Mapping Individual Differences in Intermodal Coupling in Neurodevelopment

Sarah M. Weinstein[1,†], Danni Tu[2,3,†], Fengling Hu[3], Ruyi Pan[4,5], Rongqian Zhang[4], 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[*4,11]

[1]Department of Epidemiology and Biostatistics, Temple University College of Public Health, Philadelphia, PA, USA
[2]Regeneron Pharmaceuticals, Tarrytown, NY, USA
[3]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, USA
[4]Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada
[5]Centre for Addiction and Mental Health, Toronto, ON, Canada
[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
[†]joint first authorship

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. We propose a new method, called CIDeR, which is designed to simultaneously perform hypothesis testing in a way that limits false positive results and improve detection of true positive results. 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 the utility of CIDeR 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 re-organization. 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 underlying mechanisms of risk of 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. Amid growing concerns about reproducibility in neuroimaging-based studies of neurodevelopment (Flournoy et al. 2020; Klapwijk et al. 2021; Marek et al. 2022), 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 found 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 if the relationshiop 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 to 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 (Tu et al. 2024), which can vary by age, sex, or other variables of interest. Second, in contrast to other methods utilizing symmetric measures of coupling (Hu et al. 2022), we perform spatial modeling and "adjustment" for spatial autocorrelation in each brain map, through an extension of methods described in recent work (Park and Fiecas 2022; Weinstein et al. 2022). Finally, we propose a statistical inference procedure to test individual differences in (conditional) intermodal correlations, which we illustrate achieves high power and spatial localization. We apply our method to multi-modal MRI data from the Philadelphia Neurodevelopmental Cohort (Satterthwaite et al. 2014), delineating both robust developmental effects and sex differences in intermodal coupling.

2 Methods

We propose CIDeR (**C**LEAN for testing **I**ndividual **D**iff**e**rences in ρ (**r**)), a method which extends the CLEAN framework introduced in previous work (Weinstein et al. 2022; Park and Fiecas 2022). This multi-stage model adapts existing strategies to account for spatial autocorrelation (in order to control false positive rates) while maintaining or boosting power (in order to increase true positive rates).

2.1 Notation

For participant $i = 1, \ldots, N$ and location (i.e., vertex) $v = 1, \ldots, V$, we denote brain measurements for two modalities as $y_{1i}(v)$ and $y_{2i}(v)$, respectively. Across vertices, these measurements make up two brain maps $\mathbf{y}_{1i} = \{y_{1i}(v)\}_{v=1}^{V}$ and $\mathbf{y}_{2i} = \{y_{2i}(v)\}_{v=1}^{V}$ for each participant. In this article, we are interested in coupling between $y_{1i}(v)$ and $y_{2i}(v)$ for each participant, which we define as a residual correlation $\rho_i(v)$ between $y_{1i}(v)$ and $y_{2i}(v)$ that differs by location v and subject i. The covariate of interest and the other covariates adjusted for in the model are are denoted as x_i and $\mathbf{z}_i$, respectively. For example, in our analysis of real data discussed below, when we quantify age effects (x_i) on local intermodal coupling ($\rho_i(v)$), we adjust for sex effects ($\mathbf{z}_i$).

2.2 Null hypothesis

We hypothesize that $\rho_i(v)$ for subject i is a function of a variable of interest x_i (e.g., age) and other potential confounding variables $\mathbf{z}_i$ (e.g., sex), and we use procedures described below to quantify the statistical significance of the associations. We do this by comparing associations computed using observed data to those computed using data generated under the null hypothesis (where there is no association between coupling at *any* vertex and x_i). To avoid making assumptions about the distribution of associations when the null hypothesis is true, we employ a data-driven method involving permutation to generate null data through permutation-based resampling (Section 2.3).

2.3 CIDeR

A relationship between intermodal coupling ($\boldsymbol{\rho}_i = \{\rho_i(v)\}_{v=1}^{V}$) and a covariate of interest x_i can arise from several sources: there may be a true dependence of $\boldsymbol{\rho}_i$ on x_i, or the individual modalities $\mathbf{y}_{1i}$ or $\mathbf{y}_{2i}$ may depend on x_i. In order to distinguish these, we introduce a multi-stage model that adjusts separately for the effect of x_i and $\mathbf{z}_i$ on the means and variances of the image measurements, $\mathbf{y}_{1i}$ and $\mathbf{y}_{2i}$. We also account for spatial autocorrelation, a phenomenon where vertices that are closer to each other tend to be more similar to one another due to their proximity.

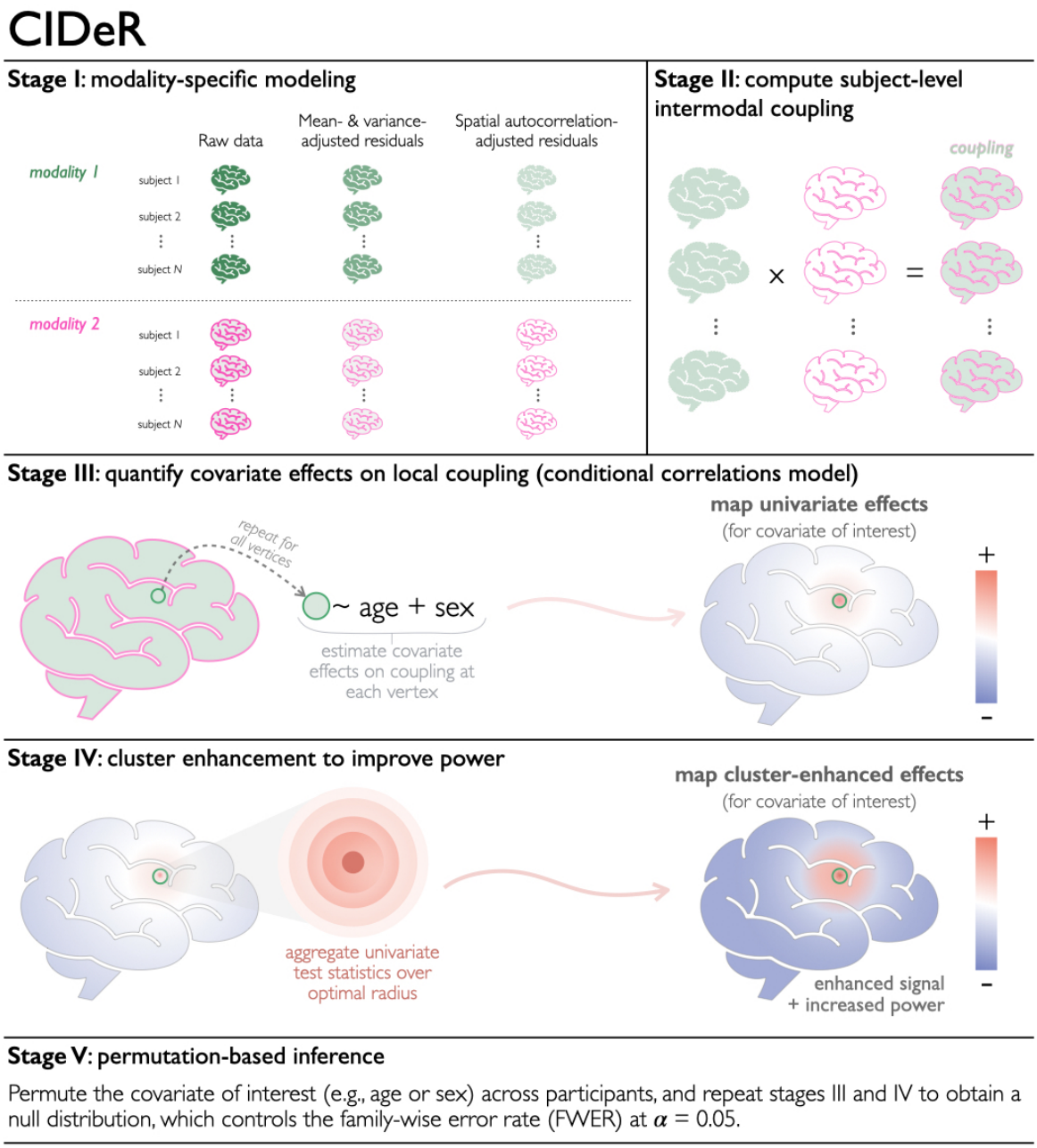

Figure 1. Overview of the proposed method, CIDeR. These steps are also outlined in Section 2.3; refer to Appendix A for additional details.

We outline the five main stages of CIDeR below. Each stage is also illustrated in Figure 1, and additional technical details can be found in Appendix A.

I. *Removing unrelated data variability.* We aim to isolate signal within each brain map that may be related to intermodal coupling by removing "nuisance" variation from each map. Such nuisance variation could be due to individual differences in the (i) mean and/or variance across participants within each modality, as well as (ii) spatial autocorrelation. The current stage is therefore comprised of the following sub-stages:

   (i) We first consider a parametric linear model at each vertex to perform *mean and variance adjustment*. We use the covariate of interest (e.g., x_i) and other nuisance variables (e.g., $\mathbf{z}_i$) to model individual differences in means and variances in each modality. We note that false positives in individual differences in intermodal coupling may arise when there exist differences in means and variance in reality, but those features are left unmodeled. Therefore, this step can help to mitigate false positive findings. For additional details, see Appendix A.1.1.

   (ii) We then perform spatial modeling and *adjustment for spatial autocorrelation*, which has been shown to reduce noise due to spatial autocorrelation on statistical inference and improve sensitivity in spatial-extent inference (Weinstein et al. 2022) involving intermodal associations. Details are described in Appendix A.1.2.

   Thus, instead of working with $\mathbf{y}_{1i}$ or $\mathbf{y}_{2i}$ directly, after implementing (i) and (ii), we work with the resulting residuals from these models, denoted by $\boldsymbol{\delta}_{1i}$ and $\boldsymbol{\delta}_{2i}$ for modalities 1 and 2 respectively. The quantities $\boldsymbol{\delta}_{1i}$ and $\boldsymbol{\delta}_{2i}$ represent the resulting brain maps when confounding due to within-modality spatial autocorrelation and nuisance covariate effects are attenuated.

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

$$\hat{\rho}_i(v) = \frac{\hat{\delta}_{1i}(v) \times \hat{\delta}_{2i}(v)}{\hat{\tau}_1(v) \times \hat{\tau}_2(v)}, \tag{1}$$

   where $\hat{\tau}_1(v)$ and $\hat{\tau}_2(v)$ are modality-specific variance estimates obtained from Step I(ii), used for standardization of the residuals. Details on why Equation (1) provides a meaningful measure of intermodal coupling are provided in Appendix A.2.

III. *Modeling intermodal coupling in terms of covariates.* We model the relationship between intermodal coupling (as quantified in Equation (1)) and subject-level covariates x_i, $\mathbf{z}_i$ as a conditional correlations model at each vertex:

$$E\left[\rho_i(v)|x_i, \mathbf{z}_i\right] = \beta_0(v) + \mathbf{z}_i\boldsymbol{\beta}_1(v) + x_i \cdot \gamma(v). \tag{2}$$

   In this formulation, conditional correlations that are derived explicitly from a data-generating model (Tu et al. 2024) offer a clear GLM-based interpretation, for which testing individual differences is straightforward. Specifically, if there is an association between $\rho_i(v)$ and x_i, then $\gamma(v) \neq 0$; otherwise, $\gamma(v) = 0$. At vertex v, we then test the null hypothesis $H_0 : \gamma(v) = 0$ against the alternative, $H_A : \gamma(v) \neq 0$. This involves computing a score test statistic at each vertex (Appendix A.4).

IV. *Cluster enhancement.* Instead of directly using the score test statistic as derived in the previous step, we perform cluster enhancement to enhance signal and improve statistical power. To do this, we compute candidate test statistics for a given vertex v, which are derived by aggregating vertex-level score test statistics estimated in Stage III across neighboring vertices across up to a maximum user-specified radius (e.g., 15mm). We have previously shown that this approach can improve statistical power in studies of intermodal associations, particularly when signals of interest are spatially localized (Weinstein et al. 2022).

V. *Determining statistical significance.* We perform inference—that is, the assessment of statistical significance at each vertex—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. For each of 1,000 (or more) permutations, Stages III-IV are repeated with x_i permuted across participants, which controls the family-wise error rate at a predetermined α (e.g., 0.05). A detailed procedure is described in Appendix A.5. Additionally, we note that alternative strategies for permutation and covariate adjustment may be suitable depending on the context, and we refer readers to Winkler et al. (2014).

2.3.1 Other approaches

In addition to CIDeR, we consider two alternative frameworks for quantifying intermodal coupling.

IMCo. Vandekar et al. (2016) used locally weighted regression 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 from within a predetermined radius around the central verrex. This procedure yields a slope at each vertex, representing local intermodal coupling for each participant. These individual-level coupling maps could then be used to test for local associations with behavioral or demographic.

In this article, we make several adjustments in our implementation of IMCo (compared with those deployed in Vandekar et al. (2016) and Baller et al. (2022), for example) to make our comparison to CIDeR as fair as possible.

1. In earlier implementations of IMCo, covariates were adjusted for only in the context of estimating individual differences (i.e., Figure 2(b)), but not when computing intermodal coupling (i.e., Figure 2(a)). Notably, such an approach could theoretically result in leakage of modality-specific covariate effects into intermodal coupling measures. To improve comparability to CIDeR and minimize risk of modality-specific covariate effect leakage, we residualize participant-level brain maps with respect to both x_i (covariate of interest) and $\mathbf{z}_i$ (confounders).

2. The regression slopes derived from earlier implementations of IMCo 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. In other words, if one were to flip the choice of dependent and independent variables, the coupling measurements would change. In contrast, CIDeR is invariant to the ordering of the two modalities. To circumvent this discrepancy, instead of locally weighted regression slopes, we estimate locally weighted *correlations* to define subject-level intermodal coupling (Figure 2(a)).

3. As illustrated in Figure 2(b), IMCo is the only method that performs spatial smoothing when computing subject-level intermodal coupling measures. Thus, it is conceivable that a larger radius used in IMCo (e.g., 15mm as in Vandekar et al. (2016) and Baller et al. (2022)) could render different biological interpretations of intermodal coupling. Therefore, in addition implementing IMCo with a 15mm radius, we also consider a 5mm radius. In this setting, the extent of spatial smoothing is much more limited, making the the definitions of coupling more comparable across the methods (note that reducing the radius to 0mm would equate to the mass-univariate approach). We also note that for CIDeR, we fixed the maximum radius to 15mm for cluster enhancement. While IMCo uses a fixed radius, CIDeR adaptively selects an optimal radius through a procedure detailed in earlier work (Weinstein et al. 2022). Again, we emphasize that CIDeR's use of varying radii for cluster enhancement does not alter how it defines intermodal coupling; rather, it serves to improve in the detection of individual differences in intermodal coupling (see Figure 1, Stage IV and Figure 2(c)-(d)).

4. Finally, while earlier implementations of IMCo used the false discovery rate (FDR) to adjust for multiple comparisons across image locations, we used permutation-based testing to control FWER, ensuring a fairer comparison across methods.

Mass-univariate. To emphasize the advantage of leveraging the spatial structure the data (e.g., through cluster enhancement), we also consider a "mass-univariate" approach. This involves estimating vertex-level test statistics measuring individual differences in intermodal coupling—but without lever-

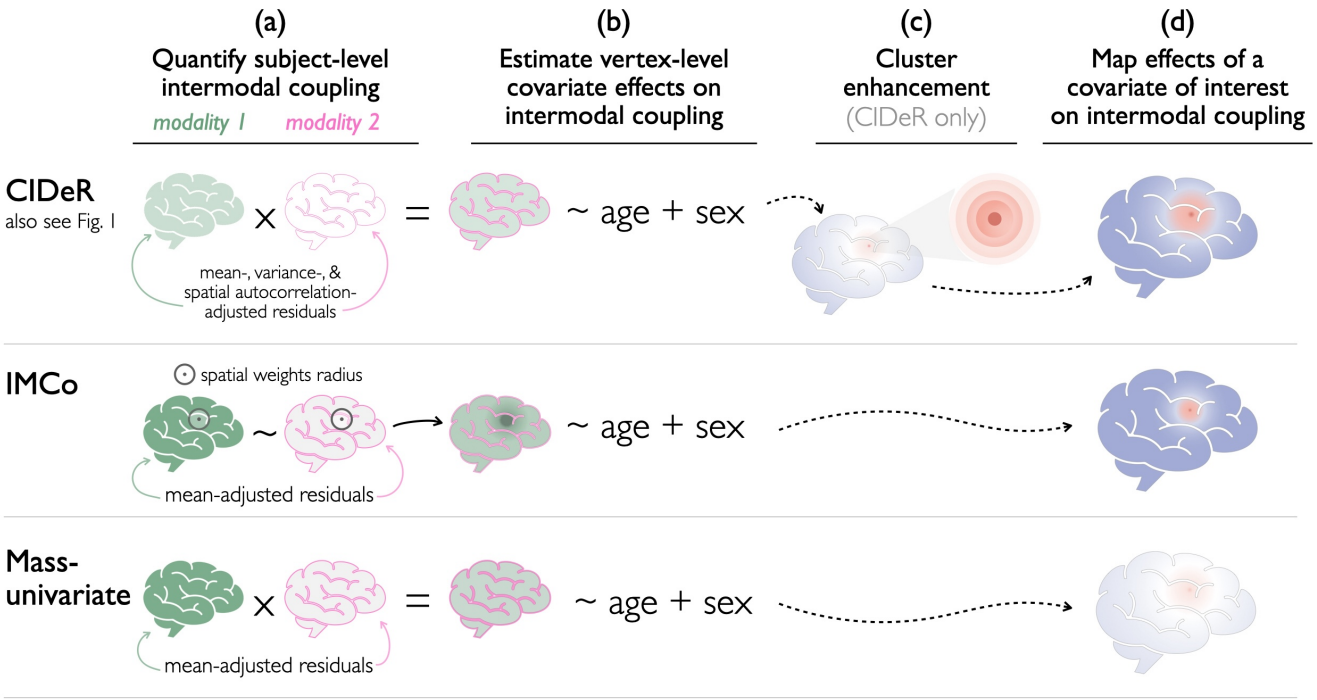

Figure 2. Overview of the proposed method, CIDeR, along with two alternative approaches to quantifying intermodal coupling (a-b) and individual differences (c-d). A more detailed illustration of CIDeR is provided in Figure 1. We perform statistical inference for all three methods using the same permutation-based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

aging neighboring information. In this article, our implementation of the mass-univariate framework is equivalent to IMCo with a radius of 0mm. (It is also equivalent to CIDeR, if both Stage I(ii) and Stage IV are excluded). While the mass-univariate approach is the most simple of the three considered in this article, we anticipate its statistical power will be much lower than its competitors given well-established advantages of incorporating spatial information for enhanced power.

Inference. While CIDeR, 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 statistical testing across the three methods. In our implementation, we randomly permute the covariate of interest, x_i across participants (after adjusting for confounders $\mathbf{z}_i$), and then repeating the steps outlined in Figure 2(b)-(d) to estimate null effects of a covariate of interest on intermodal coupling. To control FWER at α, a brain-wide critical value, t_α (achieved by the max T approach, Westfall and Young (1993)), is used to determine local statistical significance (see Park and Fiecas (2022) and Weinstein et al. (2022)). We also note that different approaches to permutation may be preferable depending on features of of the data, and extensions of CIDeR may include procedures detailed by Winkler et al. (2014).

Considerations when comparing results across methods. In practice, researchers might choose to apply any of the three methods discussed in this article to assess individual differences in intermodal coupling. Thus, when we compare CIDeR to other approaches, it is important to note that subtle differences in the approaches of these methods may lead to downstream differences in interpretation of results.

To highlight differences in the three approaches, Figure 2 provides an overview of their main features (with a more detailed illustration of CIDeR provided in Figure 1). It is important to highlight that IMCo (Vandekar et al. 2016) may be testing a different null hypothesis compared to both CIDeR and the mass-univariate approach, given that its definition of intermodal coupling is fundamentally different (with the extent of this difference depending on the radius it uses in Figure 2(a)). In contrast to IMCo, CIDeR attempts to initially *reduce* the influence of spatial autocorrelation when computing intermodal coupling (Figure 1, Stages I-II). Subsequently, CIDeR *leverages* spatial information through cluster enhancement when computing test statistics for individual differences (Figure 1, Stage IV and Figure 1(c)). While CIDeR and IMCo make use of the inherent spatial structure of brain imaging data in different ways, the mass-univariate approach does not leverage the spatial structure of the data in any way; as our results will show, this makes for much less powerful inference.

2.4 Data application: the Philadelphia Neurodevelopmental Cohort

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). Details about the PNC study and preprocessing are provided in Appendix B. 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).

2.4.1 Comparison of methods: type I error control and power

To compare our method to others, we evaluate their performance in terms of their ability to control false positives and maximize power (i.e., true positives). To do this, we conduct data-driven simulation studies by repeatedly selecting random subsets of participants from the full PNC sample and applying each method under both null and alternative settings to estimate the rate of false positives (type I error) and true positives (power), respectively.

Type I error simulations. To verify that false positive findings are limited, we use real data to emulate a setting where the null hypothesis is true. This is done by randomly shuffling a covariate of interest (age) across participants in the full sample. We then select a random subsample of 300 participants from the (null version of the) full sample and apply each method to test whether structural coupling (CT-SD) differs with respect to null covariate effects. If a significant association is found at even one vertex, then the null hypothesis is rejected for that sub-sample. We estimate the type I error rate for each method as the proportion of sub-samples for which the null hypothesis was rejected at the nominal $\alpha = 0.05$ level. For our type I error simulations, we conduct 5,000 permutations per test to obtain a null distribution.

Power simulations. Similar to the type I error simulations (but without first permuting the covariate of interest), we repeatedly draw random sub-samples of N=300 participants for each pair of modalities and apply each method to test individual differences with respect to both phenotypes of interest (age and sex) and both pairs of modalities (CT-SD and CBF-ALFF). Vertex-level power, which can be spatially mapped across vertices, is calculated as the proportion of simulations for which the test statistic for each vertex exceeds the brain-wide threshold t_α is the $(1 - \alpha) \times 100$th percentile of the FWER-controlling permutation-based null distribution.

3 Results

3.1 Results from full PNC sample

Results from our analysis of individual differences in structural (CT-SD) and neurovascular (CBF-ALFF) coupling are presented in Figure 3. For each method, we map local test statistics quantifying individual differences in coupling that exceeded the brain-level significance threshold (t_α). Thus, only the $T(v)$ values at locations with statistically significant associations are visible, and each brain map highlights regions where significant age (conditional on sex) or sex (conditional on age) differences in intermodal coupling are found. We will focus our interpretation on results from CIDeR and IMCo, as the mass-univariate approach is much less powerful, offering less to interpret. We also note that, across the different tests of associations, our two implementations of CIDeR (with and without adjustment for variance heterogeneity) perform nearly identically. Subsequent results will therefore only include the version of CIDeR without variance heterogeneity adjustment, and our interpretation below does not distinguish between the two.

Maps of age effects on structural (CT-SD) coupling suggest overlap across the different implementations of CIDeR (with and without adjustment for variance heterogeneity) and IMCo (with 5mm and 15mm spatial weights radii). Both methods show patterns of negative age effects (blue) on structural coupling in the visual cortex and positive age effects (red) in the motor cortex, consistent with Vandekar et al. (2016). Although the implementation of IMCo using a 15mm radius of spatial weights localizes significant age effects across a somewhat larger area of the cortical surface than the other implementations, this can likely be attributed to differences in how each method defines intermodal coupling (Figure 2(a)). Indeed, it is not surprising that these differences are most drastic when a larger radius is used for IMCo.

In the context of neurovascular coupling, however, CIDeR detects age effects across a larger portion of the cortical surface compared to IMCo. Age effects appear prominent in the attention networks, which aligns with results in Baller et al. (2022).

Interestingly, qualitative differences between CIDeR and IMCo are more apparent in tests of sex (female versus male) effects. For structural 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. CIDeR—but not IMCo—also reveals significant positive associations between female sex and structural coupling in the superior frontal and anterior cingulate cortices. Additionally, while IMCo (with 15mm radius) reveals patterns of a positive association between female sex and neurovascular coupling overlapping in the frontoparietal network (consistent with Baller et al. (2022)), CIDeR reveals

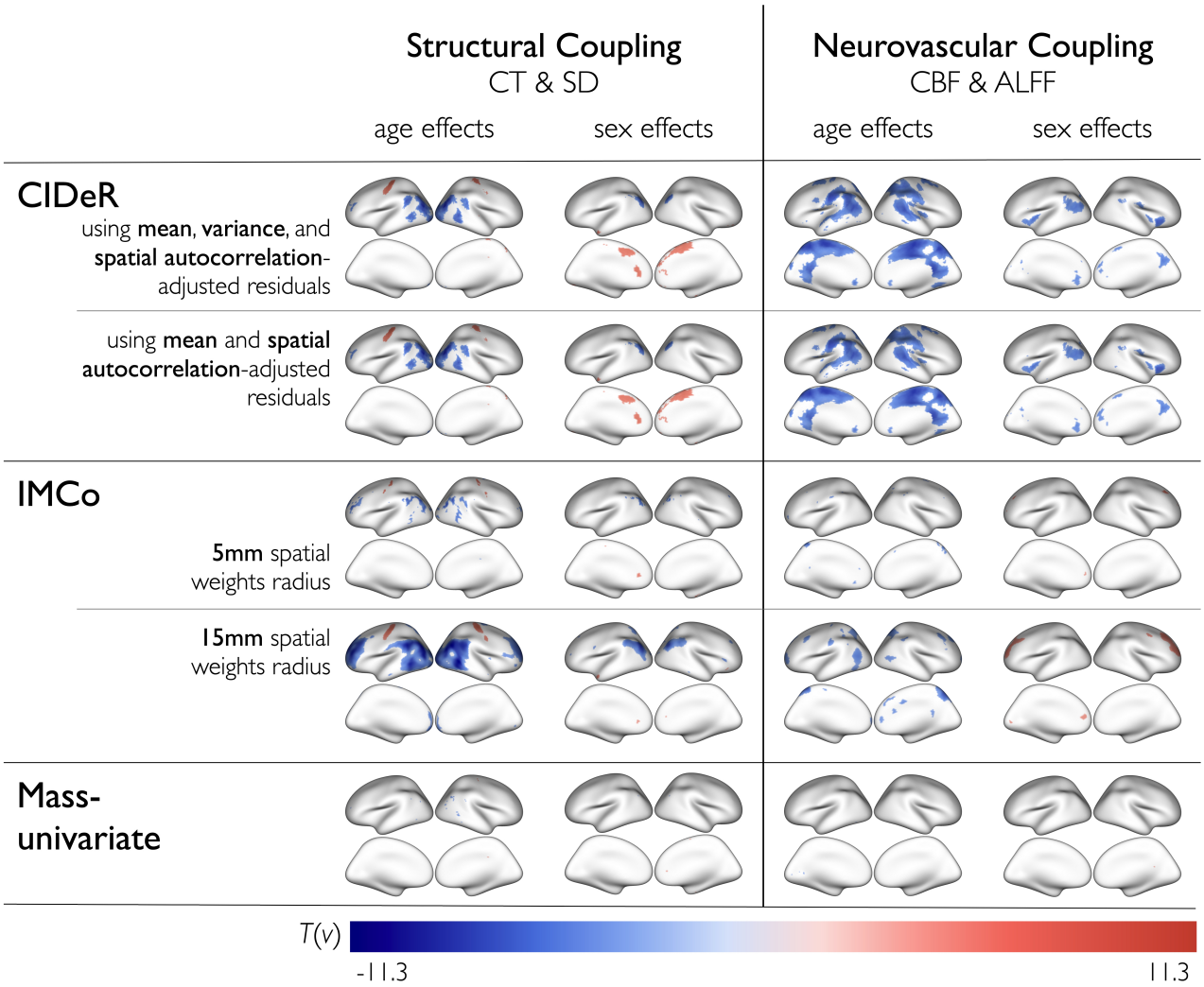

Figure 3. Full-sample results: thresholded test statistic maps and global p-values. Negative values of $T(v)$ (blue) correspond to regions where intermodal coupling measures are found to be negatively associated with a covariate of interest, while positive values of $T(v)$ (red) correspond to regions where intermodal coupling measures are positively associated with the covariate. Note that sex is coded as a binary variable with 1 for females and 0 for males; therefore, regions where we observe $T(v) > 0$ for sex effects would indicate where coupling for females was found to be higher for females versus males, while regions where we observe $T(v) < 0$ for sex effects would indicate where it is estimated to be lower for females versus males.

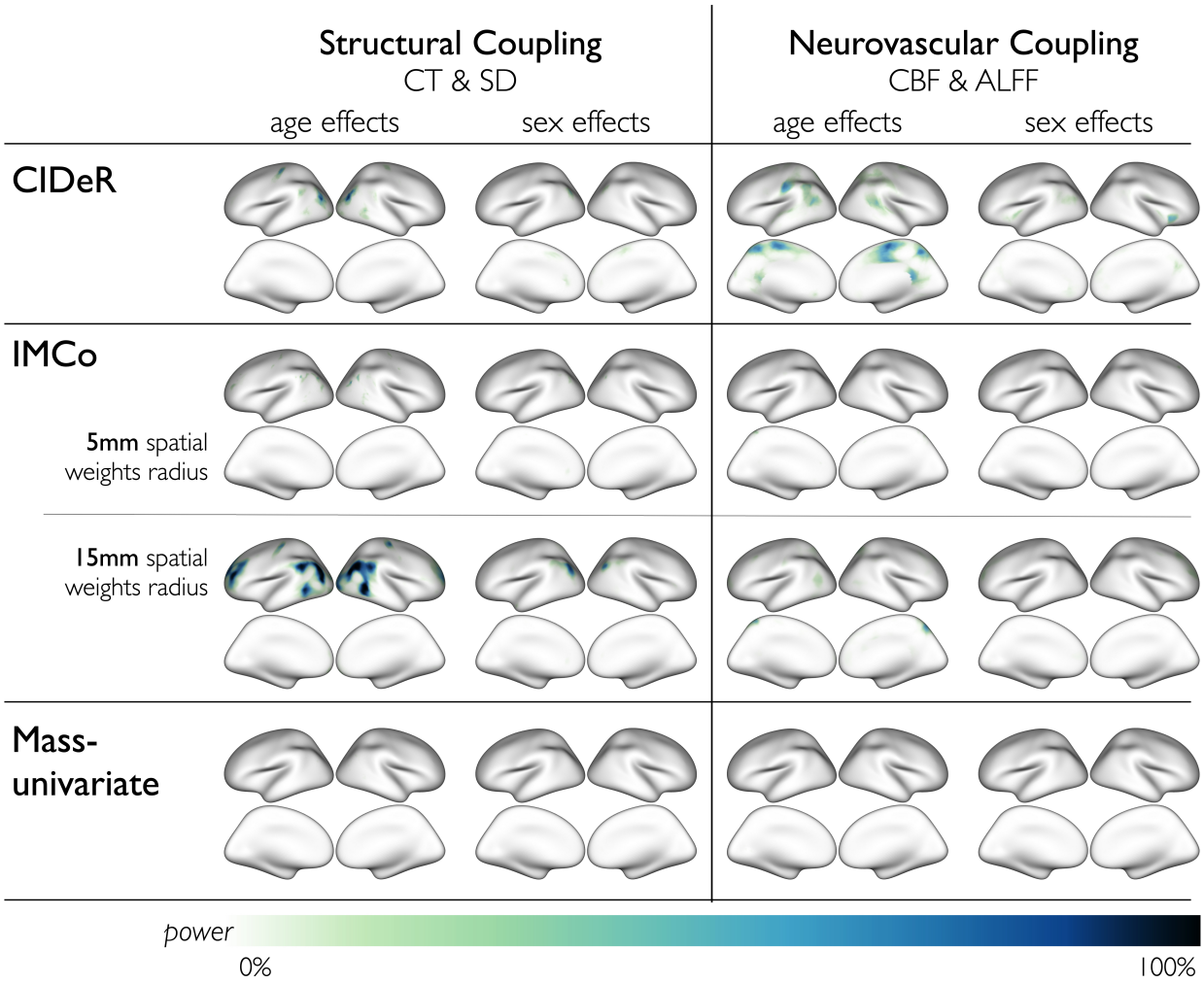

Figure 4. Power simulation results in subsamples of N=300 from the PNC study. The local power for a vertex is the rate at which the test statistic value exceeds the brain-level significance threshold (t_α) across 1,000 sub-samples drawn from the PNC data. In each subsample, we used 999 permutations to determine brain-level significance.

negative sex effects in other regions not detected by IMCo. Again, we believe that differences in the methods' performance may be attributed to differences in their definitions of intermodal coupling (see Figure 2).

3.2 Simulation results: type I error and power

False positive control. To assess false positive control of each method, we apply each method in 5,000 sub-samples of the PNC data where the null hypothesis is emulated (i.e., covariate effects on intermodal coupling are absent at all vertices). This is accomplished by permuting the covariate of interest in the full sample before randomly selecting sub-samples of 300 participants, in which each test is applied to test for (null) covariate effects on structural coupling. As expected, we find that all methods successfully control false positive findings at approximately 5%. In other words, in only 5% of simulated datasets in which covariate effects were null at *all* vertices, each method detected significant covariate effects (at at least one vertex) approximately 5% of the time.

Replicability in smaller samples. Since the sample size of the PNC is quite large ($N = 911$ for structural coupling; $N = 831$ for neurovascular coupling), it is important to consider these methods' replicability in smaller samples. Therefore, in Figure 4, we display brain maps localizing statistical power (i.e., true positive rate) after each method is repeatedly applied in random subsets of $N = 300$ from the PNC. As noted earlier, here we only present results for the version of CIDeR that includes adjustment for mean heterogeneity and spatial autocorrelation (but not adjustment for variance heterogeneity), given that results were nearly identical in both cases in Figure 3. Here, each vertex is shaded according to the proportion of simulations where the null hypothesis was rejected. Darker shades of blue indicate the highest levels of statistical power, light shades of green indicate low-to-moderate power, and white indicates negligible power. We expect the spatial patterns of these power maps to replicate those shown for our full-sample analysis (Figure 3), as the power simulations involve tests of the same hypotheses except in smaller samples.

Similar to the full-sample analysis, both CIDeR and IMCo (with 15mm radius) detect significant age effects on structural coupling in the visual and motor cortices. While our implementation of IMCo with a 15mm radius for local weights reveals higher power across a larger portion of the brain compared to CIDeR, this is likely due to the span of the spatial smoothing kernel IMCo uses to quantify coupling. When we implement IMCo using local weights from 5mm surrounding each vertex, the method's ability to replicate across sub-samples is substantially worse. For sex effects on structural coupling, both CIDeR and IMCo (15mm radius) display power localized in the visual cortex. Also consistent with our results from Figure 3, CIDeR (but not IMCo) appears to have low-to-moderate power for detecting significant sex effects on structural coupling in the frontal cortex.

For age effects on neurovascular coupling, CIDeR is by far the most successful at replicating local signal across simulations, including in the attention networks. Importantly, this is consistent with our findings from Figure 3, which involved a sample size nearly three times larger than the one considered in our power simulations. While sex effects on neurovascular coupling appear to only have moderate local power, CIDeR nevertheless outperforms IMCo. The localized sex effects on neurovascular coupling detected by IMCo in the full-sample analysis fail to replicate in our power simulations. However, the sex effects detected by CIDeR in the full sample appear to replicate with moderate-to-high power, as shown in Figure 4.

4 Discussion

Studies of intermodal coupling are an increasing focus of developmental cognitive neuroscience research. Here, we introduced a new method, called CIDeR, which may be useful in studies examining individual differences in intermodal coupling. Measures of intermodal coupling have previously been shown to be 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 address the question of statistical power and replicability, which lays the groundwork for more rigorous and generalizable research for coupling-based biomarkers in neurodevelopmental research.

To this end, we proposed CIDeR, a new statistical approach to mapping local individual differences in intermodal coupling. CIDeR provides a powerful framework for hypothesis testing that can offer replicable spatial localization in studies aiming to map brain areas in which individual differences in intermodal coupling are most prominent. Leveraging both structural and functional MRI data from the

Philadelphia Neurodevelopmental Cohort, we found significant age-related differences in coupling between both structural and neurovascular features of the developing brain.

CIDeR adapts the GLM framework commonly used in neuroimaging analyses to the conditional correlations framework described by Tu et al. (2024). This framework provides a clear statistical hypothesis for (individual differences of) intermodal coupling and includes controlling possible sources of variation in the data (e.g., confounding) that could increase false positive findings and dramatically improve power using the CLEAN framework(Park and Fiecas 2022; Weinstein et al. 2022) to perform spatial modeling and cluster enhancement. For evaluation of significant associations, we employ permutation-based hypothesis testing, which offers conservative type I error control based on the family-wise error rate (FWER). Consistent with the objectives of this article, our proposed method supports CIDeR's potential to offer more replicable and interpertable biomarkers in developmental neuroscience.

While the CIDeR and IMCo methods presented in this article are related, they differ in the biological quantities that they assess (Figure 2). Consequently, they may produce different results (e.g., Figure 3). Consistent with many recent neuroimaging methods, and based on our simulation results (Figure 4), the fact that both methods utilize spatial information (rather than assume independence across vertices) plays an important role in their power. However, IMCo utilizes spatial information when computing weighted subject-level estimates of intermodal coupling (Figure 2(a)), whereas CIDeR calculates coupling at each vertex separately but subsequently leverages spatial information when computing group-level measures of individual differences (Figure 2(c)). In other words, IMCo assumes that the association between the two modalities is similar within a neighborhood of each vertex; CIDeR instead assumes that the strength of the signal (even if the signals are different) are similar in a neighborhood of each vertex.

CIDeR has several limitations, which open several opportunities for extensions of the method. In this article, we leveraged cross-sectional data from the PNC; however, CIDeR 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 changes, in intermodal coupling differ by clinical groups or sex (Park et al. 2021; Bernal-Rusiel et al. 2013a; Bernal-Rusiel et al. 2013b). Additionally, CIDeR 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 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

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. **RP**: Methodology, Writing - review & editing.

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, Funding acquisition.

Declaration of competing interest

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

Data and code availability

Raw neuroimaging data from the PNC study are publicly available at the dbGaP (phs000607.v3.p2). All analyses were run using the R version 4.2. Code will be made available on GitHub at the time of publication.

Acknowledgments

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.

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.

Draper, Norman R and David M Stoneman (1966). "Testing for the inclusion of variables in einear regression by a randomisation technique". In: *Technometrics* 8.4, pp. 695–699. DOI: 10.1080/00401706.1966.10490413.

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.

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.

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.

Westfall, Peter H and S Stanley Young (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*. Vol. 279. John Wiley & Sons.

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.

Zhang, Rongqian, Linxi Chen, Lindsay D Oliver, Aristotle N Voineskos, and Jun Young Park (2024). "SAN: mitigating spatial covariance heterogeneity in cortical thickness data collected from multiple scanners or sites". In: *Human Brain Mapping* 45.7, e26692. DOI: 10.1002/hbm.26692.

Appendix

A CIDeR: methodological details

A.1 Stage I: Modality-specific modeling

A.1.1 Adjustment for mean and variance effects

For modality m, vertex v, and subject i, we model covariate effects on the mean and variance of the value $y_{mi}(v)$ jointly, as follows:

$$y_{mi}(v) = \alpha_{m0}(v) + \underbrace{\mathbf{z}_i\boldsymbol{\alpha}_m(v) + x_i\theta_m(v)}_{\text{Individual differences in means}} + \underbrace{f_m(\mathbf{z}_i, x_i, v) \times \epsilon_{mi}(v)}_{\text{Individual differences in variances}}, \tag{3}$$

where $\epsilon_{mi}(v)$ has unit variance. Specifically, in Equation (4), the covariates $\mathbf{z}_i$ affect the mean of $y_{mi}(v)$ through the term $\alpha_{m0}(v)+\mathbf{z}_i\boldsymbol{\alpha}_m(v)+x_i\theta_m(v)$, and the variance of $y_{mi}(v)$ through the term $f_m(\mathbf{z}_i, x_i, v)$. If there are no covariate effects on variances, then $f_m(\mathbf{z}_i, x_i, v)$ is simply a constant. Among several options for variance functions, we choose a two-component error model involving a covariate-invariant portion and a covariate-dependent portion. This is implemented in `R` via the `varConstProp` function in the `nlme` package. Once all parameters are estimated via (generalized) least squares, we obtain the scaled residuals as follows:

$$\hat{\epsilon}_{mi}(v) = \frac{y_{mi}(v) - \hat{\alpha}_{m0}(v) - \mathbf{z}_i\hat{\boldsymbol{\alpha}}_m(v) - x_i\hat{\theta}_m(v)}{\hat{f}_m(\mathbf{z}_i, x_i, v)}. \tag{4}$$

A.1.2 Adjustment for spatial autocorrelation

As different modality maps may possess distinct autocorrelation structures, we perform spatial adjustment separately using parametric spatial autocorrelation functions (SACFs). We also allow for scenarios where different SACF families are suitable for different modalities. To model modality-specific spatial autocorrelation, we first define proximity between two vertices on the cortical surface as the the geodesic distance (in mm) obtained from a triangular mesh surface. Next, we model $\epsilon_{mi}(v) = b_{mi}(v)+\delta_{mi}(v)$ where $b_{mi}(v)$ captures any spatial dependencies of an image in each modality informed by geodesic distances and $\delta_{mi}(v)$ captures non-spatial variations that would be attributed to potential intermodal correlations:

Here, $\delta_{mi}(v) \sim \mathcal{N}(0, \tau_m^2)$ is white noise, at least marginally within each image modality. $b_{mi}(v)$ is a random effect that characterizes the spatial dependence within each subject's image, which is modeled by

$$(b_{mi}(1), \ldots, b_{mi}(V)) \overset{i.i.d}{\sim} \mathcal{MVN}(\mathbf{0}_V, \sigma^2_{m,G} \cdot \boldsymbol{\Phi}_G(\phi_{m,G}) + \sigma^2_{m,E} \cdot \boldsymbol{\Phi}_G(\phi_{m,E})) \quad \text{for} \quad i = 1, \ldots, N, \tag{5}$$

where $\boldsymbol{\Phi}_G(\phi_{m,G})$ and $\boldsymbol{\Phi}_E(\phi_{m,E})$ are Gaussian and exponential SACFs. Specifically, the $(v, v^\star)$th element of $\boldsymbol{\Phi}_G(\phi_{m,G})$ is $\exp(-\phi_{m,G} \times d_{v,v^\star})$ and the $(v, v^\star)$th element of $\boldsymbol{\Phi}_E(\phi_{m,E})$ is $\exp(-\phi_{m,E} \times d_{v,v^\star})$. This formulation is equivalent to that previously introduced by Park and Fiecas 2022 and Weinstein et al. 2022. The covariance model in Equation (5) extends Park and Fiecas 2022 and Weinstein et al. 2022 by using a combination of the Gaussian and exponential SACFs (modeled by $\boldsymbol{\Phi}_G$ and $\boldsymbol{\Phi}_E$), which have shown to be flexible and robust in capturing varying degrees of spatial autocorrelation (Zhang et al. 2024).

A.2 Stage II: Quantifying intermodal coupling

Once model (5) is fit (Zhang et al. 2024; Park and Fiecas 2022), $\hat{\delta}_{mi}$ is obtained by the conditional expectation of δ_{mi} given $\hat{\epsilon}_{mi}$, which is used in Stage II. At vertex v, we obtain a realization of the intermodal correlation $\rho_i(v)$ with $\hat{\delta}_{1i}(v) \cdot \hat{\delta}_{2i}(v)/(\hat{\tau}_1(v) \cdot \hat{\tau}_2(v))$, where $\hat{\tau}_m(v)$ are the estimated variances of modality m at vertex v from the stage 1 model.

A.3 Stage III: Modelling the association between intermodal coupling and covariates of interest

For subject i and at vertex v, we model the intermodal correlation $\rho_i(v)$ as a conditional correlation through the linear model specified similarly as Tu et al. (2024):

$$E\left[\rho_i(v)|x_i, \mathbf{z}_i\right] = \tanh(\beta_0(v) + \underbrace{\mathbf{z}_i\boldsymbol{\beta}_1(v) + x_i \cdot \gamma(v)}_{\substack{\text{Individual differences} \\ \text{in correlations}}}). \tag{6}$$

Note that $\tanh(\cdot)$ is a link function ensuring that the correlations lie strictly between -1 and 1. Although tanh is not straightforward to model in terms of 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. Under this approximation the Equation (6) is reduced to a GLM formulation of

$$E\left[\rho_i(v)|x_i, \mathbf{z}_i\right] = \beta_0(v) + \mathbf{z}_i\boldsymbol{\beta}_1(v) + x_i \cdot \gamma(v). \tag{7}$$

Equation (7) is now equivalent to the setup of CLEAN (Park and Fiecas 2022), a statistical method for testing an covariate effect on a dependent variable (e.g., a realization of the conditional correlation) at vertex v. Just as CLEAN improved power and sensitivity by an explicit autocorrelation modeling and cluster enhancement, our formulation is expected to improve power and sensitivity in studying individual differences in intermodal coupling.

A.4 Stage IV: cluster enhancement

To test if $H_0 : \gamma(v) = 0$ at each vertex v, we use the score test statistic given by a linear approximation:

$$U(v) = \sum_{i=1}^{N} x_i \times \left(\frac{(\hat{\delta}_{1i}(v) \cdot \hat{\delta}_{2i}(v)}{\hat{\tau}_1(v) \cdot \hat{\tau}_2(v)} - \tilde{\beta}_0(v) - \mathbf{z}_i\tilde{\boldsymbol{\beta}}_1(v) \right),$$

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

We apply cluster enhancement to the vertex-level test statistics $U(v)$ to leverage local spatial similarity of the effect sizes and improve power (Weinstein et al. 2022). 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

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

and $N_r(v)$ is a 'candidate' cluster, i.e., the set of all vertices whose distances from the vertex v is less than

or equal to r. Note that this article uses $r_{\max} = 15$mm for a fair comparison with other methods, while existing work used $r_{\max} = 20$mm as a default.

A.5 Stage V: Permutation-based inference

Permuting the data allows us to draw samples under the null hypothesis where this association is zero. Note that it is still possible for the average image values in $\mathbf{y}_{1i}$ and $\mathbf{y}_{2i}$ to depend on x_i. Because our hypothesis of interest is the covariate effect on intermodal correlations and not on means or variances in each modality, we need only permute x_i in the stage III model to generate permuted samples, while all the other steps in stages I and II fixed. This procedure is based on the approach of Draper and Stoneman (1966). A major computational advantage of this is that there is no need to refit the stage I and stage II models in each permutation, which boosts computational efficiency. Therefore, all permutation analyses can begin with the data $\hat{\delta}_i^{(1)}(v), \hat{\delta}_i^{(2)}(v)$. As such, this permutation scheme becomes equivalent to using the CLEAN pipeline (Park and Fiecas 2022) in stage III.

B Philadelphia Neurodevelopmental Cohort

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.

B.1 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.