1. Prevailing goals of longitudinal research

At its heart, the ABCD 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. By necessity, the earliest data releases from ABCD were cross-sectional (i.e., the baseline data), but with each new wave of data collection, ABCD 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. To the extent that the study has assessed variables associated with etiological processes (e.g., the mechanisms explicating how regular use of a psychoactive substance leads to addiction in a some individuals) and the mechanisms underlying consequences of variables of interest (e.g., effects of cannabis on the development of psychosis), such longitudinal data provide the tools for evaluating the extent that the observed patterns are consistent with extant theory or suggest the need to revise our theories. Consequently, 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 assess 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. For example, while technically two waves of data are all that are needed to provide “longitudinal prediction” in one sense, with three waves of data we are better able to distinguish between-subject and within-subject associations over time. We can also start to model growth processes and succeeding waves of data allow us to model such processes with fewer statistical assumptions. That is, we can ask more nuanced questions and make stronger inferences as our number of time-ordered observations grow assuming we have assessed the “right” variables and the timings of our observations comport with the temporal dynamics of the mechanisms of interest.

*Longitudinal data and quality control.* 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. For example, a given profile of responses over time might identify an individual with extreme within-subject variability in what is presumed to be a stable trait. Also, because many measures are highly correlated over time (i.e., have high stability or test-retest correlations), we can identify individuals who show large residuals when, say, an observation made at *t+1* is regressed on *t*. Such longitudinal outliers can flag suspect data points in the same way that extreme univariate outliers can signal the need to further examine other properties of the data associated with the anomalous observation in cross-sectional data. Ostensibly aberrant observations can reflect valid phenomena or problems with subject-based invalid responding, problems with data collection and curation, or mishandling of specific observations elsewhere in the data processing pipeline. But, at a minimum, anomalous longitudinal data patterns identify those observations that warrant further scrutiny.

1. Key developmental considerations

There are a number of important concepts to consider when conducting longitudinal analyses in a developmental context. These include difference 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 and, relatedly, can ostensibly different behaviors manifested at different times represent the same underlying basic phenomenon expressing themselves differently in different developmental contexts. Appreciation of these and other issues can help to guide analysis and interpretation of data and aid translation to clinical and public health applications.

*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. For example, we can think of whether or not someone exhibits the same *level* of a trait (e.g., height) at two measurement occasions separated in time. That is, did they grow (i.e., change) or stay the same height (i.e., remain the same)? Alternatively, we often think of stability and change with respect to relative or rank-order stability with respect to ones age peers. That is, someone who is the shortest person in his class in 6th grade may grow considerably over the next two years (i.e., exhibit mean level change), but still remain the shortest person among this peers the 8th grade. That is, he is manifesting a type of rank order stability. Both types of stability and change are important. For example, 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. In contrast, rank order change (or stability) might help to quantify the extent to which certain characteristics of the individual or more trait-like. For example, in some areas of development, there is considerable mean change over the course of development (e.g., changes in Big 5 personality traits), but relatively high rank-order stability (although as might be expected, such rank-order stability tends to decrease the the interval between measurement occasions increases).

*Developmental disturbances*. Rank-order stability can be used to model interindividual stability over time as an autoregressive process,. That is, the rank-order stability can be used to model continuity of a behavior. Assuming equal intervals, the correlations between T0 and T1, and between T1 and T2 could be used to estimate the correlation between T2 and T3. However, there can be periods of life where various environmental pressures will tend to suppress this stability, disrupting this orderly stochastic process. For example, suppose there is considerable variability in how much alcohol someone drinks while in high school and that there is relatively high rank-order stability in drinking across the high-school years. Suppose further, that someone enters a fraternity or Greek house on enrollment to college and there is extreme environmental press on all members to drink at high levels, shifting the mean level of consumption up, variability down, and thus reducing autoregressivity so that drinking in the freshman year doesn’t correlate as highly with drinking in the sophomore year of college as did drinking across the high school years. Once someone leaves the “environment,” individual differences are again more freely expressed and the autoregressivity increases to levels similar to those prior to entering the environment. We call this type of situation a developmental disturbance in that the normal course of development is “disturbed” for a period of time by some time-limited process. In such cases, we might find that prediction of behavior in the period of the disturbance might be reduced and, similarly, behavior exhibited during the disturbance might have less predictive power with respect to distal outcomes compared to behavior exhibited prior to and following the disuprted period.

*Vulnerable periods.*  One of the primary scientific objectives of ABCD is to examine the effects of cannabis exposure on the development of psychosis. However, the developing brain might be more sensitive to cannabis exposure during periods of rapid brain development or remodeling. One study found that cannabis use during adolescence was associated with later psychosis to a greater degree than cannabis use initiated later in development. Note these types of associations by themselves cannot establish causation since it’s always possible that early initiation into cannabis use is a premorbid manifestation of psychosis proneness. However, it is consistent with the hypothesis and could form the basis of further analysis seeking to establish causality more formally. Similarly, experimental research on rodents has shown 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). These types of findings highlight the importance of considering developmental stage in when trying to establish the association between an exposure and an outcome.

*Developmental snares and cascade effects.* While development normatively progresses from less mature to more mature levels of functioning, different kinds of experiences can upend this idealized form of development. Various kinds of experiences (e.g., drug use) can, through various mechanisms, disrupt the normal flow of development where each stage establishes a platform for the next. For example, substance use could lead to association with deviant peers, precluding opportunities for learning various adapative skills and prosocial behaviors, in effect, creating a “snare” that retards psychosocial maturation. 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 critical 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.

*Unidirectional and transactional effects*. In studying the effects of various types of exposures (e.g., cannabis intoxication) on outcomes (e.g., psychosis), we often think of unidirectional causation; some cause at T1 results in an effect at T2. However, it is often the case that that the ostensible cause and effect transact to mutually influence each other over time. The so-called “transactional model” (e.g., Sameroff, 2009) can be used to model how certain characteristics of the individual (e.g., venturesomeness, deviance proness) could lead, through niche-seeking, to affiliation with substance using peers who in turn encourage further deviance (refsSimilarly, those prone to an outcome (e.g., have a vulnerability to develop a specific disorder) may also be prone to seek out certain types of environmental exposures which in turn aggravate the underlying vulnerability, making the unfavorable outcome more likely. One advantage of longitudinal analyses with multiple waves of data collection is that such processes can be modeled over time. Understanding transactional effects effects can be useful in devising strategies to intervene in various types of escalating cycles of violence such as child maltreatment (Cicchetti & Valentino, 2006) or the effects of intimate partner violence on child behavior problems (Chung et al., 2021). Such processes could be modeled statistically in a number of ways by relying on estimating cross-lagged relations but require at least three times of measurement in order to distinguish beween-subject and within-sources of measurement (Hamaker, xxx; see discussion below).

*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. For example, 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, 2010), 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, Ward, Mendelson, Mock, & Erlbaugh, 1961; see French & Sutton, 2010) and this phenomenon has been well documented in research on structured diagnostic interviews (Robins, 1985 ). While it is typically assumed that individuals are rescinding or telling us less information on follow-up interviews, there is reason to suspect that in some cases the initial assessment may be artefactually elevated (see Shrout et al., 2018 ).

Some designs (sprecifically, accelerated longitudinal designs) are especially well suited for discovering these effects and modelling 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 on a yearly basis, 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 number of prior assessments. However, ABCD is fundamentally a single-cohort, longitudinal design, where number of prior assessments and age are highly confounded and for, perhaps, most analyses, the possible influence of experience effects need to be kept in mind.

*Causal inference*. [I think Wes is the person best suited for this]

*Attrition.*  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 (Cottler et al., 2004; 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 design to provide the most valid inferences. Perhaps the most key concern with dealing with data missing due to attrition is assessing the degree of bias in retained variables attributable to 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 and degree of missingness and variables available to help estimate variables. [[[ need to have a quant person discuss types of estimation appropriate with MAR and other approaches such as multiple imputation, propensity scores and auxiliary variables, Bayesian methods, etc. Probably want to mention techniques to avoid including last point carried forward, listwise deletion, mean imputation, etc.]]]

*Model selection*. Not sure what was meant here but I suspect this little subheading could be a very large book.

ge population. *Biometrika* **10,** 507–521 (1915).