Imaging Neuroscience
an open access journal

Research Article

Heritability estimation of subcortical volumes in a multi-ancestry multi-site cohort study

Christian Coffman[a], Eric Feczko[b,c], Bart Larsen[b,c], Brenden Tervo-Clemmens[c,d], Gregory Conan[c], Jacob T. Lundquist[c], Audrey Houghton[c], Lucille A. Moore[c], Kimberly Weldon[c], Rae McCollum[c], Begim Fayzullobekova[c], Thomas J. Madison[c], Eric Earl[e], Oscar Miranda Dominguez[b,c], Damien A. Fair[b,c], Saonli Basu[a,c]

[a]Division of Biostatistics, University of Minnesota, Minneapolis, MN, United States
[b]Department of Pediatrics, University of Minnesota, Minneapolis, MN, United States
[c]Masonic Institute for the Developing Brain, University of Minnesota, Minneapolis, MN, United States
[d]Department of Psychology, University of Minnesota, Minneapolis, MN, United States
[e]NIMH Data Science and Sharing Team, National Institute of Health, Bethesda, MD, United States

Corresponding Author: Christian Coffman (coffm049@umn.edu)

ABSTRACT

Heritability of regional subcortical brain volumes (rSBVs) quantifies the contribution of genetic variation to individual differences in subcortical brain morphology during neurodevelopment. rSBVs are highly heritable in adults but are not well characterized in adolescents. The Adolescent Brain Cognitive Development study (ABCD), collected across 21 US sites, provides data to characterize the heritability of subcortical structures in adolescence. However, in a multi-site multi-ancestry study, such as ABCD, site-specific effects may co-occur with genetic effects, which can bias heritability estimates. Existing methods that adjust for site effects require additional steps and can lead to inconsistent estimates. We propose a random-effect model-based MOM approach to estimate heritability that is a single-step asymptotically consistent estimator even when ancestries are differentially distributed across sites. Our approach performed well under a wide range of simulations. The proposed approach yielded heritability estimates similar to previous results derived from single-site studies and studies on different age groups, with the cerebellar cortex and hippocampus showing the highest estimated heritability (>50%).

Keywords: SNP, ABCD, subcortical, heritability

1. INTRODUCTION

Quantifying the heritability of neuroanatomical traits is a cornerstone of modern neurogenetics, helping to clarify how genetic variation shapes brain structure and function. Regional subcortical brain volumes (rSBVs) are widely studied quantitative neuroanatomical phenotypes whose genetic architecture can be estimated reproducibly across cohorts. While rSBVs exhibit substantial heritability in adults (Y. Zhao et al., 2020), characterizing these structures during adolescence remains a critical frontier. This developmental period is marked by rapid neurobiological reorganization, making precise heritability estimates essential for understanding normative variation in brain morphology.

Genome-wide association studies (GWAS) have gained tremendous momentum with the availability of dense genotyping and whole-genome sequencing data. Heritability is now routinely estimated directly from genome-wide single-nucleotide polymorphisms (SNPs), commonly referred to as SNP-heritability (h^2_{SNP}). Classical

Received: 6 June 2025 Revision: 30 September 2026 Accepted: 12 October 2026 Available Online: 27 October 2026

The MIT Press
© 2026 The Authors. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

approaches estimate h^2_{SNP} by fitting variance components models using restricted maximum likelihood (REML) (Loh et al., 2015; Yang et al., 2011) or method of moments (MOM) estimators (Haseman & Elston, 1972). Recently, Lin et al. (2022) introduced Adjusted Haseman-Elston (Adj-HE), a moment-based estimator designed to produce consistent SNP-heritability estimates in the presence of population structure by incorporating principal components directly into the estimating equations. Adj-HE provides substantial computational gains over REML while improving robustness in ancestrally heterogeneous cohorts.

Prior twin studies have shown that MRI-derived subcortical volumes are substantially heritable across subcortical structures, including the thalamus, amygdala, putamen, caudate nucleus, pallidum, nucleus accumbens, and hippocampus (den Braber et al., 2013). Developmental twin studies have further shown that subcortical volumes are already highly heritable in childhood and early adolescence for most subcortical structures at ages 9 and 12 years (Swagerman et al., 2014). In older adult twins, heritability estimates for subcortical regions ranged from 0.41 for the amygdala to 0.73 for the hippocampus, supporting the persistence of genetic influences on these brain structures across the lifespan (Wen et al., 2016). Complementing these twin-based findings, large-scale genome-wide association studies have shown that common genetic variants contribute to variation in human subcortical brain volumes, including putamen and caudate volumes (Hibar et al., 2015). These SNP-heritability estimates are typically lower than twin-based figures, a discrepancy known as "missing heritability" (Matthews & Turkheimer, 2022). These differences are often attributed to factors such as rare variants not captured by arrays, gene–environment interactions, and model misspecification. Estimating SNP-heritability (h^2_{SNP}) precisely in adolescents, therefore, requires not only adequate sample size but also statistical methodology that is robust to the complexities of large-scale, multi-site neuroimaging cohorts with different ancestral groups.

Despite the advancement of large-scale studies, multi-site multi-ancestral cohorts introduce an additional layer of complexity beyond population structure. In studies such as the Adolescent Brain Cognitive Development (ABCD) study, which features a baseline cohort of over 10,000 subjects (Uban et al., 2018), scanner manufacturer differences, acquisition protocols, and site-specific processing pipelines induce non-biological "site effects" (Bayer et al., 2022). Crucially, in geographically distributed cohorts, genetic ancestry and site-specific effects are not independent. This entangles ancestry-driven genetic signals with site-driven technical noise. Standard adjustment strategies such as fixed-effect covariate correction, site-wise demeaning (SWD), and ComBat harmonization may not fully disentangle these dependencies and often introduce many additional nuisance parameters, causing a loss of efficiency in estimating the parameters of interest. In such settings, these methods potentially yield inconsistent SNP-heritability estimates that fail to converge to true values as sample sizes increase.

To address this confounding, we propose *Adjusted Haseman-Elston estimation with random site effects* (AdjHE-RE), a MOM estimator for h^2_{SNP} that explicitly accounts for site effects in the covariance terms while remaining robust to dependencies between site membership and genetic ancestry. Conceptually, AdjHE-RE augments the Haseman–Elston framework with random site effects in a second estimating equation, yielding a consistent estimator of h^2_{SNP} under multi-site designs where site and ancestry are not independent. Our development is grounded in standard estimating-equation theory for *M*-estimators (Stefanski & Boos, 2002).

In this paper, we present the statistical framework for AdjHE-RE and validate its reliability through comprehensive synthetic data simulations. We then apply the method to the multi-ancestry ABCD dataset to characterize the SNP-heritability of regional subcortical brain volumes. Our results reveal the enhanced consistency and reliability of AdjHE-RE across a wide range of scenarios, providing estimates that align with established results while resolving the unique challenges inherent in large-scale, multi-site adolescent neuroimaging. The proposed AdjHE-RE framework provides a general variance-component approach for heritability estimation in large-scale, multi-site, multi-ancestry cohorts, where genetic relatedness, population structure, and site-level heterogeneity may simultaneously contribute to phenotypic covariance. We illustrate the proposed approach using subcortical brain volume data in the ABCD study, but the same framework can be applied to other quantitative phenotypes. Our proposed approach is especially well tuned for application to large-scale, multi-site, multi-ancestry datasets.

2. MATERIALS AND METHODS

2.1. Ethics statement

This study is a secondary analysis of the publicly available Adolescent Brain Cognitive Development (ABCD) study and involved no new recruitment or new data collection with human participants; accordingly, review by an institutional review board was not required. Oversight of the original ABCD study was provided by the independent IRB at the University of California, San Diego, with

local review at participating sites. Use of the ABCD data in this work was governed by a data use agreement authorized by the University of Minnesota, signed by the investigators, and approved by the NIMH Data Archive, and all analyses were performed on de-identified data in accordance with that agreement.

2.2. Genotype, phenotype, and heritability

We consider a single quantitative phenotype ($\boldsymbol{Y}^*$) measured on n individuals from K sub-populations ($K \in \mathrm{N}$, with n_k denoting the number of observations in group k and $n = \sum_{k=1}^{K} n_k$), typed at M SNP positions. Assume a genotype matrix $\boldsymbol{X}_g$ ($n \times M$) where the imth entry represents the number of allele copies at SNP position m for subject i and takes values 0, 1, or 2. The $\boldsymbol{X}_g$ are observed from K subpopulations yielding $\boldsymbol{X}_{g_k}$ for each $k \in \{1, \ldots, K\}$. For each sub-population (k), the allelic frequency at position m is given by f_{km}. The standardized genotypes are then given by $\boldsymbol{Z}^*_{kim} = \dfrac{\boldsymbol{X}_{g_{kim}} - f_{km}}{\sqrt{2f_{km}(1-f_{km})}}$. The genetic relatedness matrix (GRM, $\boldsymbol{A}^*$) is then defined by $\boldsymbol{A}^* = \boldsymbol{Z}^* \boldsymbol{Z}'^* / M$ that captures genetic correlation among a pair of individuals. Traditional heritability estimators are typically developed for a single approximately homogeneous population, in which case the empirical GRM is estimated as $\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{Z}' / M$, where $Z_{im} = \dfrac{\boldsymbol{X}_{g_{im}} - \hat{f}_m}{\sqrt{2\hat{f}_m(1-\hat{f}_m)}}$, where $\hat{f}_m$ is the allelic frequency for the entire sample ($\hat{f}_m = \sum_{i=1}^{n} \boldsymbol{X}_{g,im} / 2n$).

We define $\boldsymbol{X}_S$ as an $n \times p_s$ site or location matrix where the (ij)th entry is 1 if subject i was observed at site j and 0 otherwise, codifying the site membership. We define $\boldsymbol{X}_d (n \times p_d)$ as a set of covariates defining demographics such as sex and age, and an intercept. The covariates can take on many forms, such as categorical for sex and continuous for age. The column-wise concatenation of the demographic variables ($\boldsymbol{X}_d$) with the site membership matrix ($\boldsymbol{X}_S$) will be denoted by $\boldsymbol{X}_{dS}$. We assume $p_d < p_s < n << M$. In addition, we assume all M SNPs are causal for phenotype $\boldsymbol{Y}^*$ and use the following linear model to capture the relationship among Y^*, X_S, X_d, and Z:

$$\begin{aligned} \boldsymbol{Y}^*_i &= \boldsymbol{Z}_i \beta_g + \boldsymbol{X}_{Si} \beta_S + \boldsymbol{X}_{di} \boldsymbol{\beta}_d + \varepsilon_i \\ E(\varepsilon_i) &= 0, \ \mathrm{Var}(\varepsilon_i) = \sigma^2_e. \end{aligned} \tag{1}$$

To quantify the genetic contribution, we assume $\boldsymbol{\beta}_g \sim N(0, \sigma^2_g \boldsymbol{I} / M)$. Then Eq. (1) reduces to

$$\boldsymbol{Y}^* \sim N(\boldsymbol{X}_{dS} \boldsymbol{\beta}_{dS}, \sigma^2_g \boldsymbol{A} + \sigma^2_e \boldsymbol{I}). \tag{2}$$

Re-expressing the model in terms of random genetic effects, and assuming that the covariance between the genetic effects and the unmodeled error (ϵ) is 0, the heritability of the trait $\boldsymbol{Y}^*$ is defined as $h^2 = \sigma^2_g / (\sigma^2_g + \sigma^2_e)$. We are interested in estimating the parameters σ^2_g and σ^2_e in model 2. If the same model is applied to K subpopulations, the underlying assumption is that the phenotype's heritability is the same across all subpopulations. The populations differ in their allele frequencies at the M markers, and these differences can be addressed by including the GRM-derived principal components as fixed effects in Eq. (1) (Lin et al., 2022). We also need to estimate β_{dS}. As discussed above, the ancestral distribution could be correlated with the site indicators $\boldsymbol{X}_S$, so the GRM-derived PCs can also be correlated with the site effects. For a condensed reference for notation see Supplementary Table S1.

2.2.1. Heritability estimation approaches

Without loss of generality, we assume that there is no site effect present. This subsection reviews the standard single-variance-component formulation used by existing GCTA and AdjHE methods; the proposed random-site extension is introduced in Section 2.4. In the absence of site effects, we have the phenotype $\tilde{Y} \sim N(\mathbf{X}_d \boldsymbol{\beta}_d, \Sigma = \mathbf{A}\sigma^2_g + \mathbf{I}\sigma^2_e)$. For Restricted Maximum Likelihood Estimation (REML), we consider the log-likelihood of the data that is given by:

$$\begin{aligned} l(\beta, \sigma^2_g, \sigma^2_e) = &\frac{N}{2}\log(2\pi) + \frac{1}{2}\log\det\Sigma \\ &- \frac{1}{2}(\tilde{\boldsymbol{Y}} - \mathbf{X}_d\beta_d)\Sigma^{-1}(\tilde{\boldsymbol{Y}} - \boldsymbol{X}_d\beta_d). \end{aligned} \tag{3}$$

The estimation of this mixed-effect model mentioned above is performed through maximum likelihood estimation. The software GCTA (Yang et al., 2011) uses the iterative restricted maximum likelihood (REML) algorithm to estimate the variance components σ^2_g and σ^2_e in the model described above and gives an estimate of heritability by $\hat{h}^2 = \hat{\sigma}^2_g / \hat{\sigma}^2_{\tilde{y}}$, where $\hat{\sigma}^2_{\tilde{y}}$ is the estimated phenotypic total variance ($\sigma^2_{\tilde{y}} = \sigma^2_g + \sigma^2_e$). For a multi-site multi-ancestry study, we assume the model in Eq. (2) for the GCTA-based REML estimation.

This REML-based estimation for a mixed-effect model can easily become computationally intractable as the sample size increases. Recently, there have been attempts to generate computationally scalable algorithms to implement mixed models on large-scale data (Fan et al., 2022; Loh et al., 2015; Smith et al., 2023). However, these approaches still encounter computational challenges in large Biobanks. Moreover, even subtle population substructure can significantly impact the heritability estimation (Conomos et al., 2016).

The MOM (Haseman–Elston regression (Haseman & Elston, 1972), LDscore regression (Bulik-Sullivan et al., 2015), MMHE (Ge et al., 2017)) approaches are another set of widely used methods for estimating heritability h^2 under Eq. (1). We will next provide a short overview of the Haseman–Elston method.

In the Haseman–Elston approach, we model the first two moments of the trait distribution for $\tilde{Y}$ as described in Eq. (2). We assume that $E\left(\tilde{\boldsymbol{Y}}\right) = \boldsymbol{X}_d \boldsymbol{\beta}_d$ and that Eq. (2) holds under our assumption of no site effects. The Haseman–Elston estimator for h^2 is obtained from the least-squares regression implied by Eq. (2), which can be written as

$$\begin{aligned} E(\mathbf{Y}_i \mathbf{Y}_{i'} \mid \mathbf{A}) &= \sigma_g^2 \mathbf{A}_{ii'} \text{ for } i \le i' \\ E(\mathbf{Y}_i^2 \mid \mathbf{A}) &= \sigma_g^2 \mathbf{A}_{ii} + \sigma_e^2, \end{aligned} \tag{4}$$

where $\mathbf{Y}$ is the residualized phenotype derived from $E\left(\tilde{\boldsymbol{Y}}\right) = \boldsymbol{X}_d \boldsymbol{\beta}_d$.

The above approaches generally assume a homogeneous population. In a sample of diverse ancestry or even in the presence of a subtle substructure, these existing methods could produce very biased estimates of heritability. The standard approach is to add the principal components (***PCs***) of the GRM $\mathbf{A}$ as covariates in Eq. (1) and perform REML estimation (Yang et al., 2011). Lin et al. (2022) proposed the AdjHE approach that proposes a product PC correction to the Haseman–Elston regression (Haseman & Elston, 1972) approach to correct for the population stratification effects in Eq. (4).

Under the AdjHE model, if we assume the K subpopulations can be completely separated by K PCs, then

$$\begin{aligned} E\tilde{Y} &= \boldsymbol{X}_d \boldsymbol{\beta}_d + \sum_{k=1}^{K} \boldsymbol{PC}_k \gamma_k, \quad \text{and} \\ Var(\boldsymbol{Y}) &= \sigma_g^2 \boldsymbol{A} + \sigma_e^2 \boldsymbol{I} + \sum_{k=1}^{K} \delta_k \boldsymbol{PC}_k \boldsymbol{PC}_k'. \end{aligned} \tag{5}$$

To avoid overloading notation used later for the random-site model, let

$$a_k := \boldsymbol{PC}_k' \boldsymbol{A} \boldsymbol{PC}_k, \qquad b_k := \boldsymbol{y}' \boldsymbol{PC}_k \boldsymbol{PC}_k' \boldsymbol{y},$$

and

$$c := (n-K)\left(\operatorname{tr}\left(\boldsymbol{A}^2\right) - \sum_{k=1}^{K} a_k^2\right) - \left(\operatorname{tr}(\boldsymbol{A}) - \sum_{k=1}^{K} a_k\right)^2.$$

The AdjHE MOM solution is then

$$\begin{aligned} \hat{\sigma}_g^2 &= \frac{(n-K)\left(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} - \sum_{k=1}^{K} a_k b_k\right) - \left(\operatorname{tr}(\mathbf{A}) - \sum_{k=1}^{K} a_k\right)\left(\boldsymbol{y}'\boldsymbol{y} - \sum_{k=1}^{K} b_k\right)}{c}, \\ \hat{\sigma}_e^2 &= \frac{-\left(\operatorname{tr}(\boldsymbol{A}) - \sum_{k=1}^{K} a_k\right)\left(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} - \sum_{k=1}^{K} a_k b_k\right) + \left(\operatorname{tr}\left(\boldsymbol{A}^2\right) - \sum_{k=1}^{K} a_k^2\right)\left(\boldsymbol{y}'\boldsymbol{y} - \sum_{k=1}^{K} b_k\right)}{c}, \end{aligned} \tag{6}$$

where $\boldsymbol{y}$ is the observed residualized phenotype value. The heritability is then estimated as $\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_e^2}$ (Lin et al., 2022).

2.3. Heritability estimation with AdjHE using existing methods to control for site effects

We have thus far assumed the absence of site effects on the phenotype. However, in multi-site neuroimaging studies, substantial site-specific variability is commonly observed due to differences in scanner hardware, field strength, acquisition protocols, software versions, and preprocessing pipelines, and demographic composition across sites (Fortin et al., 2017, 2018; Friedman et al., 2008; Jovicich et al., 2009; Pomponio et al., 2020). Such site effects may induce systematic shifts in mean structure, variance inflation, and heteroskedasticity across sites.

Failure to appropriately account for the site effects may bias the estimation of genetic variance components, distort SNP-heritability estimates, and reduce power for association testing. In particular, when the site is correlated with ancestry, recruitment strategy, or genotype distribution, unmodeled site effects may induce confounding between environmental and genetic sources of variation, thereby affecting the genetic relationship matrix (GRM)-based variance decomposition. First, we consider GCTA with sites as covariates as described in Eq. (1). Note that, if site acts only through the mean, GCTA/REML with site fixed effects is appropriate and not biased by site–ancestry association alone. However, if site contributes residual covariance beyond mean shifts, then a GCTA model with only *A* and *I* covariance components is misspecified. In addition, we consider, here, three ways to control for site effects: (a) Perform a two-step approach with residualizing the phenotype against site memberships as fixed effects using existing Site-Wise De-meaning or ComBat/CovBat approaches (Johnson et al., 2007; Reardon et al., 2021), followed by AdjHE

estimation on the residualized phenotypes; (b) adjust for site as fixed effects in the first moment equation of AdjHE estimation in Eq. (4), and (c) generate separate heritability estimates for each site and pool them in a meta-analysis.

In approach (a), we first apply existing site-adjustment procedures, including Site-Wise De-meaning (SWD) (Reardon et al., 2021) and ComBat/CovBat adjustment (Johnson et al., 2007), to remove site-related mean and variance shifts in Step 1. We then estimate SNP-heritability using the AdjHE estimator (Lin et al., 2022) based on the adjusted phenotypes. We note that additional strategies for mitigating site effects exist, including harmonization through preprocessing pipelines (Kochunov et al., 2014; Pinto et al., 2020) and nonlinear transformation-based corrections (Bayer et al., 2022). These approaches are not considered in the present work.

Under a unified MOM framework, interpreted through M-estimation and generalized estimating equation (GEE) theory, we evaluate the large-sample properties of heritability estimators derived from approaches (a), (b), and (c). In particular, we establish conditions under which these estimators are consistent. Consistency holds when residualization of site and demographic effects affects only the mean structure of the phenotype and does not alter the covariance structure of the genetic component. Formally, if the projection used to remove site effects is orthogonal to the genetic random effect space (i.e., does not distort the genetic covariance or GRM $\boldsymbol{A}$), then the resulting MOM estimator remains consistent for σ_g^2.

2.3.1. Method a1: Site-wise de-meaning (SWD) adjustment

Approach (a1) applies Site-Wise De-meaning (SWD) adjustment (Reardon et al., 2021) to remove site-specific mean effects prior to heritability estimation.

Let X_S denote the design matrix containing site indicators (and optionally other fixed covariates). The site-adjusted phenotype is obtained by solving the first-moment estimating equation

$$\psi_1(\beta_S) = \sum_{i=1}^{n} \left(y_i - X_{S,i}\beta_S\right) X_{S,i} = 0,$$

yielding the estimator $\tilde{\beta}_S$. The residualized phenotype is then defined as

$$Y_{\text{SWD}} = y - X_S\tilde{\beta}_S.$$

In the second step, SNP-heritability is estimated by applying the AdjHE estimator (Eq. (5)) to Y_{SWD}. Let $\tilde{\sigma}_g^2$ and $\tilde{\sigma}_e^2$ denote the resulting MOM estimators of the genetic and residual variance components. The estimated heritability is $\tilde{h}^2 = \dfrac{\tilde{\sigma}_g^2}{\tilde{\sigma}_g^2 + \tilde{\sigma}_e^2}$.

2.3.2. Method a2: ComBat/CovBat adjustment

ComBat employs an empirical Bayes framework to remove site-specific mean and variance effects while preserving the biological signal (Johnson et al., 2007).

Heritability estimation under this approach proceeds in two steps. First, the phenotype is harmonized using ComBat, thereby adjusting for site-related mean, variance. Let $\tilde{y}$ denote the harmonized phenotype after this adjustment. SNP-heritability is then estimated by applying the AdjHE estimator (Eq. (5)) to the harmonized phenotype $\tilde{y}$. Let $\tilde{\sigma}_g^2$ and $\tilde{\sigma}_e^2$ denote the corresponding variance component estimators. The estimated heritability is given by $\tilde{h}^2 = \dfrac{\tilde{\sigma}_g^2}{\tilde{\sigma}_g^2 + \tilde{\sigma}_e^2}$.

2.3.3. Method b: AdjHE with site as fixed effects

In Method (b), site indicators are incorporated directly as fixed effects within the AdjHE framework. This corresponds to treating site as an additional set of covariates in the mean model, analogous to Adjusted Residual Harmonization (Bayer et al., 2022).

Let X_{dS} denote the design matrix containing demographic covariates and site indicators. The first-moment condition is given by

$$E(Y) = X_{dS}\beta_{dS} + \sum_{k=1}^{K} PC_k\gamma_k,$$

where PC_k represents the k-th principal component used to adjust for population structure.

The residual phenotype is obtained as

$$Y_{\text{AdjHE}} = Y - X_{dS}\tilde{\beta}_{dS} - \sum_{k=1}^{K} PC_k\tilde{\gamma}_k.$$

Variance components are then estimated using the AdjHE estimating equation for variance (Eq. (5)). The resulting MOM estimators $\tilde{\sigma}_g^2$ and $\tilde{\sigma}_e^2$ yield the heritability estimate $\tilde{h}^2 = \dfrac{\tilde{\sigma}_g^2}{\tilde{\sigma}_g^2 + \tilde{\sigma}_e^2}$. Thus, Method (b) amounts to extending the standard AdjHE model by including site indicators as fixed covariates.

2.3.4. Method c: Meta-analysis across sites

In Method (c), heritability is estimated separately within each site using the AdjHE equations in Eq. (5), and the resulting estimates are combined using meta-analysis.

Specifically, for site j with sample size n_j, SNP-heritability $\hat{h}_j^2$ is estimated using the AdjHE estimator (Eq. (5)). Let p_s denote the total number of sites and $n = \sum_{j=1}^{p_s} n_j$ the overall sample size.

The meta-analytic estimate is computed as the sample-size weighted average

$$\hat{h}^2_{\text{meta}} = \sum_{j=1}^{p_s} \frac{n_j}{n} \hat{h}_j^2.$$

This approach effectively treats the site as a stratification variable, estimating heritability within site and combining estimates post hoc. As such, it shares limitations with fixed-effect approaches to site adjustment, including reduced efficiency and potential instability when site-specific sample sizes are small (Kelley & Kelley, 2012).

2.3.5. *Limitations of fixed-effect site adjustments in Method a and Method b*

When site indicators are included as fixed effects, and the phenotype is residualized via $P_{dS} = I - \tilde{X}_{dS}\left(\tilde{X}'_{dS}\tilde{X}_{dS}\right)^{-1}\tilde{X}'_{dS}$,, the resulting covariance becomes $\text{Cov}\left(P_{dS}Y\right) = P_{dS}\text{Cov}(Y)P_{dS}$. Thus, ideally, in the second order estimating equation in Eq. (5), the genetic relationship matrix A should be replaced by $P_{dS}AP_{dS}$. If site indicators are correlated with ancestry, using A instead of $P_{dS}AP_{dS}$ in Eq. (5) could potentially bias in the MOM heritability estimation. Formal consistency conditions and proofs are provided in Supplementary Materials Section S3.

Proposition 1. *When site membership is confounded with ancestry, two-step residualization methods such as SWD and ComBat/CovBat, as well as AdjHE with site as fixed effects, are generally based on misspecified second-moment equations. Therefore, these methods need not be consistent for SNP-heritability in multi-site multi-ancestry studies.*

2.4. New method AdjHE-RE: Adjusting for sites as a random effect

We aim to expand the settings under which SNP-heritability can be estimated using an MOM approach without bias in multi-site studies by modeling site as a random effect.

In earlier MOM approaches, site indicators were treated as fixed effects and removed via residualization. As shown in Section 2.3.5, residualization induces the covariance transformation $\text{Cov}(y) = P_{dS}\text{Cov}(Y)P_{dS}$, which replaces the GRM A by $P_{dS}AP_{dS}$. When site membership is correlated with genetic ancestry (so that the column space of site indicators aligns with population structure), this transformation can remove ancestry-aligned components of the genetic signal, leading to a biased MOM estimation.

To address this, we instead include the site as an explicit variance component. In addition to the polygenic effect, we assume a random site effect with $\beta_S \sim N\left(0, \sigma_S^2\right)$ and $\text{Cov}\left(\beta_S, \beta_g^c\right) = 0$. Equivalently, the proposed model reparameterizes site from a nuisance mean-adjustment term into a covariance component that is estimated jointly with the genetic and residual variance components. Under this model, the phenotype satisfies

$$\tilde{Y} \sim \text{N}\left(\tilde{X}_d\beta_d, \sigma_g^2 A + \sigma_S^2 S + \sigma_e^2 I\right), \tag{7}$$

where A is the GRM and

$$S = X_S X'_S$$

is the site-similarity matrix that encodes site membership. Crucially, even when A and S are correlated (e.g., $\text{tr}(AS) \neq 0$ due to site–ancestry confounding), the model in Eq. (7) remains correctly specified because the cross-structure is handled through the joint estimation of $\left(\sigma_g^2, \sigma_S^2, \sigma_e^2\right)$ rather than by removing site through projection.

We account for population stratification in the mean by including the top K ancestry PCs along with demographic fixed effects $\tilde{X}_d$ (such as age and sex):

$$E\left(\tilde{Y}\right) = \tilde{X}_d\beta_d + \sum_{k=1}^{K} PC_k\gamma_k.$$

We then work with the residualized phenotype y after regressing out these fixed effects. In AdjHE-RE, site/scanner indicators are not included as fixed effects in the first-order residualization step. Instead, site is retained through $\boldsymbol{S}$, while ancestry-related mean and covariance structure is accounted for through the leading PCs, following the AdjHE framework (Lin et al., 2022). The second-moment condition is then based on the working covariance model.

$$\text{Cov}(y) = \sigma_g^2 A + \sigma_S^2 S + \sigma_e^2 I,$$

with additional PC-based low-rank adjustments when needed.

We estimate $\left(\sigma_g^2, \sigma_S^2, \sigma_e^2\right)$ by solving the corresponding MOM normal equations (details in Supplementary Materials Section S1), yielding

$$\begin{bmatrix} \hat{\sigma}_g^2 \\ \hat{\sigma}_s^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} tr(A^2) - \sum_{k=1}^{K} t_k^2 & tr(AS) - \sum_{k=1}^{K} s_k t_k & tr(A) - \sum_{k=1}^{K} t_k \\ tr(SA) - \sum_{k=1}^{K} t_k s_k & tr(S^2) - \sum_{k=1}^{K} s_k^2 & tr(S) - \sum_{k=1}^{K} s_k \\ tr(A) - \sum_{k=1}^{K} t_k & tr(S) - \sum_{k=1}^{K} s_k & N - K \end{bmatrix}^{-1} \begin{bmatrix} y'Ay - \sum_{k=1}^{K} u_k t_k \\ y'Sy - \sum_{k=1}^{K} u_k s_k \\ y'y - \sum_{k=1}^{K} u_k \end{bmatrix}, \tag{8}$$

where N is the total sample size,

$$s_k := \mathrm{PC}_k' S \mathrm{PC}_k, \quad t_k := \mathrm{PC}_k' A \mathrm{PC}_k, \quad u_k := y' \mathrm{PC}_k \mathrm{PC}_k' y.$$

This random-site variance component formulation mitigates bias induced by site-ancestry dependence by allocating site-driven covariance to $\sigma_S^2 S$ rather than eliminating it through projection, while retaining computational efficiency via closed-form moment equations.

Theorem 1. *Assume that the residual phenotype y satisfies the correctly specified second-moment model*

$$E(yy') = \sigma_g^2 A + \sigma_S^2 S + \sigma_e^2 I + \sum_{k=1}^{K} \delta_k \mathrm{PC}_k \mathrm{PC}_k'.$$

If the associated Gram matrix of basis matrices is nonsingular, then the AdjHE-RE method-of-moments estimator is unbiased under the correctly specified second-moment model. Moreover, under standard quadratic-form concentration conditions, the estimator is consistent for $\left(\sigma_g^2, \sigma_S^2, \sigma_e^2\right)$*, and consequently* $\hat{h}^2$ *is a consistent estimator of* h^2 (Rudelson & Vershynin, 2013).

To remain consistent with the heritability definition stated above, we report site-adjusted (conditional) SNP-heritability under the random-site formulation, namely $h^2 = \sigma_g^2 / \left(\sigma_g^2 + \sigma_e^2\right)$. Derivations, consistency arguments, and the asymptotic variance expression are given in Supplementary Materials Sections S1–S4.

The major advantage of both AdjHE and AdjHE-RE approaches is their substantial gain in computational efficiency compared with the GCTA approach. AdjHE-RE is a closed-form method-of-moments estimator and does not require iterative likelihood optimization or repeated dense $N \times N$ matrix inversions, which are major computational costs in REML-based approaches such as GCTA. The primary computational cost for the AdjHE-RE approach arises from constructing the GRM $\boldsymbol{A}$ and evaluating trace inner products among the covariance basis matrices. Once $\boldsymbol{A}$ is available, the dense operations involving $\boldsymbol{A}$ scale approximately quadratically in N for the required trace/product calculations. The additional site covariance matrix $\boldsymbol{S} = \boldsymbol{X}_S \boldsymbol{X}_S'$ has rank at most the number of sites and is, therefore, highly structured. Consequently, terms such as $\mathrm{tr}(\boldsymbol{AS})$, $\mathrm{tr}\left(\boldsymbol{S}^2\right)$, and $\mathrm{tr}(\boldsymbol{S})$ can be computed without treating $\boldsymbol{S}$ as a dense unstructured matrix.

We also considered the memory requirements of AdjHE-RE. The leading memory cost for GCTA, AdjHE, and AdjHE-RE is storage of the genetic relationship matrix $\boldsymbol{A}$, which requires $O(N^2)$ memory for N individuals. AdjHE-RE introduces the site-similarity matrix $\boldsymbol{S} = \boldsymbol{X}_S \boldsymbol{X}_S'$, but this matrix does not need to be stored explicitly as a dense $N \times N$ matrix. Because $\boldsymbol{X}_S$ is an $N \times p_S$ site-indicator matrix, where p_S is the number of sites, all moments involving $\boldsymbol{S}$ can be computed using $\boldsymbol{X}_S$ directly.

2.5. Simulations

Genetic data were generated under the Balding–Nichols model (Balding & Nichols, 1995) to induce population structure. Each experiment was replicated 100 times.

For each replicate, ancestral allele frequencies (p_{0m}) were drawn independently from $\mathrm{Uniform}(0.1, 0.9)$. Ancestry-specific allele frequencies (p_{km}) were generated from

$$\mathrm{Beta}\left(\frac{p_{0m}(1-\theta)}{\theta}, \frac{(1-p_{0m})(1-\theta)}{\theta}\right),$$

with $\theta = 0.1$. Genotypes were then sampled as $\mathrm{Binomial}\left(2, p_{km}\right)$ conditional on ancestry.

A proportion (2%) of SNPs were selected as causal, with effects drawn from $N\left(0, \sigma_g^2 / C\right)$. A continuous demographic covariate was simulated to represent fixed effects such as age or sex. Phenotypes were generated according to Eq. (1).

Unless otherwise specified, simulations used $n = 2000$ individuals, $C = 20{,}000$ SNPs, $\theta = 0.1$, $\sigma_g^2 = 0.2$, $\sigma_S^2 = 0.1$, and $\sigma_e^2 = 0.8$. Under this specification, the total phenotypic variance is 1, and the site-adjusted (conditional) heritability equals $\sigma_g^2 / \left(\sigma_g^2 + \sigma_e^2\right) = 0.2$. Simulation factors were varied to evaluate robustness to (i) the number of sites (1, 2, or 25), (ii) the number of ancestries (1 or 2), (iii) the distribution of site effects, and (iv) the allocation of ancestries across sites. These scenarios were designed to assess estimator behavior under increasing degrees of site-ancestry confounding, as predicted by the theoretical results in Section 2.3.5.

Estimates were obtained using SWD (Reardon et al., 2021), ComBat (Johnson et al., 2007), AdjHE with fixed site effects (Lin et al., 2022), meta-analysis approaches (Kelley & Kelley, 2012), and the proposed random-site estimator (AdjHE-RE). Across scenarios with balanced ancestry and a normally distributed site effect, all estimators were approximately unbiased (Fig. 1). Under increasing site-ancestry confounding, bias emerged for fixed-effect and two-stage harmonization approaches, consistent with the covariance-distortion results of Section 2.3.5. AdjHE-RE demonstrated the greatest robustness across heterogeneous site configurations. When site effects departed substantially from normality, all estimators exhibited some bias, although the magnitude was smallest for AdjHE-RE (Fig. 2).

2.5.1. Simulation 0: Single site, single cluster

The first simulation involved the estimation of a homogeneous population from a single site to test the quality of the genetic simulation procedure. Since it only involved one site, it restricted estimation to GCTA or AdjHE methods. These simulations were repeated over an array of $\sigma_g^2 \in (0, 0.25, 0.5)$, which are represented as faceted column plots in Supplementary Figure S1. The true simulated heritability in this figure is represented as a horizontal black line, and the distribution of estimation results from each procedure is represented as a boxplot with estimations following outside 1.5 multiples of the interquartile range represented as points.

Both AdjHE and GCTA have median estimates that are indistinguishable from the simulated heritability and with similar distributions about the simulated heritability under all combinations of variance contributions from site, genetics, and error. These simulations affirmed the reliability of the genetic simulation process.

2.5.2. Simulation 1, homoskedastic site variance

Results are contained in Supplementary Figure S2. Simulated ancestries were equally distributed across the sites, which had equal variance (homoskedastic). The

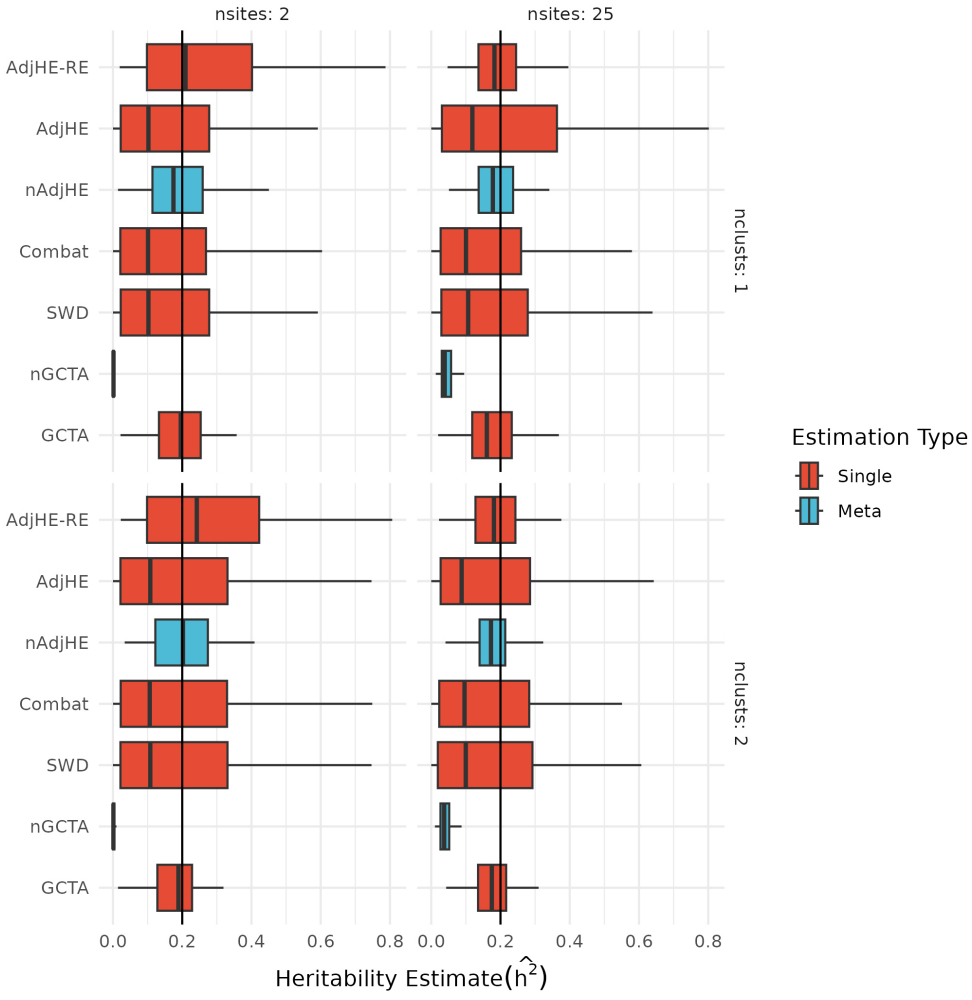


Fig. 1. Estimated heritability with an increasing number of sites (columns) and clusters (rows) where h^2 is fixed at 20% (vertical black line). The variance attributable to site effects was fixed at an equal value.

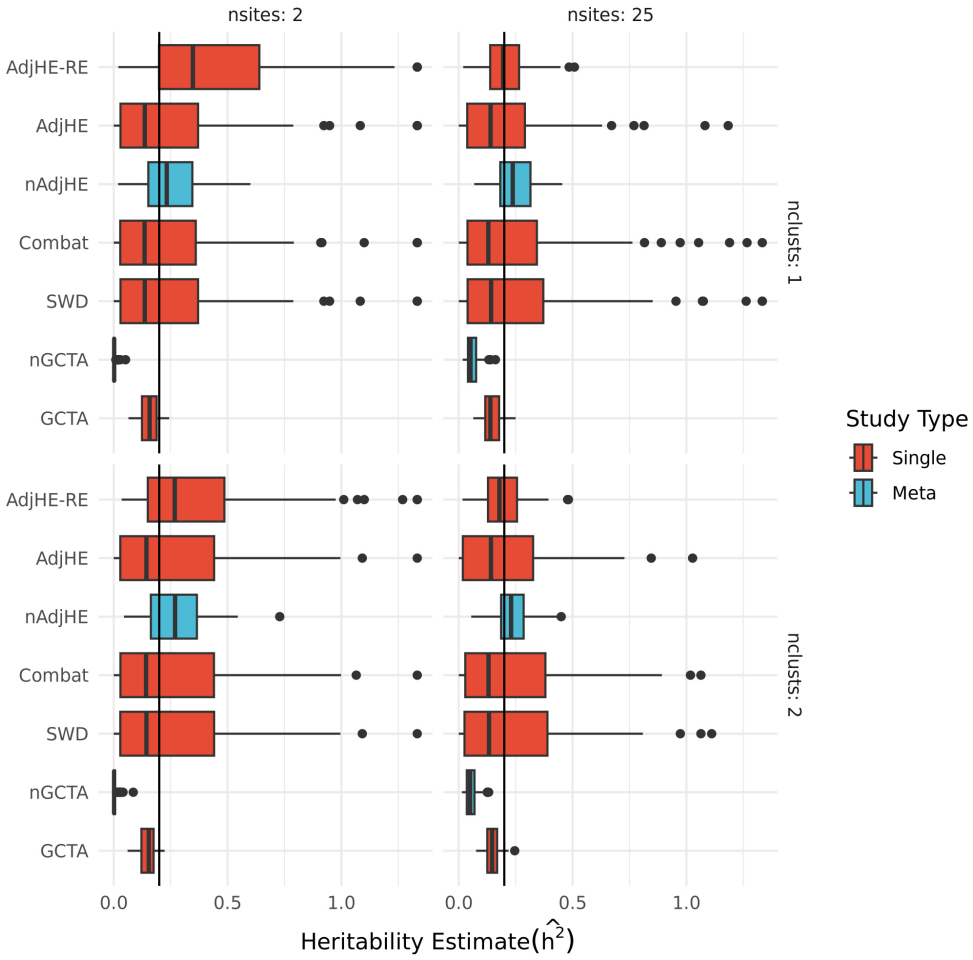


Fig. 2. Estimated heritability with an increasing number of sites (columns) and clusters (rows) where h^2 is fixed at 20% (vertical black line). The variance attributable to site effects was fixed at an equal value.

simulation was repeated for 2 and 25 sites (columns in Supplementary Fig. S2) as well as 1 and 2 simulated ancestries (rows in Supplementary Fig. S2). Note that the site effect variance $\left(\sigma_S^2\right)$ is not included in the denominator of the heritability calculation for AdjHE-RE to be comparable with the the other estimation methods.

Estimates in these plots are colored according to whether they are emulating a single study with multiple sites (red) or a meta-analysis (blue), where each estimate comprises averaging over single-site studies. The median of AdjHE-RE and GCTA are close to the true heritability. The method of moments approaches showed higher standard error for the estimates. As expected, the variability in AdjHE-RE estimates significantly improved as we added more sites and more ancestries, since it treats both ancestry and sites as random effects.

2.5.3. Simulation 2, heteroskedastic site variance

The next set of simulations involved keeping the same setup as previously discussed except with the addition that site variances (δ_j) were sampled iid from *Gamma*(4, 4). The results of these simulations are shown in Figure 1.

AdjHE-RE demonstrated unbiasedness across most conditions, exhibiting the lowest variance with large number of sites. In contrast, AdjHE performed similar across all scenarios as Simulation 1, showed bias and larger variance compared with AdjHE-RE. In fact, all Method (a) and Method (b) approaches showed biases as well as their second order estimating equations do not correct for the site effects. Notably, nGCTA displayed unbiasedness except in one scenario, while nAdjHE consistently showed bias. nAdjHE performed very well in generating precise estimates of the true heritability. The findings of this simulation highlight the robustness of AdjHE-RE in estimating heritability when site effects follow a normal distribution when the number of sites is large.

2.5.4. Simulation 3, fixed site effects

To test AdjHE-RE beyond the supposed standard situation of normally distributed site effects, the same

simulation setup as from the previous section was again repeated, but where the site effects were uniformly spaced about 0 and kept as a fixed effect across all replicates, scaled to σ_g^2. The results of this simulation setup are shown in Figure 2.

Under this simulation scenario, nGCTA showed significant bias with very small standard error AdjHE-RE again performed very well as the number of sites increased. All Method (a) and Method (b) approaches showed some bias and high variability in their estimates.

2.5.5. *Estimation time efficiency*

Comparisons of analysis time are given in Supplementary Table S3 and shown in Figure 3. For GCTA, the complexity is based on the required GRM matrix inversion. For the AdjHE methods, the most complex operations are solving the matrix multiplication steps for solving the second-moment equation. However, from profiling (Supplementary Table S3) it is apparent that the AdjHE random variable estimator remains a consistent 3–4x faster than the GCTA algorithm, and the AdjHE fixed effect model remains approximately 20x faster than GCTA. Python-based memory monitoring revealed that AdjHE-RE required an average of 5000 MB of RAM, while both GCTA and standard AdjHE required approximately 2000 MB each. This counterintuitive finding—which the moment-based methods (AdjHE and AdjHE-RE) do not inherently require less memory than REML—is attributed to the fact that all three methods must load and process the full genetic relatedness matrix (GRM) and phenotype data.

While standard REML methods (like GCTA) and moment-based approaches (AdjHE, AdjHE-RE) share the same cubic $O(N^3)$ theoretical complexity for naive matrix inversions and products, AdjHE and AdjHE-RE achieve substantial runtime speedups through a constant-factor reduction. This reduction derives from algebraic simplifications on the site-similarity matrix $S = X_S X_S'$ and the identity matrix I. Since S is block-diagonal and I is diagonal, computing the traces and products of A^2, AS, S^2 avoids the costly unstructured $N \times N$ matrix inversions required by GCTA. Additionally, as closed-form method-of-moments estimators, AdjHE and AdjHE-RE bypass the computationally expensive iterative optimization steps required by REML solvers. For large N, this translates to a constant-factor speedup even though the asymptotic scaling remains cubic.

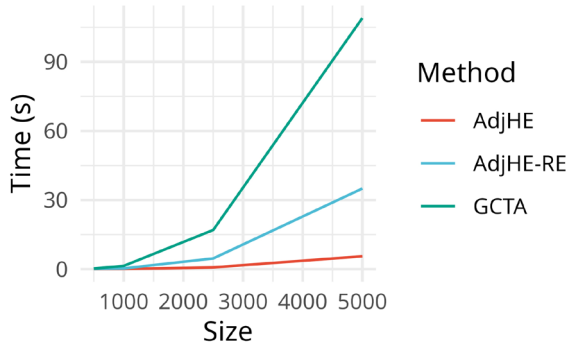


Fig. 3. Computational wall time as a function of increasing sample sizes for AdjHE estimation (red), AdjHE-RE (blue), and GCTA (green).

3. APPLICATIONS

We compare results on subcortical volumetric data derived from brain imaging data provided in the ABCD study.

3.1. ABCD data and processing

The ABCD study comprises $N = 11{,}572$ 9–10-year olds. ABCD is (parentheses describe the range across sites) 48% female (range 43–61%), 52% White (range 8–87%), 15% Black (range 0–54%), 2% Asian (range 0–13%), 20% Hispanic (range 2.3–74%), and 11% Other (range 3.6–23%) (see Supplementary Table S2).

3.1.1. *Genomics data*

All subjects were assayed using the Smokescreen™ Genotyping array (Baurley et al., 2016), which includes more than 300,000 SNPs. Genotypes were phased and imputed on the Michigan imputation server (Das et al., 2016). Automated imputation QC steps were used on the Michigan server, and additional QC steps were implemented on the imputed data using Plink (v1.90b6.24, (Purcell et al., 2007)), including filtering the minor allele frequency to 5%, removing samples with heterozygosity beyond 3 SDs, selecting only single base-pair replacement mutations, applying a Hardy–Weinberg equilibrium threshold of 1×10^{-6}, enforcing sliding-window independence (200 bp window size with 50 bp steps) of 0.25, and requiring imputation $R^2 > 0.99$. GRMs and eigendecompositions were then computed using Plink. The full dataset was filtered to remove both explicit and hidden family structures. After QC, the final sample size was $N = 7{,}234$ distributed across 21 sites, and the number of SNPs was 1,292,075.

3.1.2. *Image acquisition, preprocessing, and parcellation*

Data acquisition is fully described on the ABCD website (Casey et al., 2018). DCAN Labs ABCD-BIDS Community Collection (ABCC) MRI data (nda-3165) were accessed

through the ABCD 6.0 tabulated data release on the Deapscience portal. The T1-weighted 3D MP-RAGE sequence had 1 mm isotropic spatial resolution, TR/TE/TI = 2500 / 2.88 / 1060, and an 8° flip angle. Parcellations were carried out following the DCAN ABCD-BIDS pipeline (Feczko et al., 2021). Anatomical data are registered to an MNI template, masked, denoised, and corrected for bias. Subcortical segmentations were generated using the HCP-like FreeSurfer v5.3.0 pipeline using the aseg atlas (Feczko et al., 2021; Fischl, 2012). The ABCD release provides 45 subcortical measures, of which 6 are white matter hypointensity volumes that quantify lesion load rather than tissue volume and were, therefore, excluded, leaving the 38 regional subcortical volumes analyzed here. The phenotype file contains the subject identifiers (*src_subject_id*,*eventname*) along with subcortical volumes (*smri_t1w_scs_**) derived from the aseg segmentation atlas, but renamed to fit ABCD data curation standards. The imaging data were used as is with no additional QC metrics applied. Demographic (*ab_g_dyn__visit_age*, *ab_g_stc__cohort_sex*), scanner, and site covariates (*ab_g_dyn__design_site*) (also identified utilizing the Deapscience portal) were compiled from site membership data. For replication, the structural variables are distributed as tabulated columns prefixed *smri_t1w_scs_* (for example, *smri_t1w_scs_tplh*, *smri_t1w_scs_hpusrh*, and *smri_t1w_scs_putamenlh*), which map one-to-one onto the FreeSurfer aseg labels listed in Supplementary Table S4. Genotypes were drawn from the Smokescreen array and imputed on the Michigan imputation server, then filtered on minor allele frequency, heterozygosity, Hardy–Weinberg equilibrium, imputation accuracy, and sliding-window linkage disequilibrium as described above; these steps are applied to the full sample before relatives are removed, so the resulting unrelated sample does not depend on family ascertainment. Imaging quality control was performed entirely within the upstream ABCD curation pipeline, and we applied no additional subject-level imaging exclusions beyond requiring a valid aseg segmentation for the phenotype in question.

3.1.3. Heritability estimates

Estimators were compared on subcortical volumes derived from the full ABCD dataset ($n = 7{,}234$), including subjects of all genetic ancestries at all sites. Each estimate controlled for age, sex, 20 PCs, household income, parent education status, and 21 nonredundant site indicators, with 1 site serving as the reference level. Heritability estimates of subcortical volumes did not control for global head volume. Twin estimates of additive genetic variance (A) were modeled along with common environment (C), and individual environment (E) effects in an ACE model using the twinlm function from the mets package (version 1.3.12) in R. All ACE models adjusted for the same covariates included in our SNP-heritability specifications except genetic PCs. Additive genetic variance was then directly compared with SNP heritability (i.e., $A = h^2$) (Scheike & Holst, 2022). The ABCD study is enriched in same-sex twins (772 pairs) recruited from birth registries from four sites (Garavan et al., 2018; Maes et al., 2023). For a complete cases analysis, a twin pair is only kept when both twins have non-missing data further restricting the sample size. Here we retain 78 MZ and 107 DZ pairs for analysis. Since h^2_{twin} and h^2_{SNP} are estimated in samples that differ in size and demographic composition, special consideration during comparison is warranted. Here, we compare the rank order and relative magnitude between SNP and twin estimates. ACE model twin estimates are provided in Supplementary Table S5.

Estimates carried out for subjects of European ancestry across all sites (n = 3669) across all phenotypes are shown in Supplementary Figure S4. European heritage was chosen as it provided the largest sample size that yielded an average of 280 observations at each site with a minimum of 16 (see Supplementary Table S2). Since this subset reduced the number of reported ancestries present in the data, the estimates were carried out with 0, 5, and 10 PCs. The estimates were relatively stable across the number of PCs for all estimators. In general, nAdjHE and AdjHE-RE methods showed the largest heritability with other estimators falling lower generally between 5% and 15% heritability.

Estimates were carried out for site 16 ($n = 622$) across all reported ancestries and phenotypes. Site 16 was chosen because it had the largest sample size; the mean site sample size was approximately 329 with a standard deviation of 162. At site 16, the distribution of ancestries was 81% White, 12% Hispanic, 5.8% Other, 1.1% Black, and 0.2% Asian subjects (see Supplementary Table S2). Estimates were compared across a range of PCs (0, 10, and 20) to assess the impact of population stratification on heritability estimates. These estimates are shown in Supplementary Figure S3. Since these estimates were based on a single site, we did not perform the meta-analysis or apply the AdjHE-RE approach. GCTA estimates were similar to AdjHE estimates, but the Method (a) and Method (b) estimators were consistently higher showing some evidence of bias. The highest heritability estimate was for the cerebellar cortex.

Estimates for the full dataset (n= 7,234) across all phenotypes are shown on the brain in Figure 4. Since some regions are not possible to visualize within the standard central slices, estimates are also displayed with a scatter plot in Supplementary Figure S5 to convey the

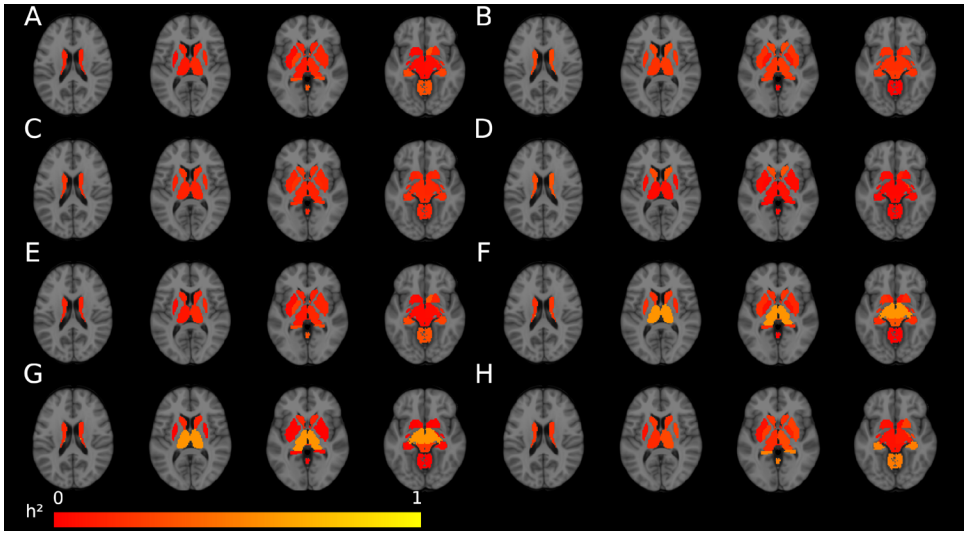


Fig. 4. Heritability on the subcortical volumes from the full ABCD dataset visualized on the Conte T1 averaged brain. Each panel comprises four medial slices from superior to inferior, taken at 10 mm intervals (z = 20, 10, 0, and -10). Heritability is reported as a heat map, with red representing low heritability and yellow representing high heritability. From left to right and top to bottom, each four image panel represents estimates from (A) AdjHE with fixed site effects, (B) GCTA with fixed site effects, (C) AdjHE meta-analysis, (D) GCTA meta-analysis, (E) SWD, (F) ComBat, (G) CovBat, and (H) AdjHE estimating site as a random effect.

magnitude of these estimates. The AdjHE methods estimate a larger heritability and the meta and ad hoc methods fall in between. Most estimates fell below 20% heritability with some (hippocampus and cerebellum cortex) being near or above 50%. Heritability estimates on the twin subset (78 MZ and 107 DZ pairs) yielded the following correlations: AdjHE, 0.37, Combat = 0.34, AdjHE-RE= 0.45, GCTA= 0.18, nAdjHE = -0.19, nGCTA = -0.32, SWD= 0.37. These correlations compare the SNP-based estimates computed in the full ABCD sample against twin heritabilities estimated in the smaller, demographically distinct twin subsample described in Section 3.1.3, and should be read as a descriptive comparison of regional rankings rather than as a formal validation. ACE estimates with standard errors and 95% confidence intervals are reported in Supplementary Table S5. As a positive control, we estimated the heritability of global head volume (0.39 SE 0.01 for AdjHE-RE, 0.15 SE 0.01 for GCTA).

4. DISCUSSION

Overall, this study shows that AdjHE-RE can provide unbiased and consistent SNP-heritability estimates under the proposed random-site model for both balanced and unbalanced multi-site designs, including settings with simulated ancestry heterogeneity. We also identified multiple rSBVs with appreciable heritability in adolescents, highlighting stable patterns of genetic contribution to subcortical brain morphology. Together with its computational efficiency, this makes AdjHE-RE well suited to phenome-wide heritability analyses in large neuroimaging cohorts (Fig. 3). AdjHE-RE, as well as a wrapper for all of the other methods compared here, is available publicly via the MASH toolbox on GitHub. The toolbox offers a command line tool for batch scripting as well as a Python interface for interactive computations.

Our work makes two contributions to imaging genetics. First, we extend AdjHE to accommodate site effects through a random-site variance component, yielding a statistically consistent estimator under the working model. This contribution is methodological and applies to any continuous phenotype measured in a multi-site cohort; the subcortical application below is a demonstration on a challenging dataset rather than a claim about subcortical neuroanatomy. Theoretical arguments and simulations support the usefulness of the estimator across a wide range of settings, including varying numbers of sites and ancestry clusters, although the current comparisons remain limited to linear transformations. AdjHE-RE remained consistent even under non-uniform site error with imbalanced sites for 1–2 population clusters and 2–25 sites (Fig. 1), and it was up to 4 times faster than REML estimation (Fig. 3).

The generalizability of AdjHE-RE beyond subcortical structures is of particular interest for multi-site neuroimaging consortia. By modeling site as a covariance component, our method can be directly applied to higher-dimensional phenotypes, such as voxel-wise structural maps or functional connectivity networks from fMRI

efficiently and without introducing a large number of modeled degrees of freedom. In these settings, site-specific scanner and acquisition protocol disparities are especially pronounced. Applying AdjHE-RE to functional and cortical morphology datasets will help characterize the genetic contribution to IDPs across neuroimaging modalities.

Second, we estimated heritability using AdjHE-RE for rSBVs in ABCD and found multiple highly heritable subcortical regions (Fig. 4). For example, the cerebellar cortex had an estimated heritability of around 60%, and the hippocampus had an estimated heritability of about 50%. SNP estimates closely track with estimates from twin studies. Highly heritable structures in our analysis—such as the hippocampus ($h^2 \approx 0.50$), amygdala ($h^2 \approx 0.40$), and thalamus ($h^2 \approx 0.20$)—agree closely with results seen in a Dutch cohort (Swagerman et al., 2014) ($h^2_{twin} = 0.55 - 0.72$ for hippocampus at 9 and 12 years old, respectively) that has been replicated (den Braber et al., 2013). Structures with lower or more variable twin estimates, such as the nucleus accumbens (h^2_{twin} 0.09 – 0.61), showed lower SNP heritability here ($h^2 \approx 0.12$). AdjHE-RE SNP estimates, therefore, reflect the heritability gap between common-variant and twin models and also preserves trends in genetic associations among subcortical structures.

These estimates also share similarities with previous findings from PING (n = 500, age range = 3–21 years) and UKBiobank (n = 9000, age range = 40–69 years) despite very large differences in sample demographics (mainly age and ethnicities) (B. Zhao et al., 2019). The largest estimates reported here, which were 60% for the cerebellar cortex and 50% for the hippocampus, largely agree with UKBiobank SNP-heritability estimates (63% and 50%, respectively) (B. Zhao et al., 2019). Additionally, the cerebellum has a major role in motor function and is thus understood to scale with height, a trait that has a well-established heritability above 60% (Dandine-Roulland et al., 2016). The estimated heritability for amygdala volume is 35% compared with UKBiobank, which was 23–42% (B. Zhao et al., 2019). GWAS estimates on the UKBiobank set a range of 20–60% heritability for T1- and T2-weighted subcortical structures (Elliott et al., 2018). Some weakly heritable traits were much lower than in previous reports. For example, estimates on the brain stem were 10% as compared with 82% from UKBiobank studies (B. Zhao et al., 2019). Differences might arise from age differences, given that UKBiobank comprises adults aged 40–69 years old, compared with the ABCD sample that is 9–10 years old. The dramatic differences in brain stem estimates are additionally affected by "midbrain shrinkage" that would be observed in UKBiobank and not in ABCD (Lambert et al., 2013; Lindquist, 2012).

In addition, we compared our SNP heritability estimates with the twin heritability estimates (h^2_{twin}) using 78 MZ and 107 DZ pairs in our sample. Twin estimates had a mean h^2 of 0.55 (IQR: 0.43 – 0.71, range: 0 – 0.92) for all of our traits, and the highest heritabilities were observed for the posterior and anterior corpus callosum, putamen, and caudate, with the hippocampus and amygdala also among the more heritable structures (Supplementary Table S5 and Supplementary Fig. S6). The AdjHE-RE approach showed the highest concordance with the twin heritability, reaching a Pearson correlation of $r = 0.45$, whereas AdjHE, SWD, and ComBat showed $r = 0.34 - 0.37$. GCTA showed weaker concordance ($r = 0.18$), while the meta-analytic approaches (nAdjHE and nGCTA) were negatively correlated ($r = -0.19$ and –0.32).

The differences observed between GCTA and AdjHE-RE in the ABCD analysis should be interpreted in the context of the study's complex multi-site and multi-ancestry design. Although simulations showed broadly similar central estimates across methods under simpler settings, the ABCD data include multiple ancestral populations that are differentially distributed across sites, as well as additional scanner- and protocol-related sources of heterogeneity. In this setting, treating site as a fixed effect may over-adjust the phenotype by removing variation that is correlated with both site and ancestry, thereby attenuating the estimated genetic contribution. By modeling site as a random variance component, AdjHE-RE may better preserve ancestry-correlated genetic signal while accounting for site-related variability, which likely explains the larger heritability estimates observed for some subcortical regions.

AdjHE-RE has methodological limitations. The proposed approach is based on a method-of-moments framework, which substantially improves computational efficiency relative to REML-based variance component estimation; however, this gain in speed may come at the cost of larger standard errors for the resulting heritability estimates. In addition, treating site effects as random effects is a simplification of scanner- and protocol-related variation, and this formulation may not capture every source of cross-site heterogeneity. In addition, the current implementation does not estimate nested confounding structures, such as scanner effects within site, which may contribute additional variability to the heritability estimates. These limitations suggest that the present method is an important step forward, but not the final word on site adjustment in multi-site imaging genetics.

Several study limitations warrant careful interpretation of these results. First, the twin subsample was recruited via a distinct mechanism across only four sites, reducing racial, socioeconomic, and geographical diversity relative

to the full sample. Consequently, these estimates derive from subsamples that differ systematically across relevant phenotypes. Second, the precision of the heritability estimates is limited; across five regions, the 95% confidence interval for h^2_{twin} spans zero (Supplementary Table S5). Third, as a secondary analysis of observational data, these findings are strictly exploratory. Fourth, the ABCD cohort is restricted to 9–10-year olds living near specialized research sites, limiting generalizability to the broader population and particularly under-representing rural communities.

As multi-site studies continue to grow, robust control of site effects will become increasingly important. Future work will extend AdjHE-RE to multivariate settings, which are already showing promise in neuroimaging studies (Ge et al., 2016), and will develop a simulation toolbox for standardized evaluation of new imaging-genetics methods that must account for site effects. Although we focused on regional subcortical brain volumes, AdjHE-RE is not limited to these phenotypes. The method provides a general variance-component framework for SNP-heritability estimation in large multi-site and multi-ancestry datasets. The same approach can be used for other imaging-derived phenotypes, including cortical thickness, surface area, cortical volume, diffusion MRI measures, and functional MRI-derived connectivity or activation traits. Moreover, the site covariance matrix considered here can be generalized to other structured covariance matrices defined by scanner manufacturer, acquisition protocol, field strength, or other study-design features. Thus, the ABCD application illustrates a broader methodological tool for scalable heritability estimation in modern imaging genetics studies.

DATA AND CODE AVAILABILITY

The ABCD imaging dataset, from fast track portal 2.0 release, used in this study is publicly available through abcdstudy.org. ABCD genomics data were obtained from the fourth genomics release version. Follow this link for directions on acquiring rights to access ABCD data. Data processing scripts are available at GDCGenomicsQC and analysis functions are also available. The MASH toolbox, which implements AdjHE-RE and all comparison methods, is available at https://github.com/umn-gdc/MASH

AUTHOR CONTRIBUTIONS

C.C. created the estimation methodology, developed the software and visualizations, wrote the original draft, and revised for later submissions. E.F. supervised the application to real-world data and aided in writing the original draft. B.L. and B.T.C. aided in writing the original draft. G.C. and L.A.M. aided in standardizing software development and reviewed the software. J.T.L. aided in data curation. A.H., L.A.M., K.W., R.M., A.J.P., B.F., T.J.M., E.E., O.M.D. aided in writing, editing, and data processing. D.A.F. aided in funding acquisition, and review and editing. S.B. aided in conceptualization, acquiring funding, supervising, method development, and all aspects of the writing process.

FUNDING

This work was supported by the National Institute of General Medical Sciences (T32GM132063; support for C.C.), the National Institute on Drug Abuse (R25DA061823), and the University of Minnesota's Masonic Institute for the Developing Brain (MIDB project M1015; support for University of Minnesota Genomic Data Commons infrastructure).

DECLARATION OF COMPETING INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

We thank the Adolescent Brain Cognitive Development (ABCD) Consortium for the design, recruitment, and collection of the study data, and the DCAN Labs at the University of Minnesota for processing the neuroimaging data and providing access to the subcortical segmentations used here. We also thank the participants and their families, whose contributions made this work possible. This work was supported by the National Institute of Mental Health.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: https://doi.org/10.1162/IMAG.a.1455#supplementary-data

REFERENCES

Balding, D. J., & Nichols, R. A. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica*, *96*(1–2), 3–12. https://doi.org/10.1007/bf01441146

Baurley, J., Edlund, C., Pardamena, C., Conti, D., & Bergen, A. (2016). Smokescreen: A targeted genotyping array for addiction research. *BMC Genomics*, *17*, 145. https://doi.org/10.1186/s12864-016-2495-7

Bayer, J. M. M., Thompson, P. M., Ching, C. R. K., Liu, M., Chen, A., Panzenhagen, A. C., Jahanshad, N., Marquand, A., Schmaal, L., & Sämann, P. G. (2022). Site effects how-to and when: An overview of retrospective

techniques to accommodate site effects in multi-site neuroimaging analyses. *Frontiers in Neurology*, *13*, 923988. https://doi.org/10.3389/fneur.2022.923988

Bulik-Sullivan, B. K., Loh, P. R., & Finucane, H. K. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211

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. *Developmental Cognitive Neuroscience*, *32*, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Conomos, M. P., Reiner, A. P., Weir, B. S., & Thornton, T. A. (2016). Model-free estimation of recent genetic relatedness. *The American Journal of Human Genetics*, *98*(1), 127–148. https://doi.org/10.1016/j.ajhg.2015.11.022

Dandine-Roulland, C., Bellenguez, C., Debette, S., Amouyel, P., Génin, E., & Perdry, H. (2016). Accuracy of heritability estimations in presence of hidden population stratification. *Scientific Reports*, *6*, 26471. https://doi.org/10.1038/srep26471

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P. R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., … Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature Genetics*, *48*(10), 1284–1287. https://doi.org/10.1038/ng.3656

den Braber, A., Bohlken, M. M., Brouwer, R. M., van 't Ent, D., Kanai, R., Kahn, R. S., de Geus, E. J. C., Hulshoff Pol, H. E., & Boomsma, D. I. (2013). Heritability of subcortical brain measures: A perspective for future genome-wide association studies. *NeuroImage*, *83*, 98–102. https://doi.org/10.1016/j.neuroimage.2013.06.027

Elliott, L. T., Sharp, K., Alfaro-Almagro, F., Shi, S., Miller, K. L., Douaud, G., Marchini, J., & Smith, S. M. (2018). Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature*, *562*(7726), 210–216. https://doi.org/10.1038/s41586-018-0571-7

Fan, C. C., Palmer, C. E., Iversen, J. R., Pecheva, D., Holland, D., Frei, O., Thompson, W. K., Hagler, D. J., Jr., Andreassen, O. A., Jernigan, T. L., Nichols, T. E., & Dale, A. M. (2022). FEMA: Fast and efficient mixed-effects algorithm for population-scale whole-brain imaging data. *bioRxiv*. https://doi.org/10.1101/2021.10.27.466202

Feczko, E., Conan, G., Marek, S., Tervo-Clemmens, B., Cordova, M., Doyle, O., Earl, E., Perrone, A., Sturgeon, D., Klein, R., Harman, G., Kilamovich, D., Hermosillo, R., Miranda-Dominguez, O., Adebimpe, A., Bertolero, M., Cieslak, M., Covitz, S., Hendrickson, T., … Fair, D. A. (2021). Adolescent Brain Cognitive Development (ABCD) community MRI collection and utilities. *bioRxiv*. https://doi.org/10.1101/2021.07.09.451638

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Fortin, J. P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., McInnis, M., Phillips, M. L., Trivedi, M. H., Weissman, M. M., & Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, *167*, 104–120. https://doi.org/10.1016/j.neuroimage.2017.11.024

Fortin, J. P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C., Gur, R. E., Schultz, R. T., Verma, R., & Shinohara, R. T. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, *161*, 149–170. https://doi.org/10.1016/j.neuroimage.2017.08.047

Friedman, L., Stern, H., Brown, G. G., Mathalon, D. H., Turner, J., Glover, G. H., Gollub, R. L., Lauriello, J., Lim, K. O., Cannon, T., Greve, D. N., Bockholt, H. J., Belger, A., Mueller, B., Doty, M. J., He, J., Wells, W., Smyth, P., Pieper, S., … Potkin, S. G. (2008). Test-retest and between-site reliability in a multicenter fMRI study. *Human Brain Mapping*, *29*(8), 958–972. https://doi.org/10.1002/hbm.20440

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. *Developmental Cognitive Neuroscience*, *32*, 16–22. https://doi.org/10.1016/j.dcn.2018.04.004

Ge, T., Chen, C. Y., Neale, B. M., Sabuncu, M. R., & Smoller, J. W. (2017). Phenome-wide heritability analysis of the UK Biobank. *PLoS Genetics*, *13*(7), e1006711. https://doi.org/10.1371/journal.pgen.1006711

Ge, T., Reuter, M., Winkler, A. M., Holmes, A. J., Lee, P. H., Tirrell, L. S., Roffman, J. L., Buckner, R. L., Smoller, J. W., & Sabuncu, M. R. (2016). Multidimensional heritability analysis of neuroanatomical shape. *Nature Communications*, *7*, 1329. https://doi.org/10.1038/ncomms13291

Haseman, J. K., & Elston, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics*, *2*(1), 3–19. https://doi.org/10.1007/bf01066731

Hibar, D. P., Stein, J. L., Renteria, M. E., Arias-Vasquez, A., Desrivières, S., Jahanshad, N., Toro, R., Wittfeld, K., Abramovic, L., Andersson, M., Aribisala, B. S., Armstrong, N. J., Bernard, M., Bohlken, M. M., Boks, M. P., Bralten, J., Brown, A. A., Chakravarty, M. M., Chen, Q., … Medland, S. E. (2015). Common genetic variants influence human subcortical brain structures. *Nature*, *520*(7546), 224–229. https://doi.org/10.1038/nature14101

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B., & Fischl, B. (2009). MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *NeuroImage*, *46*(1), 177–192. https://doi.org/10.1016/j.neuroimage.2009.02.010

Kelley, G. A., & Kelley, K. S. (2012). Statistical models for meta-analysis: A brief tutorial. *World Journal of Methodology*, *2*(4), 27–32. https://doi.org/10.5662/wjm.v2.i4.27

Kochunov, P., Jahanshad, N., Sprooten, E., Nichols, T. E., Mandl, R. C., Almasy, L., Booth, T., Brouwer, R. M., Curran, J. E., de Zubicaray, G. I., Dimitrova, R., Duggirala, R., Fox, P. T., Hong, L. E., Landman, B. A., Lemaitre, H., Lopez, L. M., Martin, N. G., McMahon,

K. L., … Glahn, D. C. (2014). Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: Comparing meta and megaanalytical approaches for data pooling. *NeuroImage*, *95*, 136–150. https://doi.org/10.1016/j.neuroimage.2014.03.033

Lambert, C., Chowdhury, R., Fitzgerald, T., Fleming, S., Lutti, A., Hutton, C., Draganski, B., Frackowiak, R., & Ashburner, J. (2013). Characterizing aging in the human brainstem using quantitative multimodal MRI analysis. *Frontiers in Human Neuroscience*, *7*, 462. https://doi.org/10.3389/fnhum.2013.00462

Lin, Z., Seal, S., & Basu, S. (2022). Estimating SNP heritability in presence of population substructure in biobank-scale datasets. *Genetics*, *220*(4), iyac015. https://doi.org/10.1093/genetics/iyac015

Lindquist, M. A. (2012). Functional causal mediation analysis with an application to brain connectivity. *Journal of the American Statistical Association*, *107*(500), 1297–1309. https://doi.org/10.1080/01621459.2012.695640

Loh, P. R., Tucker, G., Bulik-Sullivan, B. K., Vilhjálmsson, B. J., Finucane, H. K., Salem, R. M., Chasman, D. I., Ridker, P. M., Neale, B. M., Berger, B., Patterson, N., & Price, A. L. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics*, *47*(3), 284–290. https://doi.org/10.1038/ng.3190

Maes, H. H. M., Lapato, D. M., Schmitt, J. E., Luciana, M., Banich, M. T., Bjork, J. M., Hewitt, J. B., Madden, P. A., Heath, A. C., Barch, D. M., Thompson, W. K., Iacono, W. G., & Neale, M. C. (2023). Genetic and environmental variation in continuous phenotypes in the ABCD study. *Behavior Genetics*, *53*(1), 1–24. https://doi.org/10.1007/s10519-022-10123-w

Matthews, L. J., & Turkheimer, E. (2022). Three legs of the missing heritability problem. *Studies in History and Philosophy of Science*, *93*, 183–191. https://doi.org/10.1016/j.shpsa.2022.04.004

Pinto, M. S., Paolella, R., Billiet, T., Van Dyck, P., Guns, P. J., Jeurissen, B., Ribbens, A., den Dekker, A. J., & Sijbers, J. (2020). Harmonization of brain diffusion MRI: Concepts and methods. *Frontiers in Neuroscience*, *14*, 396. https://doi.org/10.3389/fnins.2020.00396

Pomponio, R., Erus, G., Habes, M., Doshi, J., Srinivasan, D., Mamourian, E., Bashyam, V., Nasrallah, I. M., Satterthwaite, T. D., Fan, Y., Launer, L. J., Masters, C. L., Maruff, P., Zhuo, C., Völzke, H., Johnson, S. C., Fripp, J., Koutsouleris, N., Wolf, D. H., … Davatzikos, C. (2020). Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage*, *208*, 116450. https://doi.org/10.1016/j.neuroimage.2019.116450

Purcell, S., Neale, B. M., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Reardon, A. M., Li, K., & Hu, X. P. (2021). Improving between-group effect size for multi-site functional connectivity data via site-wise de-meaning. *Frontiers in Computational Neuroscience*, *15*, 762781. https://doi.org/10.3389/fncom.2021.762781

Rudelson, M., & Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electronic Communications in Probability*, *18*, 1–9. https://doi.org/10.1214/ecp.v18-2865

Scheike, T. H., & Holst, K. K. (2022). A practical guide to family studies with lifetime data. *Annual Review of Statistics and Its Application*, *9*, 47–69. https://doi.org/10.1146/annurev-statistics-040120-024253

Smith, D. M., Loughnan, R., Friedman, N. P., Parekh, P., Frei, O., Thompson, W. K., Andreassen, O. A., Neale, M. C., Jernigan, T. L., & Dale, A. M. (2023). Heritability estimation of cognitive phenotypes in the ABCD study using mixed models. *Behavior Genetics*, *53*(3), 169–188. https://doi.org/10.1007/s10519-023-10141-2

Stefanski, L., & Boos, D. (2002). The calculus of M-estimation. *The American Statistician*, *56*(1), 29–38. https://doi.org/10.1198/000313002753631330

Swagerman, S. C., Brouwer, R. M., de Geus, E. J. C., Hulshoff Pol, H. E., & Boomsma, D. I. (2014). Development and heritability of subcortical brain volumes at ages 9 and 12. *Genes, Brain and Behavior*, *13*(8), 733–742. https://doi.org/10.1111/gbb.12182

Uban, K. A., Horton, M. K., Jacobus, J., Heyser, C., Thompson, W. K., Tapert, S. F., Madden, P. A. F., & Sowell, E. R. (2018). Biospecimens and the ABCD Study: Rationale, methods of collection, measurement and early data. *Developmental Cognitive Neuroscience*, *32*, 97–106. https://doi.org/10.1016/j.dcn.2018.03.005

Wen, W., Thalamuthu, A., Mather, K. A., Zhu, W., Jiang, J., de Micheaux, P. L., Wright, M. J., Ames, D., & Sachdev, P. S. (2016). Distinct genetic influences on cortical and subcortical brain structures. *Scientific Reports*, *6*, 32760. https://doi.org/10.1038/srep32760

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zhao, B., Ibrahim, J. G., Li, Y., Li, T., Wang, Y., Shan, Y., Zhu, Z., Zhou, F., Zhang, J., Huang, C., Liao, H., Yang, L., Thompson, P. M., & Zhu, H. (2019). Heritability of regional brain volumes in large-scale neuroimaging and genetic studies. *Cerebral Cortex*, *29*(7), 2904–2914. https://doi.org/10.1093/cercor/bhy157

Zhao, Y., Li, T., & Zhu, H. (2020). Bayesian sparse heritability analysis with high-dimensional neuroimaging phenotypes. *Biostatistics*, *23*(2), 467–484. https://doi.org/10.1093/biostatistics/kxaa035