerable to social influences at some ages more than others; rank order change might help to quantify the extent to which certain characteristics of the individual are more trait-like. For example, in some areas of development, there is considerable mean-level change that occurs over time (e.g., changes in Big 5 personality traits), but relatively high rank-order stability. Despite the useful information afforded by examining mean-level and rank-order change, these approaches are limited in that they provide little information about patterns of "within-individual" change and, in turn, can result in fundamental misinterpretations about substantial or meaningful changes in an outcome of interest.

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

This is illustrated well by Figure 1.

3.1.3 Model Assumptions

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

tests of independence for time series data (e.g., the Durbin-Watson statistic (Durbin and Watson 1950)), more commonly independence is assumed to be violated in study designs with repeated assessments. Therefore, an initial question to be addressed by a researcher analyzing prospective data is how to best model the covariance structure of said data.

3.1.4 Covariance Structures

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

3.1.5 Accounting for Correlated Data

As an example, one alternative structure that attempts to handle the reality that correlations between repeated assessments tend to diminish across time is the autoregressive design. As the name implies, the structure assumes a subsequent measurement occasion (e.g., assessment at Wave 2) is regressed onto (that is, is predicted by) a prior measurement occasion (e.g., assessment at Wave 1). The most common type of autoregressive design is the AR(1), where assessments at time T + 1 are regressed on assessments at Time T. Identical to compound symmetry, this model assumes the variances are homogenous across time. Diverting from compound symmetry, this model assumes the correlations between repeated assessments decline exponentially across time rather than remaining constant. For example, per the AR(1) structure, if the correlation between Time 1 and Time 2 data is thought to be .5, then the correlation between Time 1 and Time 3 data would be assumed to be .5*.5 = .25, and the correlation between Time 1 and Time 4 data would be assumed to be .5*.5*.5 = .125. As with compound symmetry, the basic AR(1) model is parsimonious in that it only requires two parameters (the variance of the assessments and the autoregressive coefficient). Notably, the assumption of constant autoregressive relations between assessments is often relaxed in commonly employed designs that use autoregressive modeling (e.g., cross-lagged panel models [CLPM]). These designs still typically assume an AR(1) process (e.g., it is sufficient to regress the Time 3 assessment onto the Time 2 assessment and is not necessary to also regress the Time 3 assessment onto the Time 1 assessment, which would result in an AR(2) process). However, the magnitude of these relations is often allowed to differ across different AR(1) pairs of assessment (e.g., the relation between Time 1 and Time 2 can be different from the relation between Time 2 and Time 3). These more commonly employed models also often relax the assumption of equal variances of the repeated assessments. Although the AR(1) structure may involve a more realistic set of assumptions compared to compound symmetry, in that the AR(1) model allows for diminishing correlations across time, the basic AR(1) model, as well as autoregressive models more generally, can also suffer from several limitations in contexts that are common in prospective designs. In particular, recent work demonstrates that if a construct being assessed prospectively across time is trait-like in nature, then autoregressive relations fail to

adequately account for this trait-like structure, with the downstream consequence that estimates derived from models based on AR structures (such as the CLPM) can be misleading and fail to adequately demarcate between- vs. within-person sources of variance (Hamaker, Kuiper, and Grasman 2015).

3.1.6 Linear vs non-linear models

Identification of optimal statistical models and appropriate mathematical functions requires an understanding of the type of data being used. Repeated assessments can be based on either continuous or discrete measures. Examples of discrete measures include repeated assessments of binary variables (e.g., past 12-month alcohol use disorder status measured across ten years), ordinal variables (e.g., a single item measuring the level of agreement to a statement on a three-point scale including the categories of "disagree", "neutral", and "agree" in an ecological momentary assessment study that involves multiple daily assessments), and count variables (e.g., number of cigarettes smoked per day across a daily diary study). In many ways, the distributional assumptions of indicators used in longitudinal designs mirror the decision points and considerations when delineating across different types of discrete outcome variables, a topic that spans entire textbooks (e.g., see Lenz 2016). For example, the Mplus manual (Muthén 2017) includes examples of a) censored and censored-inflated models, b) linear growth models for binary or ordinal variables, c) linear growth models for a count outcome assuming a Poisson model, d) linear growth models for a count outcome assuming a zero-inflated Poisson model and e) discrete- and continuous-time survival analysis for a binary outcome. Beyond these highlighted examples, other distributions (e.g., negative binomial) can be assumed for the indicators when modeling longitudinal data. These models can account for issues that can occur when working with discrete outcomes, including overdispersion (when the variance is higher than would be expected based on a given distribution) and zero-inflation [when more zeros occur than is expected based on a given distribution; see Lenz (2016)]. Models involving zero-inflation parameters are referred to as two-part models, given one part of the model predicts the zero-inflation whereas the other part of the model predicts outcomes consistent with a given distribution [e.g., Poisson distribution; see Farewell et al. (2017), for a review of two-part models for longitudinal data]. Although there exist several alternative models for discrete indicators, some more recent models that have been proposed for prospective data are only feasible in cases where indicators are assumed to be continuous rather than discrete [e.g., LCM-SR; Curran et al. (2014)]. Given the sheer breadth of issues relevant to determining better models for discrete outcomes, it is not uncommon for texts on longitudinal data analysis to only cover models and approaches that assume continuous indicators (e.g., T. D. Little 2013). However, some textbooks on categorical data analysis provide more detailed coverage of the myriad issues and modeling choices to consider when working with discrete outcomes [e.g., Lenz (2016), Chapter 11 for matched pair/two-assessment designs; Chapter 12 for marginal and transitional models for repeated designs, such as generalized estimating equations, and Chapter 13 for random effects models for discrete outcomes].

3.1.7 Marginal vs Conditional Models (?)

[section still under development…]

3.1.8 Missing Data/Attrition

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

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

More modern approaches, such as using maximum likelihood or multiple imputation to estimate missing data, are thought to avoid some of the biases of older approaches (see Enders 2010; Graham 2009). Graham (2009) noted several "myths" regarding missing data. For example, Graham notes many assume the data must be minimally MAR to permit estimating procedures (such as maximum likelihood or multiple imputation) compared to other, more traditional approaches (e.g., using only complete case data). Violations of MAR impact both traditional and more modern data estimation procedures, though as noted by Graham, violations of MAR tend to have a greater effect on older methods. Graham thus suggests that estimating missing data is a better approach compared to the older procedures in most circumstances, regardless of the model of missingness [i.e., MCAR, MAR, MNAR; see Graham (2009)].

Attrition from a longitudinal panel study such as ABCD is inevitable and represents a threat to the validity of longitudinal analyses and cross-sectional analyses conducted at later time points, especially since attrition can only be expected to grow over time. While, to date, attrition in ABCD has been minimal (some cite here), it remains an important focus for longitudinal analysis and its significance is likely to only grow as the cohort ages. Ideally, one tries to minimize attrition

through good retention practices from the outset via strategies designed to maintain engagement in the project (Cotter et al. 2005; Hill et al. 2016; Watson et al. 2018). However, even the best-executed studies need to anticipate growing attrition over the length of the study and implement analytic strategies designed to provide the most valid inferences. Perhaps the most key concern when dealing with data that is missing due to attrition is determining the degree of bias in retained variables that is a consequence of attrition. Assuming that the data are not missing completely at random, attention to the nature of the missingness and employing techniques designed to mitigate attrition-related biases need to be considered in all longitudinal analyses. Several different approaches can be considered and employed depending upon the nature of the intended analyses, the degree of missingness, and data available to help estimate missing and unobserved values.

3.1.9 Quantifying effect sizes longitudinally

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

4 Part III: Additional Section(s)

In Progress: [TBD]

Section focused on different ways of considering/grouping/deciding on which specific analysis to choose (see curran&bauer slides)

4.0.1 Subsection TBD.

xxxxxx

xxxxxx

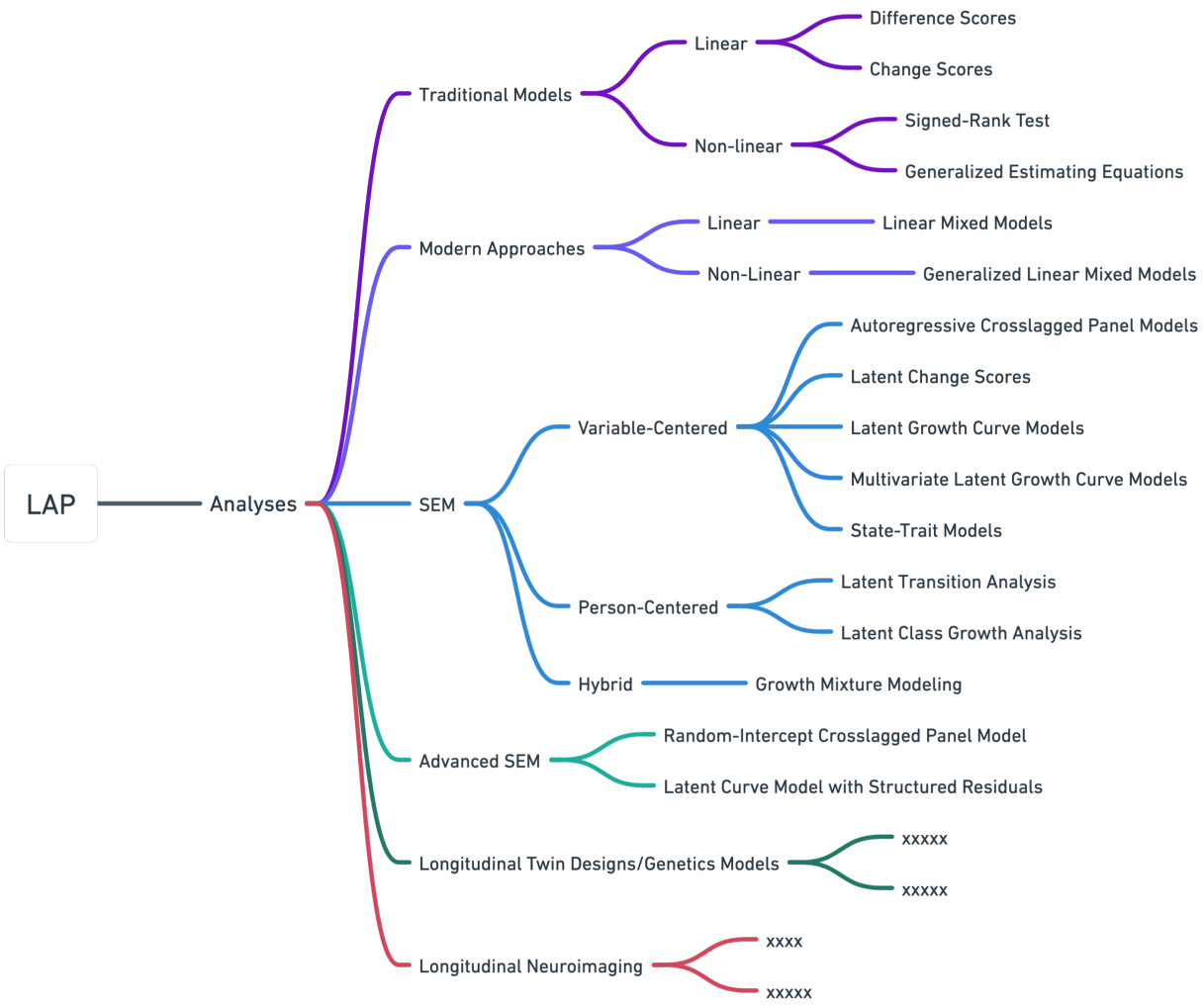

Figure 2: test

5 Part IV: Discussion

5.0.1 Subsection TBD.

Table 1: Models for the analysis of longitudinal data

Models	Form of Change	Minimum # of Timepoints	Variable-Centered vs Person-Centered	Time-varying Covariates
Difference Scores	mean-level	2+	-	none
Change Scores	rank-order	2+	-	none
Signed-Rank Test	rank-order	2+	-	none
Generalized Estimating Equations	rank-order	2+	-	allowed
Linear Mixed Models	within and between	3+	Variable-Centered	allowed
Generalized Linear Mixed-Effects Models	within and between	3+		allowed
Autoregressive Crosslagged Panel Models	rank-order	3+	Variable-Centered	allowed
Latent Change Scores Models	rank-order	2+	Variable-Centered	allowed
Latent Growth Curve Models	within and between	3+	Variable-Centered	allowed
Multivariate Latent Growth Curve Model	within and between	3+	Variable-Centered	allowed
Latent Transition Analysis	within and between	3+	Mixed	allowed
Latent Class Growth Analysis	within and between	3+	Person-Centered	allowed
Growth Mixture Models	within and between	3+	Mixed	allowed
State-Trait Models	continuous and discrete	3+	Variable-Centered	allowed
Random-Intercept Crosslagged Panel Models	within and between	4+	Variable-Centered	allowed
Latent Curve Models Structured Residuals	within and between	4+	Variable-Centered	allowed

Part V: References

Bagot, K. S., S. A. Matthews, M. Mason, Lindsay M. Squeglia, J. Fowler, K. Gray, M. Herting, A. May, Ian Colrain, and J. Godino. 2018. "Current, Future and Potential Use of Mobile and Wearable Technologies and Social Media Data in the ABCD Study to Increase Understanding of Contributors to Child Health." *Developmental Cognitive Neuroscience* 32: 121–29.

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

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

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

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

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

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

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

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

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

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

Farewell, V. T., D. L. Long, B. D. M. Tom, S. Yiu, and L. Su. 2017. "Two-Part and Related Regression Models for Longitudinal Data." *Annual Review of Statistics and Its Application* 4 (1): 283–315. https://doi.org/10.1146/annurev-statistics-060116-054131.

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

———. 2019. "Time-Varying Effect Sizes for Quadratic Growth Models in Multilevel and Latent

Growth Modeling." *Structural Equation Modeling: A Multidisciplinary Journal* 26 (3): 418–29.

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

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

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

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

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

Kincaid, C. 2005. "Guidelines for Selecting the Covariance Structure in Mixed Model Analysis, Paper 198-30 in Proceedings of the Thirtieth Annual SAS Users Group Conference." *Inc., Cary, North Carolina*.

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

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

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

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

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

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

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

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

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

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

Muthén, L. K. 2017. "Mplus User's Guide. Los Angeles: Muthén & Muthén; 1998."

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

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

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

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

Shrout, Patrick E., Gertraud Stadler, Sean P. Lane, M. Joy McClure, Grace L. Jackson, Frederick D. Clavél, Masumi Iida, Marci E. J. Gleason, Joy H. Xu, and Niall Bolger. 2018. "Initial Elevation Bias in Subjective Reports." *PNAS* 115 (1): E15–23. https://doi.org/10.1073/pnas.1712277115.

Sullivan, Edith V., Ty Brumback, Susan F. Tapert, Devin Prouty, Rosemary Fama, Wesley K. Thompson, Sandra A. Brown, Kevin Cummins, Ian M. Colrain, and Fiona C. Baker. 2017. "Effects of Prior Testing Lasting a Full Year in NCANDA Adolescents: Contributions from Age, Sex, Socioeconomic Status, Ethnicity, Site, Family History of Alcohol or Drug Abuse, and Baseline Performance." *Developmental Cognitive Neuroscience* 24: 72–83.

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

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

Volkow, Nora D., George F. Koob, Robert T. Croyle, Diana W. Bianchi, Joshua A. Gordon, Walter J. Koroshetz, Eliseo J. Pérez-Stable, et al. 2018. "The Conception of the ABCD Study: From Substance Use to a Broad NIH Collaboration." *Developmental Cognitive Neuroscience*, The Adolescent Brain Cognitive Development (ABCD) Consortium: Rationale, Aims, and Assessment Strategy, 32 (August): 4–7. https://doi.org/10.1016/j.dcn.2017.10.002.

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

Watson, Nicole, Eva Leissou, Heidi Guyer, and Mark Wooden. 2018. "Best Practices for Panel

Maintenance and Retention." In *Advances in Comparative Survey Methods*, 597–622. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118884997.ch29.

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