Population-weighted Image-on-scalar Regression Analyses of Large Scale Neuroimaging Data

Zikai Lin[a], M. Fiona Molloy[b], Chandra Sripada[c], Jian Kang[d], and Yajuan Si[e,1]

[a]Department of Biostatistics, University of Michigan, Ann Arbor, 1415 Washington Heights, Ann Arbor, MI 48109, USA,
Email: zikai@umich.edu
[b]Department of Psychiatry, University of Michigan, Ann Arbor, 4250 Plymouth Rd, Ann Arbor, MI 48104, USA,
Email: marymoll@umich.edu
[c]Departments of Psychiatry and Philosophy, University of Michigan, Ann Arbor, 4250 Plymouth Rd, Ann Arbor, MI 48104, USA,
Email: sripada@umich.edu
[d]Department of Biostatistics, University of Michigan, Ann Arbor, 1415 Washington Heights, Ann Arbor, MI 48109, USA,
Email: jiankang@umich.edu
[e]Institute for Social Research, University of Michigan, Ann Arbor, ISR 4014, 426 Thompson St, Ann Arbor, MI 48104, USA,
Email: yajuan@umich.edu
[1]**Corresponding author**, Institute for Social Research, University of Michigan, Ann Arbor, ISR 4014, 426 Thompson St, Ann Arbor, MI 48104, USA,
Email: yajuan@umich.edu

Abstract

Recent advances in neuroimaging modeling highlight the importance of accounting for subgroup heterogeneity in population-based neuroscience research through various investigations in large scale neuroimaging data collection. To integrate survey methodology with neuroscience research, we present an imaging data analysis aiming to achieve population generalizability with screened subsets of data. The Adolescent Brain Cognitive Development (ABCD) Study has enrolled a large cohort of participants to reflect the individual variation of the U.S. population in adolescent development. To ensure population representation, the ABCD Study has released the base weights. We estimated the associations between brain activities and cognitive performance using the functional Magnetic Resonance Imaging (fMRI) data from the ABCD Study's n-back working memory task. Notably, the imaging subsample exhibits differences from the baseline cohort in key child characteristics, and such discrepancies cannot be addressed simply by applying the ABCD base weights. We developed new population weights specific to the subsample and included the adjusted weights in the image-on-scalar regression model. We validated the approach through synthetic simulations and applications to fMRI data from the ABCD Study. Our findings indicate that population weighting adjustments influence association estimates between brain activities and cognition, emphasizing the importance of evaluating validity and generalizability in population neuroscience research.

Keywords: Population neuroscience, functional Magnetic Resonance Imaging (fMRI), image-on-scalar regression, weighting, active area detection.

1 Introduction

Neuroimaging research links brain activity to personal traits to explore how biological, social, and environmental factors influence brain functions. These influences could present inter-individual variation, where associations between cognitive functioning and brain activity measures may differ across demographic groups or be moderated by environmental and risk factors. Capturing this variability requires analyzing individual-level data from a broad population (Dotson and Duarte, 2020). Prior research shows that sample composition alters associations between age and brain structure measures of gray matter volume, thickness, and surface area (LeWinn et al., 2017). When the effect modifiers are unrepresented in a sample, the resulting estimates are biased. The emerging field of population neuroscience (Falk et al., 2013; Garavan et al., 2018; Si et al., 2024; Han et al., 2025) was created specifically to address this, emphasizing the role of mediation and moderation factors and advocating for analysis beyond aggregate summaries, and achieve generalizability to ensuring that neuroimaging findings are meaningful reflections of the general population.

However, generalizability in neuroscience is challenging due to the field's inherent complexities. Collecting neuroimaging data that accurately reflect a specific target population remains difficult, and most studies rely on convenience samples (e.g., college student volunteers, single-site clinical cohorts), which are non-representative due to intensive study protocols, limited geographical distributions of study sites, and other logistical barriers. Furthermore, the small effect sizes often require large sample sizes that most studies cannot afford. Data quality is susceptible to technical variance from different scanners and acquisition protocols. Most critically, the nonergodic nature of the brain means that group-level averages often fail to apply to any single person. To overcome these logistical, statistical, and biological limitations, large-scale, multi-site studies with diverse cohorts are essential.

The Adolescent Brain Cognitive Development (ABCD) Study is a unique resource addressing this problem. As the largest long-term study of child neurodevelopment and health, the ABCD has collected a demographically broad sample that allows researchers to potentially disentangle generalizable effects from sampling bias, thereby advancing reliable findings in developmental neuroscience (Garavan et al., 2018).

The ABCD Study has enrolled 11,875 children aged 9-10 from multiple sites and collected their functional magnetic resonance imaging (fMRI) data to characterize child brain development, one objective of which is to examine associations between cognitive and psychological functioning and brain activities (Hagler Jr et al., 2019). Each of the ABCD Study sites has constrained catchment areas, where eligible children must live within reasonable travel distance (e.g., 50 miles) of a major medical center or research facility where fMRI imaging could be performed. Within school districts located in the catchment areas, a probability sample of the public and private schools was selected as the basis for the recruitment of the majority of eligible children (3rd and 4th grade students as modal grade levels of 9/10 year olds) to the ABCD baseline cohort. All the children in the targeted age range in selected schools were provided with recruitment materials to take home to their families. The information directed families to the research site. The recruitment was designed to randomize potential confounders that affected sample enrollment and child development outcomes and maintained a national quota of demographic targets based on age, sex, and race/ethnicity (Garavan et al., 2018). However, the process of site selection and obtaining school cooperation and then parental consent could selectively impact the final characteristics of the sample that was actually observed (Gard et al., 2023).

As a corrective post-collection adjustment, the study constructs population weights to match the ABCD baseline cohort composition with the targeted population using an

external, highly accurate source as a representative benchmark for the population of interest (Heeringa and Berglund, 2020). The ABCD base weights are positive values assigned to each enrolled child and indicate their relative representation, i.e., the number of children with identical characteristics, in the population. Various risk factors may affect the connection between brain and cognitive abilities (Casey et al., 2018; Sternberg and Grigorenko, 2002; Sripada et al., 2022), some of which are related to the ABCD sample selection and thus the weights. Without accounting for modifiers that affect selection into the study sample, the estimated exposure effects based on the sample will be biased and cannot be generalized to the population (Rudolph et al., 2023). Due to effect modification, overall exposure effects require adjustment to account for the fact that the distribution of these modifiers in the population may be quite different in the sample, such as weighting adjustment (Cole and Stuart, 2010). While these weights are primarily recommended for estimating descriptive characteristics (e.g., means, proportions), their effect on complex model estimates is non-trivial, depending on model specifications and the inclusion of relevant confounding variables (Si et al., 2024, 2022; West et al., 2025). In this article, we apply weighting adjustment to balance the distributions of variables that affect both sample selection and analysis outcomes and assess whether population weighting adjustment would influence the mapping population-level relationships between brain activities and cognition measures.

We are interested in analyzing the working memory task because it provides critical insights into cognitive function and its neural underpinnings (Christophel et al., 2017). Processed imaging data were only available for 3,782 children, a smaller subsample compared to the baseline sample of 9,905 participants (excluding cases with missing item values of sociodemographic measures). The imaging subsample differs from the baseline sample

in child characteristics and may be subject to selection bias due to non-participation, lack of consent to fMRI data collection, or exclusion during data quality screening with non-compliant to protocols on collection or processing. Correcting for distributional imbalance is vital for population inferences. We could adjust for the subsample selection bias using the baseline sample characteristics, assuming no additional unmeasured confounders.

Image-on-scalar regression models are widely used to estimate voxel-wise associations between multidimensional brain imaging outcomes and scalar variables of interest, while accounting for the spatial dependency between voxels (Zhu et al., 2014; Yu et al., 2021; Zeng et al., 2022). These models enhance the detection of active regions compared to mass univariate analyses (Li et al., 2020; Zhang et al., 2020; Liu et al., 2017). However, most applications of the image-on-scalar regression model overlook selection bias, limiting generalizability. Previous work has found subgroup heterogeneity in the image-on-scalar associations, and the identified subgroups present different socio-demographic profiles, providing evidence that the associations are moderated by socio-demographics (Lin et al., 2024).

The presence of unmeasured variables that affect selection and outcomes will result in biased associations estimates, especially when these variables are modifiers. We hypothesize that this issue is critical in the ABCD Study, where the specific imaging subsample composition may alter associate estimates from image-on-scalar regression models. Therefore, we assert that the population weights should be accounted for in these association studies. To test this hypothesis and the effect of model specification, we adopt a comparative approach. We analyze two primary model types: one without adjustment variables (to gauge the raw impact of weighting) and one with adjustment variables (to account for measured confounders related to both selection and outcome). We hypothesize that the inclusion of population weights will have differentially influences on these two models.

Since it is impossible to verify the correctness of any observational model, our selection of control variables is guided by substantive domain expertise and the specification of a commonly used regression model. We compare weighted and unweighted estimates of both adjusted and unadjusted models. This comparative strategy allows us to explicitly assess the importance of population weighting across various analytical scenarios.

Here, we implemented inverse propensity score weighting methods based on selected baseline sample characteristics to construct the subsample-adjusted weights, as one nonresponse bias adjustment approach (Si et al., 2023, 2024). The products of the base weights and inverse propensity score weights were treated as the final weights in the population-weighted analysis with the imaging subsample. We developed statistical methods to include population weights in the image-on-scalar regression models and improve inferences and detection of active brain areas related to cognition. We validated the proposed methods with simulation studies generating synthetic data and assessed the results in the ABCD Study with cross-validation and substantive insights. Our ABCD data analysis revealed differences in several cognition-related brain regions when comparing weighted and unweighted image-on-scalar regression models.

2 Method

In this section, we describe the ABCD data and measures, new weighting construction for the imaging subsample, weighted image-on-scalar regression model estimation and inference, and simulation designs in detail.

2.1 Data and Measures

The ABCD Study aims to characterize the trajectories of child neurodevelopment and the factors influencing youth mental health and cognitive outcomes, and our analysis focuses on the connection between the brain activity and cognitive ability that may be influenced by various risk factors (Casey et al., 2018; Sternberg and Grigorenko, 2002).

Our imaging outcome variables were the 2-back and 0-back contrast maps from the fMRI data collected during a working memory task in the ABCD Study Curated Annual Release 2.0.1. The exposure variable of interest was the general cognition measure, g-factor, a composite score derived using confirmatory factor analysis of subjects' behavioral performance on six cognitive tasks, comprising *Picture Vocabulary*, *Oral Reading Recognition*, *Flanker*, *Pattern Comparison Processing*, *Picture Sequence Memory*, and *Rey Auditory Verbal Learning* (Akshoomoff et al., 2013), collected via an iPad during a 70-minute in-person session conducted outside the fMRI environment (Thompson et al., 2019). The cognitive ability assessed by the g-factor serves as a trait linked to activities observed during specific cognitive tasks, such as working memory (Waiter et al., 2009; Gray et al., 2003). Exploring the relationship between the g-factor and the differential brain activation elicited by the 2-back versus 0-back tasks may uncover neural regions crucial to general intelligence, particularly highlighting aspects of working memory's capacity and efficiency (Thompson et al., 2019; Lin et al., 2024). We included control variables: sex, race/ethnicity, household size, parental marital status, household income, and study sites.

Table 1: Comparison of sociodemographic distributions between the ABCD baseline sample (restricted to complete cases without any missing items) and the imaging subsample (UW: Unweighted, Base-W: Weighted with the base weights, Final-W: Weighted with the final subsample weights). Values are shown as counts (percentages) for unweighted columns and percentages for weighted columns.

Variable	**Baseline sample** (n = 9,905)		**Imaging subsample** (n = 3,782)		
	UW	Base-W	UW	Base-W	Final-W
Sex					
Female	4764 (48.1%)	49.1%	1925 (50.9%)	51.8%	49.4%
Male	5141 (51.9%)	50.9%	1857 (49.1%)	48.2%	50.6%
Age					
≥ 10	4699 (47.4%)	49.6%	2064 (54.6%)	56.8%	49.8%
< 10	5206 (52.6%)	50.4%	1718 (45.4%)	43.2%	50.2%
Race/Ethnicity					
Asian	208 (2.1%)	3.5%	74 (2.0%)	3.4%	3.6%
Black	1328 (13.4%)	12.1%	323 (8.5%)	7.9%	11.8%
Hispanic	1853 (18.7%)	22.0%	608 (16.1%)	19.8%	23.2%
Other	1040 (10.5%)	6.6%	362 (9.6%)	6.0%	6.8%
White	5476 (55.3%)	55.8%	2415 (63.9%)	62.9%	54.6%
Parental Marital Status					
No	2457 (24.8%)	31.4%	716 (18.9%)	25.9%	31.6%
Yes	7448 (75.2%)	68.6%	3066 (81.1%)	74.1%	68.4%
Household Income					
<50K	2815 (28.4%)	37.6%	788 (20.8%)	30.1%	37.9%
50–100K	2834 (28.6%)	31.5%	1154 (30.5%)	34.3%	31.4%
$\geq$100K	4256 (43.0%)	30.9%	1840 (48.7%)	35.6%	30.7%
Household Size					
> 3	8254 (83.3%)	81.7%	3249 (85.9%)	83.5%	80.8%
≤ 3	1651 (16.7%)	18.3%	533 (14.1%)	16.5%	19.2%

2.2 Descriptive Summaries and Subample Weights

Table 1 presents a comparison of the sociodemographic distributions between the ABCD baseline sample and imaging subsample in our analysis. We removed cases with missing items of the included variables, resulting in a baseline sample size of n =9,905. We included the unweighted and weighted summaries with the base weights and the final weights, respectively.

The ABCD base weight construction borrowed information from the American Community Survey (ACS), a large probability sample survey of U.S. households conducted annually by the Bureau of Census, as a benchmark for selected demographic and socio-economic characteristics of U.S. children ages 9 and 10 (Heeringa and Berglund, 2020). The baseline weight adjusted the under-coverage of the ABCD participants from some racial and ethnical groups and non-marital families with low household income. Applying the ABCD base weights to the baseline sample matches the sociodemographic distribution with the target population, and it is recommended to apply these weights in ABCD analyses (Gard et al., 2023; Li et al., 2025; Hawes et al., 2025). Some groups have small sample sizes, e.g., Asian, the specific estimates of which will not be stable.

As shown in Table 1, comparing to the baseline sample, the imaging subsample includes more female (50.9% vs. 48.1%), 10-year-old (54.6% vs. 47.4%), and White (63.8% vs. 55.3%) children from married families (81.1% vs. 75.2%) with more household members (Household size > 3: 85.9% vs. 83.3%) and higher household incomes (Household income $\geq$100K: 48.7% vs. 43.0%). As a result, the imaging subsample fails to retain participants with low socio-economic status.

One might anticipate that applying ABCD base weights to the imaging subsample would mitigate the discrepancy. However, Table 1 shows that this disparity persists despite the

application of ABCD base weights to the imaging subsample. Comparing to the weighted ABCD baseline samples, the weighted imaging subsample still over-represents 10-year-old (56.8% vs. 49.6%), White (62.9% vs. 55.8%), and children from married families (74.1% vs. 68.6%) with larger household sizes (Household size > 3: 83.5% vs. 81.7%) and higher household incomes (Household income $\geq$100K: 35.6% vs. 30.9%). Such discrepancy indicates the existence of sample selection bias in the ABCD imaging subsample, and cannot be addressed simply by applying the ABCD base weights.

To mitigate this discrepancy, we employ an inverse propensity score weighting approach, aimed at aligning the weighted imaging subsample distributions more closely with the base-weighted ABCD cohort, i.e., the targeted U.S. children population. We fit a multiple logistic regression model to the baseline data with the imaging subsample inclusion indicator as the binary outcome, which assigns a value of 1 to individuals in the imaging subsample and a value of 0 to the baseline participants without imaging data. The model covariates includes baseline measures: race/ethnicity, household income, household size, parental marital status, age, sex and the g-factor. We obtained the predicted inclusion probabilities of all baseline samples and calculated their inverse values as the adjustment factors. The final imaging subsample weights were the product of the ABCD baseline weight and the adjustment factors. The sequential adjustments correct for ABCD recruitment selection by using the base weights and imaging subsample screening by the inverse inclusion propensity score weighting. We compare the frequency distributions of the base and final weights in Figure 7 of the Appendix A.

The final weights, assigned to n =3,782 individuals with an average value of 1925.5 and a standard deviation of 1458.8, have larger variability than the base weights, assigned to n =9,905 individuals with an average of 650.2 and a standard deviation of 341.0. We

include the frequent distributions of the two sets of weights in the Supplement. Table 1 shows that the final weighting adjustment effectively matches the subsample distribution with the weighted ABCD baseline sample, an improvement over applying the existing base weights.

2.3 Weighed Image-on-scalar Regression Model

With the newly constructed final weights, we proceeded to fit the image-on-scalar regression model. To begin, we introduce the model notation. Let $\mathcal{V} = \{v_m \in \mathbb{R}^3\}_{m=1}^d$ denote the set of d vertices in an image, where each vertex v corresponds to a three-dimensional spatial coordinate in the image space and is represented as a vector of length 3. For individual $i = 1, \ldots, n$, the outcome $y_i(\mathcal{V}) \in \mathbb{R}^d$ is collected as a vectorized, spatially distributed image with a set of d locations. Let $\mathbf{x}_i = (x_{ij})_{j=1}^p \in \mathbb{R}^p$ be the p-dimensional exposure variables of interest and $\mathbf{z}_i = (z_{ir})_{r=1}^q \in \mathbb{R}^q$ be a vector of q control variables.

We model the association between imaging outcome $y_i(\mathcal{V})$ and scalar variables via the following image-on-scalar regression on the vertex-level:

$$y_i(v) = \beta_0(v) + \sum_{j=1}^{p} x_{ij}\beta_j(v) + \sum_{r=1}^{q} z_{ir}\eta_r(v) + \epsilon_i(v), \quad \epsilon_i(v) \sim \mathcal{GP}(0, \kappa). \tag{1}$$

Here $\beta_0(v)$ and $\{\beta_j(v)\}_{j=1}^p$ are the spatially-varying intercept function and spatially-varying coefficient (SVC) function of exposure variables $\{x_{ij}\}_{j=1}^p$, respectively. The spatially-varying coefficient functions $\{\beta_j(v)\}_{j=1}^p$ capture the spatial variability of the effect of the exposure variables $\mathbf{x}_i$ on the image outcome. The $\{\eta_r(v)\}_{r=1}^q$ are the SVCs of control variables $\mathbf{z}_i$. Since the vertex-space in image data typically exhibits spatial homogeneity, adjacent vertices tend to be correlated, therefore we assume the error term $\epsilon_i(v)$ follows a

Gaussian process ($\mathcal{GP}$) with zero mean and a kernel covariance function κ.

Figure 1 summarizes the model. We are interested in identifying the vertices in the outcome image $\mathbf{y}(\mathcal{V})$ that show significant association with the exposure variable $\mathbf{x}$ in Model (1). More specifically, we aim to account for the sampling weights and make population-based inferences regarding the coefficients $\boldsymbol{\beta}(\mathcal{V})$.

We define $\mathbf{w} = \{w_i\}_{i=1}^n$ as a vector of the sample weights, with each weight $w_i \in \mathbb{R}^+$ corresponding to the i-th observation. In our analysis, we introduce additional notation: $\boldsymbol{\beta}(\mathcal{V})$ represents the set of functions $\{\beta_j(\mathcal{V})\}_{j=0}^p \in \mathbb{R}^{(p+1)\times d}$; $\boldsymbol{\eta}(\mathcal{V}) = \{\eta_r(\mathcal{V})\}_{r=1}^q \in \mathbb{R}^{q\times d}$; $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^{n\times(p+1)}$; $\mathbf{z}$ is defined as the set $\{\mathbf{z}_i\}_{i=1}^n$; $\mathbf{y}(\mathcal{V}) = \{y_i(\mathcal{V})\}_{i=1}^n$; and $\epsilon(\mathcal{V}) = \{\epsilon_i(\mathcal{V})\}_{i=1}^n$. Given the image-on-scalar regression model presented in Model (1), we seek to obtain the estimates of $\{\beta_j(v)\}_{j=0}^p$ and $\{\eta_r(v)\}_{r=1}^q$ via maximizing the following weighted likelihood function:

$$f_n\left(\boldsymbol{\beta}(\mathcal{V}), \boldsymbol{\eta}(\mathcal{V}) \middle| \mathbf{x}, \mathbf{z}, \mathbf{y}(\mathcal{V}), \epsilon(\mathcal{V})\right) = \prod_{i=1}^n \left[\mathcal{L}\left(\boldsymbol{\beta}(\mathcal{V}), \boldsymbol{\eta}(\mathcal{V}) | \mathbf{x}_i,\ \mathbf{z}_i, y_i(\mathcal{V}), \epsilon_i(\mathcal{V})\right)\right]^{w_i}. \tag{2}$$

Here, $\mathcal{L}\left(\boldsymbol{\beta}(\mathcal{V}), \boldsymbol{\eta}(\mathcal{V}) | \mathbf{x}_i, , \mathbf{z}_i, y_i(\mathcal{V}), \epsilon_i(\mathcal{V})\right)$ denotes the multivariate Gaussian likelihood function.

2.3.1 Model Assumptions

We adopt the basis expansion approach to model the spatial varying coefficients $\{\beta_j(v)\}_{j=0}^p$ and $\{\eta_r(v)\}_{r=1}^q$ in Model (1). We consider L pre-specified basis functions $\Psi = \{\psi_1, \ldots, \psi_L\}$ that are orthonormal and satisfy $\int_{\mathcal{V}} \psi_l^2(v)dv = 1$ for $l = 1, \ldots, L$. We construct the orthonormal basis functions from the eigen-functions of a GP covariance kernel, and the choice of kernel depends on the data structure. For example, we can use the γ-exponential

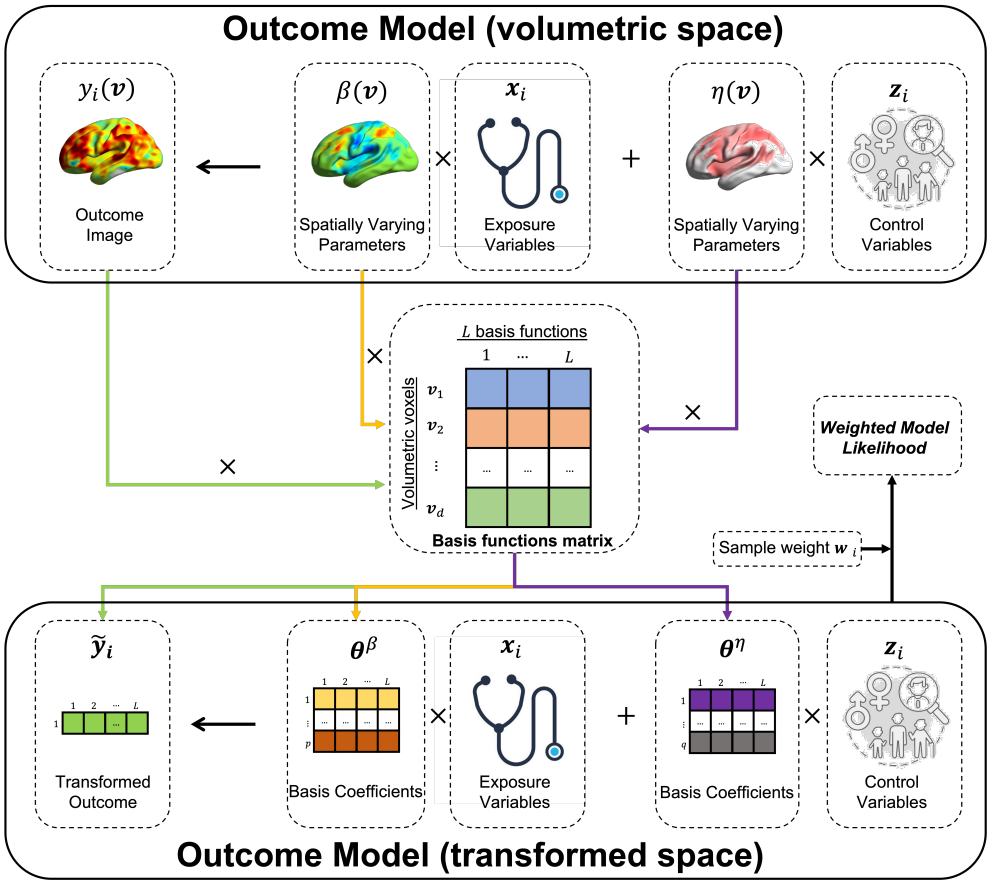

Figure 1: Graphical illustration of the weighted image-on-scalar regression model. The outcome image of individual i at vertex v, denoted as $y_i(v)$, is composed of two mean components: 1) The intercept SVCs $\beta_0(v)$ and the association SVCs $\{\beta_j(v)\}_{j=1}^p$ with the exposure variable $\mathbf{x}_i$; 2) The association SVCs $\boldsymbol{\eta}_i(v)$ with control variables $\mathbf{z}_i$. The volumetric outcome model is converted into a lower-dimensional space by approximating the SVCs using basis function expansions with L constant basis functions $\{\psi_l(v)\}_{l=1}^L$ and their corresponding coefficients: $\{\boldsymbol{\theta}_j^\beta\}_{j=0}^p$ and $\{\boldsymbol{\theta}_r^\eta\}_{r=1}^q$. The sample weights w_i are applied to the likelihood functions in the transformed space.

kernel function with the form:

$$\kappa(v_1, v_2) = \exp\left\{-(\frac{d(v_1, v_2)}{\rho})^{\gamma}\right\}, \quad 0 < \gamma \leq 2,\ \rho > 0. \tag{3}$$

For cortical-surface fMRI data that are mapped onto a spherical surface, $d(v_1, v_2)$ denotes the spherical distance between vertices v_1 and v_2, given by $d(v_1, v_2) = \arccos(v_1 \cdot v_2)$ (Williams and Rasmussen, 2006). The hyperparameter ρ controls the smoothness of GP. When γ is fixed, larger value of ρ indicates a smoother GP kernel. On the other hand, when ρ is fixed, a larger value of γ indicates a steeper decline in the kernel value with an increasing distance. Note that the γ-exponential kernel (3) reduces to the standard squared-exponential kernel when $\gamma = 2$.

Given the basis functions, we assume the SVCs are approximated by the basis functions expansion:

$$\beta_j(v) \simeq \sum_{l=1}^{L} \theta_{jl}^{\beta} \psi_l(v), \qquad \eta_r(v) \simeq \sum_{l=1}^{L} \theta_{rl}^{\eta} \psi_l(v), \quad \forall v \in \mathcal{V},$$

for $j = 0, \ldots, p$ and $r = 1, \ldots, q$. Here, the coefficients $\{\theta_{jl}^{\beta}\}_{l=1}^{L}$ and $\{\theta_{rl}^{\eta}\}_{l=1}^{L}$ are unknown basis coefficients to be estimated. We also assume that, with the same set of orthonormal basis functions $\{\psi_l(v)\}_{l=1}^{L}$, the covariance kernel κ defined in (1) can be approximated as:

$$\kappa(v_1, v_2) \simeq \sum_{l=1}^{L} \lambda_l \psi_l(v_1) \psi_l(v_2), \quad \forall v_1, v_2, \tag{4}$$

where $\{\lambda_l\}_{l=1}^{L}$ are unknown positive parameters. The positivity constraint of λ_l ensures the positive definite property of covariance kernel κ (Ghosal and Van der Vaart, 2017). With

the basis functions and basis coefficients defined above, Model (1) becomes:

$$y_i(v) = \sum_{l=1}^{L} \theta_{0l}^{\beta} \psi_l(v) + \sum_{l=1}^{L} \sum_{j=1}^{p} x_{ij} \theta_{jl}^{\beta} \psi_l(v) + \sum_{r=1}^{q} \sum_{l=1}^{L} z_{ir} \theta_{rl}^{\eta} \psi_l(v) + \epsilon_i(v). \tag{5}$$

Using a basis expansion approach to represent a SVC reduces the number of parameters from $(p+q+1) \times d$ to $(p+q+1) \times L$, where $d >> L$, increasing computational efficiency.

2.3.2 Parameter Estimation

We account for the sampling weights and estimate parameters in Models (1), (2) and (5). We first discuss how to obtain orthonormal basis functions $\{\psi_l(v)\}_{l=1}^{L}$ for dimension reduction and then describe the parameter estimation process via maximizing the weighted likelihood function.

We construct $\{\psi_l(v)\}_{l=1}^{L}$ on the set of voxels $\mathcal{V} = \{v_m\}_{m=1}^{d}$ from L eigenfunctions, denoted as $\{\tilde{\psi}_l(v)\}_{l=1}^{L}$ of a pre-specified covariance kernel function on $\mathbb{R}^3$. First, we evaluate $\{\tilde{\psi}_l(v)\}_{l=1}^{L}$ on $\mathcal{V}$ and obtain the matrix $\tilde{\Psi} = \{\tilde{\psi}_l(v_m)\}_{L \times d}$. Second we perform the singular value decomposition: $\tilde{\Psi} = \Psi \mathbf{D} \mathbf{V}^\top$. Here, Ψ is a $d \times L$ rotation matrix of the functional eigenvector and satisfies the orthonormal condition, i.e., $\Psi^\top \Psi = I_L$, where I_L is an $L \times L$ identity matrix. Finally, we use $\mathbf{\Psi}$ to transform Model (1) into a L-dimensional multivariate linear regression model with a diagonal variance structure $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_L)$. Model (5) can be represented in matrix form as:

$$y_i(\mathbf{v}) = \mathbf{x}_i \boldsymbol{\theta}^{\beta} \Psi^\top + \mathbf{z}_i \boldsymbol{\theta}^{\eta} \Psi^\top + \epsilon_i(\mathbf{v}), \quad \epsilon_i(\mathbf{v}) \sim \mathcal{N}(\mathbf{0}, \mathbf{K}), \quad \mathbf{K} = \Psi \Lambda \Psi^\top, \tag{6}$$

where $\epsilon_i(\mathbf{v}) = [\epsilon_i(v_1), \ldots, \epsilon_i(v_d)]^\top$, and $\mathbf{K} = \{\kappa(v_m, v_{m'})\}_{d \times d}$. Multiplying both sides of (6)

by $\boldsymbol{\Psi}$ transforms the model into:

$$\tilde{\mathbf{y}}_i = \mathbf{x}_i \boldsymbol{\theta}^{\beta} + \mathbf{z}_i \boldsymbol{\theta}^{\eta} + \tilde{\boldsymbol{\epsilon}}_i, \tag{7}$$

where $\tilde{\mathbf{y}}_i = y_i(\mathcal{V})\Psi \in \mathbb{R}^{1\times L}$ is the transformed response variable, and $\tilde{\boldsymbol{\epsilon}}_i = \epsilon_i(\mathcal{V})\Psi \sim \mathcal{N}(\mathbf{0}, \mathbf{K}')$ is the transformed error term with $\mathbf{K}' = \Psi^\top \Psi \Lambda \Psi^\top \Psi = \Lambda$ served as the diagonal covariance matrix. Thus, by employing Model (7), we can transform the original d-dimensional Model (1) into a more manageable L-dimensional multivariate regression problem, where we typically have $d >> L$ in real-world applications. Moreover, the transformed random error term $\tilde{\boldsymbol{\epsilon}}_i$ follows a normal distribution with zero mean and a diagonal variance structure Λ, increasing computational efficiency.

We employ maximum likelihood estimation (MLE) to estimate the parameters specified in Model (7). Let $\mathbf{m}_i = [\mathbf{x}_i, \mathbf{z}_i] \in \mathbb{R}^{1\times(p+q+1)}$ denote the concatenated predictor vector and $\mathbf{M} = \{\mathbf{m}_i\}_{i=1}^n \in \mathbb{R}^{n\times(p+q+1)}$ represent the combined design matrix. The combined basis coefficients can be represented by $\boldsymbol{\theta} = [\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\eta}] \in \mathbb{R}^{(p+q+1)\times L}$, and the transformed outcome matrix is denoted as $\tilde{\mathbf{Y}} = \{\tilde{\mathbf{y}}_i\}_{i=1}^n \in \mathbb{R}^{n\times L}$. Model (7) can be represented in the matrix form: $\tilde{\mathbf{Y}} = \mathbf{M}\boldsymbol{\theta} + \tilde{\boldsymbol{\epsilon}}$. The corresponding weighted likelihood function $f_n(\boldsymbol{\theta}|\mathbf{M}, \tilde{\mathbf{Y}})$ can be defined as:

$$f_n(\boldsymbol{\theta}|\mathbf{M}, \tilde{\mathbf{Y}}) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{(2\pi)^L |\mathbf{K}'|}} \exp\left(-\frac{1}{2} (\tilde{\mathbf{y}}_i - \mathbf{m}_i \boldsymbol{\theta}) \mathbf{K}'^{-1} (\tilde{\mathbf{y}}_i - \mathbf{m}_i \boldsymbol{\theta})^\top \right) \right]^{w_i}.$$

We use $\mathbf{W} = \mathrm{diag}(w_1, \ldots, w_n)$ to represent the weight matrix, where each diagonal element corresponds to the weight of an individual. We solve the following score equations and

estimate the population parameters $\boldsymbol{\theta}$:

$$U(\boldsymbol{\theta}) = \frac{\partial \log f_n(\boldsymbol{\theta}|\mathbf{M}, \tilde{\mathbf{Y}})}{\partial \boldsymbol{\theta}} = \mathbf{M}^\top \mathbf{W}(\tilde{\mathbf{Y}} - \mathbf{M}\boldsymbol{\theta})\mathbf{K}'^{-1} = 0,$$

and we obtain:

$$\hat{\boldsymbol{\theta}} = (\mathbf{M}^\top \mathbf{W}\mathbf{M})^{-1}(\mathbf{M}^\top \mathbf{W}\tilde{\mathbf{Y}}). \tag{8}$$

Based on the orthonormal properties of the basis functions matrix $\boldsymbol{\Psi}$ discussed above, we can derive the estimates of the SVCs, $\{\boldsymbol{\beta}(\mathcal{V}), \boldsymbol{\eta}(\mathcal{V})\}$ through a linear transformation of $\hat{\boldsymbol{\theta}}^\beta$ and $\hat{\boldsymbol{\theta}}^\eta$:

$$\hat{\boldsymbol{\beta}}(\mathcal{V}) = \hat{\boldsymbol{\theta}}^\beta \boldsymbol{\Psi}^\top \quad \text{and} \quad \hat{\boldsymbol{\eta}}(\mathcal{V}) = \hat{\boldsymbol{\theta}}^\eta \boldsymbol{\Psi}^\top.$$

2.3.3 Inference on Population Coefficients

With sample survey data, the variance estimation of coefficient estimates needs to account for the sampling variance introduced by the weights (Rust, 1985). Since the covariance matrix $\mathbf{K}'$ is diagonal, we use the delta-method (Binder, 1983) to obtain the "sandwich" form estimator for the covariance of the coefficient estimates $\hat{\boldsymbol{\theta}}$

$$V(\hat{\boldsymbol{\theta}}) = \text{diag}\Big\{V(\hat{\boldsymbol{\theta}}_1), V(\hat{\boldsymbol{\theta}}_2), \ldots, V(\hat{\boldsymbol{\theta}}_L)\Big\} \tag{9}$$

$$\begin{aligned} V(\hat{\boldsymbol{\theta}}_l) &= \left(\frac{\partial U(\boldsymbol{\theta}_l)}{\partial \boldsymbol{\theta}_l}\right)^{-1} \hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_l) \left(\frac{\partial U(\boldsymbol{\theta}_l)}{\partial \boldsymbol{\theta}_l}\right)^{-1} \\ &= (\mathbf{M}^\top \mathbf{W}\mathbf{M})^{-1} \hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_l)(\mathbf{M}^\top \mathbf{W}\mathbf{M})^{-1} \\ &= (\mathbf{M}^\top \mathbf{W}\mathbf{M})^{-1} \mathbf{M}^\top \mathbf{W} \Big(\text{diag}\{\tilde{\mathbf{Y}}_l - \mathbf{M}\hat{\boldsymbol{\theta}}_l\}\Big)^2 \mathbf{W}\mathbf{M}(\mathbf{M}^\top \mathbf{W}\mathbf{M})^{-1}, \end{aligned} \tag{10}$$

for $l = 1, \ldots, L$, where we use the White's heteroscedasticity-consistent estimator to replace $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_l)$ (White, 1980).

The following proposition gives the representation of $V\left(\hat{\beta}_j(\mathcal{V})\right)$.

Proposition 2.1. *Given the variance-covariance matrix $V(\hat{\boldsymbol{\theta}})$ for $\hat{\boldsymbol{\theta}}$, since the SVC $\hat{\boldsymbol{\beta}}(\mathcal{V})$ is a linear transformation of $\hat{\boldsymbol{\theta}}$, i.e., $\hat{\boldsymbol{\beta}}(\mathcal{V}) = \hat{\boldsymbol{\theta}}\Psi^\top$. the covariance matrix for the vectorized* $\mathrm{vec}(\hat{\boldsymbol{\beta}}(\mathcal{V}))$ *can be expressed as:*

$$V\left(\mathrm{vec}\left(\hat{\boldsymbol{\beta}}(\mathcal{V})\right)\right) = (\Psi \otimes \mathbf{I}_{p+1})\, V(\hat{\boldsymbol{\theta}})\, (\Psi \otimes \mathbf{I}_{p+1})^\top, \tag{11}$$

where $V(\hat{\boldsymbol{\theta}})$ is defined in Equation (9).

Proof. Since $\boldsymbol{\beta}(\mathcal{V}) = \boldsymbol{\theta}\Psi^T$, by using the Kronecker product property for the vectorization operator $\mathrm{vec}(\cdot)$, we have $\mathrm{vec}(\hat{\boldsymbol{\beta}}(\mathcal{V})) = \mathrm{vec}(\hat{\boldsymbol{\theta}}\Psi^T) = (\Psi \otimes \mathbf{I}_{p+1})\,\mathrm{vec}(\hat{\boldsymbol{\theta}})$. We apply this for the covariance of a Kronecker product:

$$V(\mathrm{vec}(\hat{\boldsymbol{\beta}}(\mathcal{V}))) = V((\Psi \otimes \mathbf{I}_{p+1})\,\mathrm{vec}(\hat{\boldsymbol{\theta}})) = (\Psi \otimes \mathbf{I}_{p+1}) V(\mathrm{vec}(\hat{\boldsymbol{\theta}}))(\Psi \otimes \mathbf{I}_{p+1})^\top, \tag{12}$$

where $V(\mathrm{vec}(\hat{\boldsymbol{\theta}}))$ is the estimated variance-covariance matrix of $\mathrm{vec}(\hat{\boldsymbol{\theta}})$. The variance-covariance matrix $V(\mathrm{vec}(\boldsymbol{\theta}))$ is given by

$$V(\mathrm{vec}(\hat{\boldsymbol{\theta}})) = \mathbf{I}_{p+1} \otimes \left[(\mathbf{M}^\top\mathbf{W}\mathbf{M})^{-1}\mathbf{M}^\top\mathbf{W}\hat{\boldsymbol{\Sigma}}\mathbf{W}\mathbf{M}(\mathbf{M}^\top\mathbf{W}\mathbf{M})^{-1}\right]. \tag{13}$$

Combining (12) and (13), we have

$$\begin{aligned} &V(\mathrm{vec}(\hat{\boldsymbol{\beta}}(\mathcal{V}))) = (\Psi \otimes \mathbf{I}_{p+1}) V(\mathrm{vec}(\hat{\boldsymbol{\theta}}))(\Psi \otimes \mathbf{I}_{p+1})^\top \\ &= (\Psi \otimes \mathbf{I}_{p+1}) \left\{\mathbf{I}_{p+1} \otimes \left[(\mathbf{M}^\top\mathbf{W}\mathbf{M})^{-1}\mathbf{M}^\top\mathbf{W}\hat{\boldsymbol{\Sigma}}\mathbf{W}\mathbf{M}(\mathbf{M}^\top\mathbf{W}\mathbf{M})^{-1}\right]\right\} (\Psi \otimes \mathbf{I}_{p+1})^\top. \end{aligned} \tag{14}$$

This completes the proof. □

The diagonal elements of $V\{\text{vec}(\hat{\boldsymbol{\beta}}(\mathcal{V}))\}$ are the variance estimates of $\boldsymbol{\beta}(\mathcal{V})$ evaluated at each voxel v for $j = 0, \ldots, p$. With the estimated variance $V(\hat{\beta}_j(v))$, we can test the null hypothesis $H_0 : \beta_j(v) = 0$ against $H_1 : \beta_j(v) \neq 0$ with the Wald test statistics:

$$T_j^{\beta}(v) = \frac{|\hat{\beta}_j(v)|}{\sqrt{V(\hat{\beta}_j(v))}}, \tag{15}$$

for $j = 0, \ldots, p$ and $\forall v \in \mathcal{V}$. After computing the Wald test statistics at each voxel v, we generate a p-value image. To reduce the false discovery rate (FDR), we apply multiple comparison corrections to the p-value image using the random field theory (Nichols and Hayasaka, 2003). To facilitate interpretation, we summarize the vertices presenting significant associations at the functional region level using the the Gordon 333 region cortical parcellation (Gordon et al., 2016).

2.4 Simulation Design

We describe the data generation processes in the simulation studies to examine the impact of including survey weights. We generated a synthetic population with N_{pop} =1,000,000 of individuals, from which we repeatedly drew samples and fit image-on-scalar regression models. We mimicked real-life scenarios where it is often impractical to correctly specify a model. We considered a data generation model and two fitting scenarios as the following: 1) a correctly-specified model the same as the data generation model, which includes two-way interaction terms between control variables and the exposure variable, and 2) a mis-specified

model that ignores the interaction term.

$$\begin{cases} \text{Correctly-specified:} & y_i(v) = \beta_0(v) + \beta_1(v)x_i + \eta(v)z_i + \gamma(v)(z_i \times x_i) + \epsilon_i(v) \\ \text{Mis-specified:} & y_i(v) = \beta_0(v) + \beta_1(v)x_i + \eta(v)z_i + \epsilon_i(v). \end{cases}$$

We considered univariate exposure and control variables, both of which were generated from a standard normal distribution, i.e., $x_i \sim \mathcal{N}(0,1)$ and $z_i \sim \mathcal{N}(0,1)$. We used a logistic model to calculate the inclusion probability $\Pr(\delta_i = 1|z_i)$ for each individual to be sampled, with a logit fuction, $\text{logit}(\Pr(\delta_i = 1|z_i)) = -1 - z_i$. We generated the binary inclusion indicator δ_i for each participant from a Bernoulli distribution: $\delta_i \sim \text{Bernoulli}\,(\Pr(\delta_i = 1|z_i))$. Given the covariates and coefficients, we simulated the outcome image for each subject in the population using the following data generation model (DGM):

$$y_i(v) = \beta_0(v) + \beta_1(v)x_i + \eta(v)z_i + \gamma(v)(z_i \times x_i) + \epsilon_i(v),$$

for $v \in \mathcal{V}$ and $i = 1, \ldots, N_{\text{pop}}$. The true $\beta_0(v)$ and $\beta_1(v)$ are plotted in the top-row of Figure 2. The parameter $\gamma(v)$ represents the SVC for the interaction term $(z_i \times x_i)$ and is from an approximate Gaussian process as follows:

$$\theta_l^\gamma \overset{i.i.d}{\sim} \mathcal{N}(0,1), \quad \gamma(v) = \sum_{l=1}^{L} \theta_l^\gamma \psi_l(v),$$

where $\{\psi_l(v)\}_{l=1}^L$ are orthornormal basis functions discussed in Section 2.3.2. In our simulation studies, we set $L = 500$. We generate error terms $\epsilon_i(v)$ from an approximated Gaussian process with zero mean and a covariance function defined as $\kappa(v_1, v_2) = \sum_{l=1}^L \lambda_l \psi_l(v_1)\psi_l(v_2)$. We adjust the signal-to-noise ratio by varying values of pre-specified eigenvalue λ_l, each

taking on values proportionate to the eigenvalues of the kernel covariance function κ.

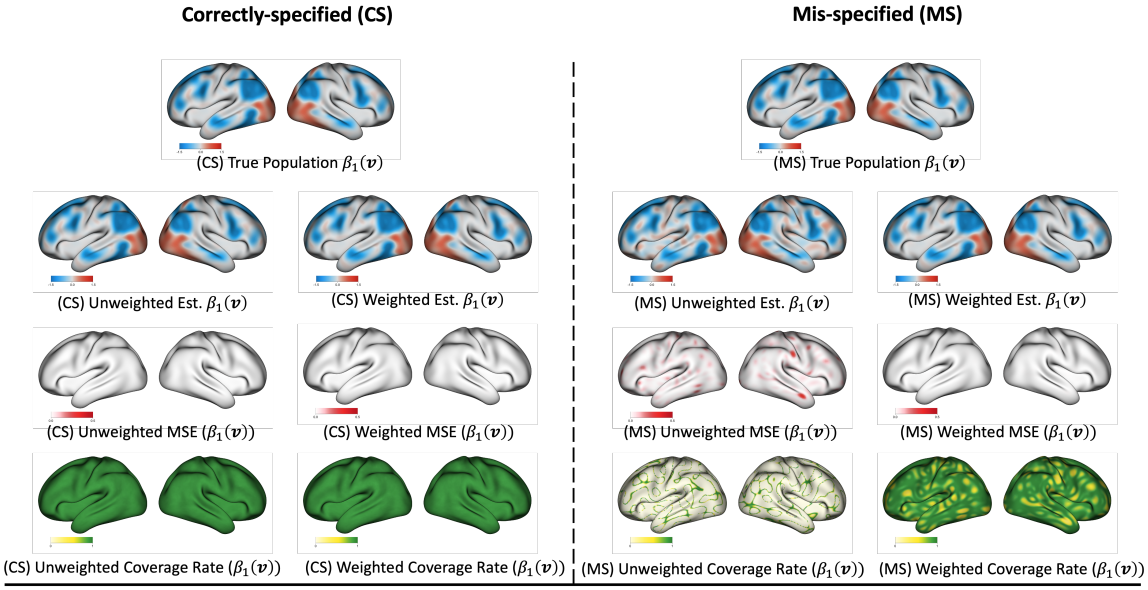

Figure 2: Visual representation of SVC slope $\beta_1(\mathbf{v})$ across different model specification. Top row: true population coefficients of slope SVC $\beta_1(\mathbf{v})$. Second row: point estimation by unweighted and weighted image-on-scalar regression. Third row: average mean squared error across replicates by unweighted and weighted methods. Bottom row: average coverage rate across replicates by unweighted and weighted methods.

We repeatedly sampled $n = 2000$ cases from the population using the selection probabilities $\Pr(\delta_i = 1|z_i)$ 500 times. The graphical illustration of the DGM and sample selection is presented in Figure 3.

We calculated the mean squared error (MSE) of the point estimates $\beta_1(v)$ and averaged their values across all voxels $v \in \mathcal{V}$ and all replicates. Additionally, to assess the robustness of the model inferences in both weighted and unweighted methods, the true coefficient coverage rate over 500 sets of the 95% confidence intervals was computed for each voxel $v \in \mathcal{V}$. We reported four primary performance metrics: bias, MSE, coverage rates, and FDR.

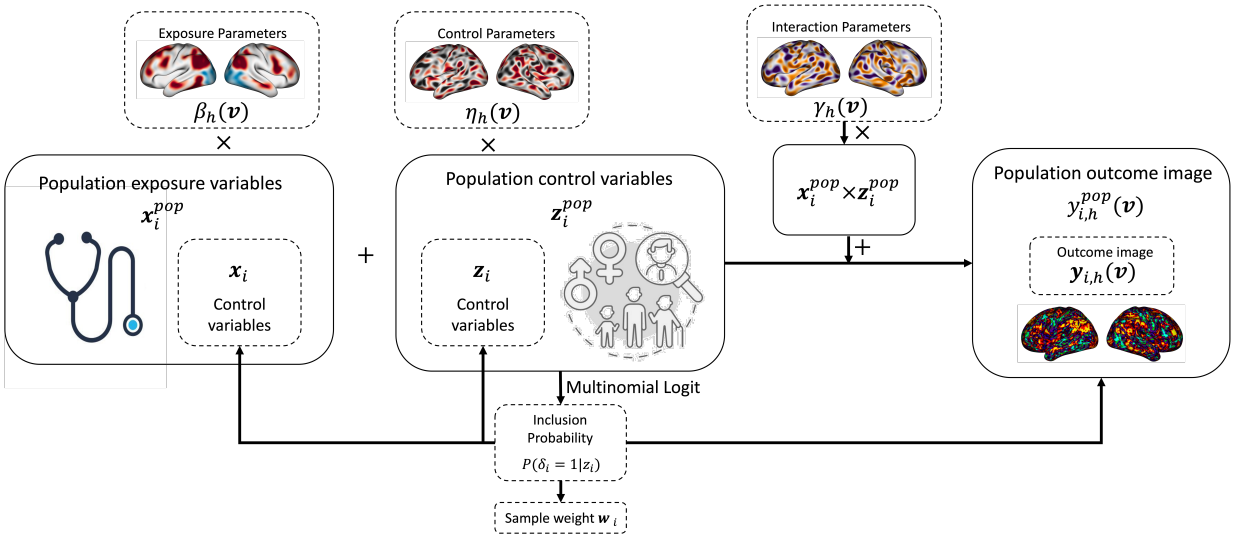

Figure 3: Illustration of the data generation model and sample selection for the simulation study: The population-simulated outcome images, denoted as $\mathbf{y}_i^{\text{pop}}(\mathcal{V})$, were generated using population exposure variables $\mathbf{x}_i^{\text{pop}}$ and their associated exposure parameters $\boldsymbol{\beta}(\mathcal{V})$, population control variables $\mathbf{z}_i^{\text{pop}}$ and their associated control parameters $\boldsymbol{\eta}(\mathcal{V})$, along with interaction terms $(\boldsymbol{\gamma}(\mathcal{V})(\mathbf{x}_i^{\text{pop}} \times \mathbf{z}_i^{\text{pop}}))$. The population control variables $\mathbf{z}_i^{\text{pop}}$ were input into a predetermined multinomial logit model to determine the inclusion probabilities for each individual. Within each replicate, subsamples were drawn from the population based on these inclusion probabilities. The sample weights $\mathbf{w}$ were computed as the inverse of the inclusion probabilities.

These metrics were determined across the vertex space for both weighted and unweighted methods, using 500 replicates.

It is important to note that when obtaining the "true coefficient" values in the misspecified model, we should refer to the coefficients estimated from the finite population. However, with a population size of N_{pop} =1,000,000, obtaining the population estimates of $\beta_0(v)$, $\beta_1(v)$, and $\eta(v)$ can be computationally demanding. This challenge becomes particularly noticeable when performing regression analysis on the transformed outcome $\tilde{\mathbf{y}}_i$ on x_i and z_i. To address this computational challenge, we employed a stochastic gradient descent (SGD) method (Diebolt and Ip, 1996) to estimate the finite population coefficients. This approach iteratively updates the estimates of the population-level coefficients using the gradient of the loss function and reduces the computational burden associated with transforming the outcome and estimating the coefficients.

The SGD method for estimating the population coefficients can be summarized as follows. We define the design matrix as $\mathbf{m}_{\text{pop}} = [\mathbf{x}_{\text{pop}}, \mathbf{z}_{\text{pop}}] \in \mathbb{R}^{N_{\text{pop}} \times q(p+1)}$. Additionally, we denote the concatenated basis coefficients as $\boldsymbol{\theta}_{\text{pop}} = [\boldsymbol{\theta}^{\beta}_{\text{pop}}; \boldsymbol{\theta}^{\eta}_{\text{pop}}] \in \mathbb{R}^{(q+p+1)\times L}$. Our goal is to minimize the loss function $\text{Loss}(\boldsymbol{\theta}_{\text{pop}}) = ||\mathbf{m}_{\text{pop}}\boldsymbol{\theta}_{\text{pop}} - \tilde{\mathbf{y}}||_2^2$. To update the coefficients, we calculate the derivative of the loss function with respect to $\boldsymbol{\theta}_{\text{pop}}$ as follows:

$$\frac{\partial \text{Loss}(\boldsymbol{\theta}_{\text{pop}})}{\partial \boldsymbol{\theta}_{\text{pop}}} = 2\mathbf{m}_{\text{pop}}^{\top}(\mathbf{m}_{\text{pop}}\boldsymbol{\theta}_{\text{pop}} - \tilde{\mathbf{y}}).$$

During the process of generating a sample $(\mathbf{m}_i, \tilde{\boldsymbol{y}}_i)$ from the population, we update the estimates of the population coefficients $\tilde{\boldsymbol{\theta}}_{\text{pop}}$ using the gradient descent algorithm:

$$\tilde{\boldsymbol{\theta}}_{\text{pop}} := \tilde{\boldsymbol{\theta}}_{\text{pop}} - 2\alpha \left[\mathbf{m}_{\text{pop},i}^{\top}(\mathbf{m}_{\text{pop},i}\tilde{\boldsymbol{\theta}}_{\text{pop}} - \tilde{\mathbf{y}}_i)\right],$$

where α represents the learning rate of the gradient descent algorithm. This iterative approach updates the population-level coefficients and allows us to estimate $\boldsymbol{\theta}_{\text{pop}}$ effectively. In comparison to the least square estimator, which has a computation complexity of $O(p^3 + p^2 N_{\text{pop}}^2 + p^2 L)$ for estimating population-level basis coefficients, SGD reduces the computation complexity to just $O(N_{\text{pop}} p^3 + n_{\text{pop}} p^2 L)$, by circumventing extensive matrix multiplications.

3 Results

In this section, we present the results from both the ABCD Study application and simulation studies.

3.1 The ABCD Data Analysis Results

We fit weighted and unweighted image-on-scalar regression models and assessed the impact of sample weights on models with and without control variables $\mathbf{z}_i$ given in Section 2.1. Our focus is making inferences of SVCs $\beta_1(\mathcal{V})$ for one exposure variable (i.e., $p = 1$), the g-factor.

Figure 4 presents brain mapping of estimated SVCs $\beta_1(\mathcal{V})$ with and without adjustment for control variables, corrected by the FDR at a 0.01 significance level. Both weighted and unweighted analyses indicate a notable presence of vertices exhibiting positive associations in clusters distributed throughout the brain, particularly in frontoparietal regions. Broadly, these vertices fall within key macroscale networks identified by Gordon et al. (2016), including the Dorsal Attention Network (DAN), Fronto-Parietal Network (FPN), and Salience Network (SN). These results suggest that, on average, a higher g-factor is associated with

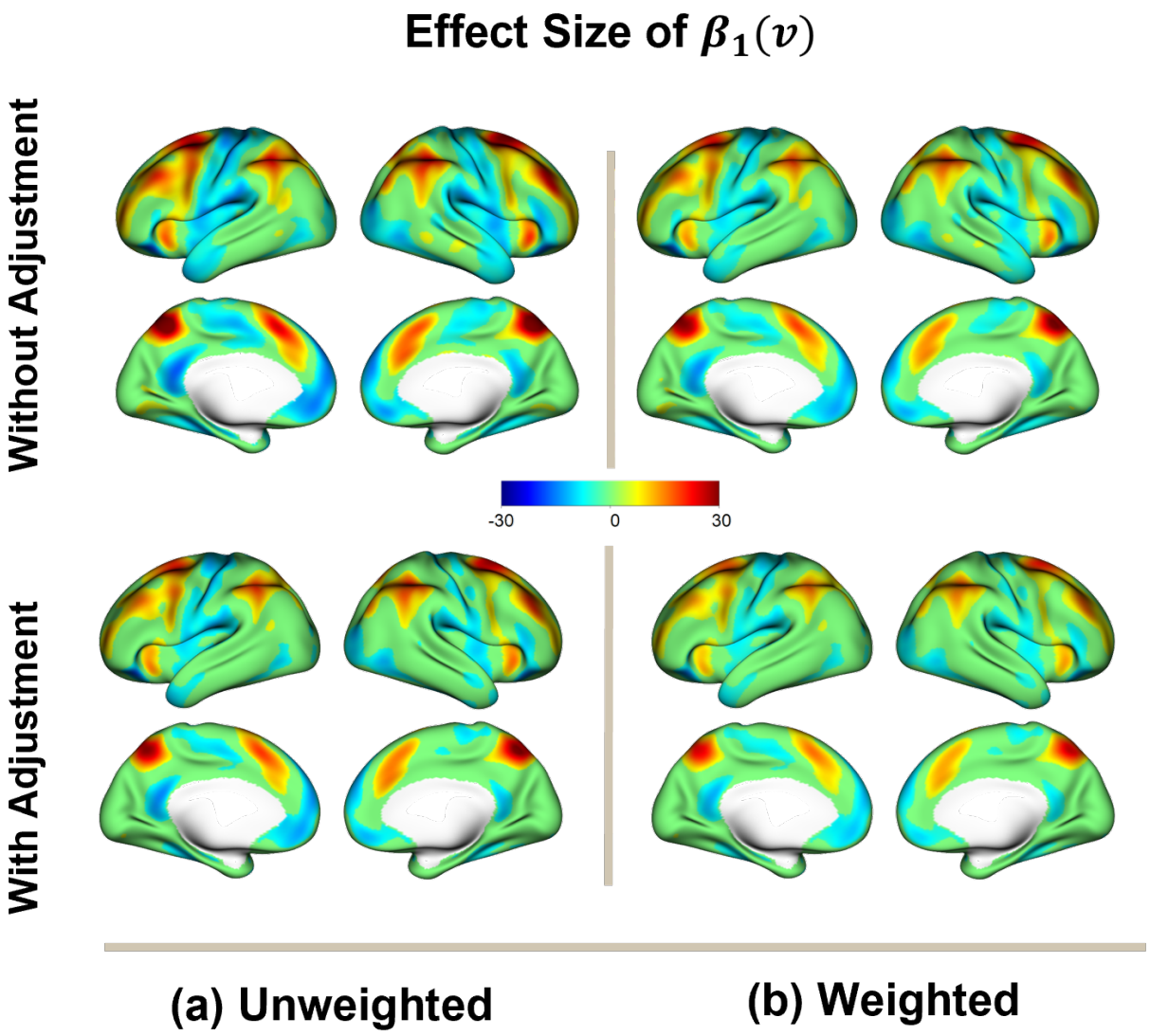

Figure 4: Brain mapping of $\beta_1(\mathcal{V})$ estimates with unweighted and weighted image-on-scalar regression models. The top row depicts results without adjusting control variables, while the bottom row illustrates results adjusting control variables. All vertices with significant associations displayed were corrected by a false discovery rate at a significance level 0.01.

increased working memory activation in these cortical areas. Conversely, we observe a large number of vertices in the Cingulo-Opercular Network (CON), Default Mode Network (DMN), and Parieto-Occipital Network (PON) showing negative associations between the g-factor and working memory.

Figure 5 compares detected vertices with significant associations between weighted and unweighted image-on-scalar regression methods, corrected by the FDR at a 0.01 significance level. Unweighted methods detect more vertices with significant associations than the weighted methods, and the differences mainly occur at the boundaries of brain areas with significant associations. Several regions significantly associated with working memory, such as the insula, middle temporal gyrus, and cingulate cortex, are only detected by the unweighted image-on-scalar regression. Moreover, regions such as the primary motor region, primary somatosensory-hand cortex, and primary auditory cortex are identified by the unweighted model but not by the weighted model. Conversely, the weighted image-on-scalar regression detects the inferior temporal gyrus, which is not identified by the unweighted model.

By comparing the left and right plots of Figure 5, we assess the impacts of population-weighting on models adjusting for control variables or not by comparing the detected areas with significant associations. Several noteworthy findings emerge from this comparison. For instance, in the left lateral occipital cortex, vertex detection between unweighted and weighted regression is nearly identical when control variables are adjusted. Conversely, without these adjustments, unweighted regression reveals a larger area of negative associations in the left lateral occipital cortex compared to weighted regression. Furthermore, in the cingulate cortex within the Frontoparietal Network, substantial differences in vertex detection persist between weighted and unweighted regression, even after adjusting for

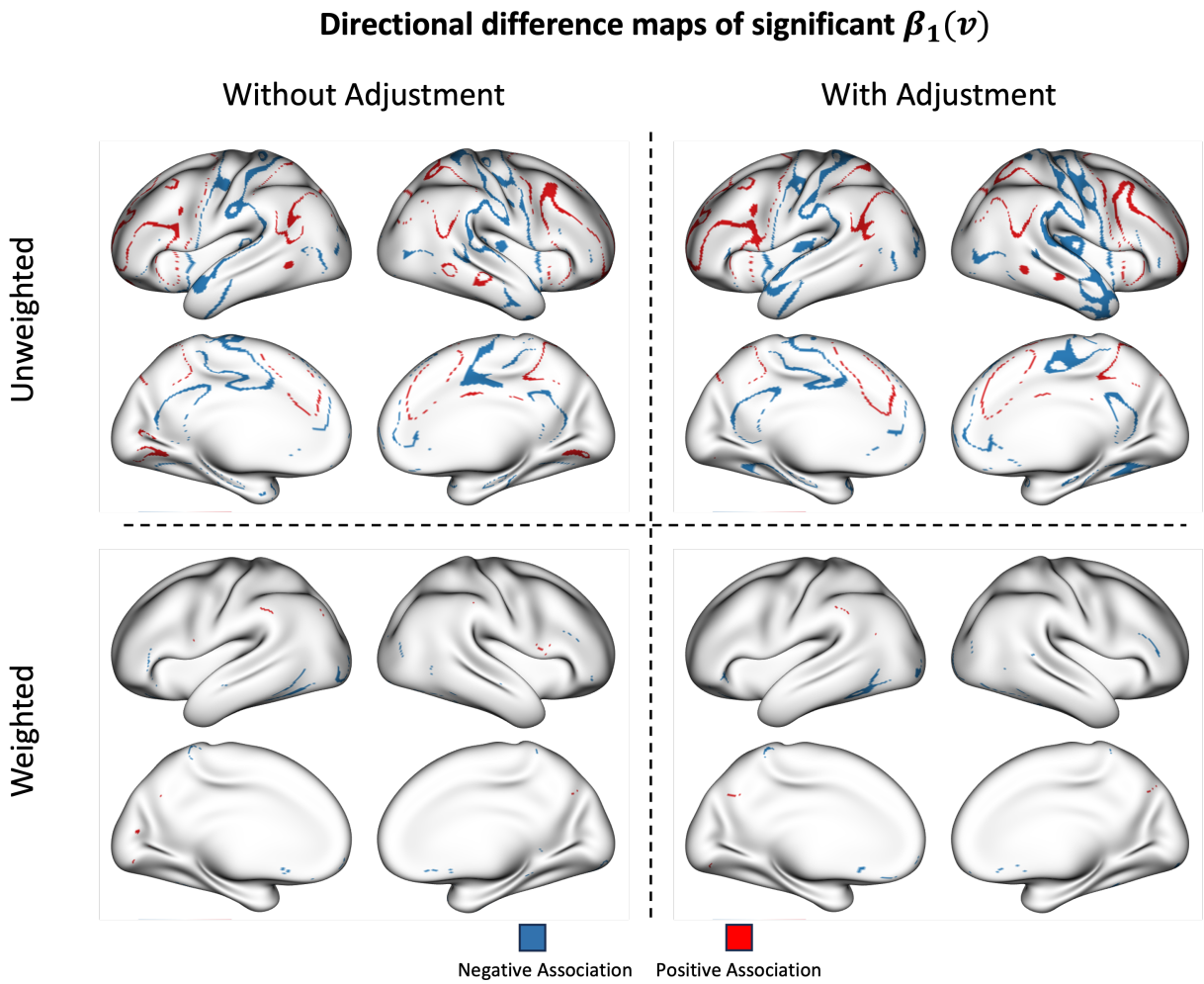

Figure 5: Directional binary maps of vertices differentially detected by weighted and unweighted image-on-scalar regressions ($\beta_1(v)$), after FDR correction at 0.01. Columns show model results without (left) and with (right) control variable adjustment (Z).

control variables.

3.2 Evaluation of the Application Study Results

We evaluated the accuracy in predictions of our fitted models to the ABCD Study data through cross-validation. Specifically, we performed 100 replicates; in each replicate, we randomly selected 80% of the data as a training dataset and used the remaining 20% as a test dataset. We fit our model to the training data and then predicted the imaging outcomes for the held-out test records. For each cortical location (vertex), we computed the predictive root mean squared error (RMSE) between the observed and predicted outcomes. We compared weighted versus unweighted regression models, both with and without control-variable adjustment, by averaging the RMSE values across 100 replicates.

Table 2 presents the average of predictive RMSE values from weighted and unweighted regression models, with and without adjustment of control variables. Without the adjustment of control variables $\mathbf{z}$, the weighted regression tends to generate smaller RMSE values than the unweighted regression at the vertex level across most networks, except PON, Somatomotor Mouth and SN. However, with the adjustment of control variables $\mathbf{z}$, the predictive RMSE values obtained from weighted regression are larger than those of unweighted image-on-scalar regression. This aligns with our finding that the impact of survey weights depends on the model specification.

3.3 Simulation Results

We conducted a simulation study and compared the weighted and unweighted image-on-scalar regression models on SVC parameter estimation accuracy when the model was correctly-specified and mis-specified. We evaluated the MSE, nominal coverage rates of

Table 2: Comparison of the predictive Root Mean Squared Error (RMSE) for each functional network between weighted (RMSE_w) and unweighted (RMSE_u) image-on-scalar regression, with and without adjustment of control variables ($\mathbf{z}_i$).

	Percentage of vertices with $\text{RMSE}_w < \text{RMSE}_u$ (%)	
Network	No adjustment	With $\mathbf{z}_i$ Adjusted
Auditory	57.1%	7.1%
Cingulo-Opercular	54.8%	3.2%
Default Mode	53.1%	3.1%
Medial Parietal	0%	0%
Dorsal Attention	69.6%	4.3%
Ventral Attention	53.8%	7.7%
Visual	56.7%	3.3%
Fronto-Parietal	53.8%	7.7%
Parieto-Occipital	0%	0%
Somatomotor Hand	58.6%	3.4%
Somatomotor Mouth	0%	0%
Salience	0%	0%

95% confidence intervals (CIs), and FDR. We expect a valid statistical model to have small MSE and FDR values, with coverage rates of 95% CIs close to 0.95.

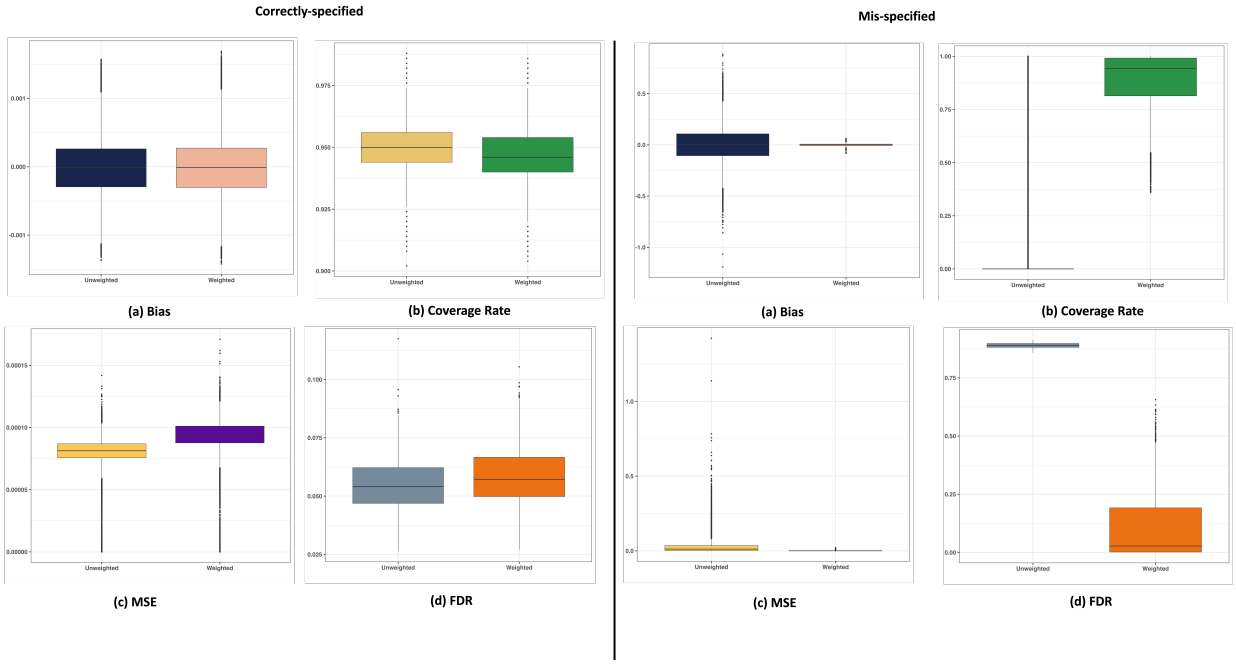

Figure 6: Simulation: Comparison of bias, mean squared error (MSE), coverage rates of 95% confidence intervals, and false discovery rates when estimating $\beta_1(v)$ using weighted and unweighted methods with both correctly-specified and mis-specified models.

Figure 6 compares bias, MSE, coverage rates of 95% CIs, and FDRs when estimating $\beta_1(v)$ using weighted and unweighted methods with both correctly-specified and mis-specified models. As expected, in the correctly specified model, the incorporation of sample weights does not affect the accuracy of the $\hat{\beta}_1(v)$ estimates. Both weighted and unweighted regression models give mostly unbiased estimates and small MSE values, where the average MSE value across vertices is 8.07×10^{-5} for the unweighted method and 9.37×10^{-5} for the weighted method. Both methods achieve 0.95 coverage rates and similar FDRs for most

estimates.

However, the results from the mis-specified model reveal substantial differences. The unweighted methods exhibit larger biases with larger variability, ranging from -1.2 to 0.88, than the weighted methods, where the biases range from -0.08 to 0.06. Additionally, the average MSE across vertices for the unweighted model is 0.026, which is higher than the weighted model's average MSE of 4.9×10^{-4}. The weighted methods also demonstrate lower FDRs than the unweighted approach. Notably, the incorporation of sample weights in a mis-specified model improves coverage rates; most vertices under the weighted approach achieve nearly 95% coverage rates, with the lowest observed rate being approximately 30%. In contrast, the unweighted method generally results in low coverage rates.

4 Discussion

This study demonstrates the importance of constructing and applying sample weights to address selection bias and subsample screening in neuroimaging analyses. By constructing new weights and including these weights into the image-on-scalar regression model, we aimed to obtain associate estimates that are generalizable to the broader population, rather than limited to the study sample. We developed new inference methods and implemented the process of accounting for weights in voxelwise analysis frameworks. Weighted estimates will be different from those unweighted when the model fails to correctly specify the correlation structure between the predictors and outcomes. When the model has omitted unmeasured modifiers, whose distributions are different between the sample and the population, the estimated effects will be biased and cannot be generalizable to the population. Our comparison between weighted and unweighted regression models highlighted

how these methodological choices affect the inference of brain–cognitive associations and their broader relevance.

The whole brain analysis results from newly weighted and unweighted image-on-scalar regression revealed valuable insights into the association between cognitive behavior and working memory activation, as well as the impact of sample weights on model inference and generalizability. In our image-on-scalar regression analysis, the association between the g-factor and the 2-back versus 0-back contrast was interpreted as how cognitive abilities (g-factor) affect brain activity in trials with high memory demands compared to those with low memory demands. For example, identifying positive associations highlighted specific brain regions, particularly the frontoparietal regions, which showed increased activation during trials requiring greater memory effort, especially in individuals with higher general intelligence. Conversely, negative associations in the sensorimotor regions and the ventromedial prefrontal cortex revealed decreased activation during trials that were less demanding in terms of memory.

Regardless of whether sample weights were applied, our vertex detection results showed that specific brain regions, primarily frontal, parietal, and cingulate regions, had a higher proportion of detected vertices with significant associations than other regions. These vertices were components of several broad-scale functional networks, including the FPN and DAN. We observed a positive association between the g-factor and the 2-back versus 0-back contrast, indicating that adolescents with higher out-of-scanner general cognitive ability exhibited increased brain activity during high versus low memory load tasks in regions overlapping with the FPN and DAN. These findings are consistent with previous research (Finn et al., 2017; Rosenberg et al., 2020; Menon and D'Esposito, 2022). Conversely, we also observed negative associations within the DMN, CON, and PON, which are consistent

with previous studies demonstrating an 'anticorrelation' between these networks in healthy adolescents, as evidenced by other analyses using ABCD data (Owens et al., 2020; Daamen et al., 2015).

In addition to the general findings on the association between working memory and cognitive ability among adolescents, our data analysis revealed the impact of applying sample weights in image-on-scalar regression, offering a deeper understanding of the methodological considerations in such analyses. Notably, the unweighted regression method tended to identify more vertices with significant associations compared to the weighted method, particularly in association cortices. There are two potential interpretations for regions detected exclusively by unweighted regression but not by weighted regression: 1) Sample-specific significance: These regions, including the insula, middle temporal gyrus, and cingulate cortex, may be related to the working memory task but are only significant within the ABCD imaging subsample. They are not significant once sample weights are incorporated, and therefore, they are not likely generalizable to the broader population. 2) Potential false discoveries: Some regions might be identified due to potential false discoveries. For instance, the primary motor region, primary somatosensory-hand cortex, and primary auditory cortex are not typically associated with working memory and cognitive tasks. On the other hand, the inferior temporal gyrus was detected only by the weighted image-on-scalar regression model and not by the unweighted model. Since the inferior temporal gyrus is commonly associated with working memory and cognitive ability according to Fiebach et al. (2007), this result suggests that incorporating sample weights in image-on-scalar regression may help identify brain regions with associations that are significant in larger populations but may be undetected when analyzing subsamples.

Addressing sociodemographic variables provides only partial mitigation for selection

bias. Our analyses showed that differences between weighted and unweighted regression persisted, even after adjusting for sociodemographic covariates, particularly in the lateral occipital and cingulate cortex. This indicates that adjusting for control variables by only including their main effects does not necessarily mitigate these differences and highlights the existence of heterogeneity in brain-covariate associations within the imaging subsamples. This aligns with the results from our previous work (Lin et al., 2024), in which we identified four subgroups within the ABCD Study that present distinct patterns of brain activities and associations with the g-factor during working memory tasks. The allocation probabilities of individuals to different subgroups were influenced by sociodemographic factors such as race, income, and parental education, demonstrating the heterogeneity in associations between cognitive function and brain activity across different sociodemographic profiles. The sociodemographic characteristics and the weights are likely to moderate the associations, where high-order interaction terms can be highly predictive and should be included in the model. Unmeasured variables that simultaneously impact sampling and outcomes can further bias estimates, especially when they serve as effect modifiers.

Simulation studies reinforced the necessity of sample weighting, particularly when models are mis-specified and do not include all relevant interactions. This is crucial because, in real-world neuroimaging studies, it is often impractical to include all potential interaction terms to account for all possible confounders. By addressing selection biases, weighted image-on-scalar regression models yield robust statistical inferences, with reasonable coverage rates of 95% CIs and low FDR values. Incorporating weights into the regression framework enhances the accuracy and reliability of parameter estimation. These advantages are most pronounced in mis-specified models that omit interactions between exposure and control variables. Results from our simulation experiments clarified the relative per-

formance of weighted and unweighted approaches under different model specifications. In scenarios with correctly specified models, both methods performed similarly with negligible differences in bias, MSE, and coverage. However, in mis-specified models, weighting markedly improved estimation robustness, reducing bias and enhancing coverage. Thus, sample weights are particularly critical in analytic contexts where the model structure may not fully reflect the underlying data.

In conjunction with our real-world analyses using the ABCD data, we observed substantial differences when utilizing weighted methods. The simulation results support the need for weighting, especially under model mis-specification. However, several limitations must be acknowledged. First, the effect of weighting depends on the model specification and thereby depends on the associations and outcomes of interest. Our substantive findings on the use of weights cannot be simply transported to other contexts. Second, selection bias adjustment does not address issues relating to data quality; for example, ABCD release 2.0.1 contains preprocessing errors and missing behavioral measures that were not resolved in our analysis due to practical constraints on data updates. Our weighting construction and outcome model did not include parental education as a predictor due to large amounts of missing values. Third, population weighting adjustment aims for overall population representative inference, but not for subgroups, especially groups with small sample sizes. Subgroup estimates would require models beyond weighting adjustments. Finally, our research centered on association estimates with external validity rather than on causal inference, which requires both internal and external validity. Achieving internal validity would necessitate analytic approaches that explicitly account for confounding factors related to both exposure assignment and outcomes, in addition to addressing selection and outcome factors for external validity Si et al. (2024).

Several directions for future research can improve our weighted image-on-scalar regression models. First, our model currently only applies to cortical surface fMRI data. For tasks related to reward circuits, subcortical regions are also important, so extending the model to accommodate subcortical data would be a meaningful development. Second, we use the g-factor as the cognitive measure, and future investigation can examine alternative measures, such as the general cognitive ability assessment. Third, our analysis uses the ABCD Study data and targets for the population of 9 or 10 year old U.S. children. Extending our work to other study cohorts with different data collection paradigms, such as the UK Biobank Study with pseudo-weights for population representation, will be crucial to advance population neuroscience research. Lastly, future work is needed to develop open-source software interface and improve the computational efficiency, scalability, and reproducibility when fitting weighted image-on-scaler regression models with practical application studies.

Author Contributions

YS: conceptualization, supervision, project administration, resources, methodology, formal analysis, validation, writing—original draft, writing—review and editing. ZL: conceptualization, software, methodology, validation, data curation, formal analysis, writing—original draft. JK: conceptualization, supervision, resources, methodology, formal analysis, validation, writing—review and editing. MFM: writing—review and editing. CS: data curation, writing—review and editing.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgments

The work was supported by the National Institutes of Health (NIH) with grants: R21HD105204, U01MD017867 and T32AA007477.

The Adolescent Brain Cognitive Development (ABCD) Study (`abcdstudy.org`) is a multisite, longitudinal study designed to recruit more than 10,000 children age 9 to 10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the NIH and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, and U24DA041147. A full list of supporters is available at `abcdstudy.org/nih-collaborators`. A listing of participating sites and a complete listing of the study investigators can be found at `abcdstudy.org/principal-investigators.html`. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in 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.

Data availability

The ABCD data used in this paper can be obtained from the NIMH Data Archive (NDA) under Collection 2573 with a data use agreement. The code is available at GitHub: `https://github.com/zikaiLin/weighted_image_on_scalar`.

References

Akshoomoff, N., J. L. Beaumont, P. J. Bauer, S. S. Dikmen, R. C. Gershon, D. Mungas, J. Slotkin, D. Tulsky, S. Weintraub, P. D. Zelazo, et al. (2013). VIII. NIH toolbox cognition battery (CB): Composite scores of crystallized, fluid, and overall cognition. *Monographs of the Society for Research in Child Development 78*(4), 119–132.

Binder, D. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 279–292.

Casey, B., T. Cannonier, M. I. Conley, A. O. Cohen, D. M. Barch, M. M. Heitzeg, M. E. Soules, T. Teslovich, D. V. Dellarco, H. Garavan, et al. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience 32*, 43–54.

Christophel, T. B., P. C. Klink, B. Spitzer, P. R. Roelfsema, and J.-D. Haynes (2017). The distributed nature of working memory. *Trends in Cognitive Sciences 21*(2), 111–124.

Cole, S. R. and E. A. Stuart (2010). Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *American Journal of Epidemiology 172*, 107–115.

Daamen, M., J. G. Bäuml, L. Scheef, C. Sorg, B. Busch, N. Baumann, P. Bartmann, D. Wolke, A. Wohlschläger, and H. Boecker (2015). Working memory in preterm-born adults: Load-dependent compensatory activity of the posterior default mode network. *Human Brain Mapping 36*(3), 1121–1137.

Diebolt, J. and E. H. Ip (1996). Stochastic EM: Method and application. In *Markov chain Monte Carlo in practice*, pp. 259–273. Springer.

Dotson, V. M. and A. Duarte (2020). The importance of diversity in cognitive neuroscience. *Annals of the New York Academy of Sciences 1464*(1), 181–191.

Falk, E. B., L. W. Hyde, C. Mitchell, J. Faul, R. Gonzalez, M. M. Heitzeg, D. P. Keating, K. M. Langa, M. E. Martz, J. Maslowsky, et al. (2013). What is a representative brain? Neuroscience meets population science. *Proceedings of the National Academy of Sciences 110*(44), 17615–17622.

Fiebach, C., A. Friederici, E. Smith, and D. Swinney (2007). Lateral inferotemporal cortex maintains conceptual—semantic representations in verbal working memory. *Journal of Cognitive Neuroscience 19*(12), 2035–2049.

Finn, A. S., J. E. Minas, J. A. Leonard, A. P. Mackey, J. Salvatore, C. Goetz, M. R. West, C. F. Gabrieli, and J. D. Gabrieli (2017). Functional brain organization of working memory in adolescents varies in relation to family income and academic achievement. *Developmental Science 20*(5), e12450.

Garavan, H., H. Bartsch, K. Conway, A. Decastro, R. Goldstein, S. Heeringa, T. Jernigan, A. Potter, W. Thompson, and D. Zahs (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience 32*, 16–22.

Gard, A. M., L. W. Hyde, S. G. Heeringa, B. T. West, and C. Mitchell (2023). Why weight? Analytic approaches for large-scale population neuroscience data. *Developmental Cognitive Neuroscience 59*, 101196.

Ghosal, S. and A. Van der Vaart (2017). *Fundamentals of nonparametric Bayesian inference*, Volume 44. Cambridge University Press.

Gordon, E., T. Laumann, B. Adeyemo, J. Huckins, W. Kelley, and S. Petersen (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral cortex 26*(1), 288–303.

Gray, J. R., C. F. Chabris, and T. S. Braver (2003). Neural mechanisms of general fluid intelligence. *Nature neuroscience 6*(3), 316–322.

Hagler Jr, D. J., S. Hatton, M. D. Cornejo, C. Makowski, D. A. Fair, A. S. Dick, M. T. Sutherland, B. Casey, D. M. Barch, M. P. Harms, et al. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage 202*, 116091.

Han, Z., G. Yang, T. Liu, S. Funahashi, X.-N. Zuo, and T. Yan (2025). Enriching population diversity in neuroscience. *Science Bulletin 70*(16), 2560–2564.

Hawes, S. W., A. K. Littlefield, D. A. Lopez, K. J. Sher, E. L. Thompson, R. Gonzalez, L. Aguinaldo, A. R. Adams, M. Bayat, A. L. Byrd, L. F. C. de Araujo, A. Dick, S. F. Heeringa, C. M. Kaiver, S. M. Lehman, L. Li, J. Linkersdörfer, T. J. Maullin-Sapey, M. C. Neale, T. E. Nichols, S. Perlstein, S. F. Tapert, C. E. Vize, M. Wagner, R. Waller, and W. K. Thompson (2025). Longitudinal analysis of the ABCD® study. *Developmental Cognitive Neuroscience 72*, 101518.

Heeringa, S. G. and P. A. Berglund (2020). A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) Study baseline data. *BioRxiv*, 2020–02.

LeWinn, K., M. Sheridan, K. Keyes, A. Hamilton, and K. McLaughlin (2017). Sample composition alters associations between age and brain structure. *Nat Communication 8*(1), 874.

Li, L., M. Bayat, T. B. Hayes, W. K. Thompson, A. M. Gard, and A. Dick (2025). Missing data approaches for longitudinal neuroimaging research: Examples from the adolescent brain and cognitive development (abcd) study. *Developmental Cognitive Neuroscience 74*.

Li, X., L. Wang, H. J. Wang, and A. D. N. Initiative (2020). Sparse learning and structure identification for ultrahigh-dimensional image-on-scalar regression. *Journal of the American Statistical Association*, 1–15.

Lin, Z., Y. Si, and J. Kang (2024). Latent subgroup identification in image-on-scalar regression. *Annals of Applied Statistics 18*(1), 468–486.

Liu, Y., B. Yan, K. Merikangas, and H. Shou (2017). Total variation regularized tensor-on-scalar regression. *arXiv preprint arXiv:1703.05264*.

Menon, V. and M. D'Esposito (2022). The role of PFC networks in cognitive control and executive function. *Neuropsychopharmacology 47*(1), 90–103.

Nichols, T. and S. Hayasaka (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research 12*(5), 419–446.

Owens, M. M., D. Yuan, S. Hahn, M. Albaugh, N. Allgaier, B. Chaarani, A. Potter, and H. Garavan (2020). Investigation of psychiatric and neuropsychological correlates of default mode network and dorsal attention network anticorrelation in children. *Cerebral Cortex 30*(12), 6083–6096.

Rosenberg, M. D., S. A. Martinez, K. M. Rapuano, M. I. Conley, A. O. Cohen, M. D. Cornejo, D. J. Hagler, W. J. Meredith, K. M. Anderson, T. D. Wager, et al. (2020). Behavioral and neural signatures of working memory in childhood. *Journal of Neuroscience 40*(26), 5090–5104.

Rudolph, J., Y. Zhong, P. Duggal, S. Mehta, and L. B (2023). Defining representativeness of study samples in medical and population health research. *BMJ Medicine 2*(1), e000399.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics 1*(4), 381–397.

Si, Y., G. Bandoli, K. M. Cole, M. Daniele Fallin, E. A. Stuart, K. K. Gurka, K. N. Althoff, and W. K. Thompson (2024). Advancing high quality longitudinal data collection: Implications for the HEALthy Brain and Child Development (HBCD) Study design and recruitment. *Developmental Cognitive Neuroscience 69*, 101432.

Si, Y., S. Lee, and S. G. Heeringa (2024). Population weighting in statistical analysis. *JAMA Internal Medicine 184*(1), 98–99.

Si, Y., R. J. Little, Y. Mo, and N. Sedransk (2023). A case study of nonresponse bias analysis in educational assessment surveys. *Journal of Educational and Behavioral Statistics 48*(3), 271–295.

Si, Y., R. J. Little, Y. Mo, and N. Sedransk (2024). Nonresponse bias analysis in longitudinal studies: A comparative review with an application to the Early Childhood Longitudinal Study. *International Statistical Review 92*(3), 383–405.

Si, Y., B. T. West, P. Veliz, M. E. Patrick, J. E. Schulenberg, D. D. Kloska, Y. M. Terry-McElrath, and S. E. McCabe (2022). An empirical evaluation of alternative approaches

to adjusting for attrition when analyzing longitudinal survey data on young adults' substance use trajectories. *International Journal of Methods in Psychiatric Research 31*(3), e1916.

Sripada, C., A. M. S. Gard, M. Angstadt, A. Taxali, T. Greathouse, K. McCurry, L. W. Hyde, A. Weigard, P. Walczyk, and M. Heitzeg (2022). Socioeconomic resources are associated with distributed alterations of the brain's intrinsic functional architecture in youth. *bioRxiv*.

Sternberg, R. J. and E. L. Grigorenko (2002). *The general factor of intelligence: How general is it?* Psychology Press.

Thompson, W. K., D. M. Barch, J. M. Bjork, R. Gonzalez, B. J. Nagel, S. J. Nixon, and M. Luciana (2019). The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery. *Developmental Cognitive Neuroscience 36*, 100606.

Waiter, G. D., I. J. Deary, R. T. Staff, A. D. Murray, H. C. Fox, J. M. Starr, and L. J. Whalley (2009). Exploring possible neural mechanisms of intelligence differences using processing speed and working memory tasks: An fMRI study. *Intelligence 37*(2), 199–206.

West, B. T., Y. Si, Y. Hu, S. E. McCabe, and P. V. and (2025). The role of weighting adjustment for attrition in longitudinal trajectory modeling: a simulation study. *Communications in Statistics - Simulation and Computation 54*(3), 866–888.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, 817–838.

Williams, C. K. and C. E. Rasmussen (2006). *Gaussian processes for machine learning.* MIT press Cambridge, MA.

Yu, S., G. Wang, L. Wang, and L. Yang (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Statistica Sinica 31*(3), 1463–1487.

Zeng, Z., M. Li, and M. Vannucci (2022). Bayesian image-on-scalar regression with a spatial global-local spike-and-slab prior. *Bayesian Analysis 1*(1), 1–26.

Zhang, D., L. Li, C. Sripada, and J. Kang (2020). Image-on-scalar regression via deep neural networks. *arXiv preprint arXiv:2006.09911*.

Zhu, H., J. Fan, and L. Kong (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association 109*(507), 1084–1098.

A Distributions of population weights

We include the frequent distributions of the two sets of weights in Figure 7.

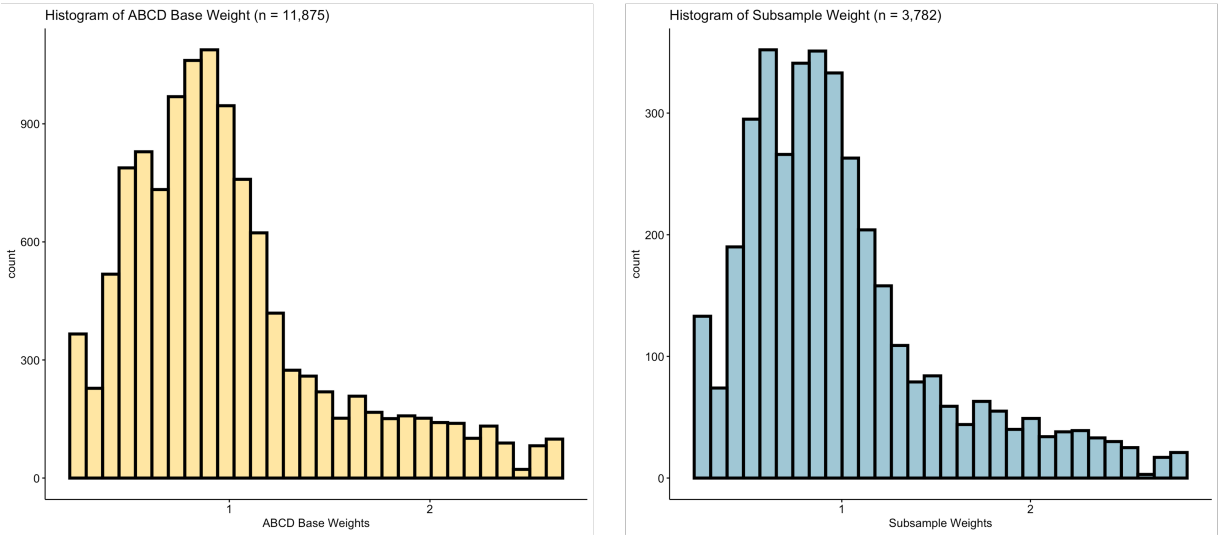

Figure 7: Frequency distributions of the base weights and the subsample weights.