Sequential Pathway Inference for Multimodal Neuroimaging Analysis

Lexin Li[1] | Chengchun Shi[2] | Tengfei Guo[3] | William J. Jagust[3]

[1]Department of Biostatistics and Epidemiology, University of California, Berkeley, CA, USA

[2]Department of Statistics, London School of Economics and Political Science, London, UK

[3]Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA

Correspondence

*Lexin Li, Department of Biostatistics and Epidemiolog, University of California, Berkeley, CA, USA. Email: lexinli@berkeley.edu

Abstract

Motivated by a multimodal neuroimaging study for Alzheimer's disease, in this article, we study the inference problem of sequential mediation analysis. The existing sequential mediation solutions mostly focus on sparse estimation, while inference is an utterly different and more challenging problem. Meanwhile, the few mediation inference solutions often ignore the potential dependency among the mediators, or cannot be applied to the sequential problem directly. We propose a statistical inference procedure to test mediation pathways when there are sequentially ordered multiple data modalities and each modality involves multiple mediators. We allow the mediators to be conditionally dependent, and the number of mediators within each modality to diverge with the sample size. We produce the explicit significance quantification and establish the theoretical guarantees in terms of asymptotic size, power, and false discovery control. We demonstrate the efficacy of the method through both simulations and an application to a multimodal neuroimaging pathway analysis of Alzheimer's disease.

KEYWORDS:

Alzheimer's disease; Boolean matrix; Directed acyclic graph; High-dimensional inference; Mediation analysis; Multimodal neuroimaging analysis.

1 | INTRODUCTION

Alzheimer's disease (AD) is an irreversible neurodegenerative disorder, characterized by progressive impairment of cognitive functions, then global incapacity and ultimately death. It is the leading form of dementia, and is currently affecting 5.8 million American adults aged 65 years or older. Its prevalence continues to grow, and is projected to reach 13.8 million by 2050 (Alzheimer's Association 2020). Multimodal neuroimaging is frequently used in AD research, where different brain characteristics are measured using different imaging technologies for a common set of subjects. Notable AD imaging biomarkers include, among others, grey matter cortical thickness measured by structural magnetic resonance imaging (MRI), amyloid-beta ($A\beta$) protein deposition measured by positron emission tomography (PET) using a variety of radiopharmaceuticals such as ^{18}F-florbetapir or ^{18}F-florbetaben tracer, and tau protein measured by PET using radiopharmaceuticals such as ^{18}F-flortaucipir tracer. While brain grey matter atrophy is a well-known measure associated with AD progression, $A\beta$ and tau are two hallmark pathological proteins believed to be part of the driving mechanism of AD. Models of AD pathophysiology hypothesize a temporal sequence in which disruptions in $A\beta$ production, clearance, or both initiate a biological cascade that leads to $A\beta$ plaque formation, then neurofibrillary tangles of tau, followed by structural atrophy and neuronal dysfunction, and ultimately clinical decline in cognitions (Jack et al. 2013 2010). These models have been continuously supported by new evidences (see, e.g., Gordon et al. 2018; Guo, Korman, Baker, Landau, & Jagust 2020; Jack et al. 2019, among others). However, many questions about the nature of this pathological cascade and its spatial distribution in the brain remain unanswered. As part of the Berkeley Aging Cohort Study, multimodal neuroimages were collected, and a key scientific question is to understand how $A\beta$ deposition affects cognitive decline through possible intermediate pathways of regional tau deposition, then cortical thinning.

This problem can be cast into the framework of *sequential mediation analysis*. Mediation analysis seeks to identify and explain the mechanism, or pathway, that underlies an observed relationship between an exposure and an outcome variable, through the inclusion of an intermediate variable known as a mediator. It decomposes the effect of exposure on the outcome into a direct effect and an indirect effect, the latter of which indicates whether the mediator is on a pathway from the exposure to the outcome. In our AD example, the $A\beta$ deposition, summarized by a single measure for each subject, serves as the exposure. The cognitive decline, measured by the decrease of memory scores at two consecutive visits about one year apart, serves as the outcome. The brain regional tau deposition and cortical thickness, each summarized by a vector of measures for multiple brain regions for each subject, serve as potential mediators.

Mediation analysis was first proposed with a single mediator (Baron & Kenny 1986), and has been extended to the setting of multivariate and high-dimensional mediators; see VanderWeele (2016) for a review and the references therein. However, most of the existing work has focused on a *single* modality of features as potential mediators. By contrast, we target the problem of mediation analysis that involves *multiple* modalities and each modality includes *multivariate* mediators. Moreover, different sets of mediators are *sequentially* ordered on the potential pathways following certain biological constraints. For instance, in our example, tau deposition precedes cortical thickness that mediates cognition.

Recently, Lai, Shih, Huang, and Wang (2020) proposed a probit model for multiple sequentially ordered mediators for a dichotomous outcome, but only considered a *univariate* mediator for each modality. It is far more challenging to analyze multi-modality of multivariate mediators. Zhao, Li, and Caffo (2020) proposed a penalized Lasso approach for two modalities of sequentially ordered and multivariate mediators. However, they focused on *estimation*, instead of *inference* for sequential mediation analysis. Even though estimation and inference are closely related, and both can in effect identify important pathways, estimation does not produce an explicit quantification of statistical significance, and does not explicitly control the false discovery. By contrast, we aim at statistical inference for sequential mediation analysis with multi-modality of multivariate mediators.

Inference for individual mediators is challenging. The key difficulty lies in the fact that the total number of potential paths that go through any mediator is super-exponential in the number of mediators, rendering almost any testing procedure ineffective. Consequently, most existing mediation inference solutions either explicitly impose that the mediators are conditionally independent given the exposure, or simply ignore any potential directed paths among the mediators (e.g., Djordjilović et al. 2019; Huang & Pan 2016; Zhang et al. 2016). Though proven useful in some applications, ignoring potential paths and interactions among the mediators seems not sensible in plenty of other scientific applications; e.g., different brain regions are generally conceived to influence each other, and different genes are expected to interact with each other. More recently, Chakrabortty, Nandy, and Li (2018) proposed a mediation test based on interventional calculus, and Shi and Li (2021) proposed a test based on logic of Boolean matrices, while both allowed interactions among the mediators. Nevertheless, none of the existing mediation inference solutions tackle multi-modality of mediators.

In this article, we propose a statistical inference procedure to test mediation pathways when there are sequentially ordered multiple modalities and each modality involves multivariate mediators. We allow the mediators to be conditionally dependent, and the number of mediators within each modality to diverge with the sample size. We produce the explicit significance quantification, and establish the theoretical guarantees, including the asymptotic size and power of the test, and a valid false discovery control. Our proposal makes several useful contributions. Scientifically, understanding mediation pathways of different imaging modalities in AD progression is a crucial but open question. Our test offers the first solution that does not have to impose any restrictive conditional independence condition and is theoretically guaranteed. Moreover, even though motivated by a multimodal neuroimaging example, our test is applicable to a wide range of multimodal analyses, e.g., the multi-omics data analysis (Richardson, Tseng, & Sun 2016), and the multimodal healthcare application (Cai, Wang, Li, & Liu 2019). Methodologically, our test extends Shi and Li (2021); however, it is far from a straightforward extension. Particularly, both our test and that of Shi and Li (2021) rely on an estimator of the directed acyclic graph (DAG) that encodes the directional relationships among all the variables. In our setup, we need to incorporate the sequential constraint to characterize the orders of multiple modalities of mediators. Naively employing the initial estimator in Shi and Li (2021) would fail this constraint, then fail the test. Actually, most existing DAG structure estimation methods neglect this constraint, and we need to devise an approach to carefully embed this constraint into DAG estimation. As such, our proposal also makes a useful addition to the toolbox of mediation inference.

The rest of the article is organized as follows. We formally specify the model and the hypotheses in Section 2. We develop the testing procedure in Section 3, and study the asymptotic properties in Section 4. We present the simulations in Section 5, and revisit the multimodal neuroimaging application in Section 6. We conclude the paper with a discussion in Section 7. We relegate all technical proofs to the online supplement.

2 | MODEL AND HYPOTHESES

We begin with a Gaussian graphical model, based on which we formulate the mediation testing problem and formally define our hypotheses. For simplicity, we only consider two sets of mediators, while our proposal can be extended to more than two modalities.

Let $Y \in \mathbb{R}$ denote the outcome variable, $E \in \mathbb{R}$ denote the exposure variable, $X_1 = (X_{11} \ldots, X_{1d_1})^\top \in \mathbb{R}^{d_1}$ and $X_2 = (X_{21} \ldots, X_{2d_2})^\top \in \mathbb{R}^{d_2}$ denote two sets of mediators. Let $U = (E, X_1^\top, X_2^\top, Y)^\top \in \mathbb{R}^{d_1+d_2+2}$ collect all variables, and suppose it follows a Gaussian graphical model that

is associated with a directed acyclic graph,

$$(U - \mu) = W_0 (U - \mu) + \varepsilon. \tag{1}$$

Here $\mu = E(U)$, $W_0 \in \mathbb{R}^{(d_1+d_2+2)\times(d_1+d_2+2)}$ encodes the directional relations among the variables in U, and $\varepsilon \in \mathbb{R}^{d_1+d_2+2}$ is a vector of random errors. We assume the error variables in ε follow a normal distribution with mean zero and variance σ^2. The constant error variance is to ensure the identifiability of W_0 and to simplify the analysis. This condition is often imposed; see, e.g., Peters and Bühlmann (2014); Yuan, Shen, Pan, and Wang (2019). Meanwhile, it is also possible to relax this condition; see Shi and Li (2021) for more discussion.

For any two variables U_i, U_j in U, if $W_{0,j,i} \neq 0$, then an arrow is drawn from U_i to U_j, i.e., $U_i \rightarrow U_j$. In this case, U_i is called a parent of U_j. A directed path between U_i and U_j is a sequence of distinct nodes from U_i to U_j: $U_i \rightarrow U_{i_1} \rightarrow \cdots \rightarrow U_{i_k} \rightarrow U_j$ for some $\{U_l\}_{1 \leq l \leq k}$, and U_i is called an ancestor of U_j. Suppose the random variables in U comply with the sequential mediation framework as illustrated in Figure 1. That is, no potential mediator in X_1 and X_2 is a parent of E, no mediator in X_2 is a parent of any mediator in X_1, and the response Y is not a parent of any other variables. Under this framework, we can decompose W_0 accordingly as

$$W_0 = \begin{pmatrix} 0 & 0_{d_1}^\top & 0_{d_2}^\top & 0 \\ W_{0,1} & W_{1,1} & 0_{d_1 \times d_2} & 0_{d_1} \\ W_{0,2} & W_{1,2} & W_{2,2} & 0_{d_2} \\ W_{0,3} & W_{1,3}^\top & W_{2,3}^\top & 0 \end{pmatrix} \in \mathbb{R}^{(d_1+d_2+2)\times(d_1+d_2+2)}, \tag{2}$$

where $W_{0,1} \in \mathbb{R}^{d_1}$, $W_{0,2} \in \mathbb{R}^{d_2}$, $W_{0,3} \in \mathbb{R}$, $W_{1,1} \in \mathbb{R}^{d_1 \times d_1}$, $W_{1,2} \in \mathbb{R}^{d_2 \times d_1}$, $W_{1,3} \in \mathbb{R}^{d_1}$, $W_{2,2} \in \mathbb{R}^{d_2 \times d_2}$, and $W_{2,3} \in \mathbb{R}^{d_2}$.

Based on model (1), we aim at testing the following pair of hypotheses:

- $H_0(q_1, q_2)$: There does *not* exist a path from the exposure E to the outcome Y that passes through some mediator X_{1,q_1} in X_1 *and* some mediator X_{2,q_2} in X_2;

- $H_1(q_1, q_2)$: There exists a path from the exposure E to the outcome Y that passes through some mediator X_{1,q_1} in X_1 *and* some mediator X_{2,q_2} in X_2,

for some $q_1 = 1, \ldots, d_1$, and $q_2 = 1, \ldots, d_2$. In Figure 1, the null hypothesis means that, at least one potential pathway denoted by (ii), (iv) and (vi) is completely missing in this diagram. We remark that, this form of null hypothesis is motivated by our AD study, as the cascading pathway of Aβ, regional tau deposition, cortical thickness shrinking and cognition is of particular interest. Meanwhile, we can test other forms of null hypothesis; see our discussion in Section 7.

To test the above null hypothesis, we observe that it can be decomposed into the union of the following three null hypotheses:

- $H_0^*(0, q_1)$: There does *not* exist a directed path from the exposure E to X_{1,q_1};
- $H_0^*(q_1, q_2 + d_1)$: There does *not* exist a directed path from X_{1,q_1} to X_{2,q_2};
- $H_0^*(q_2 + d_1, d_1 + d_2 + 1)$: There does *not* exist a directed path from X_{2,q_2} to the outcome.

To further elaborate on the relation between $H_0(q_1, q_2)$ and $H_0^*(0, q_1)$, $H_0^*(q_1, q_2 + d_1)$, $H_0^*(q_2 + d_1, d_1 + d_2 + 1)$, we consider a directed path ζ: $E \rightarrow X_{1,i_1} \rightarrow \cdots \rightarrow X_{1,i_{m_1}} \rightarrow X_{1,q_1} \rightarrow X_{2,j_1} \rightarrow \cdots \rightarrow X_{j,m_2} \rightarrow X_{j,q_2} \rightarrow X_{2,k_1} \rightarrow \cdots \rightarrow X_{2,k_{m_3}} \rightarrow Y$. The total effect of E on Y attributed to this path can be written as

$$\omega_\zeta = \underbrace{W_{0,1,i_1} \left(\prod_{t=1}^{m_1-1} W_{1,1,i_{t+1},i_t} \right) W_{1,1,q_1,i_{m_1}}}_{\omega_\zeta^{(1)}} \underbrace{W_{1,1,j_1,q_1} \left(\prod_{t=1}^{m_2-1} W_{1,2,j_{t+1},j_t} \right) W_{1,2,q_2,j_{m_2}}}_{\omega_\zeta^{(2)}} \underbrace{W_{2,2,k_1,q_2} \left(\prod_{t=1}^{m_3-1} W_{2,2,k_{t+1},k_t} \right) W_{2,3,k_{m_3}}}_{\omega_\zeta^{(3)}}, \tag{3}$$

where $W_{0,1}$, $W_{1,1}$, $W_{1,2}$, $W_{2,2}$ and $W_{2,3}$ are the submatrices defined in (2). Under $H_0(q_1, q_2)$, we have $\omega_\zeta = 0$ for any ζ. By the decomposition in (3), this is equivalent to requiring $\omega_\zeta^{(1)}$, $\omega_\zeta^{(2)}$, or $\omega_\zeta^{(3)}$ to be zero for any ζ. Note that $\omega_\zeta^{(1)}$, $\omega_\zeta^{(2)}$, and $\omega_\zeta^{(3)}$ correspond to the total effects from E to X_{1,q_1}, from X_{1,q_1} to X_{2,q_2}, and from X_{2,q_2} to Y, which in turn are characterized by $H_0^*(0, q_1)$, $H_0^*(q_1, q_2 + d)$, and $H_0^*(q_2 + d_1, d_1 + d_2 + 1)$, respectively.

By the union-intersection principle, it suffices to derive the individual p-values, $p^*(0, q_1)$, $p^*(q_1, q_2 + d_1)$ and $p^*(q_2, d_1 + d_2 + 1)$, for $H_0^*(0, q_1)$, $H_0^*(q_1, q_2 + d)$, and $H_0^*(q_2 + d_1, d_1 + d_2 + 1)$, respectively, then compute the final p-value for $H_0(q_1, q_2)$ as

$$p(q_1, q_2) = \max \{ p^*(0, q_1), p^*(q_1, q_2 + d_1), p^*(q_2, d_1 + d_2 + 1) \}.$$

Next, we propose the tests to compute $p^*(0, q_1)$, $p^*(q_1, q_2 + d_1)$, and $p^*(q_2, d_1 + d_2 + 1)$, respectively.

3 | TESTING PROCEDURE

We first develop a testing procedure to test $H_0(q_1, q_2)$ through $H_0^*(0, q_1)$, $H_0^*(q_1, q_2 + d)$, and $H_0^*(q_2 + d_1, d_1 + d_2 + 1)$, for a given pair q_1 and q_2. We then augment it with a multiple testing procedure for all $q_1 = 1, \ldots, d_1$ and $q_2 = 1, \ldots, d_2$ with false discovery control.

3.1 | Testing procedure for a given pair of mediators

We begin with a summary of the testing procedure for a given pair q_1 and q_2 in Algorithm 1. We then detail the main steps of this algorithm.

Suppose we observe n i.i.d. copies of the data, $\{u_i\}_{1 \leq i \leq n}$, where $u_i = (e_i, x_{1,i}^\top, x_{2,i}^\top, y_i)^\top \in \mathbb{R}^{d_1 + d_2 + 2}$. We first split the data $\{1, 2, \cdots, n\}$ into two equal halves $\mathcal{I}_1$ and $\mathcal{I}_2$. The purpose of data splitting is to ensure a valid type-I error control under minimal conditions. This strategy has been commonly used in high-dimensional inferences in recent years (Chernozhukov et al. 2018; Romano & DiCiccio 2019). One issue with data splitting is the potential loss of power due to the usage of only a fraction of data. In our setting, we construct two test statistics based on both halves of data, then combine them to derive the final p-value. We show the test constructed this way achieves a good power both asymptotically and numerically.

3.2 | Initial estimation of weight matrices

Next, we estimate $W_{1,1}$, $W_{1,2}$ and $W_{2,2}$ in (2) based on each half $\mathcal{I}_1$ and $\mathcal{I}_2$ of the data.

Specifically, to estimate $W_{1,1}$, we first linearly regress each variable X_{1j} in X_1 on the exposure variable E, using the data in the subset $\mathcal{I}_s$, $s = 1, 2$. This is a simple regression with a single response and a single predictor, and it yields a residual estimate $\widehat{\varepsilon}_{1,1,i,j}^{(s)}$, $i \in \mathcal{I}_s$, $j = 1, \ldots, d_1$, $s = 1, 2$. We then apply a DAG estimation method to the residual to obtain an estimate of $W_{1,1}$. There are multiple choices for this purpose, e.g., Yuan et al. (2019); Zheng, Aragam, Ravikumar, and Xing (2018). In our implementation, we have chosen the method of Zheng et al. (2018). That is, we obtain

$$\widetilde{W}_{1,1}^{(s)} = \underset{W \in \mathbb{R}^{d_1 \times d_1}}{\arg\min} \sum_{i \in \mathcal{I}_s} \|\widehat{\varepsilon}_{1,1,i}^{(s)} - W\widehat{\varepsilon}_{1,1,i}^{(s)}\|_2^2 + \lambda_{1,1} |\mathcal{I}_s| \sum_{j,k=1}^{d_1} |W_{j,k}|, \quad \text{subject to} \quad \text{trace}\{\exp(W \circ W)\} = d_1, \tag{4}$$

where $\widehat{\varepsilon}_{1,1,i}^{(s)} = (\widehat{\varepsilon}_{1,1,1,i}^{(s)}, \ldots, \widehat{\varepsilon}_{1,1,d_1,i}^{(s)})^\top \in \mathbb{R}^{d_1}$, $\lambda_{1,1}$ is a penalty parameter, $|\mathcal{I}_s|$ is the cardinality of $\mathcal{I}_s$ and equals the number of data observations in split s, $\text{trace}(A)$ is the trace of a matrix A, $\exp(A)$ is the matrix exponential of A, and $\circ$ is the Hadamard product. We tune the sparsity parameter $\lambda_{1,1}$ by cross-validation. We also remark that, we only require the resulting DAG estimator to be estimation consistent, which is much weaker than requiring it to be selection consistent. The latter means all the zero entries of the DAG estimator have to match those of the true DAG. For the method of Zheng et al. (2018), Shi and Li (2021) established its estimation consistency.

To estimate $W_{2,2}$, the idea is similar. We first regress each variable X_{2j} in X_2 on both the exposure variable E and all the variables in X_1, again using the data in the subset $\mathcal{I}_s$, $s = 1, 2$. Note that X_1 can be high-dimensional. We thus employ a penalized linear regression approach, i.e.,

$$\min \sum_{i \in \mathcal{I}_s} \left\{ x_{2,i,j} - (e_i, x_{1,i}^\top)\beta_j^{(s)} - \beta_{0,j}^{(s)} \right\}^2 + \sum_{k=1}^{d_2+1} |\mathcal{I}_s| \, p_{\lambda_{2,1}}\left(|\beta_{j,k}^{(s)}|\right),$$

where $p_{\lambda_{2,1}}$ is some penalty function with the tuning parameter $\lambda_{2,1}$. In our implementation, we have chosen the MCP penalty function of Zhang (2010), and tune $\lambda_{2,1}$ by the Bayesian information criterion (BIC). This penalized regression in turn yields a residual estimate $\widehat{\varepsilon}_{i,j,2,2}^{(s)}$, $i \in \mathcal{I}_s$, $j = 1, \ldots, d_1$, $s = 1, 2$. We then apply the DAG estimation method of Zheng et al. (2018) again to $\widehat{\varepsilon}_{i,2,2}^{(s)} = (\widehat{\varepsilon}_{i,1,2,2}^{(s)}, \ldots, \widehat{\varepsilon}_{i,d_1,2,2}^{(s)})^\top \in \mathbb{R}^{d_2}$, and obtain

$$\widetilde{W}_{2,2}^{(s)} = \underset{W \in \mathbb{R}^{d_2 \times d_2}}{\arg\min} \sum_{i \in \mathcal{I}_s} \|\widehat{\varepsilon}_{2,2,i}^{(s)} - W\widehat{\varepsilon}_{2,2,i}^{(s)}\|_2^2 + \lambda_{2,2} |\mathcal{I}_s| \sum_{j,k=1}^{d_2} |W_{j,k}|, \quad \text{subject to} \quad \text{trace}\{\exp(W \circ W)\} = d_2. \tag{5}$$

To estimate $W_{1,2}$, we first compute the residual $\widehat{\varepsilon}_{1,2,i}^{(s)} = x_{2,i} - \widetilde{W}_{2,2}^{(s)}x_{2,i} \in \mathbb{R}^{d_2}$, $i \in \mathcal{I}_s$, where $\widetilde{W}_{2,2}^{(s)}$ is obtained from (5). We then regress each variable in this residual on both the exposure variable E and all the variables in X_1, using the data in the subset $\mathcal{I}_s$, $s = 1, 2$. Again, considering the high-dimensionality of X_1, we employ a penalized linear regression approach like MCP (Zhang 2010). That is, we estimate the jth row of $W_{1,2}$ as,

$$\left(\widetilde{W}_{1,2,j}^{(s)}, w_{0,j}^{(s)}\right) = \underset{w_j^{(s)}, w_{0,j}}{\arg\min} \sum_{i \in \mathcal{I}_s} \left\{ \widehat{\varepsilon}_{1,2,i,j}^{(s)} - (e_i, x_{1,i})^\top w_j^{(s)} - w_{0,j} \right\}^2 + \sum_{k=1}^{d_2+1} |\mathcal{I}_s| \, p_{\lambda_{1,2}}\left(|w_{j,k}^{(s)}|\right), \tag{6}$$

for $j = 1, \ldots, d_2$, and $p_{\lambda_{1,2}}$ is some penalty function with the tuning parameter $\lambda_{1,2}$.

Algorithm 1 Testing procedure for a single path $H_0(q_1, q_2)$.

1: Randomly split the data $\{1, 2, \cdots, n\}$ into two halves $\mathcal{I}_1$ and $\mathcal{I}_2$.

2: Estimate $W_{1,1}$, $W_{2,2}$ and $W_{1,2}$ using (4), (5) and (6), respectively. Putting those estimators together to obtain $\widetilde{W}_{1:2}^{(s)}$, $s = 1, 2$.

3: Compute the refined estimator $\widetilde{W}^{(s)}$ for W_0 based on $\widetilde{W}_{1:2}^{(s)}$, $s = 1, 2$, then the binary matrix $\widehat{B}^{(s)}$ based on $\widetilde{W}^{(s)}$, $s = 1, 2$.

4: Compute the debiased estimator $\widehat{W}^{(s)}$ for W_0 based on $\widetilde{W}^{(s)}$ by cross-fitting, $s = 1, 2$.

5: Compute the test statistic and the associate p-value for $H_0^*(q_1, q_2)$ using bootstrap as in (10). Obtain the p-value for $H_0(q_1, q_2)$, $\widehat{p}^{(s)}(q_1, q_2)$, $s = 1, 2$, using the union-intersection principle as in (11). Reject $H_0(q_1, q_2)$ if either $\widehat{p}^{(1)}(q_1, q_2) \leq \alpha/2$, or $\widehat{p}^{(2)}(q_1, q_2) \leq \alpha/2$.

Putting these estimators together, we obtain the matrix,

$$\widetilde{W}_{1:2}^{(s)} = \begin{pmatrix} \widetilde{W}_{1,1}^{(s)} & 0_{d_1 \times d_2} \\ \widetilde{W}_{1,2}^{(s)} & \widetilde{W}_{2,2}^{(s)} \end{pmatrix} \in \mathbb{R}^{(d_1+d_2) \times (d_1+d_2)}.$$

3.3 | Refined and debiased estimation

Next, we discuss how to obtain an estimator of $W_0 \in \mathbb{R}^{(d_1+d_2+2) \times (d_1+d_2+2)}$ based on $\widetilde{W}_{1:2}^{(s)}$ from the previous step, and how to further debias this estimator.

Specifically, we begin with setting the first row of $\widetilde{W}^{(s)}$ to be a zero vector. We then obtain the $(j+1)$th row of $\widetilde{W}^{(s)}$, for $j = 1, \ldots, d_1 + d_2$, by fitting a penalized linear regression using the data in the subset $\mathcal{I}_s$, $s = 1, 2$. The response of this regression is $X_{1,j}$ when $j = 1, \ldots, d_1$, and $X_{2,j-d_1}$ when $j = d_1+1, \ldots, d_1+d_2$, the predictors come from the jth row of $\widetilde{W}_{1:2,j,k}^{(s)}$, i.e., $\widetilde{W}_{1:2,j,1}^{(s)}, \ldots, \widetilde{W}_{1:2,j,d_1+d_2}^{(s)}$, and the estimated regression coefficient is the $(j+1)$th row of $\widetilde{W}^{(s)}$. We end this step by obtaining the last row of $\widetilde{W}^{(s)}$, from the regression coefficient of fitting another penalized linear regression with Y as the response, and E, X_1, X_2 as the predictors. This step gives an estimator of the entire $(d_1 + d_2 + 2) \times (d_1 + d_2 + 2)$ matrix of W_0, and also helps improve the estimation accuracy of the initial estimator of $\widetilde{W}_{1:2}^{(s)}$.

We then compute a binary matrix $\widehat{B}^{(s)}$ of the same dimension as $\widetilde{W}^{(s)}$, and set each entry of $\widehat{B}^{(s)}$ as one if and only the corresponding entry in $\widetilde{W}^{(s)}$ is nonzero. This step serves for the screening purpose, as it helps reduce both the number of potential paths and the variance of the subsequent test statistics, which in turn helps improve the power of our test.

Given the estimator $\widetilde{W}^{(s)}$, we further employ the cross-fitting strategy to compute a debiased estimator $\widehat{W}^{(s)}$ for W_0. We first estimate the set of ancestors of the jth variable in U, $j = 1, \ldots, d_1 + d_2 + 2$. Toward that end, let B_0 denote the binary version of W_0. Let $\oplus$ denote the Boolean matrix addition operator, such that the (j_1, j_2)th entry of $A_1 \oplus A_2$ equals $\max(A_{1,j_1,j_2}, A_{2,j_1,j_2})$, and $\otimes$ denote the Boolean matrix multiplication operator such that the (j_1, j_2)th entry of $A_1 \otimes A_2$ equals $\max_k \min(A_{1,j_1,k}, A_{2,k,j_2})$. Define $A^{(k)}$ as the Boolean matrix power k of the square matrix A in a recursive fashion, such that $A^{(k)} = A^{(k-1)} \otimes A$. Define $B_{0,*} = B_0 \oplus B_0^{(2)} \oplus \ldots \oplus B_0^{(d_1+d_2+2)}$. Then, the key observation is that, finding the set of ancestors is equivalent to finding the transitive closure of the Boolean matrix B_0 (Fischer & Meyer 1971). That is, the set of ancestors of the jth variable is,

$$\mathrm{ACT}(j) = \left\{ 1 \le k \le d_1 + d_2 + 2 : B_{0,*,k,j} \neq 0 \right\}, \tag{7}$$

We note that, since B_0 is a binary matrix, the maximum and minimum operators are equivalent to the logic operators "or" and "and" in Boolean algebra. This motivates us to estimate $\mathrm{ACT}(j)$ by

$$\mathrm{ACT}\left(j, \widehat{B}^{(s)}\right) = \left\{ 1 \le k \le d_1 + d_2 + 2 : \widehat{B}_{*,k,j}^{(s)} \neq 0 \right\}. \tag{8}$$

We also note that, we always include the exposure E in the set of ancestors, and include all the potential mediators when estimating the ancestors of the outcome Y.

We next debias $\widetilde{W}_{j_1,j_2}^{(s)}$ using the other half of the data in $\mathcal{I}_s^c$, for any entry that $\widehat{B}_{j_1,j_2}^{(s)} \neq 0$,

$$\widehat{W}_{j_1,j_2}^{(s)} = \frac{\sum_{i \in \mathcal{I}_s^c} \left(\widetilde{u}_{i,j_2} - \widehat{\beta}_{j_1,j_2}^{(s)\top} \widetilde{u}_i \right) \left(\widetilde{u}_{i,j_1} - \sum_{j \neq j_2} \widetilde{u}_{i,j} \widetilde{W}_{j_1,j}^{(s)} \right)}{\sum_{i \in \mathcal{I}_s^c} \widetilde{u}_{i,j_2} \left(\widetilde{u}_{i,j_2} - \widehat{\beta}_{j_1,j_2}^{(s)\top} \widetilde{u}_i \right)}, \tag{9}$$

where $\widetilde{u}_{i,j} = u_{i,j} - \bar{u}_j$, $\bar{u}_j$ is the sample average, and $\widetilde{u}_i = (u_{i,1}, \ldots, u_{i,d_1+d+2})^\top \in \mathbb{R}^{d_1+d+2}$, $\widehat{\beta}_{j_1,j_2}^{(s)}$ is the coefficient vector with $\{\widetilde{u}_{i,j_2}\}_{i \in \mathcal{I}_s^c}$ as the "response", and $\{\widetilde{u}_i\}_{i \in \mathcal{I}_s}$ as the "predictors", and the support of β is constrained to the estimated set of ancestors as in (8). See Section 2 of the supplementary material for the detailed definition. We remark that, we have used the cross-fitting strategy in the estimation of $\widehat{W}^{(s)}$, since it is based on the complement data set $\mathcal{I}_s^c$, instead of the data set $\mathcal{I}_s$ that is used to construct $\widetilde{W}^{(s)}$. This strategy guarantees that each entry of the debiased estimator $\widehat{W}^{(s)}$ is asymptotically normal, regardless of whether the estimator $\widetilde{W}^{(s)}$ is selection consistent or not.

3.4 | Test statistic and p-value

We next compute the test statistic and the corresponding critical value for $H_0^*(q_1, q_2)$.

Similar to (7), we can show that $H_0^*(q_1, q_2)$ holds, if and only if the $(q_2 + 1, q_1 + 1)$th entry of $W_{0,*} = |W_0| \oplus |W_0|^{(2)} \oplus \ldots \oplus |W_0|^{(d_1+d_2+2)}$ equals zero, where $|A|$ denotes the matrix of the same dimension as A whose (j_1, j_2)th entry is $|A_{j_1,j_2}|$. Note that $W_{0,*}$ is a real-valued matrix. A larger value of $W_{0,*,q_1,q_2}$ implies a stronger evidence against the null hypothesis. This motivates us to define our test statistic by the $(q_2 + 1, q_1 + 1)$th entry of

the debiased version of the matrix, i.e.,

$$\sqrt{|\mathcal{I}_{\bar{s}}|}\widehat{W}_*^{(s)} = \sqrt{|\mathcal{I}_{\bar{s}}^c|}\left\{|\widehat{W}^{(s)}| \oplus |\widehat{W}^{(s)}|^{(2)} \oplus \ldots \oplus |\widehat{W}^{(s)}|^{(d_1+d_2+2)}\right\}.$$

The purpose of using the Boolean algebra here is to ensure the resulting test statistic has a tractable limiting distribution. More specifically, under $H_0^*(q_1, q_2)$, $\widehat{W}_*^{(s)}$ is stochastically dominated by the maximum of Gaussian random variables whose distribution can be well approximated via bootstrap; see the proof of Theorem 1 for more details. By contrast, the asymptotic distribution of the usual power of the matrix $|\widehat{W}^{(s)}|$ is extremely challenging to derive (Shi & Li 2021). The p-value associated with this test statistic is then given by,

$$\widehat{p}_*^{(s)}(q_1, q_2) = \frac{1}{R}\sum_{b=1}^{R}\mathbb{I}\left\{T^{(s,b)}(q_1, q_2) \geq \sqrt{\mathcal{I}_{\bar{s}}^c}|\widehat{W}_{*,q_2+1,q_1+1}^{(s)}|\right\}, \tag{10}$$

where R is the number of bootstrap replicates, $T^{(s,b)}(q_1, q_2)$ is a bootstrap sample. The explicit form of $T^{(s,b)}(q_1, q_2)$ is very similar to that the bootstrap sample outlined in Shi and Li (2021). As such, we present the detailed definition in Section 2 of the supplementary material to save space. The validity of multiplier bootstrap is ensured by the asymptotic normality of $\widehat{W}_{j_1,j_2}^{(s)}$.

We then compute the p-value, $\widehat{p}_*(q_1, q_2) = 2\min\{\widehat{p}_*^{(1)}(q_1, q_2), \widehat{p}_*^{(2)}(q_1, q_2)\}$ for $H_0^*(q_1, q_2)$. By the union-intersection principle, we obtain the p-value for the null $H_0(q_1, q_2)$,

$$\widehat{p}(q_1, q_2) = \max\left\{\widehat{p}_*(0, q_1), \widehat{p}_*(q_1, q_2 + d_1), \widehat{p}_*(q_2 + d_1, d_1 + d_2 + 1)\right\}. \tag{11}$$

For a given significance level α, we reject $H_0(q_1, q_2)$ if $\widehat{p}(q_1, q_2) \leq \alpha$.

3.5 | Multiple testing procedure

Finally, we present a multiple testing procedure for simultaneous inference of all pairs of $q_1 = 1, \ldots, d_1$ and $q_2 = 1, \ldots, d_2$. The objective is to identify all significant pairs (q_1, q_2) such that the null $H_0(q_1, q_2)$ is rejected, meanwhile to control the false discovery properly. We adopt and extend the ScreenMin procedure of Djordjilovic et al. (2019) to our setting.

Given the individual p-values $\widehat{p}^{(s)}(q_1, q_2)$, $q_1 = 1, \ldots, d_1$, $q_2 = 1, \ldots, d_2$, we begin by computing the minimum p-values,

$$\widehat{p}_{\min}^{(s)}(q_1, q_2) = \min\left\{\widehat{p}^{(s)}(0, q_1), \widehat{p}^{(s)}(q_1, q_2 + d_1), \widehat{p}^{(s)}(q_2 + d_1, d_1 + d_2 + 1)\right\}.$$

We then screen and select those mediators whose corresponding $\widehat{p}_{\min}^{(s)}(q_1, q_2)$ is smaller than a threshold $c^{(s)}$, with $c^{(s)}$ determined adaptively as $c^{(s)} = \max\left[c \in \{\alpha/(d_1 d_2), \ldots, \alpha/2, \alpha\} : c|\mathcal{H}_0^{(s)}(c)| \leq \alpha\right]$, and $\mathcal{H}_0^{(s)}(c)$ denotes the set of prescreened mediators when the threshold value is c. We next order the mediators in $\mathcal{H}_0^{(s)}$ according to $\widehat{p}^{(s)}(q_1, q_2)$ obtained from (11). We then apply the Benjamini-Hochberg procedure to the ordered mediators, and select $h^{(s)}$ pairs with the smallest p-values, where $h^{(s)} = \max\left\{i : \widehat{p}_{(i)}^{(s)} \leq (i\alpha)/\{2|\mathcal{H}_0^{(s)}|\sum_{j=1}^{|\mathcal{H}_0^{(s)}|}j^{-1}\}\right\}$, and $\widehat{p}_{(i)}^{(s)}$ denotes the ith minimum p-value of the set $\left\{\widehat{p}_{\max}^{(s)}(q_1, q_2) : 1 \leq p_1 \leq d_1, 1 \leq p_2 \leq d_2\right\}$. Letting $\mathcal{H}^{(s)}$ denote the selected important pairs for each half of the data, $s = 1, 2$, respectively. We set the final set of selected pairs as $\mathcal{H} = \mathcal{H}^{(1)} \cup \mathcal{H}^{(2)}$.

4 | THEORETICAL GUARANTEES

We first establish the consistency of our test by deriving the asymptotic size and power. We then show that our multiple testing procedure achieves a valid FDR control.

We begin with two regularity conditions. Define the limit of the estimator $\widehat{\beta}_{j_1,j_2}^{(s)}$ as

$$\beta_{0,j_1,j_2}^{(s)} = \underset{\beta:\beta_{j_2}=0,\text{supp}(\beta)\in\text{ACT}_{(j_1,\widetilde{w}^{(s)})}}{\arg\min} E(U_{j_2} - \beta^\top U)^2.$$

Let $\widetilde{W}_j^{(s)}$ and $W_{0,j}$ denote the jth row of $\widetilde{W}^{(s)}$ and W_0, respectively. For any directed path $\zeta : E = U_1 \rightarrow U_{i_1} \rightarrow \cdots \rightarrow U_{i_k} \rightarrow U_{d_1+d_2+2} = Y$, define the signal strength along this path by $\omega_\zeta^* = \min\left\{|W_{0,i_1,1}|, \min_{j\in\{1,\cdots,k-1\}}|W_{0,i_{j+1},i_j}|, |W_{0,d_1+d_2+2,i_k}|\right\}$. Under $H_0(q_1, q_2)$, we have $\omega_\zeta^* = 0$ for any path that passes through X_{1,q_1} and X_{2,q_2}. Under the $H_1(q_1, q_2)$, there exists at least one such path that $\omega_\zeta^* > 0$.

(C1) With probability approaching one, $\text{ACT}\left(j, \widetilde{W}^{(s)}\right)$ as defined in (8) contains all parents of j, for any $j = 1, \ldots, d_1 + d_2 + 2$ and $s = 1, 2$.

(C2) With probability approaching one, there exist some constants $\kappa_1, \kappa_2, \kappa_3, \kappa_4 > 0$, such that $\kappa_1 + \kappa_2 > 1/2$, $\left\|\widehat{\beta}_{j_1,j_2}^{(s)} - \beta_{0,j_1,j_2}^{(s)}\right\|_2 = O(n^{-\kappa_1})$, and $\left\|\widetilde{W}_j^{(s)} - W_{0,j}\right\|_2 = O(n^{-\kappa_2})$.

Conditions (C1) and (C2) are both mild. Condition (C1) is weaker than the sure screening condition that requires $\widetilde{W}^{(s)}$ to be selection consistent. It holds when some omega-min condition is satisfied (van de Geer & Bühlmann 2013). Condition (C2) essentially requires the oracle parameters $W_{0,j}$

and $\beta_{0,j_1,j_2}^{(s)}$ to satisfy certain sparsity constraints. The convergence rates of $\widetilde{W}_j^{(s)}$ and $\widehat{\beta}_{j_1,j_2}^{(s)}$ are satisfied for most of the commonly used penalized regression methods, including Lasso, SCAD, MCP, or Dantzig selector (Bickel, Ritov, & Tsybakov 2009; Fan & Lv 2011; Ning & Liu 2017).

We next derive the asymptotic size and power of the proposed testing procedure for $H_0(q_1, q_2)$ for a given pair of q_1 and q_2.

Theorem 1 (Asymptotic size). Suppose conditions (C1) and (C2) hold. Suppose $d_1 + d_2 = O(n^{\kappa_3})$ for some $\kappa_3 \geq 0$, and $\|W_0\|_2$ is bounded. Then, for a given significance level $0 < \alpha < 1$, and any pair (q_1, q_2) of the mediators, for $q_1 = 1, \ldots, d_1$ and $q_2 = 1, \ldots, d_2$,

$$\Pr\{H_0(q_1, q_2) \text{ is rejected} \mid H_0(q_1, q_2) \text{ holds}\} \leq \alpha + o(1).$$

Theorem 2 (Asymptotic power). Suppose the conditions in Theorem 1 hold. Suppose there exists some path ζ from E to Y that passes through X_{1,q_1} and X_{2,q_2}, such that $\omega_\zeta^* \gg \max\left(n^{-\kappa_2}, n^{-1/2}\sqrt{\log n}\right)$. Then,

$$\Pr\{H_0(q_1, q_2) \text{ is rejected} \mid H_0(q_1, q_2) \text{ holds}\} \rightarrow 1.$$

Finally, we show that our proposed multiple testing procedure asymptotically controls the false discovery rate.

Theorem 3 (Asymptotic FDR control). Suppose the conditions in Theorem 1 hold. Then the selected set of mediators $\mathcal{H}$ satisfies that $FDR(\mathcal{H}) \leq \alpha + o(1)$.

5 | SIMULATIONS

We first describe the simulation setup. We next examine the empirical size and power of testing a pair of mediators, then the empirical FDR control of multiple testing.

5.1 | Simulation setup

We simulate the data following models (1) and (2). Specifically, we set $\boldsymbol{\mu}$ to a vector of ones and the variance of the random errors $\sigma^2 = 1$. We generate the adjacency matrix W_0 as follows: We begin with a zero matrix, then replace every entry W_{0,j_1,j_2} in the lower off-diagonals by the product of two random variables $R_{j_1,j_2}^{(1)} R_{j_1,j_2}^{(2)}$, where $R_{j_1,j_2}^{(1)} \sim \text{Bernoulli}(p_1)$, if $j_2 = 0$, or $j_1 = d+1$, and $R_{j_1,j_2}^{(1)} \sim \text{Bernoulli}(p_2)$, otherwise, and $R_{j_1,j_2}^{(2)}$ is uniformly distributed on $[-1.4, -0.6] \cup [0.6, 1.4]$. Here p_1 and p_2 control the number of significant mediators, and all variables are generated independently. We consider two settings of the number of mediators, $d_1 = d_2 = 35$, and $d_1 = d_2 = 70$. For the first setting, we set $p_1 = 0.05, p_2 = 0.1$, and the sample size $n = 200, 400$, and for the second setting, we set $p_1 = 0.025, p_2 = 0.05$, and $n = 300, 500$. Figure 2 shows one instance of the generated W_0 for the two settings. In the first setting, 5 pairs of mediators have nonzero sequential mediation effects, whereas in the second setting, 11 pairs are significant.

5.2 | Empirical size and power

We first evaluate the empirical performance of our test for a single pair of mediators (q_1, q_2). We compare it with a sequential test modified from the interventional calculus-based method that was proposed by Chakrabortty et al. (2018). To adopt Chakrabortty et al. (2018) to the sequential mediation setting, we note that, by the union-intersection principle, it suffices to test $H_0^*(0, q_1)$, $H_0^*(q_1, q_2 + d_1)$ and $H_0^*(q_2 + d_1, d_1 + d_2 + 1)$. We then test each of these two hypotheses using Chakrabortty et al. (2018). On the other hand, we remark that the interventional calculus-based test does not directly target the null $H_0^*(q_1, q_2)$. Instead, it constructs a confidence interval for the aggregated causal effects along all paths from q_1 to q_2, and reject the null if zero is not covered by the confidence interval.

We evaluate the performance by the empirical rejection rate, i.e., the percentage of times the test rejects the null hypothesis at the significance level $\alpha = 5\%$ out of 500 data replications. This criterion reflects the empirical size of the test when the null hypothesis holds, and the empirical power otherwise. We also compute the average receiver operating characteristic (ROC) curves, aggregated over 500 replications, when we vary the significance level α.

Figure 3 and Figure 4 report the results for the setting of $d_1 = d_2 = 35$, and $d_1 = d_2 = 70$, respectively. We make a few observations. First, our test achieves a valid size under the null hypothesis. In the first setting, the empirical rejection rate is well below the nominal level for all cases. In the second setting, our test has a few inflated type-I errors when the sample size n is small, but the results improve with a larger sample size. Second, our test consistently achieves a larger empirical power over the method of Chakrabortty et al. (2018), which fails to identify two significant pairs of mediators in the first setting, and five significant pairs in the second setting. This decreased power may be due to that the effects calculated by Chakrabortty et al. (2018) along different paths may have different signs, and thus may cancel each other. Consequently, it may result in a zero sum,

even though there are significant positive and negative mediation effects along the paths. Finally, we see that the ROC curve of our test consistently lies above that of Chakrabortty et al. (2018) in all settings as α varies, demonstrating a competitive performance of our test.

5.3 | Multiple testing FDR control

We next evaluate the empirical performance of our multiple testing procedure. We compare it with the standard Benjamini-Yekutieli (BY) procedure. For the latter, instead of applying ScreenMin to determine the set $\mathcal{H}_0^{(s)}$, one simply sets $\mathcal{H}_0^{(s)} = \{(q_1, q_2) : 1 \leq q_1 \leq d_1, 1 \leq q_2 \leq d_2\}$, i.e., the set of all pairs of mediators. We evaluate the performance by the false discovery rate (FDR) and the true positive rate (TPR) over 500 data replications.

Figure 5 reports the results under the varying significance level α from 0.1 to 0.4. It is seen that both methods achieve a valid false discovery control, in that the FDRs are very close to zero. However, our method is more powerful than the BY procedure, as reflected by a considerably larger TPR in all cases. Actually, when $d_1 = d_2 = 35$ and $n = 200$, the BY procedure fails to identify any significant sequential mediators.

6 | MULTIMODAL AD PATHWAY ANALYSIS

We revisit the motivating example of multimodal AD pathway analysis. Amyloid-beta plaques and neurofibrillary tangles are two key hallmarks of AD, and appear 20 years or more before the presence of manifest clinical symptoms (Jack et al. 2013). Aβ and tau aggregation is evaluated using PET imaging, and the downstream consequence of neurodegeneration is examined with structural MRI. Recent studies generally support a unidirectional model of AD pathogenesis in which Aβ appears early, followed by deposition of abnormal tau aggregates, and eventually subsequent neurodegeneration and cognitive decline (Jack et al. 2010 2019) While it is generally conceived that the association between Aβ deposition and cognition may be mediated by tau deposition and cortical thickness, it remains unclear in which brain regions tau deposition and cortical thickness mediate the association of Aβ deposition and subsequent cognitive decline. In this analysis, we aim to determine the brain regions of tau aggregation and cortical thickness shrinking that are involved in subsequent cognitive decline using a dataset from the Berkeley Aging Cohort Study.

Each participant received concurrent Aβ PET imaging using the ^{18}F-florbetapir (FBP) or ^{18}F-florbetaben (FBB) tracers, tau PET imaging using the ^{18}F-flortaucipir (FTP) tracer, and 3T structural MRI, all acquired within a year. The PET data was acquired in 5-min frames from 50-70min (FBP), 90-110min (FBB), and 75-105 min (FTP) post-injection, averaged over time, and processed with Freesurfer v5.3.0. All fully preprocessed PET scans were coregistered to the structural MRI scan that was closest in time to the baseline PET. Regions of interest (ROIs) were defined on each structural MRI scan using Freesurfer, and were used to extract regional florbetapir, florbetaben, and flortaucipir uptake from the co-registered PET images. FBP or FBB standardized uptake value ratios (SUVRs) were calculated by referring regional florbetapir or florbetaben to that in the whole cerebellum. SUVR in a composit cortical area made up of frontal, cingulate, parietal and temporal regions was used to represent total Aβ deposition. This is to serve as the exposure variable E, which takes a continuous value, in our sequential mediation analysis. FTP SUVRs in 34 Freesurfer-defined cortical ROIs and amygdala were calculated by using inferior cerebellar gray matter intensity normalization (Maass et al. 2017). These regional tau deposition measures serve as the first modality of mediators, with $d_1 = 35$, in our analysis. All structural MRI scans were processed with the Freesurfer cross-sectional pipeline to derive ROIs in each subject's native space using the Desikan-Killiany atlas (Desikan et al. 2006). Cortical thickness in 34 Freesurfer-defined cortical ROIs were calculated from the Freesurfer output. These regional cortical thickness measures serve as the second modality of mediators, with $d_2 = 34$.

Each participant also received a PACC composite testing score at each of two consecutive visits with a median of 1.1 years in between. The PACC score combines tests that assess episodic memory, timed executive function, and global cognition. This score has been well established as showing sensitivity to decline in prodromal and mild dementia, and with sufficient range to detect early decline in the preclinical stages of AD (Donohue, Sperling, & Others 2014). Following Guo et al. (2020), we compute the change of the PACC score over the two time points using a linear mixed effect model including time, age at baseline scan, sex and education as covariates, and a random slope and intercept for each participant. We treat this change of the PACC score as the outcome variable Y. The dataset we analyze consists of $n = 182$ subjects who were classified as the Aβ positive group, who are biologically regarded as individuals on the AD continuum following the research framework of the National Institute on Aging and Alzheimer's Association (Jack et al. 2018).

We apply the proposed test to this data, with the false discovery level controlled at $\alpha = 0.1$. We regress out age, sex and education from both the exposure E and the potential mediators X_1 and X_2. Figure 6 shows the identified pathways, as well as the identified regions overlaid on a template brain. It is seen that, there is a pathway from Aβ to tau deposition in the inferior parietal region, then to cortical thickness in the entorhinal, inferior temporal, precuneus, inferior parietal, and superior parietal regions, then to the cognitive change as measured by the PACC score. A few other significant pathways include banks of the superior temporal sulcus tau to entorhinal cortical thickness, rostral middle frontal tau to lateral occipital cortical thickness, and superior frontal tau to transverse temporal cortical thickness. These findings are interesting and are consistent with the

literature. Particularly, the inferior parietal region is very likely the most typical tau deposition for individuals on the AD continuum, because this region captures the most of the regions of downstream neurodegeneration that leads to subsequent cognitive decline. Consistent with our findings, Bischof et al. (2016) found AD patients had higher tau deposition in inferior parietal and more hypometabolism in inferior parietal, inferior temporal and superior parietal regions than cognitively healthy elderly adults. Das et al. (2018) found high inferior parietal tau deposition was related to reduced cortical thickness in entorhinal, inferior temporal, inferior parietal Aβ positive individuals, which was partially in line with our findings. Cho et al. (2016) found higher tau deposition in inferior parietal, middle temporal, inferior temporal, and more cortical thinning in entorhinal, inferior parietal, inferior temporal, precuneus, superior parietal, middle temporal and superior temporal in late tau stages (V- VI) but not in earlier tau stages (I- IV) when compared to stage 0 individuals without tau deposition according to tau images, supporting the spatial patterns of tau and cortical thickness identified on the AD continuum in our study. Similarly, Harrison et al. (2019) also found high tau deposition in inferior parietal, and significant longitudinal gray matter loss in inferior parietal, superior parietal, precuneus and inferior temporal in AD patients.

7 | DISCUSSION

In Section 2, we have primarily focused on testing the potential pathway passing through (ii), (iv) and (vi) in Figure 1, as motivated by our AD study. Meanwhile, the proposed inferential framework can handle other forms of hypotheses as well. We discuss them in Section 1 of the supplementary material.

ACKNOWLEDGEMENTS

References

Alzheimer's Association. (2020). 2020 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *16*(3), 391-460.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bickel, P. J., Ritov, Y., & Tsybakov, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *The Annals of Statistics*, *37*(4), 1705–1732.

Bischof, G. N., Jessen, F., Fliessbach, K., Dronse, J., Hammes, J., Neumaier, B., … the Alzheimer's Disease Neuroimaging Initiative (2016). Impact of tau and amyloid burden on glucose metabolism in alzheimer's disease. *Annals of Clinical and Translational Neurology*, *3*(12), 934–939.

Cai, Q., Wang, H., Li, Z., & Liu, X. (2019). A survey on multimodal data-driven smart healthcare systems: Approaches and applications. *IEEE Access*, *7*, 133583-133599.

Chakrabortty, A., Nandy, P., & Li, H. (2018). Inference for individual mediation effects and interventional effects in sparse high-dimensional causal graphical models. *arXiv preprint arXiv:1809.10652*.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning. *The Econometrics Journal*, *21*, C1-C68.

Cho, H., Choi, J. Y., Hwang, M. S., Kim, Y. J., Lee, H. M., Lee, H. S., … Lyoo, C. H. (2016). In vivo cortical spreading pattern of tau and amyloid in the alzheimer disease spectrum. *Annals of Neurology*, *80*(2), 247–258.

Das, S. R., Xie, L., Wisse, L. E., Ittyerah, R., Tustison, N. J., Dickerson, B. C., … Wolk, D. A. (2018). Longitudinal and cross-sectional structural magnetic resonance imaging correlates of av-1451 uptake. *Neurobiology of Aging*, *66*, 49–58.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968 - 980.

Djordjilović, V., Page, C. M., Gran, J. M., Nø st, T. H., Sandanger, T. M., Veierø d, M. B., & Thoresen, M. (2019). Global test for high-dimensional mediation: testing groups of potential mediators. *Statistics in Medicine*, *38*(18), 3346–3360.

Donohue, M. C., Sperling, R. A., & Others. (2014). The Preclinical Alzheimer Cognitive Composite: Measuring Amyloid-Related Decline. *Journal of the American Medical Association: Neurology*, *71*(8), 961-970.

Fan, J., & Lv, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Transactions on Information Theory*, *57*(8), 5467–5484.

Fischer, M. J., & Meyer, A. R. (1971). Boolean matrix multiplication and transitive closure. In *12th annual symposium on switching and automata theory (swat 1971)* (pp. 129–131).

Gordon, B., Blazey, T., Su, Y., Hari-Raj, A., Dincer, A., Flores, S., … Benzinger, T. (2018). Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant alzheimer's disease: a longitudinal study. *The Lancet Neurology*, *17*(3), 241–250.

Guo, T., Korman, D., Baker, S. L., Landau, S. M., & Jagust, W. J. (2020). Longitudinal cognitive and biomarker measurements support a unidirectional pathway in alzheimerâĂŹs disease pathophysiology. *Biological Psychiatry*.

Harrison, T. M., La Joie, R., Maass, A., Baker, S. L., Swinnerton, K., Fenton, L., … Jagust, W. J. (2019). Longitudinal tau accumulation and atrophy in aging and alzheimer disease. *Annals of Neurology*, *85*(2), 229–240.

Huang, Y.-T., & Pan, W.-C. (2016). Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics*, *72*(2), 401–413.

Jack, Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., … Silverberg, N. (2018). Nia-aa research framework: Toward a biological definition of alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535–562.

Jack, Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., … Trojanowski, J. Q. (2013). Tracking pathophysiological processes in alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet. Neurology*, *12*(2), 207–216.

Jack, Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., … Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, *9*(1), 119 – 128.

Jack, Wiste, H. J., Therneau, T. M., Weigand, S. D., Knopman, D. S., Mielke, M. M., … Petersen, R. C. (2019). Associations of Amyloid, Tau, and Neurodegeneration Biomarker Profiles With Rates of Memory Decline Among Individuals Without Dementia. *JAMA*, *321*(23), 2316–2325.

Lai, E.-Y., Shih, S., Huang, Y.-T., & Wang, S. (2020). A mediation analysis for a nonrare dichotomous outcome with sequentially ordered multiple mediators. *Statistics in Medicine*, *39*(10), 1415–1428.

Maass, A., Landau, S., Baker, S. L., Horng, A., Lockhart, S. N., La Joie, R., … Jagust, W. J. (2017). Comparison of multiple tau-pet measures as biomarkers in aging and alzheimer's disease. *NeuroImage*, *157*, 448–463.

Ning, Y., & Liu, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics*, *45*(1), 158–195.

Peters, J., & Bühlmann, P. (2014). Identifiability of Gaussian structural equation models with equal error variances. *Biometrika*, *101*(1), 219–228.

Richardson, S., Tseng, G. C., & Sun, W. (2016). Statistical methods in integrative genomics. *Annual review of statistics and its application*, *3*, 181–209.

Romano, J., & DiCiccio, C. (2019). *Multiple data splitting for testing* (Tech. Rep.). Technical report.

Shi, C., & Li, L. (2021). Testing mediation effects using logic of boolean matrices. *Journal of the American Statistical Association, accepted*.

van de Geer, S., & Bühlmann, P. (2013). ℓ_0-penalized maximum likelihood for sparse directed acyclic graphs. *The Annals of Statistics*, *41*(2), 536–567.

VanderWeele, T. J. (2016). Mediation analysis: A practitioner's guide. *Annual Review of Public Health*, *37*(1), 17–32.

Yuan, Y., Shen, X., Pan, W., & Wang, Z. (2019). Constrained likelihood for reconstructing a directed acyclic Gaussian graph. *Biometrika*, *106*(1), 109–125.

Zhang. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, *38*(2), 894–942.

Zhang, Zheng, Y., Zhang, Z., Gao, T., Joyce, B., Yoon, G., … others (2016). Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics*, *32*(20), 3150–3154.

Zhang, C.-H., & Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society. Series B.*, *76*(1), 217–242.

Zhao, Y., Li, L., & Caffo, B. S. (2020). Multimodal neuroimaging data integration and pathway analysis. *Biometrics*, 1–20.

Zheng, X., Aragam, B., Ravikumar, P. K., & Xing, E. P. (2018). Dags with no tears: Continuous optimization for structure learning. In *Advances in neural information processing systems* (pp. 9472–9483).
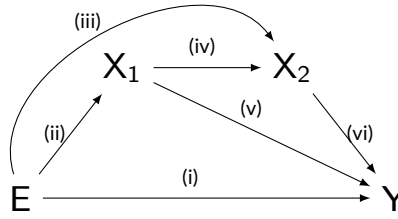
FIGURE 1 Illustration of sequential mediation pathways. Since X_1 and X_2 are multivariate, each potential pathway denoted by (ii) to (vi) can represent multiple paths.
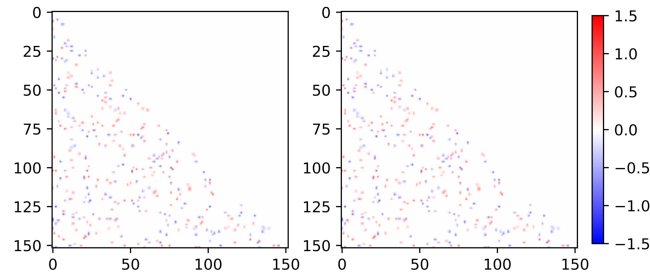

FIGURE 2 The weight adjacency matrix W_0. The left panel: $d_1 = d_2 = 35$, and the right panel: $d_1 = d_2 = 75$.
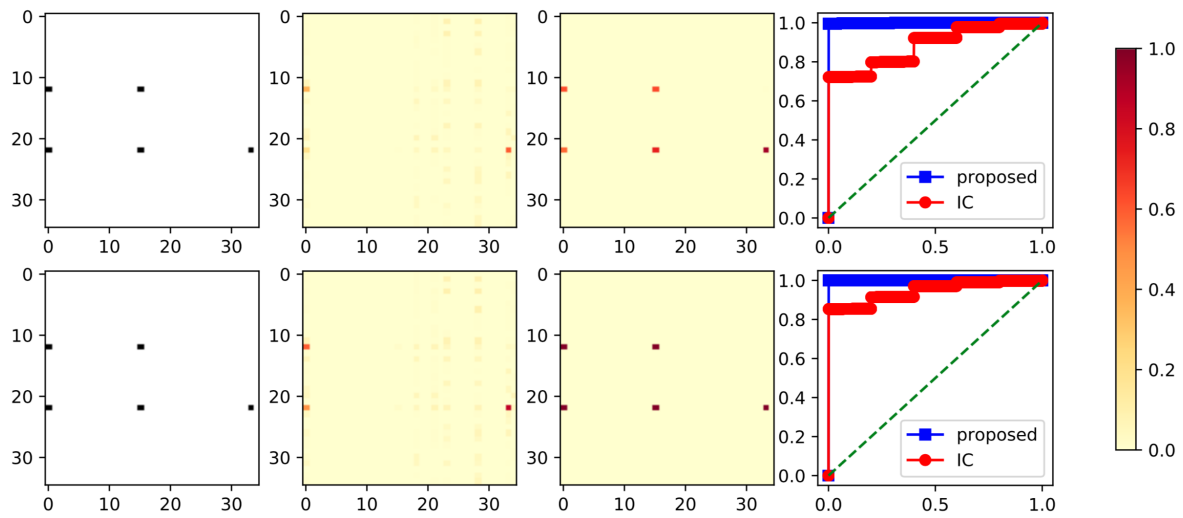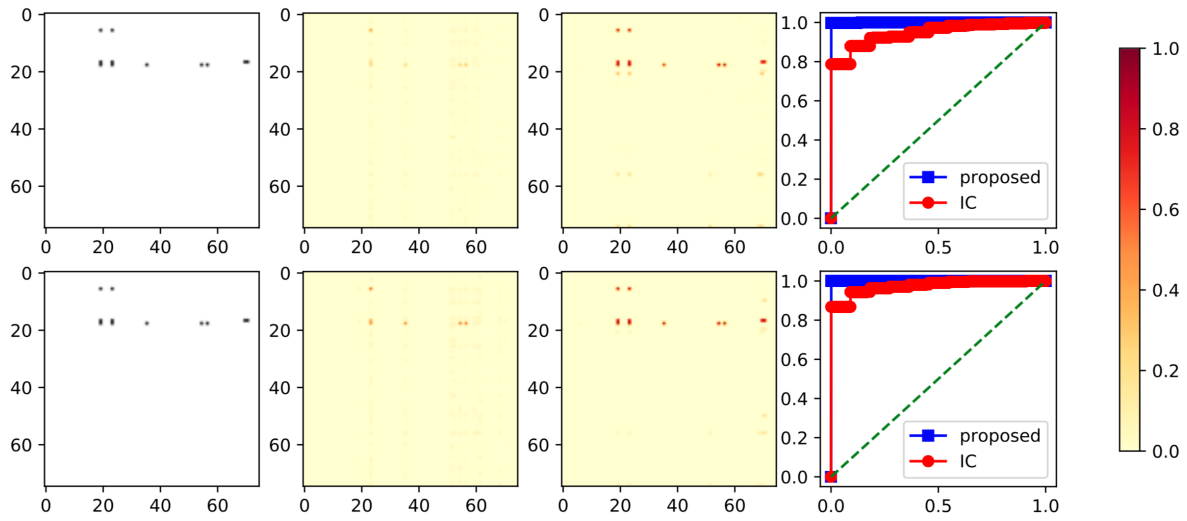

FIGURE 3 Empirical size and power when $d_1 = d_2 = 35$. First column: the vertical axis denotes the indices of the mediators in the first set, and the horizontal axis the second set. The black dots indicate the true significant mediator pairs. Second and third columns: the empirical rejection rate by the method of Chakrabortty et al. (2018), and our test, respectively. Fourth column: the average ROC curve with a varying α. First row: $n = 200$, and second row: $n = 400$.

FIGURE 4 Empirical size and power when $d_1 = d_2 = 75$. First column: the vertical axis denotes the indices of the mediators in the first set, and the horizontal axis the second set. The black dots indicate the true significant mediator pairs. Second and third columns: the empirical rejection rate by the method of Chakrabortty et al. (2018), and our test, respectively. Fourth column: the average ROC curve with a varying α. First row: $n = 300$, and second row: $n = 500$.
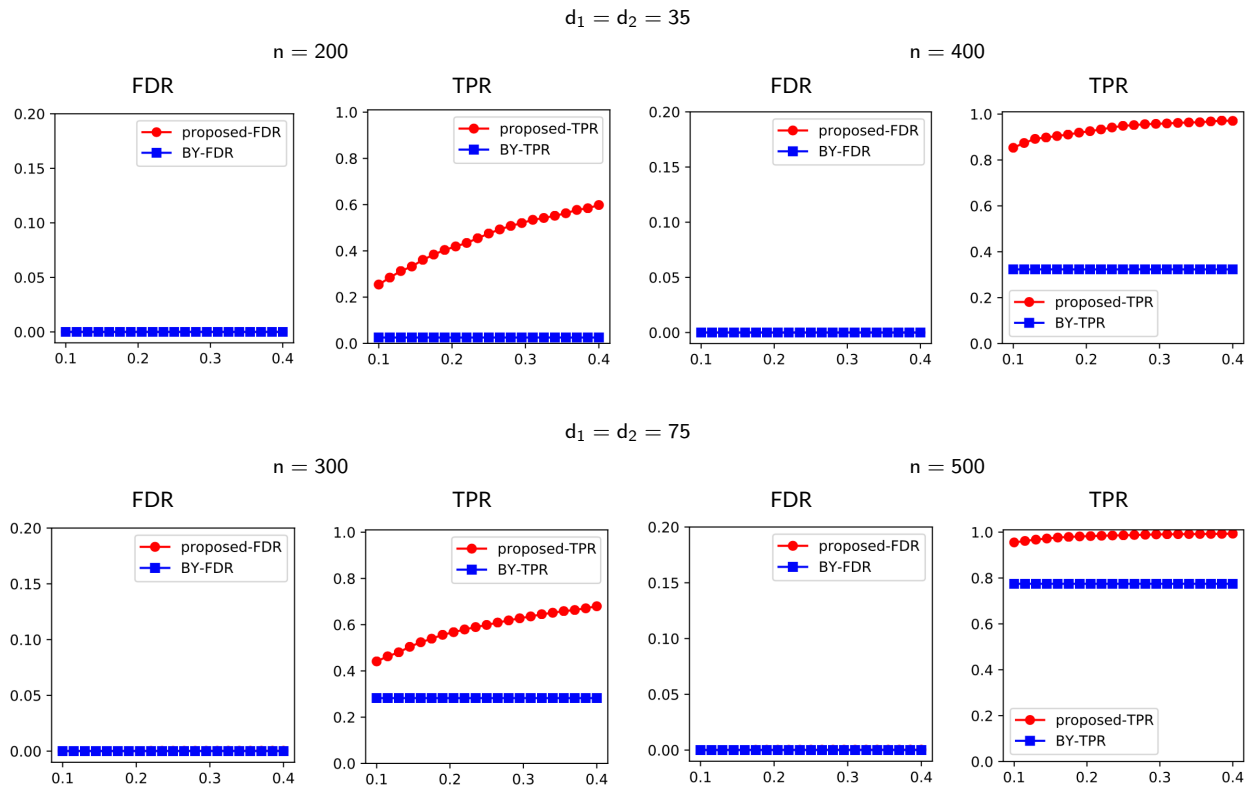

FIGURE 5 False discover rate (FDR) and true positive rate (TPR) of multiple testing. The horizontal axis denotes the varying significance level α.
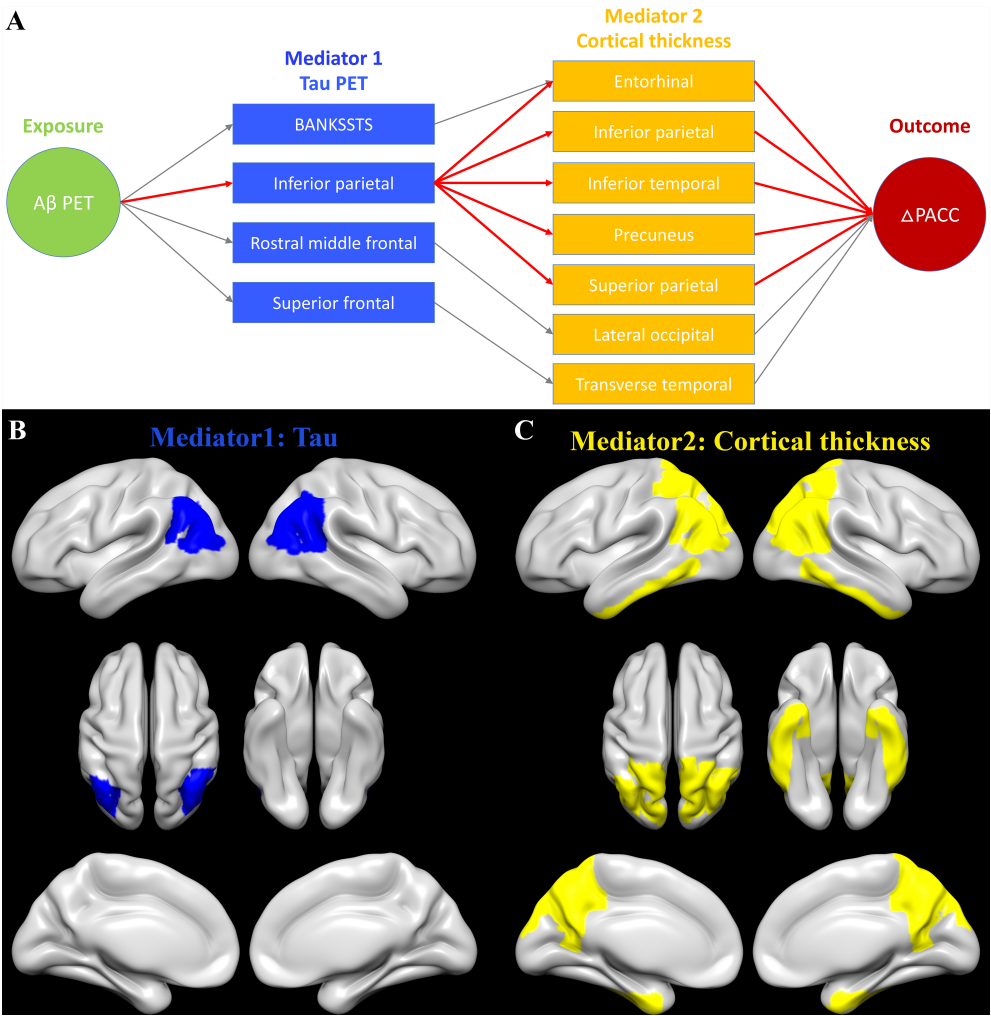
FIGURE 6 Multimodal AD analysis: the identified pathways and brain regions overlaid on a template brain.