Mapping Individual Differences in Intermodal Coupling in Neurodevelopment

Ruyi Pan[1,2,†], Sarah M. Weinstein[3,†], Danni Tu[4,5,†], Fengling Hu[5], Büşra Tanrıverdi[5], Rongqian Zhang[1], Simon N. Vandekar[6], Erica B. Baller[7,8], Ruben C. Gur[7,9], Raquel E. Gur[7,9], Aaron F. Alexander-Bloch[7,9,10], Theodore D. Satterthwaite[7,8,9], and Jun Young Park*[1,11]

[1]Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada
[2]Centre for Addiction and Mental Health, Toronto, ON, Canada
[3]Department of Epidemiology and Biostatistics, Temple University College of Public Health, Philadelphia, PA, USA
[4]Regeneron Pharmaceuticals, Tarrytown, NY, USA
[5]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, USA
[6]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA
[7]Department of Psychiatry, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA
[8]Penn Lifespan Informatics and Neuroimaging Center, Philadelphia, PA, USA
[9]Penn-CHOP Lifespan Brain Institute (LiBI), Philadelphia, PA, USA
[10]Department of Child and Adolescent Psychiatry and Behavioral Science, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[11]Department of Psychology, University of Toronto, Toronto, ON, Canada
[†]These authors contributed equally.

Abstract

Within-individual coupling between measures of brain structure and function evolves in development and may underlie differential risk for neuropsychiatric disorders. Despite increasing interest in the development of structure-function relationships, rigorous methods to quantify and test individual differences in coupling remain nascent. In this article, we explore and address gaps in approaches for testing and spatially localizing individual differences in intermodal coupling, including a new method, called CEIDR. CEIDR controls false positives in individual differences in intermodal correlations that arise from mean and variance heterogeneity and improves statistical power by adopting adaptive cluster enhancement. Through a comparison across different approaches to testing individual differences in intermodal coupling, we delineate subtle differences in the hypotheses they test, which may ultimately lead researchers to arrive at different results. Finally, we illustrate these differences in two applications to brain development using data from the Philadelphia Neurodevelopmental Cohort.

Keywords: individual differences, coupling, neurodevelopment, multimodal neuroimaging, hypothesis testing

*Corresponding author: junjy.park@utoronto.ca

1 Introduction

During adolescence, the developing brain undergoes a profound structural and functional reorganization. To understand how the brain changes during childhood, researchers increasingly use multimodal neuroimaging to measure both distinct and shared spatial patterns in features of brain anatomy and function. Measures of *intermodal coupling* have been increasingly explored in recent research (Avants et al. 2010; Devonshire et al. 2012; Ouyang et al. 2015; Phillips et al. 2016; Vandekar et al. 2016; Iadecola 2017; Baum et al. 2020; Baller et al. 2022; Hu et al. 2022). Unlike studying individual differences separately in different brain modalities, the use of intermodal coupling allows researchers to study relationships between brain features (e.g., structure-function) that are thought to be critical for mechanisms underlying risk for developmental psychopathology, cognition, and brain health across the lifespan.

Despite the emerging utility of intermodal coupling measures in developmental cognitive neuroscience research, methods for conducting statistical testing on individual differences in coupling have yet to be assessed in terms of their statistical power (i.e., replicability) and interpretability. Also, to the best of our knowledge, no clear generative model currently exists that parametrizes intermodal coupling and its individual differences, which may introduce unwanted variability in results due to limitations in the analysis methods. Amid growing concerns about reproducibility in neuroimaging-based studies of neurodevelopment (Flournoy et al. 2020; Klapwijk et al. 2021; Marek et al. 2022; Botvinik-Nezer et al. 2020), this article seeks to concretize both the formulation and evaluation of hypotheses about intermodal coupling. To this end, we use the lens of statistical hypothesis testing to formulate scientific questions about individual differences in intermodal coupling. Even in the presence of confounding and other sources of noise in imaging data, such tests can be useful for identifying both *if* and *where* individual differences in coupling exist.

To illustrate the broad applicability of our method, we leverage data from a large neuroimaging cohort to investigate how intermodal coupling evolves in the course of healthy brain development. Specifically, we first focus on intermodal coupling between cortical thickness (CT) and sulcal depth (SD), measured using structural MRI. The overall volume of the cortex generally increases in early childhood, then peaks and decreases during late childhood and adolescence (Raznahan et al. 2011). This decrease in volume is mainly driven by cortical thinning that preserves surface area (Kelly et al. 2023), although thinning does not occur uniformly across the brain, with some prior work demonstrating dissociable processes of thinning in the sulci but thickening in gyri (Vandekar et al. 2016). Therefore, a high-resolution characterization of the relationship between cortical thickness and cerebral surface anatomy may shed light on the mechanisms supporting neurodevelopment.

To demonstrate the generalizability of our method across feature types, we also evaluate whether the relationship between amplitude of low-frequency fluctuations (ALFF) and cerebral blood flow (CBF). From blood oxygen level dependent (BOLD) functional magnetic resonance imaging (fMRI), we use ALFF signals as a proxy of neuronal activity; from arterial spin labelling, we can quantify neurovascular perfusion through CBF. Thus, for each individual, each location (e.g., voxel or vertex) in the brain image is associated with a value of CBF and ALFF, yielding two brain maps. Prior imaging studies focusing on CBF and ALFF have revealed age-related changes in neurologically healthy adolescents (Satterthwaite et al. 2014; Taki et al. 2011), as well as changes between healthy adults and people with cognitive impairment (Chen 2019; Hoptman et al. 2010), suggesting that both blood oxygenation and perfusion are critical for executive function. While CBF and ALFF brain

maps offer valuable insights into brain function individually, examining the coupling between them together provides a deeper understanding of neurovascular coupling. Recent research suggests that neurovascular coupling may support cognitive functioning (Stackhouse and Mishra 2021; Phillips et al. 2016); in particular, Baller et al. 2022 demonstrated that CBF-ALFF coupling may be linked to changes in executive function in healthy adolescents, and that this coupling was sex-, age-, and region-dependent.

The overarching goal of this article is to describe statistical approaches that may be used to quantify, test, and spatially localize individual differences in intermodal coupling. We then propose several extensions to existing methodology. First, we conceptualize the symmetric relationship between two imaging modalities as a conditional correlation (Yan and Fine 2004; Tu et al. 2024), which can vary by age, sex, or other variables of interest. Finally, we propose CEIDR (**C**luster **E**nhancement for testing **I**ndividual **D**ifferences in ρ (**r**)), a statistical inference procedure to test individual differences in (conditional) intermodal correlations, which we illustrate as achieving high power and spatial localization. We apply our method to multi-modal MRI data from the Philadelphia Neurodevelopmental Cohort (PNC) (Satterthwaite et al. 2014), delineating both robust developmental effects and sex differences in intermodal coupling.

2 Methods

2.1 Notation

For subject $i = 1, \ldots, N$ and location (i.e., vertex or voxel) $v = 1, \ldots, V$, we denote brain measurements for two modalities as $y_{1i}(v)$ and $y_{2i}(v)$, respectively. And, we let x_i represent the covariate of interest and $\mathbf{z}_i$ represent the nuisance covariate vector being adjusted for subject i. Categorical variables could be dummy-coded to be used in statistical modeling in our paper.

2.2 Parametric model and the null hypothesis for conditional correlation

To clearly define subject-level intermodal coupling, we consider the following parametric model. For the mth imaging modality, we consider the following mean and variance functions depending on covariates (i.e., $\mu_m(\mathbf{z}, x_i, v)$ and $\sigma_m(\mathbf{z}_i, x_i, v)$) for a model for $y_{mi}(v)$:

$$y_{mi}(v) = \underbrace{\mu_m(\mathbf{z}_i, x_i, v)}_{\substack{\text{Individual differences} \\ \text{in means}}} + \underbrace{\sigma_m(\mathbf{z}_i, x_i, v)}_{\substack{\text{Individual differences} \\ \text{in variances}}} \times \epsilon_{mi}(v), \tag{1}$$

where the residual terms $\epsilon_{1i}(v)$ and $\epsilon_{2i}(v)$ follow the joint distribution given by:

$$\begin{pmatrix} \epsilon_{1i}(v) \\ \epsilon_{2i}(v) \end{pmatrix} \sim \mathcal{MVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho(\mathbf{z}_i, x_i, v) \\ \rho(\mathbf{z}_i, x_i, v) & 1 \end{bmatrix} \right). \tag{2}$$

Here, the residual correlation $\rho(\mathbf{z}_i, x_i, v)$ is the subject-level intermodal coupling of our interest. We hypothesize that $\rho(\mathbf{z}_i, x_i, v)$ for subject i is a function of the variable of interest x_i (e.g., age) and

other potential confounding variables $\mathbf{z}_i$ (e.g., sex), given by:

$$\rho(\mathbf{z}_i, x_i, v) = \tanh(\underbrace{\beta_0(v) + \mathbf{z}_i\boldsymbol{\beta}_1(v) + x_i \cdot \kappa(v)}_{\substack{\text{Individual differences} \\ \text{in correlations}}}) \tag{3}$$

where $\tanh(\cdot)$ is a hyperbolic tangent link function ensuring that the correlations lie strictly between -1 and 1 (Yan and Fine 2004; Tu et al. 2024).

Given our model formulations in Equations (1), (2), and (3), we consider the following null hypothesis:

$$H_0 : \kappa(v) = 0 \text{ for all } v = 1, \ldots, V. \tag{4}$$

From the model, we note that individual differences between two imaging modalities may arise in the means $\mu_m(\mathbf{z}_i, x_i, v)$ and variances $\sigma_m(\mathbf{z}_i, x_i, v)$ from each modality, as well as their 'residual' correlations $\rho(\mathbf{z}_i, x_i, v)$. Therefore, a nonzero $\kappa(v)$ indicates that intermodal coupling at vertex v is associated with the variable of interest x_i after both (i) adjusting for covariate effects on mean and variance structures and (ii) adjusting for other nuisance covariate effects from $\mathbf{z}_i$ on residual correlations.

2.3 When could false positives in intermodal correlation be inflated?

As correlation naturally depends on mean and variance, the models for mean and variance should be correctly specified to accurately specify correlation models and to prevent possibly inflated false positives. We point out two possible misspecifications in mean and variances. First, the "observed covariate effects" on mean and variance could be misspecified when, for example, there are non-linear age effects on imaging data but non-flexible models (e.g., linear age effects) are assumed in Equation (1) (Fjell et al. 2009; Ziegler et al. 2012; Tamnes et al. 2013). A relevant case would be that an interaction between covariates is present but left out in the model. Second, it is possible that some important covariates are not collected in the study design and therefore cannot be included properly in Equation (1) ("unobserved covariate effects" on mean and variance).

2.4 CEIDR

We develop a powerful method to test the null hypothesis in Section 2.2, ensuring robust model specification, high sensitivity through cluster enhancement, and proper false positive control through permutation. We outline the five main stages of CEIDR below (See Figure 1 for illustrations).

I. *Removing unrelated data variability.* To mitigate positive findings that may be driven by unrelated variability in the marginal means and variances of each modality, we propose two approaches for performing marginal mean and variance adjustment.

   I.1 *Between-subject adjustment.* We use information from the full sample to adjust for covariates like age and sex. Supported by the nonlinear brain structural dynamics found in prior neuroimaging studies (Ziegler et al. 2012; Tamnes et al. 2013), we fit the generalized additive model (GAM), which allows for flexible specification of both linear and nonlinear

covariate effects, to estimate individual means $\mu_m(\mathbf{z}_i, x_i, v)$ and variances $\sigma_m(\mathbf{z}_i, x_i, v)$ of each modality at each vertex. In this article, we consider μ_m to be modeled by a main effect of sex and a cubic-spline of age stratified by sex, and $\log(\sigma_m)$ to be modeled by the main effect of age and sex, which can be specified in the `gamlss` package in R. Once $\mu_m(\mathbf{z}_i, x_i, v)$ and $\sigma_m(\mathbf{z}_i, x_i, v)$ are estimated, we obtain the standardized residuals as follows:

$$\hat{\epsilon}_{mi}(v) = \frac{y_{mi}(v) - \hat{\mu}_m(\mathbf{z}_i, x_i, v)}{\hat{\sigma}_m(\mathbf{z}_i, x_i, v)}. \tag{5}$$

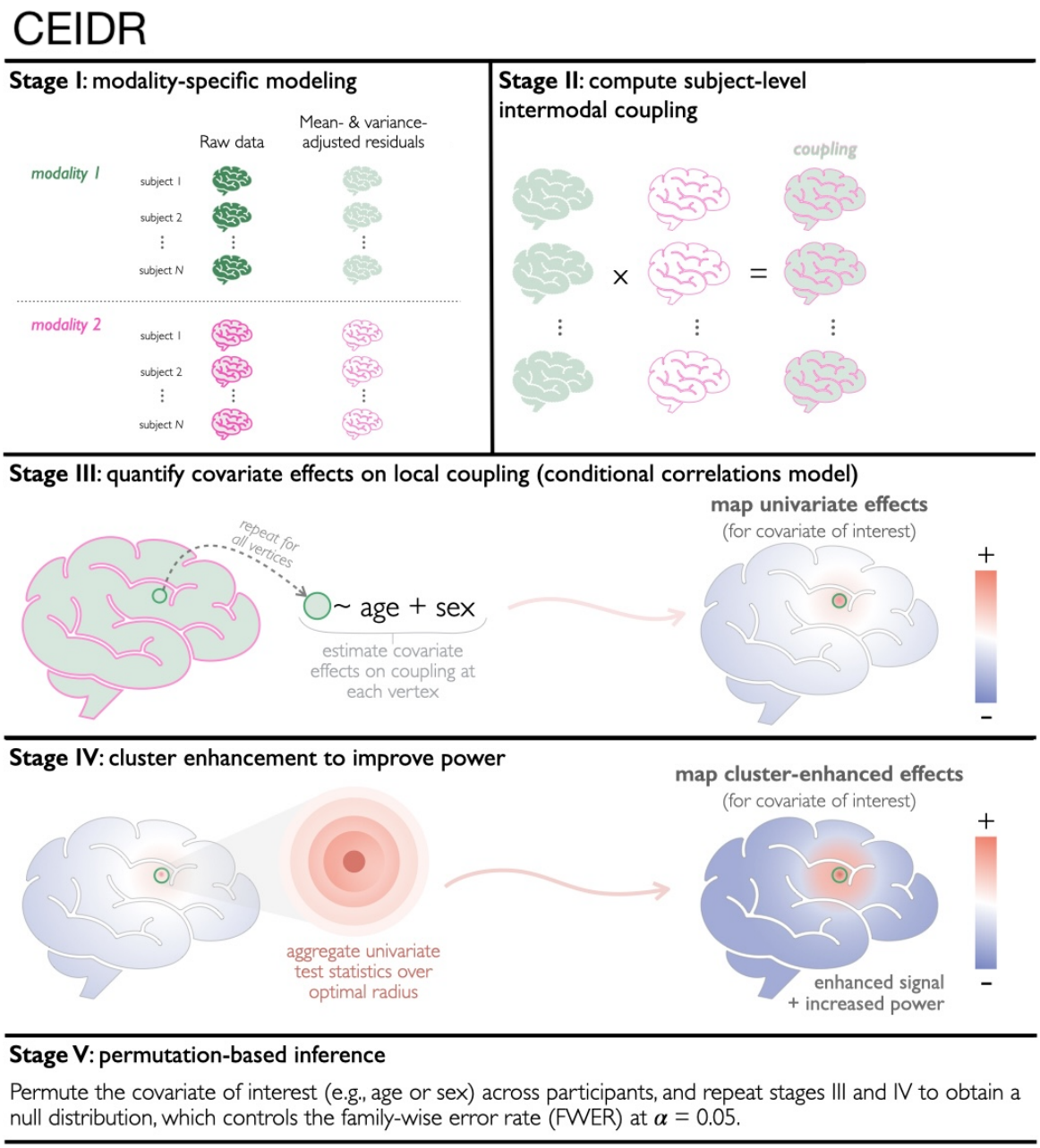


Figure 1. Overview of the proposed method, CEIDR.

I.2 *Within-subject adjustment.* One can also use local information within each subject to adjust the individual means and variances. For each subject, assuming local neighbors have the same means and variances, we use the sample mean and variance within a local region $N_r(v)$, i.e., the set of all vertices whose distances from the vertex v is less than or equal to r, to estimate the subject-specific mean and variance at vertex v and do adjustment:

$$\hat{\epsilon}_{mi}(v) = \frac{y_{mi}(v) - \bar{y}_{mi}(v)}{\sqrt{\frac{1}{|N_r(v)|-1}\sum_{v^\star \in N_r(v)}(y_{mi}(v^\star) - \bar{y}_{mi}(v))^2}}, \tag{6}$$

where $\bar{y}_{mi}(v) = \frac{1}{|N_r(v)|}\sum_{v^\star \in N_r(v)} y_{mi}(v^\star)$. We selected a radius of $r = 5$mm as the default value in this article, which is suitable for achieving robust coupling estimations. The 5mm radius remains appropriate for this resolution, as most vertices with approximately 10 neighbors (the median of $|N_5(v)|$ across all vertices in fsaverage5 is 11).

One can choose either approach depending on the study context. Additionally, the two strategies can be combined: first applying the between-subject adjustment using observed covariates, and then performing a within-subject adjustment (on $\hat{\epsilon}_{mi}(v)$ from Equation (5)) to account for potential unknown factors that may still influence the marginal mean and variance. In this paper, we use both between- and within-subject adjustments in our data analysis.

II. *Estimating subject-level intermodal coupling.* We compute subject-level measures of intermodal associations using the product of the modality-specific residuals:

$$\hat{\rho}_i(v) = \hat{\epsilon}_{1i}(v) \times \hat{\epsilon}_{2i}(v). \tag{7}$$

III. *Modeling $\hat{\rho}_i(v)$ in terms of covariates.* We assume the covariates regulate $\rho_i(v)$ through $\tanh(\cdot)$ link function specified in Equation (3). Although $\tanh(\cdot)$ is not straightforward to model using a GLM, we observe that $\tanh(t) \approx t$ when $|t|$ is small or moderate (e.g., less than 0.5), which is consistent with the magnitude of correlations often found in neuroimaging studies (see Figures 3 and 4 in Section 3). Under this approximation, we model the relationship between intermodal coupling and covariates using a GLM formulation of

$$\mathrm{E}\left[\hat{\rho}_i(v)|x_i, \mathbf{z}_i\right] = \beta_0(v) + \mathbf{z}_i\boldsymbol{\beta}_1(v) + x_i \cdot \kappa(v). \tag{8}$$

From the formulation above, the null hypothesis $H_{0v} : \kappa(v) = 0$ can be tested by fitting the GLM and extracting the corresponding p-value.

IV. *Adaptive cluster enhancement.* To improve sensitivity, we perform adaptive cluster enhancement on the vertex-level score test statistics $U(v)$ to leverage local spatial similarity of the effect sizes and improve power (Park and Fiecas 2022; Weinstein et al. 2022; Pan et al. 2024). The score test statistic at each vertex is defined as

$$U(v) = \sum_{i=1}^{N} x_i \times \left(\hat{\rho}_i(v) - \tilde{\beta}_0(v) - \mathbf{z}_i\tilde{\boldsymbol{\beta}}_1(v)\right), \tag{9}$$

where $\tilde{\beta}_0(v), \tilde{\boldsymbol{\beta}}_1(v)$ are obtained under the null model (i.e., the model without any $x_i\kappa(v)$ terms).

The cluster-enhanced test statistic, denoted by $T(v)$, is constructed by the standardized average of the vertex-level test statistics: $T(v) = \max\{T_r(v)|r = 0\text{mm}, 1\text{mm}, \ldots, r_{\max}\}$ where

$$T_r(v) = \frac{\left(\sum_{v^\star \in N_r(v)} U(v^\star)\right)^2}{\widehat{\text{Var}}_{H_0}\left(\sum_{v^\star \in N_r(v)} U(v^\star)\right)}.$$

V. *Determining statistical significance.* We perform inference using a permutation-based procedure. This helps to prevent spurious associations (i.e., false positives) when working with high-dimensional neuroimaging features (Lindquist and Mejia 2015). Permutation allows us to (i) compare the observed relationship between coupling and x_i to observations generated from the null hypothesis, and (ii) account for multiple comparisons across all vertices by using a brain-wide threshold, t_α, to establish statistical significance.

Because our hypothesis of interest is the covariate effect on intermodal correlations and not on means or variances in each modality, we need to permute x_i (or $(x_i, \mathbf{z}_i)$ when covariates are highly correlated) across subjects in stage III to generate permuted samples, while all the other steps in stages I and II are fixed. Other permutation strategies could be considered, which we refer readers to Winkler et al. (2014). A major computational advantage of this is that there is no need to redo Stage I and Stage II in each permutation, which boosts computational efficiency. Therefore, all permutation analyses can begin with the residuals $\hat{\epsilon}_{1i}(v), \hat{\epsilon}_{2i}(v)$. As such, this permutation scheme becomes equivalent to using the CLEAN pipeline (Park and Fiecas 2022) in stage III. For each of 1000 (or more) permutations, Stages III-IV are repeated with x_i or $(\mathbf{z}_i, x_i)$ permuted across subjects, which controls the family-wise error rate at a predetermined α (e.g., 0.05).

2.5 Other comparable methods

2.5.1 IMCo

Vandekar et al. (2016) used locally weighted regression with pre-defined weights to estimate subject- and vertex-specific measures of intermodal coupling. That is, at each vertex, intermodal coupling was estimated by regressing one modality on the other using brain measurements. The weight is determined by (i) the distance between v and $v^\star$ denoted by $d_{v,v^\star}$ and (ii) user-specified parameter for controlling the spatial extent of neighborhood, the full width at half maximum (FWHM), with the specific form as $w_{v,v^\star} = \exp\{-d^2_{v,v^\star}/(2\theta^2)\}$ with $\theta = \text{FWHM}/(2\sqrt{2 \times \log(2)})$. As FWHM increases, vertices that are far apart from a vertex are given more weight in fitting locally weighted regression. This procedure yields a slope at each vertex, representing local intermodal coupling for each subject. These individual-level coupling maps could then be used to test for local associations with behavioral or demographic variables.

The regression slopes derived from earlier implementations of IMCo (Vandekar et al. 2016; Baller et al. 2022) are an inherently asymmetric measure of association, since one modality must be arbitrarily chosen as the dependent variable and the other as the independent variable (Hu et al. 2022). To circumvent this discrepancy, instead of locally weighted regression slopes, we estimate locally weighted *correlations* to define subject-level intermodal coupling (Figure 2(a)).

2.5.2 Mass-univariate

To emphasize the advantage of leveraging the spatial structure of the data (cluster enhancement), we also consider a "mass-univariate" approach. This involves estimating vertex-level test statistics measuring individual differences in intermodal coupling—but without leveraging neighboring information. In this article, our implementation of the mass-univariate framework is equivalent to CEIDR, when both Stage I.2 and Stage IV are excluded.

While CEIDR, IMCo, and the mass-univariate method each takes a different approach to *quantifying* subject-level intermodal coupling, we use the same permutation-based procedure for controlling FWER across the three methods. That is, given the estimate of $\hat{\rho}_i(v)$ from each method for Equation (8), we randomly permute the covariate of interest, x_i across subjects (after adjusting for confounders $\mathbf{z}_i$), and then we repeat the steps outlined in Figure 2(b)-(d) to estimate null effects of a covariate of interest on intermodal coupling.

2.6 Remarks

2.6.1 Comparing between-subject and within-subject adjustments

Although the between- and within-subject adjustments pursue the same goal, they rely on different assumptions and offer different advantages and limitations. The between-subject adjustment benefits from utilizing the full sample (e.g., more than 800 subjects in our real data analysis), which

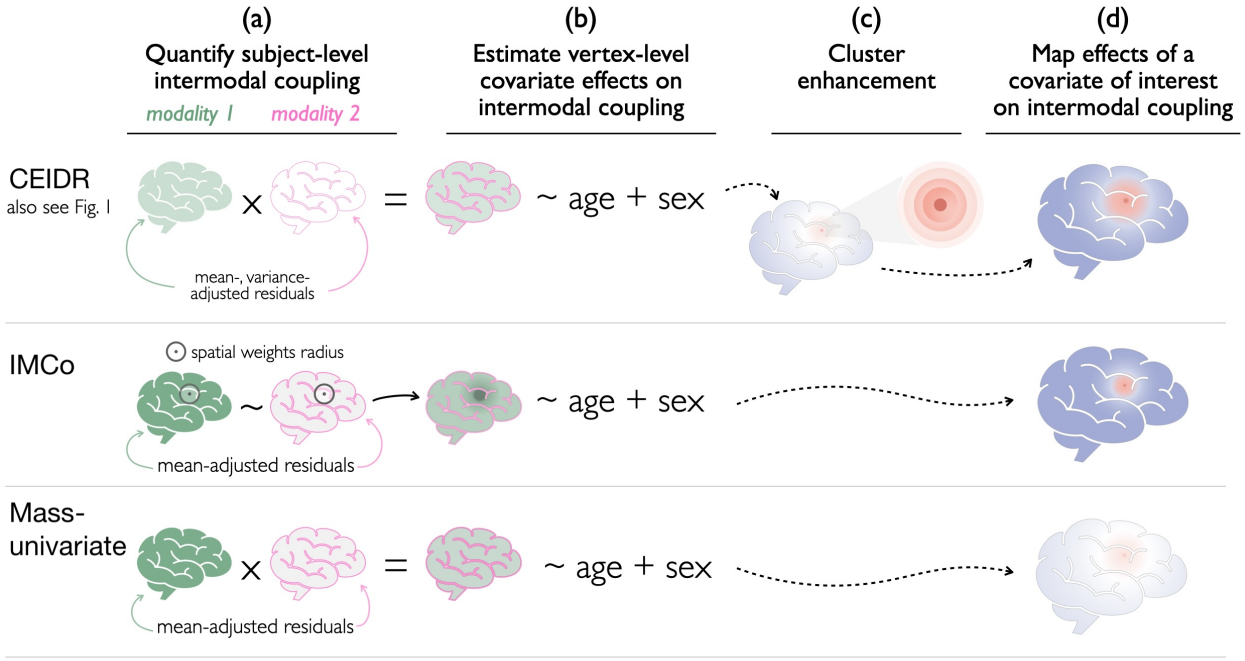


Figure 2. Overview of the proposed method, CEIDR, along with two alternative approaches to quantifying intermodal coupling (a-b) and individual differences (c-d). A more detailed illustration of CEIDR is provided in Figure 1. We perform statistical inference for all three methods using the same permutation-based procedure, which uses a brain-wide threshold for statistical significance in order to control the family-wise error rate (FWER) at $\alpha = 0.05$.

reduces variance of estimates. However, we note that GAM could still be misspecified and it does not address unknown covariate effects. The within-subject approach avoids reliance on covariate information, but its validity depends on the assumption that the mean and variance functions are the same within a specified neighborhood, which may be violated in practice, especially when neighbor size is large. Therefore, we recommend using a small radius to specify neighborhood (e.g., 5mm) to prevent false positives, which is used in CEIDR as a default. Although the number of vertices within a 5mm radius is small in fsaverage5 (median: 11 vertices in the pial surface), we note that adopting a higher-resolution would be helpful at the expense of increased burden in multiple comparison.

2.6.2 IMCo is a special case of CEIDR

In the Supplementary Material I, we derive that IMCo can be interpreted as a special case of CEIDR under a specific setup. When equal weights are assigned to all vertices within a predefined region $N_{r^\star}(v)$ and zero weights are assigned outside this region, IMCo becomes equivalent to CEIDR with the within-subject adjustment (i.e., without between-subject adjustment) in Step I and $T(v) = T_r^\star(v)$ in Step IV (i.e., without *adaptive* cluster enhancement). This explains when IMCo would suffer from inflated false positives or reduced power. Specifically, when the specified full-width at half-maximum (FWHM) is large (e.g., 15mm), it would lead to biased adjustment of mean and variance (see Section 2.6.1), and power loss is expected when signal region is smaller. Similarly, when the specified FWHM is small (e.g., 5mm), false positive rate would remain appropriately controlled, but power loss is expected when signal region is larger. Therefore, the ability to separately define (i) radius used to adjust mean and variance and (ii) radii used to conduct adaptive cluster enhancement speaks to the practical utility of CEIDR.

3 Data analysis

3.1 The Philadelphia Neurodevelopmental Cohort study

The Philadelphia Neurodevelopmental Cohort is an open-source and multi-modality investigation of brain development in children aged 8 to 21 years (Satterthwaite et al. 2016). The PNC dataset includes structural neuroimaging, functional neuroimaging (resting-state and task-based fMRI), psychiatric evaluations, and a neurocognitive battery for each individual. This study is a collaboration between the Children's Hospital of Philadelphia and the Brain Behavior Laboratory at the University of Pennsylvania, and is funded by the National Institutes of Mental Health (NIMH). Informed consent was obtained from all participating adults and guardians of underage participants, and the study protocol received approval from the Institutional Review Board at both the University of Pennsylvania and the Children's Hospital of Philadelphia.

We apply our proposed method to multimodal neuroimaging data from the Philadelphia Neurodevelopmental Cohort (Satterthwaite et al. 2016), mapping age and sex differences in intermodal coupling between (i) cortical thickness (CT) and sulcal depth (SD) and (ii) cerebral blood flow (CBF) and amplitude of low-frequency fluctuations (ALFF). Individual differences in structural (CT-SD) and neurovascular (CBF-ALFF) coupling were also examined in previous studies involving the PNC data (Vandekar et al. 2016; Baller et al. 2022; Hu et al. 2022). Using similar exclusion criteria as in previous work, we include $N = 911$ participants in our analyses of CT-SD coupling

(Weinstein et al. 2021) and $N = 831$ participants in our analyses of CBF-ALFF coupling (Baller et al. 2022; Weinstein et al. 2022).

3.2 Imaging parameters

Resting-state fMRI data was gathered from $N = 1601$ participants using the same scanner, a single 3T Siemens TIM Trio whole-body scanner equipped with a 32-channel head coil. Images were collected using a whole-brain, single-shot, multi-slice, gradient-echo echoplanar sequence. The sequence parameters were as follows (124 volumes): time repetition = 3000 ms, time echo = 32 ms, field of view = 192 $\times$ 192 mm, matrix size = 64 $\times$ 64, number of slices = 46, slice thickness = 3 mm, slice gap = 0 mm, flip angle = 90 degrees, and voxel resolution = 3 $\times$ 3 $\times$ 3 mm.

Brain perfusion was imaged with a 3D-encoded spin-echo pseudocontinuous arterial spin labeling sequence. The sequence parameters for perfusion were as follows (80 volumes): time repetition = 4000 ms, time echo = 15 ms, field of view = 220 $\times$ 220 mm, matrix size = 96 $\times$ 96, number of slices = 20, slice thickness = 5mm; slice gap = 1mm, and resolution 2.3 $\times$ 2.3 $\times$ 6 mm. More detailed information on these modalities and image acquisition can be found in Satterthwaite et al. 2014.

Pre-processing of BOLD signal involved field inhomogeneity correction, registration, linear and quadratic denoising, and mitigation of motion artifacts through a confound regression model, as described in Ciric et al. 2017 and Weinstein et al. 2022. The resulting BOLD signal was filtered using a first-order Butterworth filter with a pass-band of 0.01 to 0.08 Hz in order to remove trends and noise associated with physiological processes (e.g., breathing). Then, ALFF was calculated as the sum of the amplitudes of the power spectrum over all bins in the 0.01 - 0.08 Hz band. As previously described in Baller et al. 2022, CBF was calculated using parameters derived from arterial spin labelling.

The brain measurements were mapped onto the fsaverage5 pial surface, which contains 10,242 vertices in each of the left and right hemispheres. The full set of vertices is used to compute the geodesic distance matrices separately for each hemisphere. During model fitting, we exclude the medial wall vertices—888 in the left hemisphere and 881 in the right hemisphere—resulting in 9,354 vertices in the left hemisphere and 9,361 vertices in the right hemisphere for analysis.

3.3 Analysis results and comparisons

We compared CEIDR to IMCo (with FWHM=15mm and FWHM=5mm) and the mass-univariate approach outlined in Section 2.5. In our analysis, the permutation approach outlined in step V of CEIDR was used in all methods considered to promote fair comparisons, although previous implementations of IMCo (Vandekar et al. (2016) and Baller et al. (2022)) controlled false discovery rate (FDR). We set FWER controlled at 5%. We also made $r_{\max} = 15$mm in step IV of CEIDR for a fairer comparison with IMCo with FWHM=15mm. We fitted and compared each model with two versions of Step I: (i) no adjustment made in mean and variance functions and (ii) sex-specific non-linear effects are modeled by splines and a log-linear model is applied for variance. For a single test, the total runtime for applying CEIDR on one hemisphere (e.g., left hemisphere) is around 12 minutes on MacBook Pro 2019 (2.3 GHz 8-core Intel Core i9 and 32GB RAM) without parallel computing: adjustment for mean and variance effects takes around 7 minutes, and permutation testing with 2,000 permutations takes roughly 5 minutes.

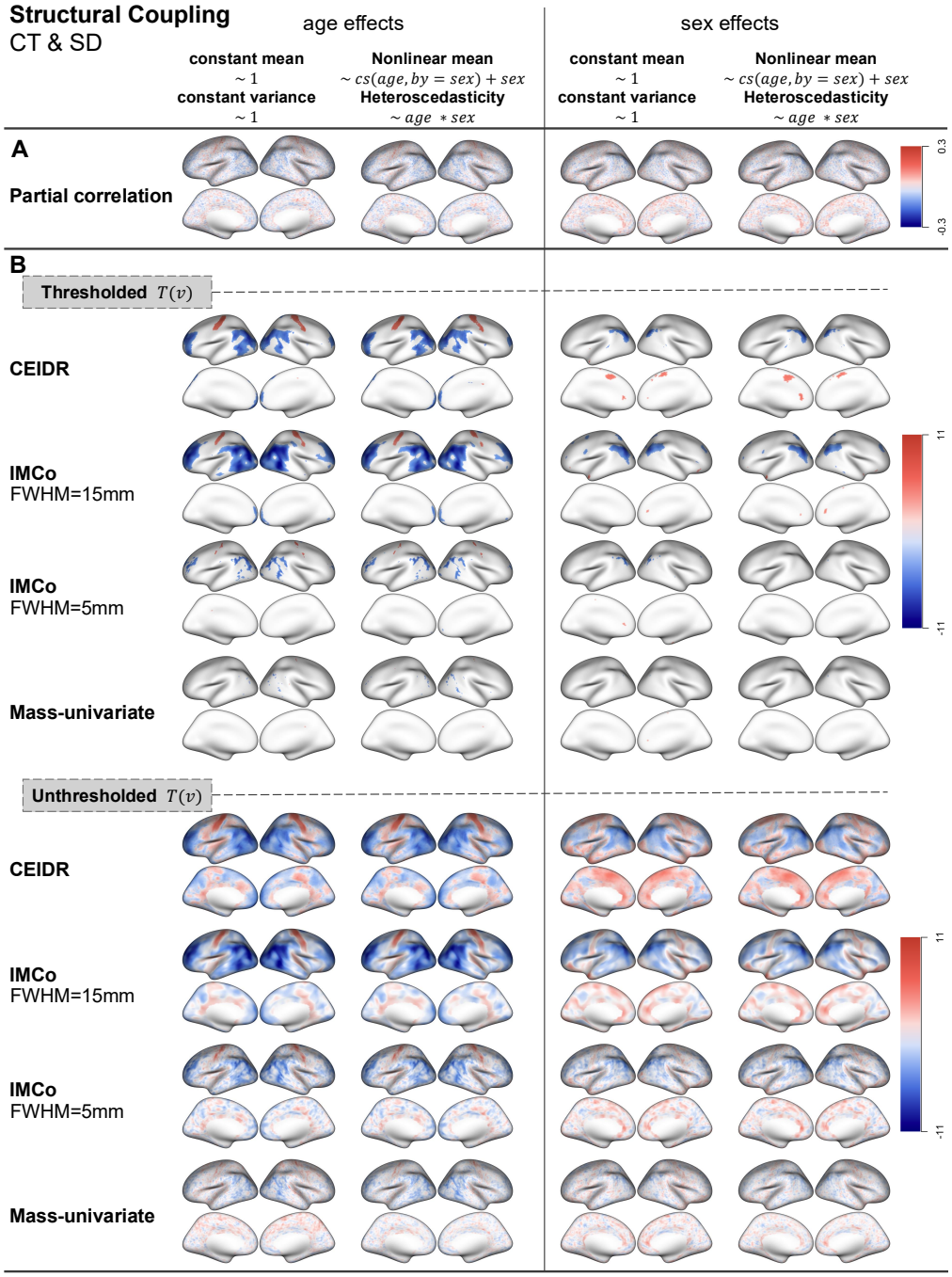


Figure 3. Full-sample results for CT and SD. (A) effect size maps, represented as a partial correlation of $\hat{\rho}_i(v)$ and x_i (e.g., age) given $\mathbf{z}_i$ (e.g., sex). (B) thresholded and unthresholded test statistic maps. Negative values of test statistics or partial correlation (blue) imply that intermodal coupling measures are found to be negatively associated with a covariate of interest, while positive values imply that intermodal coupling measures are positively associated. Note that sex is coded as a binary variable with 1 for females and 0 for males; therefore, regions where we observe positive values for sex effects would indicate where coupling was found to be higher in females than in males, while regions where we observe negative values for sex effects would indicate where it is estimated to be lower for females versus males.

Results from our analysis of individual differences in structural (CT-SD) and neurovascular (CBF-ALFF) coupling are presented in Figures 3 and 4. Panel A of each figure shows the effect size represented by the partial correlation between $\hat{\rho}_i(v)$ and the covariate of interest after adjusting for other covariates. In panel B, for each method, we map FWER-thresholded local test statistics $T(v)$ quantifying individual differences in coupling. The unthresholded test statistics are also presented to aid interpretation and enhance reproducibility (Taylor et al. 2023). We will focus on interpret-

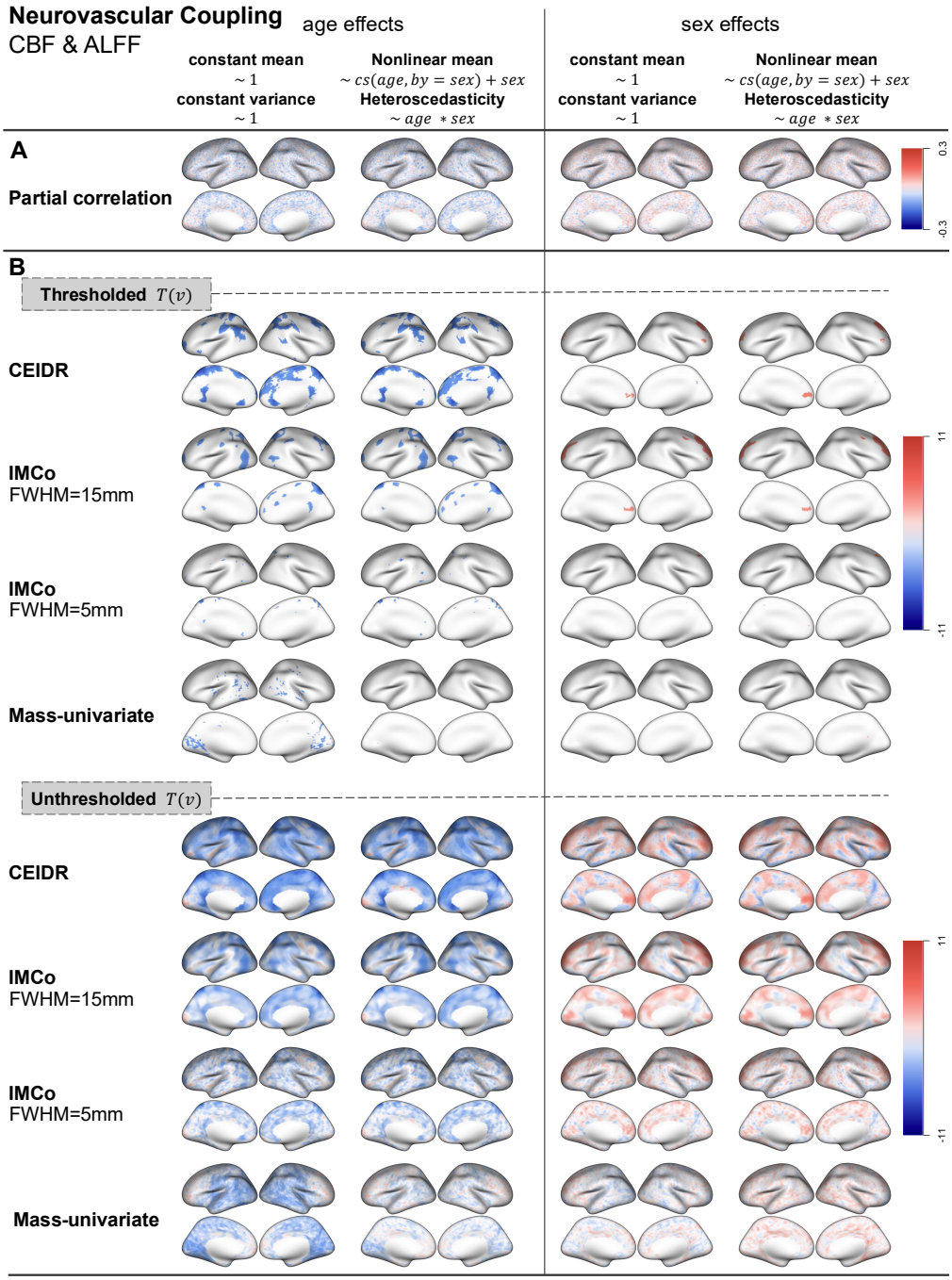


Figure 4. Full-sample results for ALFF and CBF, with the description analogous to Figure 3.

ing the results for individual differences, although the estimated coupling maps are provided in Supplementary material II.

The unthresholded maps reflect similar patterns of partial correlation, larger test statistics shown in the area with higher partial correlations and smaller test statistics shown in the area with lower partial correlations. Overall, the effect size is small in both modality pairs we considered and, as a result, the mass-univariate approach yields very limited areas of statistical significance even with a relatively large sample size used in the PNC study. As expected, CEIDR and IMCo that use spatial information provide more interpretable results on individual differences in coupling, while IMCo (with FWHM=5mm) produce much less significant regions, which could be explained by the inefficiency of using such a narrow smoothing kernel in capturing broader spatial patterns in our applications. Thus, we mainly compare the results of CEIDR and IMCo with 15mm FWHM.

Maps of age effects on CT-SD coupling in Figure 3 suggest an overlap across the different implementations of CEIDR and IMCo. Both methods show patterns of negative age effects (blue) on CT-SD coupling in the visual cortex and positive age effects (red) in the motor cortex, which is consistent with the analysis conducted by Vandekar et al. (2016). Although IMCo localizes significant age effects across a larger area of the cortical surface than the other implementations, this can partially be attributed to the benefits of using a large radius (FWHM=15mm) when signal clusters are large and contiguous. In the context of CBF-ALFF coupling, however, CEIDR detects age effects across a larger portion of the cortical surface compared to IMCo. This highlights the greater effectiveness of adaptive cluster enhancement in CEIDR compared to fixed-radius enhancement in IMCo (see Section 2.6.2), particularly when signals are more scattered and form smaller clusters. Age effects appear prominent in the attention networks, which aligns with results in Baller et al. (2022).

Interestingly, qualitative differences between CEIDR and IMCo are more apparent in tests of sex effects. For CT-SD coupling, both methods detect significant negative associations between female sex and structural coupling in the parietal cortex, which is also consistent with Vandekar et al. (2016)'s results. CEIDR also reveals significant positive associations between female sex and structural coupling in the superior frontal and anterior cingulate cortices. Additionally, the two different implementations of CEIDR and IMCo consistently reveal patterns of a positive association between female sex and neurovascular coupling overlapping in the frontoparietal network, which is consistent with Baller et al. (2022).

In our analyses, no significant differences are observed between implementations using a simple between-subject adjustment model (constant mean and variance) and those employing a more complex model (nonlinear mean and heteroscedasticity). Since its effect is not clearly understood without ground truth, we next conduct a brain-level simulation study in Section 4 to evaluate false positive control and power under controlled data-generating mechanisms.

4 Simulation study

4.1 Simulation design

We conducted a simulation study to examine CEIDR's ability to control the Type I error rate and its robustness and power. Specifically, we aim to highlight the impact of correct (vs incorrect) adjustment of marginal means and variances in mitigating false-positive rates, and to show how

it relates to statistical power. As done in PNC data analysis, we considered CEIDR, IMCo (with FWHM=15mm and FWHM=5mm) and the mass-univariate approach with FWER controlled by permutation.

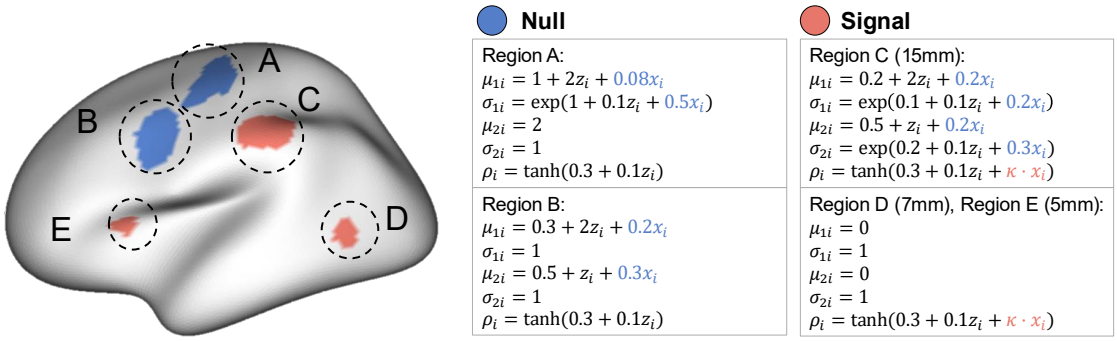


Figure 5. Figure for our simulation designs. Notations μ_{mi}, σ_{mi}, and ρ_i in the figure denote $\mu_m(z_i, x_i, v)$, $\sigma_m(z_i, x_i, v)$, and $\rho(z_i, x_i, v)$ respectively. Regions with blue color exhibit individual differences in means, variances, or correlations but correlations do not depend on x_i. Regions with red color indicate the existence of the covariate effect of x_i in correlations when $\kappa \neq 0$ regardless of the mean or variance structures. All remaining regions on the cortical surface are set as null regions with $\mu_{mi} = 0, \sigma_{mi} = 1$, and $\rho_i = 0$.

Data were simulated using the parametric model described in Section 2.2, with each simulated dataset consisting of $N = 200$ subjects. We first generated two covariates, $x_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$ (our interest) and $z_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$ (nuisance) independently, and used them to generate brain measurements $y_{mi}(v)$ following Equations (1), (2), and (3). We generated various types of regions on the left hemisphere of the brain using the fsaverage5 pial surface, which is described in detail in Figure 5. These regions were designed to reflect different combinations of individual differences in marginal means, variances, and inter-modal couplings (correlations). We included two null regions, A and B, which exhibit individual differences in marginal distributions but no differences in coupling (set to $\rho(z_i, x_i, v) = \tanh(0.3 + 0.1z_i)$). Specifically, region A displayed variability across individuals in the mean and variance of modality 1, while modality 2 remained homogeneous. Region B exhibited individual differences in the means of both modalities but maintained constant variance. In addition, we defined three signal regions, C, D, and E with varying sizes, with spatial extents of 15mm, 7mm, and 5mm, respectively. In these regions, there is individual difference in intermodal correlations, which was set $\rho(z_i, x_i, v) = \tanh(0.3 + 0.1z_i + \kappa \cdot x_i)$. Region C exhibits individual differences in marginal means, variances, and coupling, whereas regions D and E exhibited individual differences in coupling only, with fixed means and variances. All remaining regions on the surface were defined as null, with constant marginal means and variances, and zero inter-modal coupling ($\rho(z_i, x_i, v) = 0$).

In our simulations, we varied κ across the values $0, 0.01, \ldots, 0.09$. When $\kappa = 0$, regions C, D, and E reduced to null regions as well, allowing for the evaluation of family-wise error rate. When $\kappa \neq 0$, these regions exhibit nonzero covariate effects in couplings, enabling the assessment of statistical power.

4.2 Simulation results

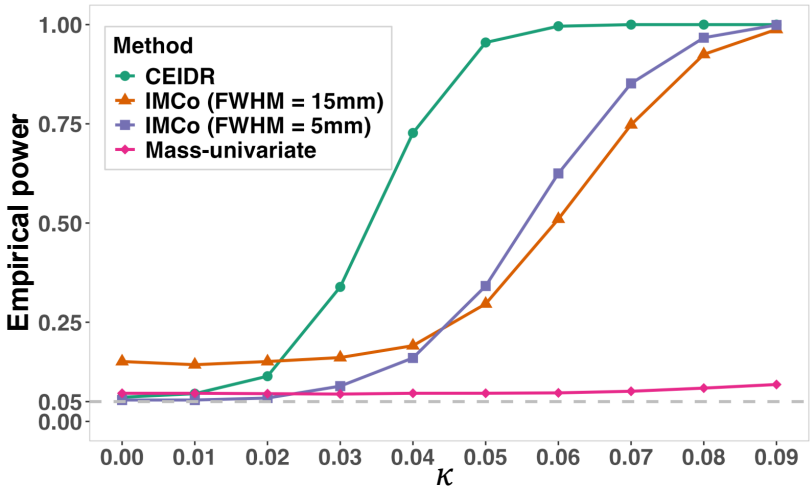


Figure 6. Empirical power of CEIDR, IMCo (FWHM=15mm), IMCo (FWHM=5mm), and mass-univariate from simulation study. The gray dashed line in the figure is the nominal FWER=0.05. When $\kappa = 0$, the corresponding value on the y-axis denotes empirical FWER.

Figure 6 summarizes the average performance across 1000 simulations. When $\kappa = 0$, the empirical FWER is defined as the proportion of simulations (out of 1000) in which at least one vertex was identified as significant. We first note that IMCo (with FWHM=15mm) had an inflated FWER. It can be explained by the presence of individual differences in means and variances in regions A, B, and C, in which IMCo's within-subject adjustment provided biased estimates in means and variances via over-specified FWHM. In particular, outside A, B, and C exhibit different means or variances compared to the values within the regions. As a result, IMCo (with FWHM=15mm) failed to adequately adjust for these differences, particularly at the boundaries of Regions A, B, and C. In contrast, IMCo (with FWHM=5mm) still controlled the FWER, likely due to its higher weighting of smaller local neighborhoods, leading to more accurate local correlation estimates compared to FWHM=15mm. We also note that the mass univariate approach also controlled FWER accurately because it specifically modeled covariate effects using the between-subject adjustment. Altogether, these explain how CEIDR, which incorporates both between-subject and within-subject adjustments (with a small neighbor size), accurately controlled FWER.

As κ increased, the power of all three methods increased, although mass-univariate approach showed minimal increases in power as κ increased, which is expected. Compared to the mass-univariate approach, both CEIDR and IMCo (with FWHM=15mm and FWHM=5mm) achieved higher power through cluster enhancement. Notably, CEIDR outperformed IMCo with FWHM=5mm at all κ values, and even IMCo with FWHM=15mm when $\kappa > 0.03$. It can be explained by the varying sizes of the signal regions because the fixed clustering widths used by IMCo are either too wide (15 mm) to capture small signals effectively, or too narrow (5 mm) to capture larger signals adequately, while CEIDR adaptively captures them.

5 Discussion

Studies of intermodal coupling are an increasing focus of neuroscience research. Measures of intermodal coupling have previously been implicated in healthy brain development, sex differences in development, cognition, and psychopathology (Avants et al. 2010; Devonshire et al. 2012; Ouyang et al. 2015; Phillips et al. 2016; Iadecola 2017; Baum et al. 2020). However, prior methodology in this area has focused primarily on the *quantification* of subject-level intermodal coupling—and less on the power of statistical tests using these measures to establish their relevance in neurodevelopment. In this article, we sought to assess power and replicability in intermodal coupling analyses, and lay the groundwork for more rigorous and generalizable research on coupling-based biomarkers in neurodevelopmental research.

Here, we introduced a new method, called CEIDR, which may be useful in studies examining individual differences in intermodal coupling. The methodological contributions of CEIDR are twofold. First, we extend the General Linear Model (GLM) framework, commonly applied in neuroimaging data analysis, to account for conditional correlations Tu et al. (2024). In doing so, we aim to define and parametrize 'intermodal correlation' and its individual differences, providing a clear null hypothesis. This model-based approach enables us to mitigate potential confounding effects arising from modality-specific individual differences in means and variances that would inflate false positive findings. The family-wise error rate is rigorously controlled through permutation testing. Utilizing both structural and functional MRI data from the Philadelphia Neurodevelopmental Cohort, we identified significant age- and sex-related differences in the coupling between structural and neurovascular features in the developing brain.

In this paper, we specifically focused on analyzing neurovascular (i.e., functional) coupling between CBF and ALFF and structural coupling between CT and SD, but not structure-function coupling (e.g., CT and ALFF). While our framework is in principle applicable to structure-function analyses, our primary focus was on neurovascular coupling in the context of this neurodevelopmental cohort. We also acknowledge that several studies (e.g., Weinstein et al. (2021) and Weinstein et al. (2022)) have shown limited evidence of the 'existence' of coupling at the group level (e.g., N-back activation and CT), which might imply limited effect sizes for individual differences in their coupling. Although structure-function coupling has not been extensively explored in this dataset, there may be other contexts or modalities where such coupling is more plausible. Future applications of our framework, potentially in combination with multivariate modeling approaches (e.g., Spisak et al., 2023), may offer more insight into individual differences in structure-function coupling in modalities where such effects are expected.

In this paper, we built a methodological connection between CEIDR and IMCo from the perspective of cluster enhancement and adjustment of means and variances in each modality, and discuss some issues related to false positives. IMCo naturally depends on the user-specified FWHM and kernel function, and its spurious findings were reported in Vandekar et al. (2016) that direction (sign) of age effects on coupling change as FWHM changes. From the perspective of CEIDR, it is attributed to the risk of overspecifying FWHM in mean-variance adjustments, which could lead to false positives. On the other hand, CEIDR provides a model-based definition of coupling and their individual differences, in which evaluation of sensitivity and specificity becomes straightforward. Therefore, we believe CEIDR provides a good methodological perspective on extending intermodal coupling research at the group level.

CEIDR has several limitations, which open several opportunities for extensions of the method. In this article, we used cross-sectional data from the PNC; however, CEIDR can be naturally extended to longitudinal data, including test-retest and repeated measures analyses. Longitudinal neuroimaging studies, such as the Adolescent Brain Cognitive Development (ABCD) study, have been instrumental in advancing the understanding of adolescent neurodevelopment by allowing the modelling of individual trajectories of imaging phenotypes over time (e.g., Romer et al. (2023) and Holm et al. (2023)). Such longitudinal data would allow us to explore whether changes, and the rate of change, in intermodal coupling differ by clinical groups or sex (Park et al. 2021; Bernal-Rusiel et al. 2013a; Bernal-Rusiel et al. 2013b). Additionally, CEIDR could be extended to analyze the test-retest reliability of individual differences in intermodal coupling, an important part of validating biomarkers from brain imaging (Pan et al. 2024). Another interesting direction is to model and test non-linear age trajectories in coupling (Vandekar et al. 2016; Baller et al. 2022). Lastly, we focused on analyzing the coupling between two imaging modalities. Extending this to pairwise comparisons across multiple modalities may be critical for a more comprehensive biological understanding and could facilitate whole-brain generative modeling (Hu et al. 2022).

Another limitation of the present work is that our interpretations of localized individual differences in intermodal coupling were based on visual assessments of the brain maps presented in Figures 3 and 4. To substantiate future interpretations, we plan to integrate methods for testing spatial specificity-or "enrichment"-of brain-phenotype associations within functional networks and other regions of interest. Methods for such analyses have been recently adapted from genetics (Subramanian et al. 2005) to neuroimaging (Weinstein et al. 2024). In the context of mapping individual differences in intermodal coupling, enrichment analyses would add to the rigor and replicability of assessments of spatial specificity in multimodal neuroimaging research.

CRediT authorship contribution statement

RP: Methodology, Software, Formal analysis, Validation, Visualization, Writing – original draft, Writing - review & editing. **SMW**: Conceptualization, Methodology, Software, Formal analysis, Validation, Visualization, Writing – original draft. **DT**: Conceptualization, Methodology, Validation, Visualization, Writing – original draft. **FH**: Software, Formal analysis, Writing - review & editing. **BT**: Formal analysis. **RZ**: Methodology, Writing - review & editing. **SNV**: Methodology, Validation, Writing - review & editing. **EBB**: Data curation, Software, Writing – review & editing. **RCG**: Resources, Writing – review & editing, Funding acquisition. **REG**: Resources, Writing – review & editing, Funding acquisition. **AAB**: Writing – review & editing, Funding acquisition. **TDS**: Resources, Methodology, Writing – review & editing, Funding acquisition. **JYP**: Conceptualization, Methodology, Validation, Visualization, Writing – original draft, Writing - review & editing, Funding acquisition, Supervision.

Declaration of competing interest

AAB receives consulting income from Octave Bioscience and holds equity in Centile Bioscience.

Ethics statement

The study procedures of the Philadelphia Neurodevelopment Cohort (PNC) were approved by the Institutional Review Boards (IRBs) of the University of Pennsylvania and the Children's Hospital of Philadelphia.

Data and code availability

Raw neuroimaging data from the PNC study are publicly available at the dbGaP https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v3.p2. ALFF, CT, and SD data from PNC, as well as other data types, is also publicly available for download through the Reproducible Brain Chart at https://reprobrainchart.github.io/ (Shafiei et al. 2025). All analyses were run using the R version 4.2. CEIDR is currently available for implementation at https://github.com/ruyipan/CEIDR as an R package. Runtime for applying CEIDR on one hemisphere in the PNC study ($N = 831$ subjects) is roughly 12 minutes on MacBook Pro 2019 (2.3 GHz 8-core Intel Core i9 and 32GB RAM) including permutation without parallel computing.

Acknowledgments

We thank three reviewers for their insightful comments and suggestions, which significantly improved the quality of the manuscript. This work was supported by NIH R01MH112847 (TDS). Additional support was provided by NIH R01MH113550 (TDS), R01MH120482 (TDS), R01EB022573 (TDS), R37MH125829 (TDS), R01MH119219 (RCG, REG), K23MH133118 (EBB), R01MH123563 (SNV), R01MH132934 (AAB), R01MH133843 (AAB), and the Brain and Behavior Research Foundation NARSAD #31319 (EBB). JYP was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-2022-04831), the University of Toronto's Data Science Institute (Catalyst Grant), McLaughlin Centre (Accelerator Grant), and the Connaught fund. DT contributed to this article as an employee of the University of Pennsylvania and the views expressed do not necessarily represent the views of Regeneron Pharmaceuticals Inc.

References

Avants, Brian B, Philip A Cook, Lyle Ungar, James C Gee, and Murray Grossman (2010). "Dementia induces correlated reductions in white matter integrity and cortical thickness: a multivariate neuroimaging study with sparse canonical correlation analysis". In: *NeuroImage* 50.3, pp. 1004–1016. DOI: 10.1016/j.neuroimage.2010.01.041.

Baller, Erica B., Alessandra M. Valcarcel, Azeez Adebimpe, Aaron Alexander-Bloch, Zaixu Cui, Ruben C. Gur, Raquel E. Gur, Bart L. Larsen, Kristin A. Linn, Carly M. O'Donnell, Adam R. Pines, Armin Raznahan, David.R. Roalf, Valerie J. Sydnor, Tinashe M. Tapera, M. Dylan Tisdall, Simon Vandekar, Cedric H. Xia, John A. Detre, Russell T. Shinohara, and Theodore D. Satterthwaite (2022). "Developmental coupling of cerebral blood flow and fMRI fluctuations in youth". en. In: *Cell Reports* 38.13, p. 110576. ISSN: 22111247. DOI: 10.1016/j.celrep.2022.110576.

Baum, Graham L, Zaixu Cui, David R Roalf, Rastko Ciric, Richard F Betzel, Bart Larsen, Matthew Cieslak, Philip A Cook, Cedric H Xia, Tyler M Moore, et al. (2020). "Development of structure–function coupling in human brain networks during youth". In: *Proceedings of the National Academy of Sciences* 117.1, pp. 771–778. DOI: 10.1073/pnas.1912034117.

Bernal-Rusiel, Jorge L, Douglas N Greve, Martin Reuter, Bruce Fischl, Mert R Sabuncu, Alzheimer's Disease Neuroimaging Initiative, et al. (2013a). "Statistical analysis of longitudinal neuroimage data with linear mixed effects models". In: *NeuroImage* 66, pp. 249–260. DOI: 10.1016/j.neuroimage.2012.10.065.

Bernal-Rusiel, Jorge L, Martin Reuter, Douglas N Greve, Bruce Fischl, Mert R Sabuncu, Alzheimer's Disease Neuroimaging Initiative, et al. (2013b). "Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data". In: *NeuroImage* 81, pp. 358–370. DOI: 10.1016/j.neuroimage.2013.05.049.

Botvinik-Nezer, Rotem, Felix Holzmeister, Colin F Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A Mumford, R Alison Adcock, et al. (2020). "Variability in the analysis of a single neuroimaging dataset by many teams". In: *Nature* 582.7810, pp. 84–88. DOI: 10.1038/s41586-020-2314-9.

Chen, J. Jean (2019). "Functional MRI of brain physiology in aging and neurodegenerative diseases". In: *NeuroImage* 187, 209–225. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2018.05.050.

Ciric, Rastko, Daniel H. Wolf, Jonathan D. Power, David R. Roalf, Graham L. Baum, Kosha Ruparel, Russell T. Shinohara, Mark A. Elliott, Simon B. Eickhoff, Christos Davatzikos, Ruben C. Gur, Raquel E. Gur, Danielle S. Bassett, and Theodore D. Satterthwaite (2017). "Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity". In: *NeuroImage* 154, pp. 174–187. DOI: 10.1016/j.neuroimage.2017.03.020.

Devonshire, Ian M, Nikos G Papadakis, Michael Port, Jason Berwick, Aneurin J Kennerley, John EW Mayhew, and Paul G Overton (2012). "Neurovascular coupling is brain region-dependent". In: *NeuroImage* 59.3, pp. 1997–2006. DOI: 10.1016/j.neuroimage.2011.09.050.

Fjell, Anders M, Lars T Westlye, Inge Amlien, Thomas Espeseth, Ivar Reinvang, Naftali Raz, Ingrid Agartz, David H Salat, Doug N Greve, Bruce Fischl, et al. (2009). "High consistency of regional cortical thinning in aging across multiple samples". In: *Cerebral Cortex* 19.9, pp. 2001–2012. DOI: 10.1093/cercor/bhn232.

Flournoy, John C, Nandita Vijayakumar, Theresa W Cheng, Danielle Cosme, Jessica E Flannery, and Jennifer H Pfeifer (2020). "Improving practices and inferences in developmental cognitive neuroscience". In: *Developmental Cognitive Neuroscience* 45, p. 100807. DOI: 10.1016/j.dcn.2020.100807.

Holm, Madelene C, Esten H Leonardsen, Dani Beck, Andreas Dahl, Rikka Kjelkenes, Ann-Marie G de Lange, and Lars T Westlye (2023). "Linking brain maturation and puberty during early adolescence using longitudinal brain age prediction in the ABCD cohort". In: *Developmental Cognitive Neuroscience* 60, p. 101220. DOI: https://doi.org/10.1016/j.dcn.2023.10122.

Hoptman, Matthew J., Xi-Nian Zuo, Pamela D. Butler, Daniel C. Javitt, Debra D'Angelo, Cristina J. Mauro, and Michael P. Milham (2010). "Amplitude of low-frequency oscillations in schizophrenia: A resting state fMRI study". In: *Schizophrenia Research* 117.1, 13–20. ISSN: 0920-9964. DOI: 10.1016/j.schres.2009.09.030.

Hu, Fengling, Sarah M Weinstein, Erica B Baller, Alessandra M Valcarcel, Azeez Adebimpe, Armin Raznahan, David R Roalf, Timothy E Robert-Fitzgerald, Virgilio Gonzenbach, Ruben C Gur, S Vandekar, JA Detre, KA Linn, A Alexander-Bloch, TD Satterthwaite, and RT Shinohara (2022). "Voxel-wise intermodal coupling analysis of two or more modalities using local covariance decomposition". In: *Human Brain Mapping* 43.15. Publisher: John Wiley & Sons, Inc. Hoboken, USA, pp. 4650–4663. ISSN: 1065-9471. DOI: 10.1002/hbm.25980.

Iadecola, Costantino (2017). "The neurovascular unit coming of age: a journey through neurovascular coupling in health and disease". In: *Neuron* 96.1, pp. 17–42. DOI: 10.1016/j.neuron.2017.07.030.

Kelly, Claire E, Deanne K Thompson, Chris L Adamson, Gareth Ball, Thijs Dhollander, Richard Beare, Lillian G Matthews, Bonnie Alexander, Jeanie L Y Cheong, Lex W Doyle, Peter J Anderson, and Terrie E Inder (2023). "Cortical growth from infancy to adolescence in preterm and term-born children". In: *Brain* 147.4, 1526–1538. ISSN: 1460-2156. DOI: 10.1093/brain/awad348.

Klapwijk, Eduard T, Wouter van den Bos, Christian K Tamnes, Nora M Raschle, and Kathryn L Mills (2021). "Opportunities for increased reproducibility and replicability of developmental neuroimaging". In: *Developmental Cognitive Neuroscience* 47, p. 100902. DOI: 10.1016/j.dcn.2020.100902.

Lindquist, Martin A. and Amanda Mejia (2015). "Zen and the Art of Multiple Comparisons". In: *Psychosomatic Medicine* 77.2, 114–125. ISSN: 0033-3174. DOI: 10.1097/psy.0000000000000148.

Marek, Scott, Brenden Tervo-Clemmens, Finnegan J Calabro, David F Montez, Benjamin P Kay, Alexander S Hatoum, Meghan Rose Donohue, William Foran, Ryland L Miller, Timothy J Hendrickson, et al. (2022). "Reproducible brain-wide association studies require thousands of individuals". In: *Nature* 603.7902, pp. 654–660. DOI: 10.1038/s41586-022-04492-9.

Ouyang, Xin, Kewei Chen, Li Yao, Bin Hu, Xia Wu, Qing Ye, Xiaojuan Guo, Alzheimer's Disease Neuroimaging Initiative, et al. (2015). "Simultaneous changes in gray matter volume and white matter fractional anisotropy in Alzheimer's disease revealed by multimodal CCA and joint ICA". In: *Neuroscience* 301, pp. 553–562. DOI: 10.1016/j.neuroscience.2015.06.031.

Pan, Ruyi, Erin W Dickie, Colin Hawco, Nancy Reid, Aristotle N Voineskos, and Jun Young Park (2024). "Spatial-extent inference for testing variance components in reliability and heritability studies". In: *Imaging Neuroscience* 2, pp. 1–15. DOI: 10.1162/imag_a_00058.

Park, Jun Young and Mark Fiecas (2022). "CLEAN: Leveraging spatial autocorrelation in neuroimaging data in clusterwise inference". In: *NeuroImage* 255, p. 119192. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2022.119192.

Park, Jun Young, Mark Fiecas, Alzheimer's Disease Neuroimaging Initiative, et al. (2021). "Permutation-based inference for spatially localized signals in longitudinal MRI data". In: *NeuroImage* 239, p. 118312. DOI: 10.1016/j.neuroimage.2021.118312.

Phillips, Aaron A, Franco HN Chan, Mei Mu Zi Zheng, Andrei V Krassioukov, and Philip N Ainslie (2016). "Neurovascular coupling in humans: physiology, methodological advances and clinical implications". In: *Journal of Cerebral Blood Flow & Metabolism* 36.4, pp. 647–664. DOI: 10.1177/0271678X15617954.

Raznahan, Armin, Phillip Shaw, Francois Lalonde, Mike Stockman, Gregory L Wallace, Dede Greenstein, Liv Clasen, Nitin Gogtay, and Jay N Giedd (2011). "How does your cortex grow?" en. In: *J. Neurosci.* 31.19, pp. 7174–7177. DOI: 10.1523/JNEUROSCI.0054-11.2011.

Romer, Adrienne L, Boyu Ren, and Diego A Pizzagalli (2023). "Brain structure relations with psychopathology trajectories in the ABCD study". In: *Journal of the American Academy of Child & Adolescent Psychiatry* 62.8, pp. 895–907. DOI: 10.1016/j.jaac.2023.02.002.

Satterthwaite, Theodore D., John J. Connolly, Kosha Ruparel, Monica E. Calkins, Chad Jackson, Mark A. Elliott, David R. Roalf, Ryan Hopson, Karthik Prabhakaran, Meckenzie Behr, Haijun Qiu, Frank D. Mentch, Rosetta Chiavacci, Patrick M.A. Sleiman, Ruben C. Gur, Hakon Hakonarson, and Raquel E. Gur (2016). "The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth". In: *NeuroImage* 124, pp. 1115–1119. DOI: 10.1016/j.neuroimage.2015.03.056.

Satterthwaite, Theodore D., Russell T. Shinohara, Daniel H. Wolf, Ryan D. Hopson, Mark A. Elliott, Simon N. Vandekar, Kosha Ruparel, Monica E. Calkins, David R. Roalf, Efstathios D. Gennatas, Chad Jackson, Guray Erus, Karthik Prabhakaran, Christos Davatzikos, John A. Detre, Hakon Hakonarson, Ruben C. Gur, and Raquel E. Gur (2014). "Impact of puberty on the evolution of cerebral perfusion during adolescence". In: *Proceedings of the National Academy of Sciences* 111.23, 8643–8648. ISSN: 1091-6490. DOI: 10.1073/pnas.1400178111.

Shafiei, Golia, Nathalia B Esper, Mauricio S Hoffmann, Lei Ai, Andrew A Chen, Jon Cluce, Sydney Covitz, Steven Giavasis, Connor Lane, Kahini Mehta, et al. (2025). "Reproducible Brain Charts: An open data resource for mapping brain development and its associations with mental health". In: *bioRxiv*, pp. 2025–02. DOI: 10.1101/2025.02.24.639850.

Stackhouse, Teresa L. and Anusha Mishra (2021). "Neurovascular Coupling in Development and Disease: Focus on Astrocytes". In: *Frontiers in Cell and Developmental Biology* 9. ISSN: 2296-634X. DOI: 10.3389/fcell.2021.702832.

Subramanian, Aravind, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, et al. (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles". In: *Proceedings of the National Academy of Sciences* 102.43, pp. 15545–15550. DOI: 10.1073/pnas.0506580102.

Taki, Yasuyuki, Hiroshi Hashizume, Yuko Sassa, Hikaru Takeuchi, Kai Wu, Michiko Asano, Kohei Asano, Hiroshi Fukuda, and Ryuta Kawashima (2011). "Correlation between gray matter density-adjusted brain perfusion and age using brain MR images of 202 healthy children". In: *Human Brain Mapping* 32.11, 1973–1985. ISSN: 1097-0193. DOI: 10.1002/hbm.21163.

Tamnes, Christian K, Kristine B Walhovd, Anders M Dale, Ylva Østby, Håkon Grydeland, George Richardson, Lars T Westlye, J Cooper Roddey, Donald J Hagler Jr, Paulina Due-Tønnessen, et al. (2013). "Brain development and aging: overlapping and unique patterns of change". In: *Neuroimage* 68, pp. 63–74. DOI: 10.1523/JNEUROSCI.3302-16.2017.

Taylor, Paul A, Richard C Reynolds, Vince Calhoun, Javier Gonzalez-Castillo, Daniel A Handwerker, Peter A Bandettini, Amanda F Mejia, and Gang Chen (2023). "Highlight results, don't hide them: Enhance interpretation, reduce biases and improve reproducibility". In: *NeuroImage* 274, p. 120138. DOI: 10.1016/j.neuroimage.2023.120138.

Tu, Danni, Bridget Mahony, Tyler M Moore, Maxwell A Bertolero, Aaron F Alexander-Bloch, Ruben Gur, Dani S Bassett, Theodore D Satterthwaite, Armin Raznahan, and Russell T Shinohara (2024). "CoCoA: conditional correlation models with association size". In: *Biostatistics* 25.1, pp. 154–170. DOI: 10.1093/biostatistics/kxac032.

Vandekar, Simon N., Russell T. Shinohara, Armin Raznahan, Ryan D. Hopson, David R. Roalf, Kosha Ruparel, Ruben C. Gur, Raquel E. Gur, and Theodore D. Satterthwaite (2016). "Subject-level measurement of local cortical coupling". In: *NeuroImage* 133, pp. 88–97. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2016.03.002.

Weinstein, Sarah M., Simon N. Vandekar, Azeez Adebimpe, Tinashe M. Tapera, Timothy Robert-Fitzgerald, Ruben C. Gur, Raquel E. Gur, Armin Raznahan, Theodore D. Satterthwaite, Aaron F. Alexander-Bloch, and Russell T. Shinohara (2021). "A simple permutation-based test of intermodal correspondence". en. In: *Human Brain Mapping* 42.16, pp. 5175–5187. ISSN: 1097-0193. DOI: 10.1002/hbm.25577.

Weinstein, Sarah M, Simon N Vandekar, Erica B Baller, Danni Tu, Azeez Adebimpe, Tinashe M Tapera, Ruben C Gur, Raquel E Gur, John A Detre, Armin Raznahan, Aaron F Alexander-Bloch, Theodore D Satterthwaite, Russell T Shinohara, and Jun Young Park (2022). "Spatially-enhanced clusterwise inference for testing and localizing intermodal correspondence". In: *NeuroImage* 264. Publisher: Academic Press, p. 119712. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2022.119712.

Weinstein, Sarah M, Simon N Vandekar, Bin Li, Aaron F Alexander-Bloch, Armin Raznahan, Mingyao Li, Raquel E Gur, Ruben C Gur, David R Roalf, Min Tae M Park, Mallar Chakravarty, et al. (2024). "Network Enrichment Significance Testing in Brain-Phenotype Association Studies". In: *Human Brain Mapping*. DOI: 10.1002/hbm.26714.

Winkler, Anderson M, Gerard R Ridgway, Matthew A Webster, Stephen M Smith, and Thomas E Nichols (2014). "Permutation inference for the general linear model". In: *NeuroImage* 92, pp. 381–397. DOI: 10.1016/j.neuroimage.2014.01.060.

Yan, Jun and Jason Fine (2004). "Estimating equations for association structures". In: *Statistics in Medicine* 23.6, pp. 859–874. DOI: 10.1002/sim.1650.

Ziegler, Gabriel, Robert Dahnke, Lutz Jäncke, Rachel Aine Yotter, Arne May, and Christian Gaser (2012). "Brain structural trajectories over the adult lifespan". In: *Human Brain Mapping* 33.10, pp. 2377–2389. DOI: 10.1002/hbm.21374.

Supplementary materials

I. Revisiting IMCo as a special case of CEIDR

The subject-level intermodal coupling estimated by IMCo using locally weighted correlation, denoted by $\phi_i(v)$ here, is

$$\phi_i(v) = \frac{\sum_{v^\star=1}^{V} \; w_{v,v^\star} \times (y_{1i}(v^\star) - \tilde{y}_{1i}(v)) \cdot (y_{2i}(v^\star) - \tilde{y}_{2i}(v))}{\sqrt{\sum_{v^\star=1}^{V} \; w_{v,v^\star} \times (y_{1i}(v^\star) - \tilde{y}_{1i}(v))^2} \times \sqrt{\sum_{v^\star=1}^{V} \; w_{v,v^\star} \times (y_{2i}(v^\star) - \tilde{y}_{2i}(v))^2}}, \tag{S1}$$

where $\tilde{y}_{1i}(v)$ and $\tilde{y}_{2i}(v)$ are local weighted means, computed by

$$\tilde{y}_{mi}(v) = \frac{\sum_{v^\star=1}^{V} \; w_{v,v^\star} \times y_{mi}(v^\star)}{\sum_{v^\star=1}^{V} \; w_{v,v^\star}},$$

and $w_{v,v^\star}$ is the weight determined by the distance $d_{v,v^\star}$ and prespecified FWHM as described in Section 2.5.1 of the paper. To build methodological connection, we consider a special case of $w_{v,v^\star}$ that

$$w_{v,v^\star} = \begin{cases} 1 & \text{if } v^\star \in N_r(v) \\ 0 & \text{otherwise.} \end{cases} \tag{S2}$$

Then, $\tilde{y}_{mi}(v) = \bar{y}_{mi}(v)$ (sample mean of $y_{mi}(v)$ within $N_r(v)$) defined in Section 2.4 (I.2) of the paper and Equation (S1) can be re-written as

$$\begin{aligned}
\phi_i(v) &= \frac{\sum_{v^\star=1}^{V} \; w_{v,v^\star} \times (y_{1i}(v^\star) - \bar{y}_{1i}(v)) \cdot (y_{2i}(v^\star) - \bar{y}_{2i}(v))}{\sqrt{\sum_{v^\star=1}^{V} \; w_{v,v^\star} \times (y_{1i}(v^\star) - \bar{y}_{1i}(v))^2}\sqrt{\sum_{v^\star=1}^{V} \; w_{v,v^\star} \times (y_{2i}(v^\star) - \bar{y}_{2i}(v))^2}} \\
&= \sum_{v^\star=1}^{V} w_{v,v^\star} \times \frac{y_{1i}(v^\star) - \bar{y}_{1i}(v)}{\sqrt{\sum_{v^\star=1}^{V} \; w_{v,v^\star} \times (y_{1i}(v^\star) - \bar{y}_{1i}(v))^2}} \times \frac{y_{2i}(v^\star) - \bar{y}_{2i}(v)}{\sqrt{\sum_{v^\star=1}^{V} \; w_{v,v^\star} \times (y_{2i}(v^\star) - \bar{y}_{2i}(v))^2}} \\
&= \sum_{v^\star=1}^{V} w_{v,v^\star} \times \frac{y_{1i}(v^\star) - \bar{y}_{1i}(v)}{\sqrt{\sum_{v^\star \in N_r(v)} (y_{1i}(v^\star) - \bar{y}_{1i}(v))^2}} \times \frac{y_{2i}(v^\star) - \bar{y}_{2i}(v)}{\sqrt{\sum_{v^\star \in N_r(v)} \; (y_{2i}(v^\star) - \bar{y}_{2i}(v))^2}} \\
&= \frac{1}{|N_r(v)| - 1} \times \sum_{v^\star \in N_r(v)} \hat{\epsilon}_{1i}(v^\star) \times \hat{\epsilon}_{2i}(v^\star) \\
&= \frac{1}{|N_r(v)| - 1} \times \sum_{v^\star \in N_r(v)} \hat{\rho}_i(v^\star),
\end{aligned}$$

where $\hat{\epsilon}_{1i}(v)$ and $\hat{\epsilon}_{2i}(v)$ are defined in Section 2.4 (I.2) and $\hat{\rho}_i(v)$ is defined in Section 2.4 (II). This result implies that IMCo's subject-level coupling ($\phi_i(v)$) is proportional to the sum of CEIDR's subject-level coupling over $N_r(v)$ after applying the within-subject adjustment using vertices of $N_r(v)$.

Once $\phi_i(v)$ is obtained, IMCo uses GLM to compute p value for individual differences. Suppose that we define (i) $\mathbf{Z}$ as matrix of $(1, \mathbf{z}_i)$ stacked in row across N subjects (ii) $\mathbf{x}$ as a column vector of

x_i, and (iii) $\boldsymbol{\phi}(v)$ as a column vector of $\phi_i(v)$ collected across N subjects. Also, let $\mathbf{I}_N$ be a $N \times N$ identity matrix. Then the score test statistic for the association between $\phi_i(v)$ and x_i adjusting for $\mathbf{z}_i$ is written as

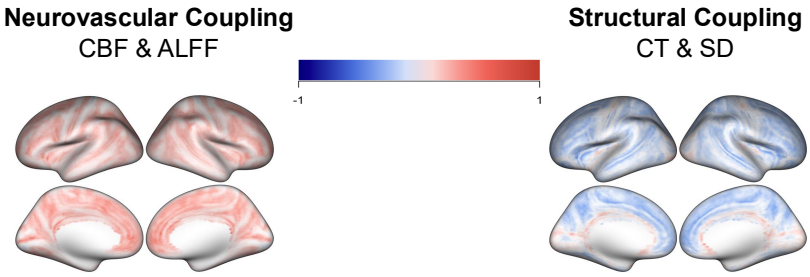
$$
\begin{aligned}
T_{\text{IMCo}}(v) &= \mathbf{x}^T(\mathbf{I}_N - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)\boldsymbol{\phi}(v) \\
&\propto \mathbf{x}^T(\mathbf{I}_N - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)\left(\sum_{v^\star \in N_r(v)} \hat{\boldsymbol{\rho}}(v^\star)\right) \\
&= \sum_{v^\star \in N_r(v)} \left(\mathbf{x}^T(\mathbf{I}_N - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)\hat{\boldsymbol{\rho}}(v^\star)\right) \\
&= \sum_{v^\star \in N_r(v)} \left(\sum_{i=1}^{N} x_i \times (\hat{\rho}_i(v^\star) - \tilde{\beta}_0(v^\star) - \mathbf{z}_i\tilde{\boldsymbol{\beta}}(v^\star)\right) \\
&= \sum_{v^\star \in N_r(v)} U(v^\star) \\
&\propto T_r(v).
\end{aligned}
$$

This result implies that the p value of IMCo at vertex v is equivalent to the fixed-radius cluster enhancement using score test statistics from CEIDR. When permutation to control FWER is used instead of FDR, Step V completes statistical inference done in CEIDR.

Altogether, our derivation demonstrates that, when specific weights $w_{v,v^\star}$ are given as Equation (S2), IMCo implemented with weighted correlation and FWER control reduces to a special case of CEIDR that (i) IMCo adjusts for means and variances in each modality by using the *within-subject adjustment* approach (Step I.2) and performs cluster enhancement with a fixed radius (Step IV).

II. Coupling maps

In our data analysis, we computed the average intermodal coupling at each vertex by averaging $\hat{\rho}_i(v)$ in Stage II across subjects. The resulting spatial map is presented in Figure S7.



Minimum	Q1	Median	Q3	Maximum
-0.141	0.073	0.147	0.215	0.693

Minimum	Q1	Median	Q3	Maximum
-0.571	-0.181	-0.108	-0.036	0.412

Figure S7. Average intermodal coupling map across subjects. For each vertex, the value represents the average of subject-specific coupling estimates computed in Stage II.