



Why weight? Analytic approaches for large-scale population neuroscience data

Arianna M. Gard^{a,b,*}, Luke W. Hyde^{a,c}, Steven G. Heeringa^a, Brady T. West^a, Colter Mitchell^a

^a Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

^b Department of Psychology, Neuroscience and Cognitive Neuroscience Program, University of Maryland, College Park, MD, USA

^c Department of Psychology, University of Michigan, Ann Arbor, MI, USA

ARTICLE INFO

Keywords:

Population neuroscience
ABCD Study®
Generalizability
convenience sampling
probability sampling

ABSTRACT

Population-based neuroimaging studies that feature complex sampling designs enable researchers to generalize their results more widely. However, several theoretical and analytical questions pose challenges to researchers interested in these data. The following is a resource for researchers interested in using population-based neuroimaging data. We provide an overview of sampling designs and describe the differences between traditional model-based analyses and survey-oriented design-based analyses. To elucidate key concepts, we leverage data from the Adolescent Brain Cognitive Development™ Study (ABCD Study®), a population-based sample of 11,878 9–10-year-olds in the United States. Analyses revealed modest sociodemographic discrepancies between the target population of 9–10-year-olds in the U.S. and both the recruited ABCD sample and the analytic sample with usable structural and functional imaging data. In evaluating the associations between socioeconomic resources (i.e., constructs that are tightly linked to recruitment biases) and several metrics of brain development, we show that model-based approaches over-estimated the associations of household income and under-estimated the associations of caregiver education with total cortical volume and surface area. Comparable results were found in models predicting neural function during two fMRI task paradigms. We conclude with recommendations for ABCD Study® users and users of population-based neuroimaging cohorts more broadly.

1. Introduction

Population neuroscience (Falk et al., 2013; Paus, 2010) arose out of a desire to increase statistical power in neuroimaging studies, capture the substantial inter-individual variability in environmental exposures and outcomes, and improve the generalizability of neuroimaging research. Recent neuroimaging studies containing hundreds to thousands of participants (e.g., Human Connectome Project (Van Essen et al., 2013), UK Biobank (Miller et al., 2016), IMAGEN (Schumann et al., 2010) have been enormously successful in increasing statistical power. Moreover, these studies have become important resources for the scientific community via accessible data sharing (Nichols et al., 2017). Less clear is the degree to which these studies have been helpful in improving the representation and generalizability of neuroimaging studies more broadly (Falk et al., 2013; Nielsen et al., 2017).

One study at the intersection of these issues is the Adolescent Brain Cognitive Development™ Study (ABCD Study®), which is both very large and adopted a complex probability sampling design intended to

improve generalizability over typical convenience samples (Garavan et al., 2018). Through the combination of open access data, a large sample size, and probability sampling, the ABCD Study® is poised to support new scientific discoveries. However, several foundational and analytical questions pose challenges to researchers interested in using these data. The current paper aims to provide neuroscientists with a brief introduction to analytic tools and concepts vital to making population inferences from complex probability samples – specifically, the components of complex sampling designs and choosing an analytic approach that matches researcher goals. We provide empirical examples using the ABCD Study® and end with recommendations for researchers interested in using population-based neuroimaging datasets.

1.1. Background

1.1.1. Sampling approaches

When designing a sampling and recruitment plan, there are tradeoffs between representation and cost. Most neuroimaging studies adopt

* Correspondence to: 4094 Campus Drive. 3123G Biology-Psychology, College Park, MD 20742, USA.

E-mail address: arigard@umd.edu (A.M. Gard).

convenience sampling methods, wherein participants are recruited in community settings (e.g., clinics, university subject pools, registry databases) through flyers, online advertisements, and word-of-mouth (Nielsen et al., 2017). Convenience sampling, and other forms of non-probability sampling, are the most cost-effective methods to recruit participants into research studies (Taherdoost, 2016). Consequently, this approach dominates many fields of study (e.g., medicine (Harris et al., 2005), epidemiology (Cheung et al., 2017), psychology (Nielsen et al., 2017), linguistics (Andringa and Godfroid, 2020).

Unfortunately, non-probability sampling approaches can be prone to selection bias (Oswald et al., 2013; Taylor, 2004), sometimes severe, in the form of under-representation of large segments of the population to which researchers are trying to generalize (i.e., the “target population”, see Table 1) (Dotson and Duarte, 2020; Nielsen et al., 2017). This is not always the case: recruitment strategies that target the population of interest (e.g., hiring “cultural insiders”, extensive outreach activities, including participants in study design (Rowley and Camacho, 2015; Yancey et al., 2006) have been enormously successful in recruiting diverse samples to participate in neuroimaging studies (Brody et al., 2017; Habibi et al., 2015; Hein et al., 2018). Yet even within samples that include participants from diverse sociodemographic backgrounds, selection biases remain. For example, in an analysis of the Pediatric Imaging, Neurocognition, and Genetics (PING) study, a non-probability sample of 1493 youth aged 3–20 years, LeWinn and colleagues adjusted the demographics of PING to match that of the U.S. population of youth aged 3–18 years (LeWinn et al., 2017). Compared to the target population, PING was over-represented by Hispanic youth, youth from high income families ($\geq \$100,000$ annually), and caregivers¹ with college degrees or higher. These discrepancies in sociodemographic composition impacted research questions of interest: in the unweighted PING sample, age associations with cortical and subcortical volume and cortical surface area were largely quadratic, following U-shaped patterns. In contrast, in the weighted sample, age effects were predominantly cubic, following S-shaped patterns. Thus, even when studies use convenience sampling to recruit a sample that contains sociodemographic diversity, unweighted analyses may still yield conclusions that are biased, leading to assumptions in the field about basic science questions (e.g., how the structure of the brain changes across development). Although arguments for larger samples sizes abound in neuroimaging and genetics research (Marek et al., 2022), it is important to note that increasing statistical power through larger sample sizes will not necessarily eliminate sample selection biases (Bradley et al., 2021; Meng, 2018).

Probability samples, by contrast, give every member of a specific target population (i.e., the population to which a study is trying to generalize) a non-zero chance of selection into the sample (Heeringa et al., 2017). For example, in a convenience sampling approach that recruits participants through flyers posted in medical offices, only individuals who see the flyers (e.g., those with access to the medical facility during the time that the flyers are posted) are given the opportunity to participate. In a probability sampling approach, researchers create a sampling frame (i.e., a list) of all potential participants from a specific target population (Table 1). In this example, the sampling frame might be a list of all patients who attend specific medical clinics during a specific year. As a result, every individual in the sampling frame is given the opportunity to participate. From the sampling frame, researchers randomly select a subset of individuals (i.e., a sample) and then attempt to recruit those individuals (and only those individuals) into the study. Of course, there will always be some degree of selection bias in the form of unobserved differences between the sampling frame and the target population (coverage error), the sample and the sampling frame (sampling error) and respondents and the selected sample

Table 1

Key Terms and Concepts.

Target population. The most abstract conceptualization of the population to be studied. For example, all 10-year-olds living in the United States (US).

Sampling frame. The set of people within the target population who have a non-zero chance of being selected into the study. For example, an administrative list of all 10-year-olds attending public, private, or charter schools in the US. A supplementary frame can be added to target under-covered populations (e.g., 10-year-olds who are home-schooled). *Coverage error* occurs when the sampling frame does not perfectly map to the target population (e.g., 10-year-olds who are not enrolled in school, or students who newly entered the US school system but are not yet registered in administrative records).

Probability sample. A sample wherein every individual in the sampling frame has a non-zero probability of being selected into the study.

Simple random sampling. A probability sampling approach that gives every individual an equal and independent probability of selection (e.g., randomly selecting names from a nationwide list of 10-year-olds).

Complex sampling designs. A probability sampling approach that leverage procedures such as stratification and cluster sampling to increase target population representation and reduce data collection costs. For example, a two-stage area probability sample of 10-year-olds in the US might identify clusters as schools (randomly selecting schools), followed by clusters as classrooms (randomly selecting classrooms), and ultimately sample students within classrooms. As individuals in a complex sampling design are not independently selected (e.g., they are nested within clusters or strata), users must account for this nesting in variance estimation, and sample sizes must increase to have similar statistical power as in a simple random sample.

Cluster Sampling. A feature of complex sampling designs wherein groups of units (e.g., students in schools) are sampled simultaneously. Clustering can function to (a) make data collection more efficient (i.e., researchers travel to schools rather than to individual households), and (b) enable the construction of a sampling frame (e.g., individual schools will maintain administrative list of students, but there may not be one nationwide administrative list of all 10-year-olds who attend public, private, or charter schools in the United States).

Stratification. A feature of complex sampling designs wherein the sampling frame is partitioned into homogeneous subpopulations based on specific characteristics (e.g., socioeconomic resources, urbanicity), generally to increase target population representation – e.g., schools stratified by per pupil spending to ensure equal representation of students attending low, medium, and high-resourced schools.

Sample. The set of individuals (e.g., 10-year-olds) from which measurements will be attempted. *Sampling error* may come in the form of *sampling bias* – wherein certain types of individuals from the sampling frame are given a reduced chance of selection and this error is systematic – or *sampling variance* – which is a function of stratification (decrease in variance), cluster sampling (increase in variance), and sample size (decrease in variance).

Respondents. Individuals who are successfully measured in the sample. *Nonresponse error* occurs when inferences based only on respondent data differ from those based on the entire sample. *Nonresponse bias* reflects both the nonresponse rate and differences in sample means between respondents and nonrespondents.

Analytic sample. The set of observations that contribute data to the analysis. In a neuroimaging study, for example, the analytic sample is often restricted to individuals with usable imaging data. A missing values analysis compares the analytic sample to the excluded sample on a variety of other sociodemographic and clinical features.

Survey Weights. In the context of a complex sample design, survey weights adjust for unequal probabilities of selection (i.e., sample selection weight), sample nonresponse (i.e., nonresponse weight), and a post-survey adjustments (e.g., post-stratification weight). Post-survey adjustments are used to match the characteristics of the sample to a known population. Several methods exist, including *post-stratification or calibration* – which relies on propensity-based matching of sample membership to target population membership – and *raking* – which relies on iterative proportional fitting of each variable (e.g., household income) separately.

(continued on next page)

¹ Throughout the manuscript, we use the term “caregiver” instead of “parent/parental” as a more inclusive descriptor.

Table 1 (continued)

The simultaneous incorporation of all three weight components in a complex sample design highlights the distinction between post-stratification weights applied to non-probability samples and survey weights applied to probability samples. For example, as LeWinn and colleagues (2019) leveraged a non-probability sample, the survey weights only included post-survey adjustments to the target population – sample selection and non-response weights were not included in the overall survey weight because these features were unknown, as in all non-probability samples.

Variance estimation. Complex sample designs (e.g., featuring clustering of students within schools) necessitate alternative methods for estimating the standard deviations of the sampling distributions of estimates based on many hypothetical samples using the same design (i.e., standard errors); otherwise *estimates will appear more precise than they are in reality*. Several methods have been developed to estimate the standard error of a point estimate, including Taylor Series Linearization, balanced repeated replication, and jackknife repeated replication.

Design-based analyses. Statistical inferences from sample data are based on the distribution of all possible samples that could have been chosen under the specified probability sampling design. Design-based analyses are sometimes labeled “nonparametric” or “distribution free” because they rely only on the known probability that a given sample was chosen (e.g., in a simple random sampling design, every individual has an equal and independent probability of selection into the sample).

Model-based analyses. Statistical inferences from sample data are based on a probability distribution for the variables of interest rather than the probability distribution for the sample selection. Model-based analyses rely on a correctly-specified model and the associated parametric distributional assumptions to produce unbiased parameters that generalize. Most neurodevelopmental studies implement model-based approaches to make population inferences.

(non-response error) (Table 1).

The key difference between selection bias in a probability sample versus a non-probability sample is that survey weights can effectively adjust for these biases (i.e., non-probability of selection, non-response adjustment, poststratification adjustment to the target population; see Table 1) and permit estimation of the variance of the sampling distribution of survey estimates that would be generated from many hypothetical samples following the exact sample design. That is, in a probability sample, the identification of a target population and a sampling frame can be used to assess who is missing and at what rate, allowing researchers to adjust estimates for sampling biases. The random selection of elements from the sampling frame based on known probabilities of selection (e.g., oversampling households in rural areas) allows for the estimation of sampling variance. In a convenience sample without an explicit target population or sampling frame, researchers may have no way of knowing how respondents are different than the target population, thus making it difficult or impossible to adjust for multiple forms of selection bias or estimate sampling variance (Kish, 1965). In sum, in a probability sample, everyone in the sampling frame has a chance of being selected and we have a good idea of who is missing in various ways. In a convenience sample, we simply don't know the target population or sampling frame, nor do we know who and why certain groups of people are missing in our sample. The latter case can directly impact our ability to generalize to a broader (and explicitly-defined) population.

Clear downsides to probability sampling are cost and feasibility, issues that are magnified by the expense of neuroimaging protocols. Depending on the population to which a researcher is trying to generalize (e.g., 10-year-old children in the U.S. versus 10-year-old children attending one public school district in one urban U.S. city), the cost of collecting data from potentially anywhere in the U.S. – from Capitan, NM to Yakima, WA – let alone finding a participating MRI facility, would be prohibitive. Moreover, if complex sampling designs leverage cluster sampling and/or stratification (Table 1) to reduce the costs of recruitment, the sample size must be larger to account for the non-independence among responding individuals due to the cluster sampling. Lastly, if the target population is more heterogeneous (e.g., the U.S.

vs. one public school), this would also necessitate a larger sample size to ensure adequate statistical power to estimate reliable effects.

Collectively, this means that designing a probability sample, particularly in neuroimaging research, may not be feasible for individual scientists. However, fueled by calls for open-access data (Nichols et al., 2017) and statistically well-powered studies (Button et al., 2013), multi-site collaborative studies are becoming more common (e.g., Generation R; [White et al., 2013]; Human Connectome Project in Development; (Somerville et al., 2018); ABCD Study®; Garavan et al., 2018). With encouragement from survey methodologists (e.g., Falk et al., 2013), some of these studies have implemented complex sampling designs or added neuroimaging to existing studies with complex sampling designs (e.g., ABCD Study®; the Michigan Twins and Neurogenetics Study, (Tomlinson et al., 2020); the Study of Adolescent Neurodevelopment, (Hein et al., 2018).

As these types of data become publicly accessible, it becomes more and more important that data users know how to leverage these studies to generate scientific conclusions. Thus, for non-survey methodologists, how can researchers take advantage of the sampling design to generalize findings to the broader target population (i.e., beyond the recruited sample)?

1.1.2. Analytic approaches for complex sampling designs

Approaches to making statistical inferences about populations based on samples selected from those populations may be *model-based* or *design-based* (Table 1). Most neurodevelopmental studies have used model-based approaches (Fisher, 1955) to make population inferences (e.g., linear mixed effects modeling, ordinary least squares regression, analysis of variance). Take, as an example, a researcher interested in examining the associations between socioeconomic resources (SER) and amygdala volume (AV), a region thought to support salience detection, emotional learning, and threat processing (Janak and Tye, 2015; LeDoux, 2000). In a model-based approach, the researcher adopts a statistical model (e.g., $AV_i = \beta_0 + SERx_i + \varepsilon_i$), makes parametric distributional assumptions (e.g., the errors in the regression model denoted by ε_i are independently and identically distributed with mean 0 and variance σ^2), and accounts for sample clustering (e.g., students in the sample are clustered within schools) and selection biases (e.g., the analytic sample differs from the excluded sample on known variables, which are included as covariates). The model-based approach is flexible in that it can be applied to data from non-probability or probability sampling designs.

However, like all statistical models, if any of these assumptions/conditions are not met (e.g., distributional assumptions, model fit, identification of constructs that predict usable data), parameter estimates may be biased (Sterba, 2009) and/or standard errors may be too small (Kish and Frankel, 1974), leading to incorrect conclusions about the association between SER and AV or the statistical significance of this association. By contrast, a design-based approach relies on the probability sampling design (and thus can only be implemented in probability samples) to make inferences from the recruited sample to the target population. Design-based analyses guard against the assumptions required in a model-based framework by leveraging survey weights to account for selection probabilities, non-response, and poststratification adjustments to estimate true population parameters.

Design-based analyses also use information about the weights, stratification (if applicable), and cluster sampling (if applicable) to correctly estimate the variance of such parameter estimates (Heeringa et al., 2017). As a result, standard errors derived from design-based analyses are unbiased (or nearly unbiased) estimates of the variance of a parameter estimate given a complex probability sampling design (Heeringa et al., 2017). Similarly, applications of survey weights will yield population-generalizable parameter estimates that are unbiased – whether these parameter estimates are similar to those produced by model-based approaches depends on the information about the variables

of interest contained in the weights (i.e., are the variables of interest correlated with the weights – either through selection probabilities, non-response, or post-stratification to the target population). It is critical to note that in model-based frameworks that account for the nested structure of the data (i.e., as in the ABCD Study®, wherein individuals are nested within study sites), survey weights must be partitioned at each level of analysis (e.g., level-1 weights for participants, level-2 weights for study site). If multi-level survey weights are not yet available (i.e., as in the case of ABCD Study®), users are suggested to include the variables used to construct the survey weights as covariates.

Although design-based approaches were originally restricted to descriptive inference (e.g., estimating the proportion of the target population that meets criteria for a neurodevelopmental disorder), statistical advances have made design-based estimation of regression models possible (Binder, 1983; Kish and Frankel, 1974) (e.g., associations between household income and the diagnostic likelihood of a neurodevelopmental disorder). A benefit of these approaches is that even if the statistical model is mis-specified, estimated parameters will still be generalizable to the target population in a probability sampling design. Further, advances in survey statistics have allowed these more “analytic” design-based approaches to be applied to a wide variety of analytic frameworks (e.g., structural equation modeling, Bayesian analyses, general linear mixed models), though this is an active area of research (Heeringa et al., 2017). A variety of packages for implementing design-based analyses have been developed in many commonly-used statistical packages (i.e., R Statistical Software, STATA, Mplus, SAS, SPSS), making the design-based framework feasible for the data user more familiar with model-based approaches.

1.1.3. Threats to generalizability in complex sampling designs

When does inference from the sample to the target population break down in the context of probability sampling designs? At each stage in the design and recruitment process, multiple types of error threaten population inference. Fig. 1 displays the lifecycle of a complex sampling design, using the ABCD Study® as an example. The ABCD Study® is a longitudinal population-based neuroimaging study of 11,878 eligible 9–10-year-olds in the U.S. (Garavan et al., 2018). The target population is 9–10-year-old children living in the U.S., born between 2006 and 2008 (Fig. 1). To identify children in the target population, a sampling frame was constructed – a list of public and private elementary schools in districts wherein at least 50% of the schools were located within the catchment area of one of 21 ABCD recruitment sites where the necessary neuroimaging equipment was present/chosen through grant applications. There is necessarily *coverage error* in the sampling frame – it does not include, for example, elementary schools outside of a catchment area (e.g., not near the participating academic medical centers) or home-school settings.²

From the sampling frame, researchers selected a sample of schools to contact. *Sampling error* arises when, by chance, the selected sample is different in some way (e.g., on demographic characteristics) than the non-selected sample. Of schools selected by the ABCD Study® survey methods team, not all consented to participate, contributing to *non-response error*. Further, once a school agreed to participate in recruitment by facilitating contact with all students, children had to be eligible, and their parents or guardians had to consent to participate in the study. Some of the exclusion criteria included MRI contraindications (e.g., cardiac pacemakers, metal implants), inability to speak or understand English, uncorrected vision, hearing or sensorimotor impairments, birth

weight < 1200 g, and gestational age < 28 weeks. Thus, eligibility restrictions and differences in who consented to participate constitute additional sources of *non-response error*.

Such biases in coverage error, sampling error, and non-response error can be accounted for by postsurvey adjustments (i.e., adjustment for non-probability of selection, non-response, and calibration to the target population through survey weights). However, in the context of a neuroimaging study wherein data loss due to poor data quality is common (e.g., due to movement, low behavioral performance, falling asleep), there may be substantial missing data in the sample of respondents used in analyses. Critically, when design features such as survey weights are correlated with missing data, target population inference cannot be achieved without attention to missing data (e.g., multiple imputation, full-information maximum likelihood) (Groves et al., 2009; Kish and Frankel, 1974). Moreover, empirical analyses have shown that the influence of survey weights on point estimates and standard errors may be larger when the weights themselves are correlated with the variables of interest (Heeringa et al., 2017; Spencer, 2000). That is, when missing data is related to the outcome being studied, survey weights will have a bigger impact on conclusions. This issue is a particular challenge for developmental neuroscientists because data loss in neuroimaging (e.g., through movement) may be highly related to a construct of interest (e.g., ADHD symptoms) (Kong et al., 2014; Satterthwaite et al., 2012).

1.1.4. The salience of socioeconomic resources for generalizable neuroimaging research

One goal of population neuroscience is to understand how environmental inputs shape brain structure and function to support cognition, behavior, and health in large representative samples (Falk et al., 2013; Paus, 2010). Owing to decades of behavioral research highlighting the centrality of socioeconomic resources (SER) in shaping environmental opportunity, a rich literature in human neuroscience has sought to explore how SER sculpts structural and functional brain development (Farah, 2017). Studies have examined structure and function across the entire cortex and within specific brain regions (Johnson et al., 2016; Rakesh and Whittle, 2021) that underlie salience detection (Maren et al., 2013), working memory (Jonides et al., 2008), and cognitive control (Fuster, 2001; Hampshire et al., 2010). Yet non-response constitutes a major threat to the generalizability of this research. Neuroimaging studies historically under-represent participants from low socioeconomic backgrounds and from marginalized identities who are more likely to be subjected to adversity due to structural inequalities and previous experiences of mistreatment in research and medical settings (Dotson and Duarte, 2020; Falk et al., 2013). Thus, postsurvey adjustments are likely to yield survey weights that are correlated with socioeconomic resources (i.e., lower-SER individuals will be assigned larger weights to maintain target population generalizability), and in analyses that examine the associations between SER and brain development, parameter estimates (i.e., point estimates and standard errors) may be highly sensitive to the decision of whether to adjust for survey weights or not (Heeringa et al., 2017; Spencer, 2000).

2. Empirical demonstration

We now provide an empirical demonstration using data from the ABCD Study®. The ABCD Study® is an important dataset within which to evaluate sampling biases and analytic options because it is a well-sampled, large, and a publicly-available resource likely to be utilized extensively over the coming decades. We had three primary research questions:

- **How does the target population of 9–10-year-olds in the U.S. differ from the ABCD Study® respondent sample?** For this aim, we examined several sociodemographic variables (e.g., race-ethnicity, household income, caregiver education, child sex), many of which

² Although home-schooled children were not included in the original sampling frame because of the school-based recruitment strategy, home-schooling was not a strict exclusion criterion. Home-schooled children were given the opportunity to participate through alternative recruitment strategies (e.g., media, outreach), which comprised less than 10% of the final ABCD Study® sample

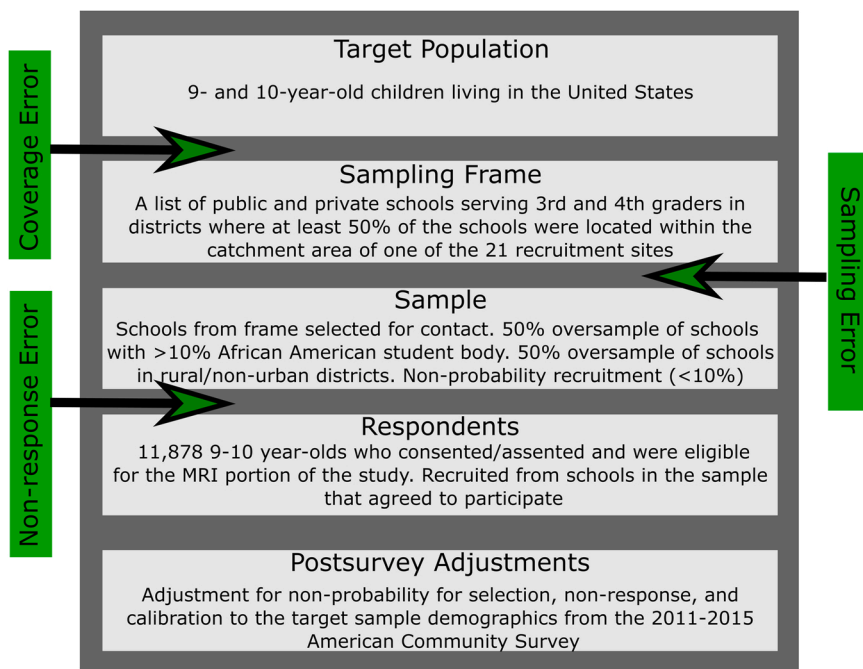


Fig. 1. Lifecycle of the ABCD Complex Sampling Design. Note. Lifecycle of the ABCD Study® complex sampling design, from identification of the target population to postsurvey adjustments before analyses. Note that the ABCD Study® cohort also includes an oversample of twins from four of the 21 study sites (150 – 250 twin pairs per site), sampled from state registries (Garavan et al., 2018). Thus, the sampling design is different from the rest of the ABCD Study® sample. For simplicity, we do not make a distinction between the twin sample and the general population sample. In addition, less than 10% of the final study sample was recruited through alternative strategies (e.g., media, outreach), which allowed for the recruitment of children who otherwise would have been excluded based on the sampling frame (e.g., home-schooled children). Figure adapted from (Groves et al., 2009).

were used to construct the ABCD Study® survey weights (see Methods). Based on previous reports of participant characteristics in neuroimaging and cognitive neuroscience studies (Dotson and Duarte, 2020; Rowley and Camacho, 2015), we hypothesized that children from low-SER households (i.e., household income below poverty line, caregivers with a high school degree) and racial-ethnically marginalized children would be under-represented in the ABCD Study® respondent sample compared to the target population. We note the issue of “population representation” in the ABCD Study® was first raised in a commentary by Compton et al. (2019). Here, we explore this question empirically.

- **How does the analytic sample in structural MRI (sMRI) and task-based functional MRI (fMRI) analyses differ from the ABCD Study® respondent sample?** To evaluate this research question, we identified usable sMRI data and fMRI data from two tasks (the Emotional N-back [EN-back] and the Stop Signal Task [SST]) using inclusion recommendations from the ABCD Study® release 3.0 documentation. We examined the EN-back and SST tasks given the wealth of previous research (Farah, 2017; Johnson et al., 2016; Rakesh and Whittle, 2021) highlighting associations between socioeconomic resources (i.e., a primary construct of interest in our evaluation of sampling biases) and processes captured by these tasks (i.e., EN-back: salience and emotion processing; SST: response inhibition, impulsivity). To comprehensively characterize predictors of missing data across imaging modalities, we examined constructs across several domains: (1) demographic (e.g., race-ethnicity, household income, caregiver education, child sex), (2) social-contextual (e.g., neighborhood safety, school quality, lead exposure risk), (3) visit characteristics (e.g., number of scanning sessions), and (4) individual-level survey weights. In addition, we leveraged participant feedback (e.g., “the consent form was clear”, “participants felt comfortable with staff”) from an ABCD Study® subsample to probe whether missing data was associated with participant perceptions of study characteristics. Although demographic differences between participants with and without task-based fMRI data were included in a recent report (Chaarani et al., 2021), we expand this investigation to include social-contextual variables, visit characteristics, survey weights, and participant feedback. Our goal was to holistically characterize the

representativeness of participants in the analytic versus recruited samples.

- **What is the impact of analytic method on parameter estimates when evaluating the associations between socioeconomic resources and brain structure and function?** Although associations between individual measures of socioeconomic resources (i.e., neighborhood disadvantage, household income-to-needs, and caregiver education) and imaging metrics have already been examined using ABCD data (e.g., Rakesh et al., 2021; Taylor, et al., 2020), these investigations have been limited to model-based analyses, most often within a multi-level modeling (MLM) framework that accounts for nesting of families within study site; less commonly occurring have been model-based analyses that do not account for the nested structure of the data (e.g., Vargas et al., 2020). Here, we compared parameter estimates (i.e., point estimates and standard errors) across analytic approaches – design-based and two model-based methods (i.e., “Model-Based MLM” and unnested ordinary least squares, or “Model-Based OLS”). Model-based approaches did not account for the survey weights in estimation because multi-level weights are not currently available for ABCD Study®; rather, we controlled for the constructs used in the generation of the survey weights (Heeringa and Berglund, 2019). Consistent with decades of research in survey methodology (Heeringa et al., 2017), we hypothesized that the clustering of participants by study site and the application of survey weights would result in larger standard errors in design-based analyses than model-based analyses, which do not account for survey weights. That is, design-based analyses which account for the clustered sampling design will necessarily increase the sampling variance in estimates – an expected tradeoff of increased generalizability in design-based analyses. We also hypothesized that point estimates (i.e., the estimated coefficients representing SER associations with brain metrics) would change across modeling frameworks as a result of the application of informative survey weights (Bollen et al., 2016; Heeringa et al., 2017; Korn and Graubard, 1995; LeWinn et al., 2017). As this was the first analysis of its kind, we made no predictions about the direction or magnitude of changes in point estimates across modeling frameworks.

3. Method

3.1. Sample

The Adolescent Brain Cognitive DevelopmentSM Study (ABCD Study®) is a longitudinal population-based neuroimaging study of 11,878 eligible 9–10-year-olds in the United States (Garavan et al., 2018; Fig. 1). To construct the sample, the study employed a clustered probability sample to recruit eligible children from a comprehensive list of public and private schools within 21 catchment areas near study sites. Multiple children from each household were eligible to participate. Participating caregivers were administered informed consent and provided caregiver consent for participating minors. The baseline ABCD Study® cohort also included an oversample of twins from four of the 21 study sites (150 – 250 twin pairs per site), sampled from state registries (Garavan et al., 2018). For simplicity, we do not make a distinction between the twin sample and the general population sample in the current empirical demonstration. Behavioral data was downloaded from the second public release (version 2.0.1, released July 2019; <https://doi.org/10.15154/1504041>) and neuroimaging and additional behavioral data was downloaded from the third public release (version 3.0, released October 2020; <https://doi.org/10.15154/1519007>). All measures were downloaded directly from the NIMH Data Archive collection #2573 (DUA #3067).

3.2. Measures

3.2.1. Demographic

Demographic- and individual-level constructs included youth sex (1 = male, 2 = female), race-ethnicity (Asian, Biracial or Multiracial, Black, Hispanic or Latino/a, White, and [due to small sample sizes] Native American/Alaskan/Hawaiian or Pacific Islander or other race not specified), pubertal development (Potter et al., 2020), number of traumatic brain incidents (TBI), caregiver-reported externalizing and internalizing behavior total scores, caregiver marital status (married, divorced, cohabitating, never married, separated, widowed), and primary caregiver employment status (employed, not in labor force, unemployed, other or refused).

Household income, caregiver education, and neighborhood disadvantage were used to index socioeconomic resources (Bradley and Corwyn, 2002; Conger et al., 2010). Annual household income reported by primary caregivers was measured semi-continuously on an ordinal scale from less than \$5,000 to more than \$200,000. We recoded this variable to reflect income-to-needs relative to the U.S. federal poverty line (FPL) for a family of four, which in 2019 was ~\$25,000 (Semega et al., 2019): 1 = < 100% FPL, 2 = 100–200% FPL, 3 = 200 – 400% FPL, and 4 = > 400% FPL. Primary caregiver and partner years of education were each recoded into six-level categorical variables: 1 = less than high school, 2 = high school degree or equivalent, 3 = some college, 4 = associates or occupational degree, 5 = college degree, 6 = masters or professional degree, and then combined into one variable. The highest education level of the primary caregiver or partner was used to index caregiver education. Neighborhood disadvantage was measured using a sum score of nine U.S. Census tract-level indicators, based on a recent analysis in the ABCD Study® baseline sample (Taylor et al., 2020). See the [Supplemental Methods](#) for more details.

3.2.2. Social-contextual and visit characteristics

Social-contextual characteristics included Census-derived neighborhood disadvantage (Taylor et al., 2020) and lead exposure risk (Marshall et al., 2020), caregiver-reported neighborhood safety, and youth-reported school quality and family conflict. We also examined visit characteristics as correlates of missingness to guide future protocols that might maximize retention. Visit characteristics included the number of scanning sessions (i.e., across how many sessions was the imaging data collected at the baseline visit; 92% one versus two sessions), and

the number of available runs for task-based fMRI (see [Supplemental Methods](#) for more detail). In addition, a small subset of ABCD Study® participants was invited to participate in the ABCD Study® Social Development Study (ABCD-SD; $N = 989$). ABCD-SD (Hoffman et al., 2019) collected more detailed phenotypic and environmental measures. Primary caregivers were also asked to provide feedback on several components of the protocol (e.g., “The staff explained the study clearly”, “I did not understand some of the questions”, “The level of compensation seemed appropriate”) by rating each item on a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). We examined caregiver feedback on 16 prompts that were administered to families who completed the ABCD-SD protocol on either the same day or a different day as the ABCD Study® baseline protocol.

3.2.3. Survey weights and primary sampling units

Survey weights were developed for the ABCD Study® baseline cohort that account for (1) quasi-probability of initial selection into the study, (2) conditional probabilities of study participation, and (3) calibration to external population controls (Heeringa and Berglund, 2019). Inverse propensity weighting (Elliott and Valliant, 2017; Kim, 2022) was used to benchmark ABCD Study® baseline weights to estimated features of the target population based on 2011 – 2015 American Community Survey (ACS) demographic and socioeconomic estimates for U.S. children ages 9 and 10 ($n = 376,370$). The variables used in the weighting procedure included: child age in years, child sex, child race-ethnicity, family income, family type, household size, caregiver employment status, and Census region. For more details about the ABCD Study® baseline survey weights, see Heeringa and Berglund (2019).

The primary sampling units (PSUs) in the ABCD Study® complex sample design are the 21 study sites. Although the study sites were based on the locations of neuroimaging research centers rather than conventional probability sampling of PSUs (i.e., the sites were not chosen at random), study sites were distributed throughout the U.S. and largely captured the range of demographic and socioeconomic diversity of the target population (Garavan et al., 2018; Heeringa and Berglund, 2019).³

3.3. Imaging data

3.3.1. Structural MRI

Details on MRI acquisition, preprocessing, and quality control can be found elsewhere (Casey et al., 2018; Hagler et al., 2019) and are summarized in the [Supplemental Methods](#).

We relied on imaging inclusion recommendations from the ABCD Study® release 3.0 documentation (Anon, 2020) (`imgincl_t1w_include = 1`, from the `abcd_imgincl01.txt` instrument) to create our final subject list. Several measures of structural brain development were used as dependent variables, all previously shown to be associated with household socioeconomic resources (Johnson et al., 2016; Rakesh and Whittle, 2021): (1) bilateral hippocampal volume (i.e., sum of left and right hippocampal volume), (2) whole brain cortical surface area, and (3) whole brain cortical volume.

3.3.2. Task-based functional MRI

Details on fMRI data acquisition, preprocessing, and quality control can be found elsewhere (Casey et al., 2018; Hagler et al., 2019), and are briefly summarized in the [Supplemental Methods](#). fMRI data from two tasks were used in the current analytic demonstration: Emotional N-Back (EN-back) and Stop Signal Task (SST). The EN-back task is used to elicit neural function during emotional salience (i.e., by contrasting

³ Note that there were originally 22 study sites. After the baseline wave, one site closed, and participating families were recruited into one of the remaining 21 sites. Therefore, for participants from site 22, we used their assigned study site at the 6-month or 1-year follow-up wave to maximize PSU completeness. Five participants could not be reassigned a PSU.

activation during happy or fearful face versus neutral face trials) and working memory (i.e., by contrasting activation during 2-back versus 0-back blocks). The SST engages neural function during impulse control (i.e., correct “stops”) and impulsivity (i.e., failed “stops”).

We relied on imaging inclusion recommendations from the ABCD Study® release 3.0 documentation (Anon, 2020) (`imgincl_nback_include = 1` and `imgincl_sst_include = 1`, from the `abcd_imgincl01.txt` instrument). In addition, we removed participants with more than 20% censored volumes (e.g., for EN-back, calculated as `tfmri_nback_ab_subthnvols/ tfmri_nback_all_beta_nvols` from the `nback_bwroi02.txt` instrument).

Multiple regions of interest were used as dependent variables in our empirical analysis. Although several regions of the prefrontal cortex are activated during salience processing, the EN-back does not include an explicit regulatory component; previous research suggests a rostral/ventral functional distinction in the mPFC wherein rostral regions underlie the cognitive rather than regulatory components of salience and emotion processing (Etkin et al., 2011). Thus, for the EN-back task, we examined the rostral anterior cingulate (rACC) and medial orbitofrontal cortex (mOFC), and amygdala reactivity to negative facial expressions versus neutral faces. We also examined hippocampal activation during the 2-back versus 0-back contrast of the EN-back task to capture short-term memory processes (Stark and Okado, 2003). For the SST, we examined rACC and OFC activation during the correct stop > correct go contrast, regions activated during successful inhibition during the SST (Casey et al., 2018).

3.4. Statistical Analysis

All analyses and graphics were conducted using the *survey* (Lumley, 2010), *lme4* (Bates et al., 2015), *ggplot2* (Wickham, 2016), and base packages in R Statistical Software version 3.6.3 (R Core Team, 2020). Reproducible code is publicly available at (<https://osf.io/d94kq/>). Outliers were winsorized to ± 3 SD from the mean. All dependent

variables were normally distributed (Supplemental Fig. 1).

First, to evaluate the generalizability of the recruited ABCD Study® sample to the target population, we compared unadjusted sample demographic proportions to design-based estimates of population demographic proportions. Unadjusted proportions ignore the complex design features of the study (i.e., study sites as PSUs) and do not adjust for the survey weights. Design-based estimates of the population proportions account for the PSUs and survey weights, implementing Taylor Series Linearization (Binder, 1983) for variance estimation. We highlight demographic groups where the sample and population proportions are significantly different, indicated by non-overlapping 95% confidence intervals. Significant differences in sample versus population demographic proportions would implicate sampling biases due to coverage error, sampling error, and/or selection bias (Table 1).

Second, to examine the effects of selection bias on in the analytic sample versus recruited sample, we conducted missing data analyses comparing participants with and without imaging data (sMRI, task-based EN-back and SST) on demographic, social-contextual, and study-level constructs. Welch’s two-sample t-tests were implemented for continuous variables, and chi-square difference tests were implemented for categorical variables. As most statistical tests will be significant in large sample sizes such as the ABCD Study® (Dick et al., 2021), we used the *effectsize* package (Ben-Schacar et al., 2021) to create measures of effect size for continuous (Cohen’s $d = \frac{M2 - M1}{SD_{pooled}}$) and categorical (Cramers $\Phi = \sqrt{\frac{\chi^2}{(N)(df_{smaller})}}$) variables. Cohen’s conventions were used to interpret the size of the effect for Cramer’s ϕ and Cohen’s d (Aron et al., 2013; Cohen, 1988), where effect size judgements are dependent on degrees of freedom for Cramer’s ϕ (Fig. 3).

Third, multivariable regression models were used to evaluate the associations between three measures of SER (i.e., caregiver education, household income, and neighborhood SER) and brain structure and function. We accounted for several confounding variables: youth sex, pubertal development, youth race-ethnicity, primary caregiver

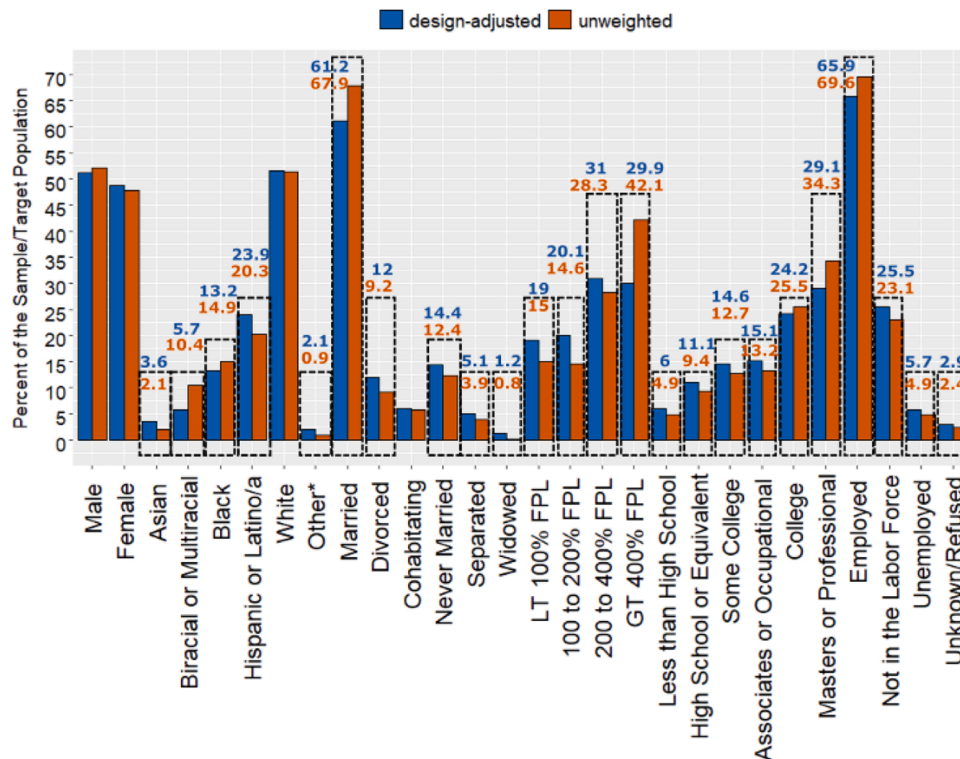


Fig. 2. Sample and Target Population Proportions of Sociodemographic Constructs in the ABCD Study® Note. $N = 10,846$ – $11,860$. Fully-design-adjusted models cluster standard errors at site and apply population weights derived from the American Community Survey. Dashed boxes indicate estimates with non-overlapping confidence intervals. Measure information can be found in the Methods section and Supplemental Materials.

Demographic- and Individual-Level Constructs										
	Sex	Race-Ethnicity	Caregiver Marital Status	Household Income	Caregiver Education	Caregiver Employment	Pubertal Development	# TBI Incidents	Youth EXT Behaviors	Youth INT Behaviors
sMRI Excluded (5%)	ns	Black	Never Married Cohabiting	< 100% FPL	< HS	ns	ns	Fewer TBI	ns	ns
EN-back Excluded (47%)	Male	Black	Never Married	< 100% FPL	High School or Less	Unemployed	Advanced Pubertal Development	Fewer TBI	Greater EXT Behaviors	Greater INT Behaviors

Social-Contextual Constructs						Visit Characteristics		
	Lead Exposure Risk	Neighborhood Safety	School Quality	Family Conflict	Neighborhood Disadvantage	# of Scanning Sessions	# Runs	Survey Weight
sMRI Excluded (5%)	Greater Lead Exposure Risk	Lower Neighborhood Safety	Lower School Quality	ns	Greater Neighborhood Disadvantage	More Scanning Sessions	not applicable	Larger Weight
EN-back Excluded (47%)	Greater Lead Exposure Risk	Lower Neighborhood Safety	Lower School Quality	Greater Family Conflict	Greater Neighborhood Disadvantage	ns	Fewer Runs	Larger Weight

Effect Size Key	Cohen's d	Cramer's ϕ (df = 1)	Cramer's ϕ (df = 3)	Cramer's ϕ (df = 5)
	Not significant (ns)			
Small	< 0.20	< 0.10	< 0.06	< 0.04
Medium	0.50	0.30	0.17	0.13
Large	> 0.80	> 0.50	> 0.29	> 0.22

Fig. 3. Missing Data Patterns in ABCD imaging data. Note. $N = 11,162$ – $11,860$. Missing data analyses compared youth with and without valid imaging data on the measures listed above. For continuous measures (e.g., pubertal development), independent samples t-tests were implemented and Cohen's d was calculated as a measure of effect size. For categorical measures (e.g., caregiver education), χ^2 difference tests were implemented and Phi ϕ was calculated as a measure of effect size. See [Supplemental Table 1](#) for reasons for data loss. All results, including for data loss on the Stop Signal Task, can be found in [Supplemental Tables 2 and 3](#).

employment status, and scanner platform (1 = GE [$n = 2962$], 2 = Phillips [$n = 1512$], 3 = Siemens [$n = 7278$]). Youth race-ethnicity is a social, not a biological, construct, which we adjust for because it is tightly linked to SER in the U.S. due to historical and current structural inequalities (McLoyd, 1998; Neblett Jr, 2019; Wilson, 2012) and, like the other control variables, is also related to the missing data. Structural MRI analyses additionally adjusted for total intracranial volume (subcortical regions only) and T1-weighted gray matter intensity (Pagliaccio et al., 2019).

All multivariable models are presented in three forms: (1) model-based OLS linear regression, (2) model-based MLM, and (3) design-based linear regression. OLS models ignore the nested structure of the ABCD datasets and, instead of accounting for survey weights, include variables that were used to construct the survey weights as covariates. MLM models were estimated using maximum likelihood estimation in the *lme4* package (Bates et al., 2015) to cluster SEs by study site; as in the model-based OLS models, variables used in the survey weighting procedure were included as covariates. Again, multi-level weights (i.e., which are not currently available in ABCD Study®) would be needed to appropriately weight MLM models (Pfeffermann et al., 1998). The design-based models account for complex sample design features (i.e., study sites as PSUs) and survey weights by implementing weighted least squares for point estimation and Taylor Series Linearization (Binder, 1983) for variance estimation, using the *survey* package (Lumley, 2010).

Subpopulation (i.e., youth with usable imaging data) analyses were implemented in the design-based framework to ensure that all survey design elements were used in variance estimation (West et al., 2008). To account for cases in which multiple children per family contributed usable imaging data, we used a genetically-independent sample (i.e., by selecting one random child per family) in all multivariable analyses (Supplemental Table 1), consistent with previous studies (e.g., Li et al.,

2021). Missing data was handled through listwise deletion.

Of the original ABCD Study® sample of 11,878, our maximum analytic sample size was 11,860.³ Five participants could not be assigned a PSU (critical to design-based analyses). In addition, three participants noted their sex as “other”, and 10 participants were missing data on sex. As point estimation for a group of $n = 3$ (i.e., for “sex = other”) would be unreliable, we opted to remove these participants. Thus, fifteen participants were removed from all analyses. Owing to missing imaging data and a small proportion of contextual and behavioral data, the analytic sample sizes were $N = 10,846$ – $11,860$ for sample demographic proportions and missing data analyses, and $N = 5215$ to 8107 for multivariable models investigating the associations between socioeconomic resources and structural and functional imaging metrics.

4. Results

4.1. How does the target population of 9–10-year-olds in the U.S. differ from the ABCD Study® respondent sample?

Coverage, sampling, and non-response errors contributed to demographic differences between ABCD participants and the target population of 9–10-year-old children living in the U.S. (Fig. 1). Fig. 2 presents the unweighted sample proportions and design-based population proportions (with adjustment for the complex sampling design and survey weights based on the American Community Survey) for several observed demographic characteristics. Most notably, the ABCD Study® sample, when unweighted, is over-represented by children of married caregivers (ABCD-67.9% versus ABCD-weighted-61.2%), high-income households earning > 400% of the federal poverty line (ABCD-42.1% versus ABCD-weighted -29.9%), caregivers with a Masters, PhD, or other professional (e.g., JD, MD) degree (34.3% versus 29.1%), Biracial

or Multiracial children (10.4% versus 5.7%), and Black children (14.9% versus 13.2%). By contrast, the ABCD Study® sample is under-represented by children from low-income households earning < 100% of the FPL (15% versus 19%) or 100–200% of the FPL (14.6% versus 20.1%), Hispanic or Latino/a children (20.3% versus 23.9%), and children of divorced (9.2% versus 12%) or never married (12.4% versus 14.4%) caregivers. The implication of these sample – target population demographic differences are that the survey weights for children from lower income families, for example, will be larger and more variable than the weights for children from higher income families (Heeringa et al., 2017; Spencer, 2000).

4.2. How does the analytic sample in sMRI and fMRI analyses differ from the ABCD Study® respondent sample?

4.2.1. Selection Biases

Demographic discrepancies between the sample and the target population are expected in survey research; these discrepancies can be adjusted for through the application of survey weights. However, this approach assumes among otherwise complete cases that item-level missingness for variables of interest is uncorrelated with the case-level survey weights that adjust for inclusion probabilities and survey unit nonresponse (Groves et al., 2009; Kish and Frankel, 1974). Unfortunately, neuroimaging research is notorious for extensive data loss (e.g., discomfort with the scan may result in participants ending the session early or not participating at all; image sensitivity to movement may result in image artifacts and the removal of “high motion” participants). Supplemental Table 1 presents the flow of participants from initial recruitment through data analysis for each of several neuroimaging modalities. Most ABCD Study® participants attempted the structural MRI (sMRI) scan (97%), with an additional 7.8% lost during quality control. By comparison, more significant data loss was observed in the fMRI EN-back and SST tasks: 19–25% of the original participants had no fMRI image series, and an additional 24 – 29% were removed during quality control, leaving 61–53% of the original participants with usable task-based fMRI data. Understanding how this missing data is correlated with demographic, social-contextual, and study-level characteristics is key to understanding how individuals with usable imaging data differ from the overall sample and, thus, the target population.

4.2.2. Missing data patterns

We next evaluated how participants with and without usable imaging data differed on demographic (i.e., child sex and race-ethnicity, caregiver marital status and education, household income, number of traumatic brain incidents, youth internalizing and externalizing behaviors), social-contextual (i.e., lead exposure risk, neighborhood disadvantage, neighborhood safety, school quality, family conflict), and visit (i.e., number of scanning sessions and/or task runs, survey weight) characteristics (Fig. 3). Summary results from the sMRI and fMRI EN-back analyses are presented in the main text; results are available in table format for all analyses in Supplemental Tables 2 and 3. Participants without usable sMRI and fMRI EN-back data were more likely to identify as Black or African American, live in low-income households, have caregivers with a high school degree or lower, have unmarried caregivers, live in a Census tract marked by greater lead exposure risk and neighborhood disadvantage, and report lower neighborhood safety and school quality (Fig. 3). Group differences in having usable task-based fMRI data were generally larger than group differences in having valid sMRI data, likely owing to greater fMRI data loss (Supplemental Table 1). Youth without usable fMRI EN-back data also had greater caregiver-reported internalizing and externalizing behaviors, were more pubertally-advanced, more likely to be male, and were more likely to have a primary caregiver who was unemployed (Fig. 3). For fMRI data loss, the largest observed effect sizes were for youth race-ethnicity, caregiver marital status and education, household income, youth externalizing behaviors, lead exposure risk, and neighborhood

disadvantage (Fig. 3). Lastly, participants without usable sMRI data had more scanning sessions (i.e., their MRI data was collected over two versus one scanning session), whereas participants without valid fMRI EN-back data had fewer task runs (both medium-sized effects). Collectively, our missing data analyses revealed that socioeconomically-disadvantaged youth were less likely to have usable structural and functional imaging data than their peers.

Given the relatively extensive fMRI EN-back and SST data loss, we also examined how youth with and without usable fMRI data differed on youth emotion ratings from pre- and post-scan questionnaires (Supplemental Table 4). That is, might how children feel before and after the scan help to explain why some youth provided usable data or not? Before and after the MRI protocol, youth rated the degree to which they felt different emotions (e.g., relaxed, upset, happy) on a 5-point Likert scale (1 = not at all; 5 = very much). Here, we highlight significant group differences that were at least of small effect (Cohen's $d > 0.20$, for continuous variables). After the MRI scanning session, youth without valid EN-back or SST data reported feeling more scared, upset, and sad ($0.21 < d < 0.30$). There were no meaningful (i.e., $d > 0.20$) group differences in pre-scan ratings (Supplemental Table 4).

Our last attempt to understand missing fMRI data patterns focused on associations with participant-reported study features within the ABCD-SD subsample ($N = 989$). Using Welch's two-sample t-tests, we examined whether caregivers' perceptions of the study protocol were linked to whether their child(ren) provided usable imaging data at baseline (Supplemental Table 5). Among families who participated in ABCD-SD on a different day as the baseline ABCD visit, caregivers of youth without imaging data were more likely to rate the consent form as unclear ($d = 0.25$), that they felt uncomfortable with the staff ($d = 0.24$), and that the questions asked of them were uncomfortable ($d = 0.25$).

Results from these missing data analyses suggest that youth with and without valid imaging data differ on a variety of demographic, social-contextual, affective, and feedback-based measures. Therefore, it is important to ask whether systematic missingness impacts our ability to make population-level inferences. A straightforward way to evaluate this question is to examine whether missingness is associated with the survey weight. In the baseline ABCD Study® imaging data, youth without usable neuroimaging data had larger population weights than youth with valid neuroimaging data; this pattern was observed for sMRI data ($t[711.26] = 2.42$, $p < 0.05$), fMRI EN-back data ($t[10811.99] = 4.90$, $p < 0.001$), and fMRI SST data ($t[9588.08] = 4.11$, $p < 0.001$). Despite these significant group differences, however, the effects were small in magnitude ($d = 0.10$), suggesting that, *at least in the sample identified here*, missing data approaches (e.g., imputation) may not be necessary to generate population-based estimates.

4.3. Comparison of analytic approaches

In addition to instances where missing data is associated with survey weights, a researcher's analytic approach may impact inferences to a target population. To compare analytic approaches, we examined a research question of great interest to developmental neuroscientists: how are socioeconomic resources (SER) associated with youth brain structure and function?

4.3.1. sMRI

Table 2 displays parameter estimates for household income (3 dummy codes: > 400% of the federal poverty line as the reference group), caregiver education (5 dummy codes: graduate or professional degree as the reference group), and neighborhood disadvantage (continuous measure; Taylor et al., 2020) in models predicting bilateral hippocampal volume, total cortical volume, and total cortical surface area. Across modeling frameworks, household family income and caregiver education were stronger predictors of sMRI metrics than neighborhood disadvantage. As expected, the standard errors of each parameter were generally *larger* in design-based analyses than in

Table 2

Comparison of parameter estimates in design-based and model-based analyses of SES – sMRI associations.

	Model-Based OLS		Model-Based MLM		Design-Based	
	B	SE	B	SE	B	SE
Hippocampal volume						
< 100% FPL	-29.98*	13.81	-30.23*	13.90	-36.26	12.56
100 – 200% FPL	-24.00*	12.19	-24.85*	12.27	-32.97	12.09
200 – 400% FPL	-9.59	8.74	-9.72*	8.79	-8.08	10.46
Less than High School	-36.14	21.29	-40.14	21.39	-33.63	14.79
High School or Equivalent	-26.85	15.49	-29.44	15.52	-30.51	12.98
Some College	-3.74	12.70	-4.72	12.72	-0.009	10.70
Associates or Occupational	-8.02	12.18	-9.27	12.20	-9.90	11.26
College	-16.43	8.87	-16.62	8.89	-19.83	7.48
Neighborhood disadvantage			Removed from model; did not degrade model fit			
Total cortical volume						
< 100% FPL	-11789.17***	2177.38	-12623.06***	2241.72	-9269.80	3751.60
100 – 200% FPL	-5531.05**	1892.16	-5989.68**	1937.08	-4767.90	2846.20
200 – 400% FPL	-2899.23*	1352.79	-3338.09*	1373.38	-3606.30	2265.10
Less than High School	-11606.03***	3201.11	-11192.27***	3368.60	-13738.30	4136.70
High School or Equivalent	-13920.08***	2360.64	-12909.50***	2423.41	-14183.80*	2325.90
Some College	-7835.98***	1949.11	-7197.18***	1976.02	-8475.30	2759.60
Associates or Occupational	-10560.95***	1860.56	-9715.94***	1886.20	-10366.90*	1604.20
College	-5490.06***	2360.64	-5463.23***	1369.46	-5274.20	1335.00
Neighborhood disadvantage	-275.36**	93.45	-245.43**	97.90	-335.10	122.30
Total cortical surface area						
< 100% FPL	-3379.73***	695.83	-3390.08***	699.67	-2705.88	1006.10
100 – 200% FPL	-1707.19**	601.28	-1692.14**	604.58	-1441.71	888.38
200 – 400% FPL	-1021.19*	426.75	-1009.19*	428.67	-1215.95	696.60
Less than High School	-3906.24***	1048.51	-3353.30**	1051.45	-4259.32	1481.15
High School or Equivalent	-3834.17***	758.68	-3439.68***	756.57	-3852.77	678.07
Some College	-2337.87***	619.10	-2164.10***	616.92	-2540.75	736.86
Associates or Occupational	-2919.19***	590.11	-2664.99***	588.84	-2929.89	477.92
College	-1342.37**	428.31	-1278.93**	427.52	-1330.64	374.06
Neighborhood disadvantage	-74.16*	29.82	-74.21**	30.52	-89.97	34.10

Note. N = 7,698 – 8,107. Asterisks denote the statistical significance of a given parameter estimate against the null hypothesis (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). All models controlled for child sex, race-ethnicity, scanner type, and T1-weighted gray matter intensity. Hippocampal volume models also controlled for total intracranial volume. Covariates that were not significantly associated with the outcome were probed for removal by comparing model fit of nested models (with or without the covariate) using one-way ANOVAs. In all models, the removal of primary caregiver employment status did not degrade model fit ($ps > 0.10$). Pubertal development was similarly removed as a non-significant predictor in models predicting total cortical volume, but was included in models predicting hippocampal volume and total cortical surface area. Reference categories were specified as the most common group: >400% FPL (household income), graduate or professional degree (caregiver education). Estimated random effects of the model-based multilevel models: Hippocampal volume: between-site variance (SD) = 221.80 (14.89), residual variance (SD) = 91595.40 (302.65); Total cortical volume: between-site variance (SD) = 7561000.00 (8695.00), residual variance (SD) = 2063000000.00 (45425.00); Total cortical surface area: between-site variance (SD) = 5012471.00 (2239.00), residual variance (SD) = 201180284.00 (14184.00).

model-based analyses (i.e., either in OLS or MLM models). Exceptions were in models predicting hippocampal volume, where some standard error estimates were smaller in design-based analyses; this may indicate that the post-stratification adjustments reduced the sampling variance of these particular estimates.

Notably, model-based analytic approaches (i.e., OLS or MLM) overestimated the coefficients for household income and under-estimated the coefficients representing the associations of caregiver education on total cortical volume and total cortical surface area, compared to the point estimates in design-based analyses (Table 2). In models predicting hippocampal volume, the point estimates for household income dummy variables were larger in design-based analyses, whereas model-based analyses (i.e., OLS, MLM) overestimated the coefficients of caregiver education (i.e., Less than high school, compared to Graduate or professional degree). The implication of these results is that the decision to leverage model-based versus design-based approaches might change a researcher's interpretation of which component of SER more strongly relates to metrics of structural brain development.

Collectively, these results highlight discrepancies in the magnitude of the associations between socioeconomic resources and sMRI metrics across modeling frameworks. That parameter estimates were similar across model-based approaches – unweighted OLS models (which ignore the sampling design and do not apply sampling weights) and multi-level models (which cluster standard errors by study site and similarly do not

apply sampling weights) – suggests that model-based multi-level models may not sufficiently account for the sampling design and selection biases through clustering standard errors by site. MLMs also do not account for survey weights.

4.3.2. Task-based fMRI

Associations between SER and neural activation during the EN-back and SST tasks were much weaker and largely non-significant (Supplemental Tables 7 and 8). Model-based analyses (i.e., OLS and MLM) revealed significant associations between having a caregiver with less than a high school degree (compared to having a caregiver with a graduate or professional degree) and greater rACC and mOFC reactivity to negative versus neutral facial expressions in the EN-back task, and less rACC reactivity to Correct Stops versus Correct Gos in the SST task. The point estimates were similar or identical in magnitude in model-based MLMs and model-based OLS models. Compared to model-based analyses, point estimates in design-based models were sometimes larger (e.g., associations between caregiver education and amygdala to negative versus neutral facial expressions in the EN-back task) and sometimes similar in magnitude (e.g., associations between household income-to-needs and IOFC activation in the 2-back versus 0-back EN-back condition). Thus, at least for the fMRI measures examined using publicly-available tabulated ABCD Study® imaging data, there were inconsistencies in whether implementing model-based (without survey

weights; clustering SEs by site in the case of MLM) versus design-based analyses (specifying site as PSUs and adjusting for survey weights) would alter the interpretation of study findings.

5. Discussion

Leveraging large, population-based datasets to study the developing brain raises several theoretical and analytical challenges for developmental neuroscientists. In the current paper, we review distinctions between probability and non-probability sampling and several forms of selection bias that may influence population generalizability. Further, we examine the influence of different analytic approaches (i.e., design-based versus more commonly implemented model-based MLM or model-based OLS models) in evaluating the associations between socioeconomic resources (SER) and brain structure and function in the ABCD Study®, a clustered probability sample of 11,878 9–10-year-olds in the United States. Our empirical analyses show that:

1. There is sampling bias in the recruited ABCD Study® sample versus the target population of 9–10-year-olds in the United States, particularly with respect to caregiver education and household income;
2. Missing data (i.e., youth with usable versus not usable sMRI and task-based fMRI data) is meaningfully associated with demographic and social-contextual constructs such that marginalized youth are under-represented in analytic samples;
3. Although missing data in the analytic sample identified here was related to the ABCD Study® survey weights, the effect sizes were relatively small, suggesting that listwise deletion may not undermine population generalizability (again, at least in the analytic sample identified in this empirical demonstration);
4. Design-based and model-based (OLS, MLM) analytic approaches differed in coefficients of SER-associations with brain structure and function, but it depended on the variable of interest and the outcome examined.

5.1. Population Representation and Systematic Missing Data

The ABCD Study® sample proportions by race-ethnicity, marital status, household income, education, and employment were significantly different from the proportions in the target population of 9–10-year-olds in the U.S. For example, ABCD Study® respondents in the baseline survey were over-represented by youth from high-income, highly-educated (i.e., masters or professional degree), and/or married households. There was also a larger proportion of Black and Biracial or Multiracial youth, owing to purposeful oversampling in the original sampling design (Garavan et al., 2018), but fewer Hispanic and Asian youth. Sociodemographic discrepancies between respondents and the target population reflect error (in some cases, purposeful error, due to oversampling) at multiple stages of study design and recruitment (Table 1): between the sampling frame and the target population (i.e., coverage error), between the sampling frame and the selected sample (i.e., sampling error), and between the sample and the respondents (i.e., non-response error). Because the ABCD Study® implemented probability sampling, however, the complex sampling design (i.e., with study sites as PSUs and constructed survey weights) can be leveraged to recapitulate the sociodemographic characteristics of the target population. In non-probability convenience samples, by contrast, coverage error, sampling error, and non-response error are unknown, making it difficult or impossible to account for differences between the observed sample and the population to which a researcher is trying to generalize. Adoption of a complex probability sampling design stands as one of the greatest strengths of the ABCD Study®.

Discrepancies between the target population and the respondent sample, however, may be magnified in the context of missing data. Missing data analyses revealed that ABCD Study® youth respondents

without usable imaging data (both structural and task-based functional MRI) at baseline were more likely to identify as Black or African American, live in low-income households, have caregivers with less than a high school degree, live in more socioeconomically-disadvantaged neighborhoods, and rate lower neighborhood safety and school quality. Thus, there was systematic missingness in the sample of usable imaging data, suggesting that parameter estimates using listwise deletion may be biased without inclusion of these specific constructs as covariates (Schafer and Graham, 2002). In the context of a complex sampling design such as that employed by the ABCD Study®, however, missing data presents additional threats to generalizability. Chiefly, if the survey weights themselves are associated with missingness, parameter estimates are likely to be biased and additional missing data solutions are needed (Groves et al., 2009; Kish and Frankel, 1974).

Our analyses revealed that youth without usable imaging data had larger survey weights than youth with usable imaging data. That is, youth who were underrepresented in the ABCD Study® baseline sample versus the target population (and, thus, were assigned larger population weights) were also more likely to be underrepresented in the analytic sample with usable imaging data. The magnitude of these effects was relatively small (Cohen's $d < 0.20$), suggesting that population generalizability remained intact using listwise deletion, at least in the analytic sample identified here. It may appear counterintuitive that some variables (e.g., household income) were meaningfully associated with missingness while the survey weight was not (Fig. 3). However, because the survey weight reflects an individual's combined propensity of belonging to the ABCD Study® sample versus the target population across multiple sociodemographic variables, the magnitude of systematic missingness by survey weight may differ from that of individual variables. Given that researchers examine constructs with varying levels of missingness, use different quality control criteria for neuroimaging data inclusion, and the likelihood of missing data varies by imaging modality (e.g., resting-state fMRI versus sMRI), we emphasize that researchers must empirically test whether both the survey weight and individual sociodemographic variables are associated with missingness in service to evaluating threats to population generalizability, rather than assume listwise deletion.

5.2. Analytic approach may impact study interpretation

Our last aim was to evaluate the influence of design-based and model-based analytic approaches on the associations between socioeconomic resources and multiple measures of brain structure and function in the ABCD Study® baseline sample. Compared to design-based analyses (i.e., which cluster SEs around the PSUs/study sites, and apply survey weights), model-based analyses (i.e., either using OLS models or MLM [cluster SEs by study site]; both include controls for constructs included in the survey weights but do not adjust for the survey weights themselves) under-estimated the coefficients of caregiver education and over-estimated the coefficients of household income in models predicting total cortical volume and total cortical surface area. That is, although the directions of the estimated associations were consistent, the interpretation of which measure of SER exerted the largest effect differed across analytic frameworks. This could pose a challenge for researchers and policy-makers relying on large datasets like the ABCD Study® to weigh investments in varied social interventions (e.g., unconditional cash transfers versus educational and occupational training programs). Less consistency was found in the comparison of analytic approaches in models linking socioeconomic resources to task-based fMRI.

The multi-level modeling strategy is consistent with the dominant analytic approach among recent ABCD Study® data users (e.g., Paul et al., 2021; Rakesh et al., 2021). Yet, as demonstrated here in the example of SER-associations with sMRI metrics, multi-level modeling without multi-level weights was not sufficient to recapitulate the unbiased target population parameters observed in design-based analyses.

This could be due to inaccurate model assumptions in the MLM framework, the inability to include informative survey weights, or both. As previously outlined in the ABCD Study® sampling guide (Heeringa and Berglund, 2019), the publicly-available survey weights cannot be used in the multi-level modeling framework. To “weight” a multi-level model, weights are required at each level of the hierarchical data structure (Pfeffermann et al., 1998; Rabe-Hesketh and Skrondal, 2006). The ABCD Study® survey weights currently available reflect the aggregated weight (i.e., both the individual level-1 and the site level-2 weight are combined). Early comparisons of the hybrid/model-based multi-level analyses with design-based analyses revealed few differences (Heeringa and Berglund, 2019), but these investigations were based on behavioral data with few missing observations. As the ABCD Study® sites were not chosen at random, the construction of 2-level survey weights requires additional assumptions that need to be evaluated before releasing disaggregated weights to the public. Moreover, releasing disaggregated multi-level survey weights presents additional concerns for participant identification. With these considerations, we outline recommendations for ABCD Study® users below.

5.3. Recommendations

Based on these findings and the critical connection between sampling, analytic approach, and generalizability, we have several recommendations for analyzing data and interpreting research findings for researchers using the ABCD Study® and other population-based datasets

Table 3
Current and Future Directions.

Increasing Generalizability in Non-Probability Samples
Inclusive Recruitment and Retention Practices. Leverage long-standing strategies in developmental science used to increase the representation and voice of marginalized communities in research. Examples include: (1) building collaborations with and hiring staff who are culturally engaged and embedded within target communities, (2) ensuring continuity in staff over time, (3) communicating researcher motivations and demonstrating shared goals that address community needs, (4) including community members in study design, recruitment, and retention, and (5) implementing face-to-face recruiting strategies.
Recommended Reading: Yancey et al., (2006). <i>Effective recruitment and retention of minority research participants</i> . Medin et al., (2017). <i>Systems of (non-)diversity</i> . Habibi et al., (2015). <i>Developmental brain research with participants from underprivileged communities: Strategies for recruitment, participation, and retention</i> . Rowley & Camacho, (2015). <i>Increasing diversity in cognitive developmental research: Issues and solutions</i> .
Weighting Non-Probability Samples. Recent developments in survey methodology suggest it may be possible to calibrate a non-probability sample to a target population through the construction of survey weights using a “reference” probability sample (e.g., LeWinn et al., 2017).
Recommended Reading: Elliott & Valliant, (2017). <i>Comparing alternatives for estimation from nonprobability samples</i> . Zhang, (2019). <i>On valid descriptive inference from non-probability sample</i> . Yang et al., (2020). <i>Doubly robust inference when combining probability and non-probability samples with high dimensional data</i> . Rueda et al., (2020). <i>The R package NonProbEst for estimation in non-probability surveys</i> .
Addressing Longitudinal Retention
Construction of Longitudinal Survey Weights. Attrition from the baseline cohort is a potential source of added selection bias in longitudinal data analysis for population neuroscience studies. Statistical adjustment for complete “loss to follow-up” of baseline participants is typically performed by applying a further attrition weighting adjustment to the baseline weight – or otherwise recalibrating the retained sample to baseline sample characteristics (endogenous) or to external population benchmarks (exogenous). As an alternative to longitudinal weighting, various methods can be employed to impute the longitudinal missing data.
Recommended Reading: Kalton & Flores-Cervantes, (2003). <i>Weighting methods</i> . Schmidt & Woll, (2017). <i>Longitudinal dropout and weighting against its bias</i> .

with complex sampling designs (Table 3).

5.3.1. Conduct missing data analyses

First, researchers must conduct and report missing data analyses by comparing their analytic sample to the original sample of 11,878 ABCD Study® respondents (Table 3). Due to variation in preprocessing pipelines and behavioral data exclusions (Botvinik-Nezer et al., 2020), missing data patterns may differ from what we have presented here. Complete and incomplete data should be examined with respect to sociodemographic constructs (e.g., child sex, race-ethnicity, socioeconomic resources, behavioral and/or cognitive outcomes), the survey weight, and other variables of interest. Group differences should be interpreted using effect size metrics rather than *p*-values, to avoid overestimating “meaningful” effects (Dick et al., 2021; Owens et al., 2021). If survey weights are meaningfully associated with missing data, researchers must be explicit in the Methods and Discussion that this impacts the population generalizability of their results. In this case, researchers can either (1) proceed with their analyses and include covariates predictive of missingness but interpret results only with respect to the sociodemographic characteristics of analytic sample (i.e., be constrained and acknowledge that results may not generalize beyond the analytic sample), or (2) address missing data (e.g., imputation, full maximum-likelihood estimation) before proceeding with weighted analyses that can be generalized to the target population. If the survey weights are not associated with missing data, analyses can proceed as planned with less concern for limitations on target population generalizability (though, again, this may differ from analysis to analysis depending on the degree and pattern of missing data).

5.3.2. Implement design-based analyses for descriptive inference and compare parameter estimates from design-based and model-based approaches in multivariable analyses

For users interested in reporting population means and/or proportions (e.g., incidence of traumatic brain injuries in 9–10-year-olds in the United States), only design-based analyses should be implemented. This includes any research question designed to report the “population prevalence” of a given construct in the ABCD Study® sample (e.g., incidental MRI findings: Li et al., 2021; substance use patterns: Lisdahl et al., 2021). For multivariable analyses, users are suggested to compare results in design-based models and model-based frameworks that adjust SEs around study sites (Heeringa et al., 2017). Model-based analyses that do not cluster SEs by study site (e.g., as in the OLS approach evaluated here) should rarely if ever be implemented, as these approaches ignore statistical dependence among observations. Comparisons between model- and design-based analyses should focus on point estimates rather than statistical significance because design-based analyses generally increase standard errors (as would be expected when accounting for complex sampling designs and the sampling variance that they introduce).

In cases where point estimates from design-based and model-based approaches differ, users should adopt the design-based results. One of the many advantages of using the ABCD Study® and other large population-based datasets is its ability to generalize to a broader population, with implications for policy and intervention. By relying on the random probability sampling design for population inference, design-based analyses will always produce point estimates that are unbiased with respect to the sampling design. Should point estimates from design-based and model-based approaches converge, users can report the model-based results with greater confidence in the generalizability of their point estimates.

5.3.3. Adopt caution in interpretation of population generalizability

“Representative” is a strong adjective to apply to any dataset (Compton et al., 2019; Heeringa and Berglund, 2019). In addition to biases in recruitment and sampling (Table 1), the descriptor “representative” will vary by variable (e.g., how much missing data is there),

by subpopulation (e.g., is this subpopulation accurately reflected in the usable data), and by the extent to which the weighting methodology or model covariates capture factors that truly impact the outcome of interest (both in terms of the variables and their functional form with the outcome). All approaches to statistical estimation and inference make assumptions. No study gets an uncontested stamp of approval on the unbiasedness of its survey estimates. Design-based and model-based analyses may produce remarkably similar interpretations of study findings, and some researchers may choose to implement model-based analyses for particular reasons. The position that we take here is that researchers must empirically evaluate the degree of population inference that is possible given their analytic sample, research question, and statistical methodology.

Thoughtful attention to probability sampling in the ABCD Study® is a dramatic improvement to the population generalizability and socio-demographic representation of more traditional convenience sampling approaches (Nielsen et al., 2017). However, readers should note that there were necessary biases in the ABCD Study® sampling frame that must be considered in describing the sample: the study did not include children who were home-schooled, children attending school districts outside of one of the 21 catchments sites, or children with certain exclusion criteria (e.g., children born premature or of extreme low birth weight, children with medical or psychiatric conditions that would affect their ability to complete the assessments, children who already met criteria for an alcohol or substance use disorder). Thus, researchers should refrain from implying study findings from using ABCD Study® data (regardless of whether the analytic approach is design-based or model-based) are “representative” to “all children”; rather, we encourage ABCD Study® users to be specific about who the results are designed to generalize to (e.g., 9- and 10-year-olds in the United States who meet certain eligibility criteria).

6. Future directions and conclusions

The advent of Population Neuroscience (Falk et al., 2013; Paus, 2010) and the inclusion of probability sampling in large-scale neuroimaging studies (Garavan et al., 2018; Tomlinson et al., 2020; White et al., 2013) holds great promise for increasing sociodemographic representation in and population generalizability of human neuroscience research. We conclude by highlighting several ongoing areas of research that complement our discussion of population-based neuroscience (Table 3). First, we highlight existing recruitment, retention, and analytic practices shown to increase sociodemographic diversity even in the context of convenience sampling designs. This includes the long history of inclusive recruitment and retention practices in developmental science that focuses on elevating the experiences of racial-ethnic marginalized communities in research. We also point readers to a growing literature evaluating the performance of survey weighting in non-probability samples. Second, an active area of inquiry in the ABCD Study® is the construction of longitudinal survey weights. Future data releases will undoubtedly include weights that account for sample retention and drop-out, information necessary for maintaining the population-level generalizability as the study sample is tracked over the next decade. Lastly, there is a growing discussion around what constitutes a “meaningful” effect size (Dick et al., 2021; Owens et al., 2021). Although we use this term and reference existing benchmarks in the field (e.g., Cohen’s *d*), more research is needed, particularly in the context of large population-based studies.

Population-based probability sampling in the ABCD Study® represents a major innovation in Population Neuroscience (Paus, 2010) by bringing together survey methodologists and developmental neuroscientists to increase the generalizability of human neuroscience research (Falk et al., 2013). Inclusion of children and families historically under-represented in human neuroimaging research is a major step forward and away from historical marginalization in brain and biomedical research (Dotson and Duarte, 2020; Qu et al., 2021). At the

same time, without thoughtful use of this open-access dataset, studies may not be leveraging its strengths, and could come to erroneous conclusions. In broadening beyond these issues, Simmons et al. (2021) highlighted two considerations for ABCD Study® data users: acknowledgment of the broader social context in which development occurs, with attention to oppression and inequality that intersects with identity (Neblett Jr, 2019); and the recognition that resilience and adaptation are common (Masten, 2001), and can be reflected at multiple levels of analysis (physiology, brain, behavior). We echo these considerations and add that sample composition and generalizability similarly factor into responsible use of open-access data from ABCD Study®.

CRedit authorship contribution statement

Arianna M. Gard: Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization. **Luke W. Hyde:** Conceptualization, Methodology, Visualization, Writing – review & editing. **Brady T. West:** Methodology, Writing – review & editing. **Steven G. Heeringa:** Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Colter Mitchell:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data and code is publicly-available and listed in the Data Statement and Manuscript.

Acknowledgements

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (<https://abcdstudy.org>), held in the NIMH Data Archive (NDA). This is a multi-site, 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. Additionally, we thank the families who participated in the ABCD Study and the research staff who coordinated data collection and made this open-access resource possible. The title for the current paper was inspired by Solon, Hader, and Wooldridge’s (2015) paper titled “What are we weighting for?”, published in the *Journal of Human Resources*.

Data Statement

Data used in this manuscript is publicly-available through the ABCD NIMH Data Archive collection. Behavioral data was downloaded from the second public release (version 2.0.1, released July 2019; <https://doi.org/10.1016/j.dcn.2019.101196>).

org/10.15154/1504041) and neuroimaging and additional behavioral data was downloaded from the third public release (version 3.0, released October 2020; <https://doi.org/10.15154/1519007>). All measures were downloaded directly from the ABCD NIMH Data Archive collection #2573 (DUA #3067).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.dcn.2023.101196](https://doi.org/10.1016/j.dcn.2023.101196).

References

- Andringa, S., Godfroid, A., 2020. Sampling bias and the problem of generalizability in applied linguistics. *Annu. Rev. Appl. Linguist.* 40, 134–142. <https://doi.org/10.1017/S0267190520000033>.
- AnonNIMH Data Archive. (2020). Adolescent Brain Cognitive Development Study (ABCD) – release 3.0 (11878). 10.15154/1520591.
- Aron, A., Coups, E.J., Aron, E.N., 2013. *Making Sense of Statistical Significance: Decision Errors, Effect Size, and Statistical Power*. Statistics for Psychology, 6th ed. Pearson Education, Inc, pp. 177–225.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67 (1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Ben-Schacar, M.S., Makowski, D., Ludecke, D., Patil, I., Kelley, K., & Stanley, D. (2021). effectsize: Indices of effect size and standardized parameters. (0.4.3) [Computer software]. <https://CRAN.R-project.org/package=effectsize>.
- Binder, D.A., 1983. On the variances of asymptotically normal estimators from complex surveys (JSTOR). *Int. Stat. Rev. / Rev. Int. De. Stat.* 51 (3), 279–292. <https://doi.org/10.2307/1402588>.
- Bollen, K.A., Biemer, P.P., Karr, A.F., Tueller, S., Berzofsky, M.E., 2016. Are survey weights needed? A review of diagnostic tests in regression analysis. *Annu. Rev. Stat. Its Appl.* 3 (1), 375–392. <https://doi.org/10.1146/annurev-statistics-011516-012958>.
- Botvinik-Nezer, R., Holzmeister, F., Camerer, C.F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J.A., Adcock, R.A., Avesani, P., Baczkowski, B.M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., Schonberg, T., 2020. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 582 (7810), 84–88. <https://doi.org/10.1038/s41586-020-2314-9>.
- Bradley, R.H., Corwyn, R.F., 2002. Socioeconomic status and child development. *Annu. Rev. Psychol.* 53 (1), 371–399. <https://doi.org/10.1146/annurev.psych.53.100901.135233>.
- Bradley, V.C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X.-L., Flaxman, S., 2021. Unrepresentative big surveys significantly overestimated US vaccine uptake. *Nature* 600 (7890), 695–700. <https://doi.org/10.1038/s41586-021-04198-4>.
- Brody, G.H., Gray, J.C., Yu, T., Barton, A.W., Beach, S.R.H., Galván, A., MacKillop, J., Windle, M., Chen, E., Miller, G.E., Sweet, L.H., 2017. Protective prevention effects on the association of poverty with brain development. *JAMA Pediatr.* 171 (1), 46–52. <https://doi.org/10.1001/jamapediatrics.2016.2988>.
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14 (5), 365–376. <https://doi.org/10.1038/nrn3475>.
- Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., Orr, C.A., Wager, T.D., Banich, M.T., Speer, N.K., Sutherland, M.T., Riedel, M.C., Dick, A.S., Bjork, J.M., Thomas, K.M., Dale, A.M., 2018. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* 32, 43–54. <https://doi.org/10.1016/j.dcn.2018.03.001>.
- Chaarani, B., Hahn, S., Allgaier, N., Adise, S., Owens, M.M., Julianio, A.C., Yuan, D.K., Loso, H., Ivanciu, A., Albaugh, M.D., Dumas, J., Mackey, S., Laurent, J., Ivanova, M., Hagler, D.J., Cornejo, M.D., Hatton, S., Agrawal, A., Aguinaldo, L., ABCD Consortium, 2021. Baseline brain function in the preadolescents of the ABCD Study. *Nat. Neurosci.* 24 (8), 1176–1186. <https://doi.org/10.1038/s41593-021-00867-9>.
- Cheung, K.L., Klooster, P.M. ten, Smit, C., Vries, H. de, Pieterse, M.E., 2017. The impact of non-response bias due to sampling in public health studies: a comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health* 17 (1), 1–10. <https://doi.org/10.1186/s12889-017-4189-8>.
- Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*, 2nd ed. Lawrence Erlbaum Associates.
- Compton, W.M., Dowling, G.J., Garavan, H., 2019. Ensuring the best use of data: the adolescent brain cognitive development study. *JAMA Pediatr.* 173 (9), 809–810. <https://doi.org/10.1001/jamapediatrics.2019.2081>.
- Conger, R.D., Conger, K.J., Martin, M.J., 2010. Socioeconomic status, family processes, and individual development. *J. Marriage Fam.* 72 (3), 685–704. <https://doi.org/10.1111/j.1741-3737.2010.00725.x>.
- R. Core Team. (2020). R: A language and environment for statistical computing (3.6.3) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Dick, A.S., Lopez, D.A., Watts, A.L., Heeringa, S., Reuter, C., Bartsch, H., Fan, C.C., Kennedy, D.N., Palmer, C., Marshall, A., Haist, F., Hawes, S., Nichols, T.E., Barch, D.M., Jernigan, T.L., Garavan, H., Grant, S., Pariyadath, V., Hoffman, E., Thompson, W.K., 2021. Meaningful associations in the adolescent brain cognitive development study. *NeuroImage* 239, 118262. <https://doi.org/10.1016/j.neuroimage.2021.118262>.
- Dotson, V.M., Duarte, A., 2020. The importance of diversity in cognitive neuroscience. *Ann. N. Y. Acad. Sci.* 1464 (1), 181–191. <https://doi.org/10.1111/nyas.14268>.
- Elliott, M.R., Valliant, R., 2017. Inference for nonprobability samples. *Stat. Sci.* 32 (2), 249–264. <https://doi.org/10.1214/16-STS598>.
- Etkin, A., Egner, T., Kalisch, R., 2011. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn. Sci.* 15 (2), 85–93. <https://doi.org/10.1016/j.tics.2010.11.004>.
- Falk, E.B., Hyde, L.W., Mitchell, C., Faul, J., Gonzalez, R., Heitzeg, M.M., Keating, D.P., Langa, K.M., Martz, M.E., Maslowsky, J., others, 2013. What is a representative brain? Neuroscience meets population science. *Proc. Natl. Acad. Sci.* 110 (44), 17615–17622.
- Farah, M.J., 2017. The Neuroscience of Socioeconomic Status: Correlates, Causes, and Consequences. *Neuron* 96 (1), 56–71. <https://doi.org/10.1016/j.neuron.2017.08.034>.
- Fisher, R., 1955. Statistical methods and scientific induction. *J. R. Stat. Soc. Ser. B (Methodol.)* 17 (1), 69–78. <https://doi.org/10.1111/j.2517-6161.1955.tb00180.x>.
- Fuster, J.M., 2001. The prefrontal cortex—an update: time is of the essence. *Neuron* 30 (2), 319–333.
- Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R.Z., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., Zahs, D., 2018. Recruiting the ABCD sample: design considerations and procedures. *Dev. Cogn. Neurosci.* 32, 16–22. <https://doi.org/10.1016/j.dcn.2018.04.004>.
- Groves, R.M., Fowler Jr., F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R., 2009. *Survey Methodology*, 2nd ed., Wiley.
- Habibi, A., Sarkissian, A.D., Gomez, M., Ilari, B., 2015. Developmental brain research with participants from underprivileged communities: strategies for recruitment, participation, and retention. *Mind, Brain, Educ.* 9 (3), 179–186. <https://doi.org/10.1111/mbe.12087>.
- Hagler, D.J., Hatton, SeanN., Cornejo, M.D., Makowski, C., Fair, D.A., Dick, A.S., Sutherland, M.T., Casey, B.J., Barch, D.M., Harms, M.P., Watts, R., Bjork, J.M., Garavan, H.P., Hilmer, L., Pung, C.J., Scat, C.S., Kuperman, J., Bartsch, H., Xue, F., Dale, A.M., 2019. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *NeuroImage* 202, 116091. <https://doi.org/10.1016/j.neuroimage.2019.116091>.
- Hampshire, A., Chamberlain, S.R., Monti, M.M., Duncan, J., Owen, A.M., 2010. The role of the right inferior frontal gyrus: inhibition and attentional control. *NeuroImage* 50 (3), 1313–1319. <https://doi.org/10.1016/j.neuroimage.2009.12.109>.
- Harris, P.A., Lane, L., Biaggioni, I., 2005. Clinical research subject recruitment: the volunteer for vanderbilt research Program www.volunteer.mc.vanderbilt.edu. *J. Am. Med. Inform. Assoc.: JAMIA* 12 (6), 608–613. <https://doi.org/10.1197/jamia.M1722>.
- Heeringa, S., Berglund, P.A., 2019. A Guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) study baseline data. *BioRxiv*. <https://doi.org/10.1101/2020.02.10.942011>.
- Heeringa, S.G., West, B.T., & Berglund, P.A. (2017). *Applied survey data analysis* (2nd ed.). CRC Press, Taylor & Francis Group.
- Hein, T.C., Mattson, W.I., Dotterer, H.L., Mitchell, C., Lopez-Duran, N., Thomason, M.E., Peltier, S.J., Welsh, R.C., Hyde, L.W., Monk, C.S., 2018. Amygdala habituation and uncinate fasciculus connectivity in adolescence: a multi-modal approach. *NeuroImage* 183, 617–626. <https://doi.org/10.1016/j.neuroimage.2018.08.058>.
- Hoffman, E.A., Clark, D.B., Orendain, N., Hudziak, J., Squaglia, L.M., Dowling, G.J., 2019. Stress exposures, neurodevelopment and health measures in the ABCD study. *Neurobiol. Stress* 10, 100157. <https://doi.org/10.1016/j.ynstr.2019.100157>.
- Janak, P.H., Tye, K.M., 2015. From circuits to behaviour in the amygdala. *Nature* 517 (7534), 284–292. <https://doi.org/10.1038/nature14188>.
- Johnson, S.B., Riis, J.L., Noble, K.G., 2016. State of the art review: poverty and the developing brain. *Pediatrics* 137 (4). <https://doi.org/10.1542/peds.2015-3075>.
- Jonides, J., Lewis, R.L., Nee, D.E., Lustig, C.A., Berman, M.G., Moore, K.S., 2008. The mind and brain of short-term memory. *Annu. Rev. Psychol.* 59 (1), 193–224. <https://doi.org/10.1146/annurev.psych.59.103006.093615>.
- Kalton, G., Flores-Cervantes, I., 2003. Weighting methods. *J. Off. Stat.* 19 (2), 81–97.
- Kim, J.K., 2022. A gentle introduction to data integration in survey sampling. *Surv. Stat.* 85, 19–29.
- Kish, L., 1965. Sampling organizations and groups of unequal sizes. *Am. Sociol. Rev.* 30 (4), 564–572. <https://doi.org/10.2307/2091346>.
- Kish, L., Frankel, M.R., 1974. Inference from complex samples. *J. R. Stat. Soc. Ser. B (Methodol.)* 36 (1), 1–22. <https://doi.org/10.1111/j.2517-6161.1974.tb00981.x>.
- Kong, X., Zhen, Z., Li, X., Lu, H., Wang, R., Liu, L., He, Y., Zang, Y., Liu, J., 2014. Individual differences in impulsivity predict head motion during magnetic resonance imaging. *PLOS ONE* 9 (8), e104989. <https://doi.org/10.1371/journal.pone.0104989>.
- Korn, E.L., Graubard, B.I., 1995. Examples of differing weighted and unweighted estimates from a sample survey. *Am. Stat.* 49 (3), 291–295. <https://doi.org/10.1080/00031305.1995.10476167>.
- LeDoux, J.E., 2000. Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23 (1), 155–184. <https://doi.org/10.1146/annurev.neuro.23.1.155>.
- LeWinn, K.Z., Sheridan, M.A., Keyes, K.M., Hamilton, A., McLaughlin, K.A., 2017. Sample composition alters associations between age and brain structure. *Nat. Commun.* 8 (1), 874. <https://doi.org/10.1038/s41467-017-00908-7>.
- Li, Y., Thompson, W.K., Reuter, C., Nillo, R., Jernigan, T., Dale, A., Sugrue, L.P., ABCD Consortium, 2021. Rates of incidental findings in brain magnetic resonance imaging

- in children. *JAMA Neurol.* 78 (5), 578–587. <https://doi.org/10.1001/jamaneurol.2021.0306>.
- Lisdahl, K.M., Tapert, S., Sher, K.J., Gonzalez, R., Nixon, S.J., Feldstein Ewing, S.W., Conway, K.P., Wallace, A., Sullivan, R., Hatcher, K., Kaiver, C., Thompson, W., Reuter, C., Bartsch, H., Wade, N.E., Jacobus, J., Albaugh, M.D., Allgaier, N., Anokhin, A.P., Heitzeg, M.M., 2021. Substance use patterns in 9–10 year olds: baseline findings from the adolescent brain cognitive development (ABCD) study. *Drug Alcohol Depend.* 227, 108946 <https://doi.org/10.1016/j.drugalcdep.2021.108946>.
- Lumley, T., 2010. *Complex Surveys*. John Wiley & Sons, Inc.
- Marek, S., Tervo-Clemmens, B., Calabro, F.J., Montez, D.F., Kay, B.P., Hatoum, A.S., Donohue, M.R., Foran, W., Miller, R.L., Hendrickson, T.J., Malone, S.M., Kandala, S., Feckzo, E., Miranda-Dominguez, O., Graham, A.M., Earl, E.A., Perrone, A.J., Cordova, M., Doyle, O., Dosenbach, N.U.F., 2022. Reproducible brain-wide association studies require thousands of individuals. *Nature* 1–7. <https://doi.org/10.1038/s41586-022-04492-9>.
- Maren, S., Phan, K.L., Liberzon, I., 2013. The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nat. Rev. Neurosci.* 14 (6), 417–428. <https://doi.org/10.1038/nrn3492>.
- Marshall, A.T., Betts, S., Kan, E.C., McConnell, R., Lanphear, B.P., Sowell, E.R., 2020. Association of lead-exposure risk and family income with childhood brain outcomes. *Nat. Med.* 26 (1), 91–97. <https://doi.org/10.1038/s41591-019-0713-y>.
- Masten, A.S., 2001. Ordinary magic: resilience processes in development. *Am. Psychol.* 56 (3), 227. <https://doi.org/10.1037/0003-066X.56.3.227>.
- McLoyd, V.C., 1998. Socioeconomic disadvantage and child development. *Am. Psychol.* 53 (2), 185–204.
- Medin, D., Ojalehto, B., Marin, A., Bang, M., 2017. Systems of (non-)diversity. *Nat. Hum. Behav.* 1 (5), 1–5. <https://doi.org/10.1038/s41562-017-0088>.
- Meng, X.-L., 2018. Statistical paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. In: *The Annals of Applied Statistics*, 12, pp. 685–726.
- Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L.R., Griffanti, L., Douaud, G., Okell, T.W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., Smith, S.M., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* 19 (11), 1523–1536. <https://doi.org/10.1038/nn.4393>.
- Nebbett Jr., E.W., 2019. Racism and health: challenges and future directions in behavioral and psychological research. *Cult. Divers. Ethn. Minor. Psychol.* 25 (1), 12–20. <https://doi.org/10.1037/cdp0000253>.
- Nichols, T.E., Das, S., Eickhoff, S.B., Evans, A.C., Glattard, T., Hanke, M., Kriegeskorte, N., Milham, M.P., Poldrack, R.A., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D.C., White, T., Yeo, B.T.T., 2017. Best practices in data analysis and sharing in neuroimaging using MRI. *Nat. Neurosci.* 20 (3), 299–303. <https://doi.org/10.1038/nn.4500>.
- Nielsen, M., Haun, D., Kärtner, J., Legare, C.H., 2017. The persistent sampling bias in developmental psychology: a call to action. *J. Exp. Child Psychol.* 162, 31–38. <https://doi.org/10.1016/j.jecp.2017.04.017>.
- Oswald, L.M., Wand, G.S., Zhu, S., Selby, V., 2013. Volunteerism and self-selection bias in human positron emission tomography neuroimaging research. *Brain Imaging Behav.* 7 (2), 163–176. <https://doi.org/10.1007/s11682-012-9210-3>.
- Owens, M.M., Potter, A., Hyatt, C.S., Albaugh, M., Thompson, W.K., Jernigan, T., Yuan, D., Hahn, S., Allgaier, N., Garavan, H., 2021. Recalibrating expectations about effect size: a multi-method survey of effect sizes in the ABCD study. *PLoS ONE* 16 (9), e0257535. <https://doi.org/10.1371/journal.pone.0257535>.
- Pagliaccio, D., Alqueza, K.L., Marsh, R., Auerbach, R.P., 2019. Brain volume abnormalities in youth at high risk for depression: adolescent brain and cognitive development study. *J. Am. Acad. Child Adolesc. Psychiatry.* <https://doi.org/10.1016/j.jaac.2019.09.032>.
- Paul, S.E., Hatoum, A.S., Fine, J.D., Johnson, E.C., Hansen, I., Karcher, N.R., Moreau, A. L., Bondy, E., Qu, Y., Carter, E.B., Rogers, C.E., Agrawal, A., Barch, D.M., Bogdan, R., 2021. Associations between prenatal cannabis exposure and childhood outcomes: results from the ABCD study. *JAMA Psychiatry* 78 (1), 64–76. <https://doi.org/10.1001/jamapsychiatry.2020.2902>.
- Paus, T., 2010. Population neuroscience: why and how. *Hum. Brain Mapp.* 31 (6), 891–903. <https://doi.org/10.1002/hbm.21069>.
- Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H., Rasbash, J., 1998. Weighting for unequal selection probabilities in multilevel models. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 60 (1), 23–40. <https://doi.org/10.1111/1467-9868.00106>.
- Potter, A., Dube, S., Allgaier, N., Loso, H., Ivanova, M., Barrios, L.C., Bookheimer, S., Chaarani, B., Dumas, J., Feldstein-Ewing, S., Freedman, E.G., Garavan, H., Hoffman, E., McGlade, E., Robin, L., Johns, M.M., 2020. Early adolescent gender diversity and mental health in the Adolescent Brain Cognitive Development study. *J. Child Psychol. Psychiatry.* <https://doi.org/10.1111/jcpp.13248>.
- Qu, Y., Jorgensen, N.A., Telzer, E.H., 2021. A call for greater attention to culture in the study of brain and development. *Perspect. Psychol. Sci.* 16 (2), 275–293. <https://doi.org/10.1177/1745691620931461>.
- Rabe-Hesketh, S., Skrondal, A., 2006. Multilevel modelling of complex survey data. *J. R. Stat. Soc. Ser. A (Stat. Soc.)* 169 (4), 805–827. <https://doi.org/10.1111/j.1467-985X.2006.00426.x>.
- Rakesh, D., Whittle, S., 2021. Socioeconomic status and the developing brain – a systematic review of neuroimaging findings in youth. *Neurosci. Biobehav. Rev.* 130, 379–407. <https://doi.org/10.1016/j.neubiorev.2021.08.027>.
- Rakesh, D., Zalesky, A., Whittle, S., 2021. Similar but distinct - effects of different socioeconomic indicators on resting state functional connectivity: findings from the Adolescent Brain Cognitive Development (ABCD) Study®. *Dev. Cogn. Neurosci.* 51, 101005 <https://doi.org/10.1016/j.dcn.2021.101005>.
- Rowley, S.J., Camacho, T.C., 2015. Increasing diversity in Cognitive Developmental Research: issues and solutions. *J. Cogn. Dev.* 16 (5), 683–692. <https://doi.org/10.1080/15248372.2014.976224>.
- Rueda, M., Ferri-Garcia, R., Castro, L., 2020. The R package NonProbEst for estimation in non-probability surveys. *R. J.* 12 (1), 408–418.
- Satterthwaite, T.D., Wolf, D.H., Loughhead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *NeuroImage* 60 (1), 623–632. <https://doi.org/10.1016/j.neuroimage.2011.12.063>.
- Schafer, J.L., Graham, J.W., 2002. Missing data: our view of the state of the art. *Psychol. Methods* 7 (2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>.
- Schmidt, S.C.E., Woll, A., 2017. Longitudinal drop-out and weighting against its bias. *BMC Med. Res. Methodol.* 17 (1), 164. <https://doi.org/10.1186/s12874-017-0446-x>.
- Schumann, G., Loh, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P.J., Dalley, J.W., 2010. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol. Psychiatry* 15, 1128–1139.
- Semega, J.L., Kollar, M.A., Creamer, J., & Mohanty, A. (2019). Income and poverty in the United States: 2018. U.S. Government Printing Office, Washington, DC. <https://www.census.gov/content/dam/Census/library/publications/2019/demo/p60-266.pdf>.
- Simmons, C., Conley, M.I., Gee, D.G., Baskin-Sommers, A., Barch, D.M., Hoffman, E.A., Huber, R.S., Iacono, W.G., Nagel, B.J., Palmer, C.E., Sheth, C.S., Sowell, E.R., Thompson, W.K., Casey, B.J., 2021. Responsible use of open-access developmental data: the Adolescent Brain Cognitive Development (ABCD) study. *Psychol. Sci.* 32 (6), 866–870. <https://doi.org/10.1177/09567976211003564>.
- Somerville, L.H., Bookheimer, S.Y., Buckner, R.L., Burgess, G.C., Curtiss, S.W., Dapretto, M., Elam, J.S., Gaffrey, M.S., Harms, M.P., Hodge, C., Kandala, S., Kastman, E.K., Nichols, T.E., Schlaggar, B.L., Smith, S.M., Thomas, K.M., Yacoub, E., Van Essen, D.C., Barch, D.M., 2018. The lifespan human connectome project in development: a large-scale study of brain connectivity development in 5–21 year olds. *NeuroImage* 183, 456–468. <https://doi.org/10.1016/j.neuroimage.2018.08.050>.
- Spencer, B.D., 2000. An approximate design effect for unequal weighting when measurements may correlate with selection probabilities. *Surv. Methodol.* 26, 137–138.
- Stark, C.E.L., Okado, Y., 2003. Making memories without trying: medial temporal lobe activity associated with incidental memory formation during recognition. *J. Neurosci.* 23 (17), 6748–6753. <https://doi.org/10.1523/JNEUROSCI.23-17-06748.2003>.
- Sterba, S.K., 2009. Alternative model-based and design-based frameworks for inference from samples to populations: from polarization to integration. *Multivar. Behav. Res.* 44 (6), 711–740. <https://doi.org/10.1080/00273170903333574>.
- Taherdoost, H., 2016. Sampling methods in research methodology; how to choose a sampling technique for research. *Int. J. Acad. Res. Manag.* 5 (2), 18–27. <https://doi.org/10.2139/ssrn.3205035>.
- Taylor, A., 2004. The consequences of selective participation on behavioral-genetic findings: evidence from simulated and real data. *Twin Res. Hum. Genet.* 7 (5), 485–504. <https://doi.org/10.1375/twin.7.5.485>.
- Taylor, R.L., Cooper, S.R., Jackson, J.J., Barch, D.M., 2020. Assessment of neighborhood poverty, cognitive function, and prefrontal and hippocampal volumes in children. <https://doi.org/10.1001/jamaneurology.2020.23774>.
- Tomlinson, R.C., Burt, S.A., Waller, R., Jonides, J., Miller, A.L., Gearhardt, A.N., Peltier, S.J., Klump, K.L., Lumeng, J.C., Hyde, L.W., 2020. Neighborhood poverty predicts altered neural and behavioral response inhibition. *NeuroImage*, 116536. <https://doi.org/10.1016/j.neuroimage.2020.116536>.
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., 2013. The WU-Minn human connectome project: an overview. *NeuroImage* 80, 62–79. <https://doi.org/10.1016/j.neuroimage.2013.05.041>.
- Vargas, T., Damme, K.S.F., Mittal, V.A., 2020. Neighborhood deprivation, prefrontal morphology and neurocognition in late childhood to early adolescence. *NeuroImage* 220, 117086. <https://doi.org/10.1016/j.neuroimage.2020.117086>.
- West, B.T., Berglund, P., Heeringa, S.G., 2008. A closer examination of subpopulation analysis of complex-sample survey data. *Stata J.* 8 (4), 520–531. <https://doi.org/10.1177/1536867x0800800404>.
- White, T., Marroun, H.E., Nijs, I., Schmidt, M., van der Lugt, A., Wielopolski, P.A., Jaddoe, V.W.V., Hofman, A., Krestin, G.P., Tiemeier, H., Verhulst, F.C., 2013. Pediatric population-based neuroimaging and the Generation R Study: the intersection of developmental neuroscience and epidemiology. *Eur. J. Epidemiol.* 28 (1), 99–111. <https://doi.org/10.1007/s10654-013-9768-0>.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. <https://ggplot2.tidyverse.org>.
- Wilson, W.J. (2012). *The truly disadvantaged: The inner city, the underclass, and public policy* (2nd ed.). University of Chicago Press. <https://www.press.uchicago.edu/ucp/books/book/chicago/T/bo13375722.html>.
- Yancey, A.K., Ortega, A.N., Kumanyika, S.K., 2006. Effective recruitment and retention of minority research participants. *Annu. Rev. Public Health* 27 (1), 1–28. <https://doi.org/10.1146/annurev.publhealth.27.021405.102113>.
- Yang, S., Kim, J.K., Song, R., 2020. Doubly robust inference when combining probability and non-probability samples with high dimensional data. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 82 (2), 445–465. <https://doi.org/10.1111/rssb.12354>.
- Zhang, L.-C., 2019. On valid descriptive inference from non-probability sample. *Stat. Theory Relat. Fields* 3 (2), 103–113. <https://doi.org/10.1080/24754269.2019.1666241>.