


A NEW MULTIPLE-MEDIATOR MODEL MAXIMALLY UNCOVERING THE MEDIATION PATHWAY: EVALUATING THE ROLE OF NEUROIMAGING MEASURES IN AGE-RELATED COGNITIVE DECLINE

BY HWIYOUNG LEE^{1,a} , CHIXIANG CHEN^{2,c}, PETER KOCHUNOV^{3,d}, L. ELLIOT HONG^{3,e}, AND SHUO CHEN^{1,2,b}

¹Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore,,
^ahwiyoun.lee@som.umd.edu; ^bshuochen@som.umd.edu

²Division of Biostatistics and Bioinformatics, Department of Epidemiology and Public Health, University of Maryland School of Medicine, ^cchixiang.chen@som.umd.edu

³Department of Psychiatry and Behavioral Science, University of Texas Health Science Center Houston,
^dpeter.kochunov@uth.tmc.edu; ^eL.Elliot.Hong@uth.tmc.edu

Aging changes brain functions and structures in a downward trajectory and consequently leads to a decline in neurocognitive performance. Our research is motivated by understanding whether and to what extent the age-effect on cognitive decline can be explained by neuroimaging measures. We consider a new mediation model with age as an independent variable, while treating neuroimaging data and cognitive function as the multiple mediators and outcome, respectively. Given that the brain is the primary organ responsible for cognitive function, it is neurobiologically intuitive that the age-related decline in cognition is largely mediated through neuroimaging measures. Additionally, cognitive function is localized to certain regions of the brain rather than being a function of the entire brain. Taking these factors into account, we propose a novel mediation model with multiple mediators that aims to maximally uncover the mediation pathway while simultaneously identifying active neuroimaging mediators by imposing an ℓ_1 penalty and ℓ_2 constraint. We develop a computationally efficient algorithm to handle the nonconvex optimization problem of penalized mediation proportion maximization. We apply our method to a data example of 37,441 participants of UK Biobank with cortical gray-matter thickness and white-matter integrity measures and cognitive performance scores. Our results show that the mediation effect of brain-imaging variables can explain 97% of age-related cognitive decline.

1. Introduction. Research into brain aging has garnered increasing interest due to the rapid growth of the elderly population throughout the world (Park and Reuter-Lorenz, 2009). Among all the symptoms of neurodegeneration, cognitive decline is the hallmark of brain aging (Morrison and Baxter, 2012). Understanding the underlying changes in the central nervous system related to cognitive decline has become an imperative task. Neuroimaging has a central role in capturing brain functional and structural changes related to aging and cognitive functions (Liem et al., 2017; Grady, 1998, 2000). Recent research has established brain charts for the human lifespan that characterize brain metrics at all stages of life (Bethlehem et al., 2022). In addition, neuroimaging-based brain-aging calculators have been widely used to identify risk factors associated with abnormal brain aging (Cole and Franke, 2017; Franke and Gaser, 2019; Niu et al., 2020; Butler et al., 2021). On the other hand, numerous studies have identified the brain locations that are correlated with different cognitive functions (Kaup et al., 2011; Näslund et al., 2000). However, to date, no comprehensive model has been developed that integrates age, neuroimaging data, and cognitive function to elucidate the impact of aging on cognitive function by altering the underlying neurobiological basis, the brain.

Keywords and phrases: Aging, Maximal Mediation Effect, Neuroimaging, Optimization.

In the current research, we consider a general neurobiological pathway in which aging changes brain functions and structures. These changes consequently lead to age-related cognitive decline. This pathway is neurobiologically sensible because (i) aging causes changes in the central nervous system, (ii) the brain is the primary organ that determines cognitive performance, and (iii) the reverse causal links in (i) and (ii) are not valid (Salthouse, 2011). Therefore, we can represent the aforementioned causal pathway by a mediation model (Baron and Kenny, 1986; VanderWeele and Vansteelandt, 2014) in which brain-imaging variables are multiple mediators.

In recent years, advanced statistical mediation models (Zhang et al., 2016; Chén et al., 2018; Zhao, Lindquist and Caffo, 2020; Zhao, Li and Caffo, 2021; Zhao and Luo, 2022) for brain-imaging data have been developed and successfully applied to various aspects of neuroscience providing biologically meaningful results. These methods leverage high-dimensional statistical techniques such as regression shrinkage and low-rank factorization to identify the mediation effects while accounting for the high-throughput variables. However, existing multivariate mediation methods are not tailored for our study’s purpose, because our objective is to develop a neurobiologically plausible mediation model that achieves two goals simultaneously: (i) fully uncovering the mediation effect of neuroimaging data for the aging effect on cognitive decline and (ii) identifying brain regions involved in the cognitive aging process. For example, under the linear structural equation model (LSEM) framework, regression shrinkage methods are employed by Zhao, Li and Caffo (2021); Zhao and Luo (2022) to select mediation pathways with parsimony. However, these methods are not specifically designed to fully uncover the mediation proportion. It is worth noting that direct shrinkage on the regression coefficients in LSEM may lead to a biased estimation of mediation effects. Consequently, the mediation proportion for multivariate mediation analysis can be largely attenuated. In many applications, the attenuation of mediation proportions can be misleading, as it may result in an underestimation of the importance of mediators in the underlying process. For example, in our application that investigates the mediation role of the brain in age-related cognitive decline, the attenuation in the mediation effect may lead to an underestimation of the brain’s role in explaining age-induced cognitive decline. To the best of our knowledge, the methodological gap we identified has not been closed. Therefore, we propose a new mediation model with multiple mediators that maximally uncovers the mediation pathway (MMP).

Built upon the commonly used LSEM framework, MMP introduces a novel objective function that implicitly integrates both mediator and outcome regression models in the LSEM (Chén et al., 2018; Zhao and Luo, 2022). This integrative formulation enables the direct maximization of the mediation effect. MMP adopts a widely used strategy to handle multivariate data in complex statistical analysis frameworks: maximization. For example, canonical correlation analysis (CCA) identifies and highlights the underlying associations between two sets of multivariate variables by finding linear combinations that *maximize* the correlation between them. The associations uncovered through this approach may be overlooked or disregarded by other statistical analysis methods (e.g., performing dimension reduction separately for the two sets of variables or maximizing likelihood with regression shrinkage). Following this rationale, our objective is to maximally uncover the mediation pathway, thereby revealing the three-way associations among the exposure, mediator, and outcome variables. Furthermore, to handle multiple neuroimaging mediators and enhance estimation stability, we incorporate both the ℓ_1 penalty and ℓ_2 constraint in the objective function. These regularization terms are well-suited for parsimonious mediator selection from high-dimensional correlated neuroimaging variables. Therefore, MMP provides an accurate estimate of the aggregate mediation effect of multiple mediators that maximally uncovers the mediation pathway, while effectively avoiding the overinflation of the estimated mediation proportion (see Section 4).

The rationale behind maximizing the MP is grounded in the distinctive characteristics of the brain, which plays a pivotal role in governing cognitive functions. Since the brain is the exclusive primary organ responsible for these functions, the pursuit of maximizing MP in this context is not only neurobiologically logical but also critical for gaining in-depth understanding of the mechanism of age-induced cognitive decline through the brain. The brain regions that are involved in maximizing MP can be considered to play the most influential roles in mediating cognitive aging. Consequently, by targeting the regions identified by the method, it contributes to the development of more effective and tailored intervention strategies for cognitive health.

To implement our method, we devise a novel optimization algorithm by coupling the alternating direction method of multipliers with semidefinite relaxation to address the challenges posed by the nonconvexity of the optimization function and the multiple constraints in the objective function. The proposed algorithm has demonstrated robust performance and efficient computation both in simulations and in real data examples.

Our method makes at least three novel contributions. First, MMP provides a new strategy to handle a special type of multiple mediator data that is exposure \rightarrow complex biosystem \rightarrow clinical outcome, where the complex biosystem (mediators) is mainly responsible for the exposure effect on the clinical outcome. In these applications, prior knowledge has been established (e.g., by biology) that the mediation effect plays a prominent role. Built on prior knowledge, MMP assesses the mediating role of the complex biosystem by prioritizing active mediators that maximize the mediation effect under regularization (overcoming the inflated estimate of the mediation effect). This approach is critical for enhancing our understanding of the complex biosystem that drives a biological process, such as age-related cognitive decline. Secondly, we have developed a computationally efficient and scalable algorithm to maximize the mediation proportion while performing active mediator selection with a penalty and constraints. Last but not least, our findings by applying MMP to a large sample of UK biobank (UKBB) data reveals that age-related cognitive decline is explained by the decrease in white matter integrity, accounting for 97% of the effect, rather than the cortical thickness of grey matter.

The rest of this paper is organized as follows. We develop our mediation model and describe the numerical algorithm in [Section 2](#). The proposed method is applied to the UKBB dataset in [Section 3](#) to understand the progressive loss of cognitive function with aging. A simulation study is presented in [Section 4](#), which demonstrates the improved performance of the proposed method. We conclude in [Section 5](#) with a discussion.

2. Method.

2.1. Background. We consider a mediation model with a univariate exposure (age), multiple mediators of neuroimaging variables, and a univariate outcome (e.g., cognitive function). Let vectors $\mathbf{X} = (X_1, \dots, X_n)^\top$ and $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ be the age and outcome of a study with n participants, respectively. We use an $n \times p$ matrix $\mathbf{M} = (\mathbf{M}_1, \dots, \mathbf{M}_p)$ to represent p neuroimaging variable mediators for the n participants. A vector $\mathbf{M}_j = (M_j^1, \dots, M_j^n)^\top$ represents the j th mediator for all participants, and a vector \mathbf{M}^i contains the p imaging variables for subject $i = 1, \dots, n$. Without loss of generality, we assume that all variables $(\mathbf{X}, \mathbf{M}, \mathbf{Y})$ are normalized to have a mean of 0 and a variance of 1. We first present the model without considering confounders, and then introduce an extended model adjusting for confounders in [Section 2.2](#).

Under the linear structural equation modeling (LSEM) framework, the multiple mediation model can be written in the following form ([Zhao, Li and Alzheimer’s Disease Neuroimaging](#)

Initiative, 2022; Zhao and Luo, 2022):

$$\begin{aligned}
 \mathbf{Y} &= \mathbf{X}\tau + \mathbf{e}, \\
 \mathbf{M}_j &= \mathbf{X}\alpha_j + \boldsymbol{\epsilon}_j, \quad \text{for } j = 1, \dots, p, \\
 \mathbf{Y} &= \mathbf{X}\gamma + \sum_{j=1}^p \beta_j \mathbf{M}_j + \boldsymbol{\epsilon},
 \end{aligned}
 \tag{1}$$

where the first equation is the linear regression model, which measures the total effect τ of \mathbf{X} on \mathbf{Y} .

The total effect can be decomposed into a direct effect γ and an indirect (mediation) effect $\sum_{j=1}^p \alpha_j \beta_j$. The conventional mediation analysis focuses on the estimation and inference of the direct, indirect, and total effects. The mediation proportion (MP), defined as the ratio of the indirect effect to the total effect, (i.e., $\sum_{j=1}^p \alpha_j \beta_j / \tau$), has been widely used to assess the mediation effect size. With a fixed total effect $\tau = \sum_{j=1}^p \alpha_j \beta_j + \gamma$, the MP captures how much of the effect of an exposure on an outcome is transmitted through a mediator, which also evaluates the relative importance of the mediating pathway over the direct pathway. Therefore, maximally uncovering the effect of the mediation pathway is equivalent to maximizing the MP. However, in a mediation analysis with multiple mediators, robustly estimating the MP is challenging because the effects of high-dimensional mediators can be interactive. While advanced multivariate mediation models have been successfully developed to handle multiple mediators in recent years (Chén et al., 2018; Zhao and Luo, 2022; Zhang et al., 2016, among many others), their primary focus is not on maximally uncovering the mediation pathway. Therefore, these methods may not maximally explain the age effect on the cognitive outcome through neuroimaging mediators. This motivated us to develop the MMP method.

2.2. A mediation model with multiple mediators that maximally uncovers the mediation pathway. In order to model a neurobiological pathway of cognitive aging (i.e., the normal aging process causes changes in multiple brain areas resulting in an aggregate impact on brain functions, such as cognitive function), we maximize the MP with a new objective function. This approach enables us to emphasize the mediating role of neuroimaging data on age-induced changes in brain function.

We build the MMP approach based on the LSEM framework. Since the mediation proportion of (1) is determined by the aggregate indirect effect of multiple mediators $\sum_{j=1}^p \alpha_j \beta_j$, we consider a more flexible aggregate mediation factor as a combination of multiple brain-imaging mediators to maximize the mediation proportion, i.e., $\mathbf{M}_* = \mathbf{M}\boldsymbol{\omega}$, where $\boldsymbol{\omega} \in \mathbb{R}^p$ are the parameters of interest (See Figure 1). The use of \mathbf{M}_* allows for the integration of brain regions that contribute to the brain mediation effect with different weights $\boldsymbol{\omega}$, which ultimately results in the maximization of the mediation proportion (MP). In addition, we make the sparsity assumption on $\boldsymbol{\omega}$ to deal with the high-dimensional nature of brain-imaging data as well as to ensure the robustness of the estimation. Assuming that \mathbf{M}_* is normalized without loss of generality (i.e., \mathbf{M}_* has a mean of 0 and a variance of 1), the mediation model with the aggregate mediation factor \mathbf{M}_* has the following form:

$$\begin{aligned}
 \text{Aggregate mediator: } \mathbf{M}_* &= \mathbf{M}\boldsymbol{\omega}, \\
 \text{Mediation model: } \mathbf{M}_* &= \mathbf{X}\alpha + \boldsymbol{\epsilon}, \\
 \mathbf{Y} &= \mathbf{X}\gamma + \beta \mathbf{M}_* + \boldsymbol{\epsilon}.
 \end{aligned}
 \tag{2}$$

Under (2) and given the total effect $\tau = \alpha\beta + \gamma$ is fixed as in (1), maximizing the mediation proportion $\alpha\beta / (\alpha\beta + \gamma)$ becomes straightforward. Here, we show that maximizing the MP

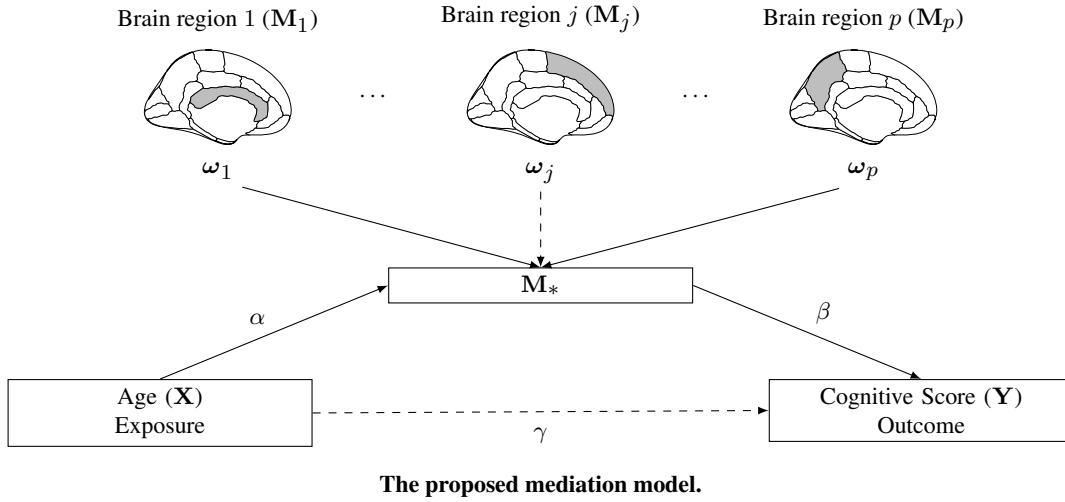
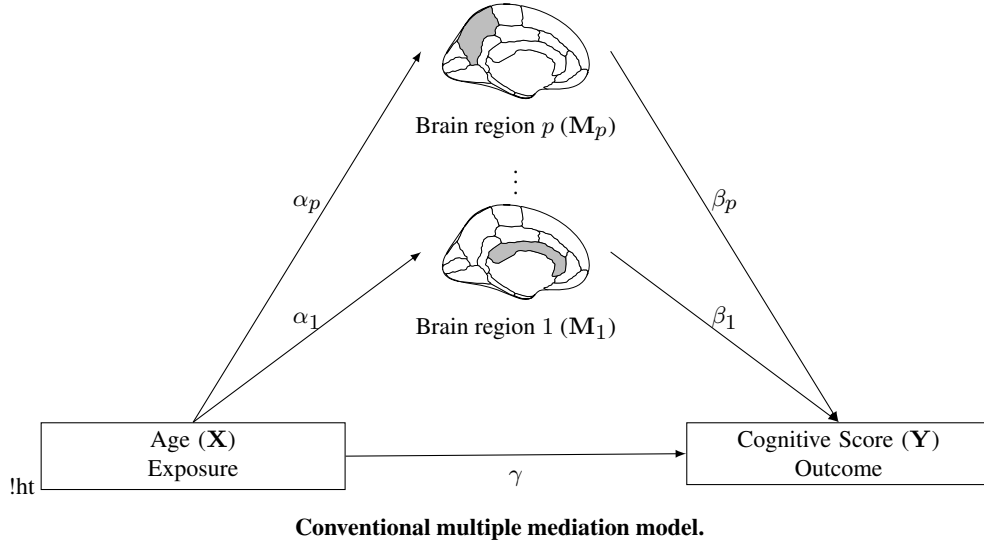


Fig 1: Graphical illustrations of mediation models using neuroimaging measures as multiple mediators. Top: The conventional multivariate mediation model (i.e., LSEM) includes multiple mediation pathways through each neuroimaging mediator, where the mediation proportion is determined by the sum of pathways $\sum_{j=1}^p \alpha_j \beta_j / \tau$ (pathways may be correlated). Bottom: The proposed MMP approach utilizes an aggregate mediator M_* , which directly computes the mediation proportion ($\alpha\beta/\tau$). Note that M_* is calculated based on a parsimonious set of mediators (e.g., M_1, M_p) that are connected to the M_* with the solid line). In our application, MMP maximally explains the effect of age on external phenotype (i.e., cognitive function) using a few neuroimaging mediators. Maximization of the MP results in the direct effect (γ) being minimized (dashed line)."

can be achieved by optimizing ω . We assume that (i) the signs of total, direct, and indirect effects are consistent, and (ii) without loss of generality in our application, we assign a negative sign to the effects since cognitive function declines with aging. Generally, in the context of mediation analysis, MP can fall outside the $[0, 1]$ range when the direct and indirect ef-

fects have opposite signs, a phenomenon referred to as inconsistent mediation (MacKinnon, Krull and Lockwood, 2000). In our application and many other studies (e.g., brain and cognitive function development in adolescents), the inconsistent mediation is not neurobiologically plausible. Therefore, we exclude inconsistent mediation for our method development. Therefore, instead of directly maximizing the aggregate MP, we maximize $-\alpha\beta$. Note that while we make the aforementioned assumptions, the MMP method can be easily adapted to general scenarios, provided that the total effect remains fixed. Furthermore, we have $\alpha \propto \mathbf{X}^\top \mathbf{M}_*$, and β proportional to the partial correlation between \mathbf{M}_* and the outcome given age, i.e., $\beta \propto \mathcal{E}_{\mathbf{M}_*|\mathbf{X}}^\top \mathcal{E}_{\mathbf{Y}|\mathbf{X}}$. However, \mathbf{M}_* is unknown and determined by multiple mediators and estimating parameters $\boldsymbol{\omega}$. We can calculate $\mathcal{E}_{\mathbf{M}_*|\mathbf{X}} = \mathbf{M}_* - \mathbf{X}\alpha = (\mathbf{I}_n - \mathbf{P})\mathbf{M}\boldsymbol{\omega}$, and $\mathcal{E}_{\mathbf{Y}|\mathbf{X}} = \mathbf{Y} - \mathbf{X}\tau$, where $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$ represents the projection matrix. Therefore, we transform the objective of maximizing the MP into optimizing the following quadratic form with regard to the estimating parameters $\boldsymbol{\omega}$ and input data:

$$\begin{aligned} \text{Maximize } \left(\frac{\alpha\beta}{\alpha\beta + \gamma} \right) &\stackrel{\text{(i)}}{=} \text{Maximize } |\alpha\beta| \stackrel{\text{(ii)}}{=} \text{Minimize } \mathbf{M}_*^\top \mathbf{X} \mathcal{E}_{\mathbf{Y}|\mathbf{X}}^\top \mathcal{E}_{\mathbf{M}_*|\mathbf{X}} \\ &= \text{Minimize } \frac{1}{2} \boldsymbol{\omega}^\top \mathbf{M}^\top \mathbf{X} \mathcal{E}_{\mathbf{Y}|\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{P}) \mathbf{M} \boldsymbol{\omega}. \end{aligned}$$

To handle the massive neuroimaging mediators, we further regularize the quadratic form by introducing penalty terms. Therefore, our primary objective function is:

$$\begin{aligned} (3) \quad &\underset{\boldsymbol{\omega} \in \mathbb{R}^p}{\text{argmin}} \frac{1}{2} \boldsymbol{\omega}^\top \mathbf{M}^\top \mathbf{X} \mathcal{E}_{\mathbf{Y}|\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{P}) \mathbf{M} \boldsymbol{\omega} + \lambda \|\boldsymbol{\omega}\|_1 \\ &\text{subject to } \|\mathbf{M}\boldsymbol{\omega}\|_2^2 = 1 \text{ and } \langle \mathbf{M}\boldsymbol{\omega}, \mathbf{1} \rangle = 0. \end{aligned}$$

The ℓ_1 norm penalty ($\|\cdot\|_1$) imposes the sparsity on the solution of $\boldsymbol{\omega}$ and selects only a small set of brain-imaging variables with nonzero $\boldsymbol{\omega}$ with a tuning parameter λ (Tibshirani, 1996). Moreover, because reparametrization of the indirect effect $\alpha\beta$ with respect to $\boldsymbol{\omega}$ is valid under the assumption that the aggregate mediator \mathbf{M}_* is normalized (unit variance, and centralized), we need two additional constraints i) $\|\mathbf{M}_*\|^2 = \|\mathbf{M}\boldsymbol{\omega}\|^2$, which is equivalent to the ℓ_2 norm of $\boldsymbol{\omega}$ associated with $p \times p$ matrix $\mathbf{M}^\top \mathbf{M}$ (i.e., $\|\boldsymbol{\omega}\|_{\mathbf{M}^\top \mathbf{M}}^2$) ensures that the aggregate mediator has a unit variance, ii) the inner product constraint $\langle \mathbf{M}\boldsymbol{\omega}, \mathbf{1} \rangle = 0$, which is equivalent to $\sum_{i=1}^n \mathbf{M}_{*i} = 0$, is designed to centralize \mathbf{M}_* , ensuring its mean is zero. Without these constraints, the proportionality of the regression coefficients does not hold. The joint application of ℓ_1 regularization and ℓ_2 constraint in (3) allows the simultaneous selection of correlated imaging variables (Zou and Hastie, 2005). This property is desirable for spatially and functionally dependent neuroimaging variables. In addition to facilitating mediator selection, this regularization is also designed to prevent the overinflation of the estimated mediation proportion by constraining the objective function. Therefore, selecting an appropriate tuning parameter is crucial. We adopt the commonly used Bayesian information criterion (BIC) for objective λ selection (Zhao, Li and Alzheimer's Disease Neuroimaging Initiative, 2022). This criterion effectively balances both the goodness of fit of the model (i.e., likelihood) and the model's complexity (i.e., the number of mediators included), ensuring a well-calibrated trade-off between these important aspects. Our application of the ℓ_1 penalty differs from that used in (Zhao and Luo, 2022), where the ℓ_1 penalty is directly applied to shrink the indirect effects, which can lead to the underestimation of the Mediation Proportion (MP). In contrast, the ℓ_1 penalty in our approach is imposed on the weight vector for obtaining the aggregate mediator factor \mathbf{M}_* . After obtaining \mathbf{M}_* , the MP is derived using the single-mediator LSEM without penalty terms. Thus, our approach enables the regularization of high-dimensional mediators without introducing an underestimation bias in the MP.

Moreover, employing LSEM with \mathbf{M}_* as a single mediator enables the straightforward application of conventional methods (such as Bootstrap or Sobel's test) for inference on mediation effects.

One can consider applying a generalized lasso penalty (Tibshirani and Taylor, 2011) to ω by using a prespecified penalty matrix $D \in \mathbb{R}^{m \times p}$ (i.e., $\|D\omega\|_1$) to account for spatial smoothness or dependencies across brain regions, which could be particularly beneficial in neuroimaging studies. However, in this paper, we have chosen to employ the standard lasso penalty to preserve the simplicity of the model. The heuristic derivation of the algorithm that utilizes the generalized lasso penalty is provided in the Supplementary Material (Lee et al., 2024).

Our objective function can be further extended to accommodate confounding covariates. Let $\mathbf{Z} \in \mathbb{R}^{n \times q}$ denote the q confounding covariates. We account for the confounding covariates with orthogonalization, i.e., $\beta \propto \mathcal{E}_{(\mathbf{M}_*, \mathbf{X})|\mathbf{Z}}^\top \mathcal{E}_{(\mathbf{Y}, \mathbf{X})|\mathbf{Z}}$, where $\mathcal{E}_{(\mathbf{Y}, \mathbf{X})|\mathbf{Z}} = \mathcal{E}_{\mathbf{Y}|\mathbf{X}} - \mathbf{Z}\eta$ and $\mathcal{E}_{(\mathbf{M}_*, \mathbf{X})|\mathbf{Z}} = \mathcal{E}_{\mathbf{M}_*|\mathbf{X}} - \mathbf{Z}\psi = (\mathbf{I}_n - \mathbf{Q})(\mathbf{I}_n - \mathbf{P})\mathbf{M}\omega$, and where $\mathbf{Q} = \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top$ is the projection matrix. Our parameter estimation procedure is compatible with this extension because the optimization procedure can straightforwardly incorporate the adjusted projection matrix.

Causal mediation assumptions. Next, we will briefly provide some definitions and assumptions that are necessary to establish a causal interpretation of the proposed mediation model. Following the counterfactual framework (Rubin, 1974), we denote by $\mathbf{M}_*^i(x)$ and $Y_i(x, \mathbf{M}_*^i(x))$ the potential outcome of the aggregate mediator and the cognitive score (outcome) of the i -th subject, respectively, that would have been realized when the exposure was set to $X_i = x$. Within the LSEM framework and our model, these quantities can be expressed as $\mathbf{M}_*^i(x) = \alpha x + \epsilon_i$, and $Y_i(x, \mathbf{M}_*^i(x)) = \gamma x + \beta \mathbf{M}_*^i(x) + \epsilon_i$, respectively. Then the total effect (TE) of x against x' , which is defined as $\text{TE} = Y_i(x, \mathbf{M}_*^i(x)) - Y_i(x', \mathbf{M}_*^i(x'))$ can be decomposed into the summation of two distinct components: the natural indirect effect (NIE), and the natural direct effect (NDE), which can be expressed as follows:

$$\text{TE} = \underbrace{Y_i(x, \mathbf{M}_*^i(x)) - Y_i(x, \mathbf{M}_*^i(x'))}_{\text{NIE}} + \underbrace{Y_i(x, \mathbf{M}_*^i(x')) - Y_i(x', \mathbf{M}_*^i(x'))}_{\text{NDE}}.$$

Because $Y_i(x, \mathbf{M}_*^i(x'))$ cannot be directly observed when $x \neq x'$, the individual NIE, and NDE are not identifiable. However by making use of the commonly imposed causal identification assumption (Imai, Keele and Tingley, 2010),

$$(A1) \{Y_i(x', \mathbf{m}_*), \mathbf{M}_*^i(x)\} \perp\!\!\!\perp X_i | Z_i$$

$$(A2) Y_i(x', \mathbf{m}_*) \perp\!\!\!\perp \mathbf{M}_*^i(x) | X_i = x, Z_i,$$

the average natural indirect effect, i.e., $\text{AIE} = \mathbb{E}[Y_i(x, \mathbf{M}_*^i(x)) - Y_i(x, \mathbf{M}_*^i(x^*))] = \alpha\beta(x - x^*)$, and the average natural direct effect, i.e., $\text{ADE} = \mathbb{E}[Y_i(x, \mathbf{M}_*^i(x^*)) - Y_i(x^*, \mathbf{M}_*^i(x^*))] = \gamma(x - x^*)$ can be identified. Note that \mathbf{m}_* , and Z_i denote the controlled aggregate mediator, and the confounder, respectively. As discussed by Imai, Keele and Yamamoto (2010), the cross-world independence assumption (A2) is often considered strong and unrealistic. This can be relaxed by employing the linear structural equation model (LSEM) with independent errors (Andrews and Didelez, 2021), which enables the identification of the NIE and NDE even when (A2) is violated. Moreover, as highlighted by Chén et al. (2018) in cases where the causal mediation claims cannot be made, we can still use $\alpha\beta(x - x^*)$ to quantify the mediation effect in exploratory mediation analysis (Serang et al., 2017).

2.3. *Estimation.* In this subsection, we focus on the optimization of the objective function (3) used for ω estimation. Note that even though we have simplified the problem of maximizing the MP to a constrained quadratic minimization problem, directly solving (3) remains an NP-hard problem, because $\mathbf{M}^\top \mathbf{X} \mathcal{E}_{\mathbf{Y}|\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{P}) \mathbf{M}$ is not guaranteed to be positive semi-definite. Therefore, algorithms commonly utilized for convex optimization are not directly applicable. To address this issue, we develop a tailored algorithm to circumvent the potential nonconvexity via semidefinite relaxation (Luo et al., 2010).

Instead of directly optimizing (3) with respect to ω , we introduce an auxiliary variable $\mathbf{u} \in \mathbb{R}^p$ analogous to the alternating direction method of multipliers framework (ADMM, Boyd et al. (2011)). This allows us to alternately updating $\{\omega, \mathbf{u}\}$ by sequentially solving a set of simple optimization problems. We begin by formulating the first line of the objective function in the following scaled augmented Lagrangian form:

$$(4) \quad \mathcal{L}_\rho(\omega, \mathbf{u}, \mathbf{v}) = \frac{1}{2} \omega^\top \mathbf{M}^\top \mathbf{X} \mathcal{E}_{\mathbf{Y}|\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{P}) \mathbf{M} \omega + \lambda \|\mathbf{u}\|_1 + \frac{\rho}{2} \|\omega - \mathbf{u} + \mathbf{v}\|_2^2 - \frac{\rho}{2} \|\mathbf{v}\|_2^2,$$

where $\mathbf{v} \in \mathbb{R}^p$ and $\rho \geq 0$ denote the dual variable and augmented Lagrangian penalty.

When $\mathbf{M}^\top \mathbf{X} \mathcal{E}_{\mathbf{Y}|\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{P}) \mathbf{M}$ is not positive semi-definite, we optimize the scaled augmented Lagrangian function with semidefinite relaxation. This procedure is different from the commonly used ADMM algorithm because updating a primal variable requires solving a nonconvex optimization problem. Moreover, the additional constraints on ω related to the normalization of \mathbf{M}_* also have to be taken into account in the optimization. Specifically, we update the primal variable ω in (4) while preserving the normalization constraints,

$$(5) \quad \omega^{t+1} = \underset{\omega \in \mathbb{R}^p}{\operatorname{argmin}} \mathcal{L}_\rho(\omega, \mathbf{u}^t, \mathbf{v}^t) \text{ s.t. } \|\mathbf{M}_*\|_2^2 = 1, \langle \mathbf{M}\omega, \mathbf{1} \rangle = 0,$$

where $t \geq 0$ denotes the iteration number in the optimization. By introducing an additional auxiliary variable $\nu \in \mathbb{R}$, (5) can be expressed as a homogeneous quadratically constrained quadratic program of $\tilde{\omega}^\top = (\omega^\top, \nu)$ by:

$$(6) \quad \underset{\tilde{\omega} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \tilde{\omega}^\top \mathbf{C}_t \tilde{\omega} \\ \text{s.t. } \tilde{\omega}^\top \mathbf{A}_1 \tilde{\omega}^\top = 1, \quad \tilde{\omega}^\top \mathbf{A}_2 \tilde{\omega}^\top = 1, \quad \tilde{\omega}^\top \mathbf{A}_3 \tilde{\omega}^\top = 0.$$

The $(p+1) \times (p+1)$ real symmetric matrices \mathbf{C}_t , \mathbf{A}_1 , \mathbf{A}_2 , and \mathbf{A}_3 are given by:

$$\mathbf{C}_t = \begin{pmatrix} \left\{ 2\rho \mathbf{I}_p + \mathbf{M}^\top \left(\mathbf{X} \mathcal{E}_{\mathbf{Y}|\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{P}) + (\mathbf{I}_n - \mathbf{P}) \mathcal{E}_{\mathbf{Y}|\mathbf{X}} \mathbf{X}^\top \right) \mathbf{M} \right\} / 4 \rho (\mathbf{v}^t - \mathbf{u}^t) / 2 & \\ \rho (\mathbf{v}^t - \mathbf{u}^t)^\top / 2 & 0 \end{pmatrix},$$

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} \mathbf{M}^\top \mathbf{M} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{pmatrix}, \quad \mathbf{A}_3 = \begin{pmatrix} \mathbf{0} & \mathbf{M}^\top \mathbf{1} / 2 \\ \mathbf{1}^\top \mathbf{M} / 2 & 0 \end{pmatrix},$$

where $\mathbf{0}$ and $\mathbf{1}$ denote the zero- and one-column vectors of order p and \mathbf{O} is the $p \times p$ zero matrix. Let $\Omega = \tilde{\omega} \tilde{\omega}^\top$. Then by dropping the rank constraint $\operatorname{rank}(\Omega) = 1$, we can convexify it to a form of semidefinite program (Vandenberghe and Boyd, 1996):

$$(7) \quad \hat{\Omega}_{t+1} = \underset{\Omega \in \mathbb{S}^{p+1}}{\operatorname{argmin}} \operatorname{Tr}(\mathbf{C}_t \Omega), \\ \text{s.t. } \operatorname{Tr}(\mathbf{A}_1 \Omega) = 1, \quad \operatorname{Tr}(\mathbf{A}_2 \Omega) = 1, \quad \operatorname{Tr}(\mathbf{A}_3 \Omega) = 0, \\ \Omega \succeq \mathbf{0},$$

where \mathbb{S}^{p+1} and $\Omega \succeq \mathbf{0}$ denote the space of $(p+1) \times (p+1)$ real symmetric matrices and the positive semidefiniteness of Ω , respectively. Given the predefined tolerance level ϵ , the SDR

problem in (7) converges to the optimum with a bounded convergence rate $\mathcal{O}(\max(3, p + 1)^4 \sqrt{p+1} \log(1/\epsilon))$ (Luo et al., 2010). Suppose $\mathbf{\Omega}_{t+1}$ is the optimal solution to (7) and that it also satisfies $\text{rank}(\mathbf{\Omega}_{t+1}) = 1$. Let $\widehat{\mathbf{\Omega}}_{t+1} = \mathbf{Q}_{t+1} \mathbf{\Lambda}_{t+1} \mathbf{Q}_{t+1}^\top$, where $\mathbf{Q}_{t+1} = (\mathbf{q}_1^{t+1}, \dots, \mathbf{q}_{p+1}^{t+1})$ and $\mathbf{\Lambda}_{t+1} = \text{diag}(\lambda_1^{t+1}, \dots, \lambda_{p+1}^{t+1})$, which is the eigendecomposition of $\widehat{\mathbf{\Omega}}_{t+1}$. Since the rank 1 approximation of $\mathbf{\Omega}_{t+1}$ from the obtained $\widehat{\mathbf{\Omega}}_{t+1}$ is given by $\widetilde{\mathbf{\Omega}}_{t+1} = \lambda_1^{t+1} \mathbf{q}_1^{t+1} \mathbf{q}_1^{t+1 \top}$, we set $\widetilde{\boldsymbol{\omega}}_{t+1} = \sqrt{\lambda_1^{t+1}} \mathbf{q}_1^{t+1}$ as a potential solution to (6). The updating rule for the auxiliary variable \mathbf{u} is straightforward because it can be obtained by evaluating the proximal map (Parikh and Boyd, 2014), i.e., $\mathbf{u}^{t+1} = \mathbf{prox}_{\|\cdot\|_1, \lambda/\rho}(\boldsymbol{\omega}^{t+1} + \mathbf{v}^t)$. The proximal map of the ℓ_1 norm is the elementwise soft-thresholding operator, i.e., $S_{\lambda/\rho}(\boldsymbol{\omega}^{t+1} + \mathbf{v}^t) = \{\text{sgn}(\omega_j^{t+1} + v_j^t)(|\omega_j^{t+1} + v_j^t| - \lambda/\rho)_+\}_{j=1}^p$, where $(|\omega_j^{t+1} + v_j^t| - \lambda/\rho)_+ = \max(|\omega_j^{t+1} + v_j^t| - \lambda/\rho, 0)$. The numerical algorithm for solving (3) is summarized in Algorithm 1, and the detailed derivation is provided in S1 of the Supplementary Material (Lee et al., 2024).

Algorithm 1 Algorithm for MMP

Require: Given the results from the t th step, then for the $t + 1$ th step,

- 1: Update $\boldsymbol{\omega}^{t+1} = \text{argmin}_{\boldsymbol{\omega}} \mathcal{L}_\rho(\boldsymbol{\omega}, \mathbf{u}^t, \mathbf{v}^t)$
 - i) Implement convex optimization, when $\mathbf{M}^\top \mathbf{X} \mathcal{E}_{\mathbf{Y}|\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{P}) \mathbf{M} \succeq 0$
 - ii) Implement SDR optimization, when $\mathbf{M}^\top \mathbf{X} \mathcal{E}_{\mathbf{Y}|\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{P}) \mathbf{M} \not\succeq 0$

$$\widehat{\mathbf{\Omega}}_{t+1} = \text{argmin}_{\mathbf{\Omega} \in \mathbb{S}^{p+1}} \text{Tr}(\mathbf{C}\mathbf{\Omega}) \quad \text{s.t.} \quad \text{Tr}(\mathbf{A}_1\mathbf{\Omega}) = 1, \text{Tr}(\mathbf{A}_2\mathbf{\Omega}) = 1, \text{Tr}(\mathbf{A}_3\mathbf{\Omega}) = 0, \mathbf{\Omega} \succeq 0$$

$$= \mathbf{Q}_{t+1} \mathbf{\Lambda}_{t+1} \mathbf{Q}_{t+1}^\top,$$

where $\mathbf{Q}_{t+1} = (\mathbf{q}_1^{t+1}, \dots, \mathbf{q}_{p+1}^{t+1})$ and $\mathbf{\Lambda}_{t+1} = \text{diag}(\lambda_1^{t+1}, \dots, \lambda_{p+1}^{t+1})$

$$\widetilde{\boldsymbol{\omega}}^{t+1} = \sqrt{\lambda_1^{t+1}} \mathbf{q}_1^{t+1} = \begin{pmatrix} \omega^{t+1} \\ \nu^{t+1} \end{pmatrix}$$

- 2: Update $\mathbf{u}^{t+1} = \text{argmin}_{\mathbf{u}} \mathcal{L}_\rho(\boldsymbol{\omega}^{t+1}, \mathbf{u}, \mathbf{v}^t)$

- 3: Update $\mathbf{v}^{t+1} = \mathbf{v}^t + \boldsymbol{\omega}^{t+1} - \mathbf{u}^{t+1}$
-

THEOREM 2.1. *Let $(\boldsymbol{\omega}^*, \mathbf{u}^*)$ be the optimal solution of \mathcal{L}_0 in (4) satisfying $\boldsymbol{\omega}^* - \mathbf{u}^* = 0$, \mathcal{L}_0^* be the corresponding optimal value, and \mathbf{v}^* be an optimal solution of the dual problem. Suppose that $\mathbf{M}^\top \mathbf{X} \mathcal{E}_{\mathbf{Y}|\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{P}) \mathbf{M}$ is positive semidefinite, then for any constant $\gamma \geq 2\|\mathbf{v}^*\|_2$, we have*

$$(8) \quad \mathcal{L}_0(\boldsymbol{\omega}^{(t)}, \mathbf{u}^{(t)}) - \mathcal{L}_0^* \leq \frac{\|\mathbf{u}^* - \mathbf{u}^0\|_{\rho\mathbf{I}}^2 + (\gamma + \|\mathbf{v}^0\|_2)^2/\rho}{2(t+1)},$$

$$(9) \quad \|\boldsymbol{\omega}^{(t)} - \mathbf{u}^{(t)}\|_2 \leq \frac{\|\mathbf{u}^* - \mathbf{u}^0\|_{\rho\mathbf{I}}^2 + (\gamma + \|\mathbf{v}^0\|_2)^2/\rho}{\gamma(t+1)},$$

where $\boldsymbol{\omega}^{(t)} = \frac{1}{t+1} \sum_{n=0}^t \boldsymbol{\omega}^{n+1}$, $\mathbf{u}^{(t)} = \frac{1}{t+1} \sum_{n=0}^t \mathbf{u}^{n+1}$.

In Theorem 2.1, we show that our algorithm converges to the optimum with a rate of convergence of $\mathcal{O}(1/t)$ which is a typical rate for most convex optimization algorithms (Beck, 2017). Theorem 2.1 ensures that our algorithms can accurately estimate weight parameters

to achieve the maximal MP. The proof is provided in S2 of the Supplementary Material (Lee et al., 2024).

3. Application.

3.1. *Sample and materials.* Our motivation was to understand the progressive loss of cognitive function throughout the aging process by studying age-related changes in the human brain. We used a MMP to investigate the neurophysiology of cognitive decline during brain aging. We utilized the UKBB data in the investigation because this cohort includes a large sample of healthy older participants with multimodal brain-imaging data, demographic variables, and comprehensive cognitive test results.

The analysis in this study focused on a subset of the UKBB dataset consisting of 37,441 healthy participants having access to both cognitive test results and brain imaging data. Subjects in the sample were aged between 40 and 70 (mean 54.18 and standard deviation 7.40). Moreover, 45.14% of them were male and 54.86% female. We focused on two types of brain measures from the imaging data: (1) the microstructural integrity of white matter, derived from diffusion tensor imaging (DTI) data and (2) cortical thicknesses calculated using MRI T1 data. We processed the DTI data by following the ENIGMA-DTI analysis pipeline (Kochunov et al., 2015; Zhao et al., 2021). We calculated the fractional anisotropy of 40 white matter tracts to measure the condition of microstructures (i.e., myelin sheaths and axonal cell membranes) in local brain white matter. In addition, we used FreeSurfer in a cortical reconstruction (<https://surfer.nmr.mgh.harvard.edu>) to extract cortical thicknesses from 34 cortical gray matter regions in each hemisphere defined according to the DesikanKilliany atlas (Desikan et al., 2006).

The cognitive function of the study participants was assessed by the various domains of cognitive tests, including processing speed, memory, perceptual reasoning, executive function, and fluid intelligence. It has been established in the neuroscience literature that the g factor, a measure of general intelligence (calculated by a factor analysis of multiple cognitive tests), has the strongest association with brain aging (Hoogendam et al., 2014). Therefore, we considered the g factor to be the outcome variable and adjusted confounding covariates, including sex, body mass index, and education level.

3.2. *MMP analysis results.* We applied MMP to the aforementioned dataset. We optimized the objective function in (3) to estimate ω and \mathbf{M}_* ($\lambda = 0.00097$). Figure 2 shows the regularization paths for neuroimaging variables. The results indicate that the nonzero ω values, which reflect the contributions of neuroimaging measures to the age-related cognitive decline, are all derived from white-matter integrity measures. The MP obtained by the estimated \mathbf{M}_* was 96.93%, which suggests that age-related cognitive decline can be almost completely explained by white-matter integrity measures ($p < 0.001$).

Our results show that 30 brain regions of white matter integrity measures contribute to age-related cognitive decline. The identified brain regions are displayed and listed in the Supplementary Material (Lee et al., 2024). The selected white matter integrity regions have been frequently discussed in the literature on cognitive neuroscience. For example, previous research has demonstrated a correlation between the cingulum and cognitive function (Lin et al., 2014). Cingulum is a white matter tract that conveys information from the cingulate gyrus, a brain region involved in cognitive and emotional processes, to the hippocampus, which primarily governs memory and learning (Bettio, Rajendran and Gil-Mohapel, 2017). Consistent with this, the MMP method identifies the cingulum in both the cingulate gyrus regions (CGC.L, CGC.R) and the hippocampal regions (CHG.L, CHG.R). Moreover, all brain regions (Body, Genu, and Splenium) in the corpus callosum (CC) were selected. Even though

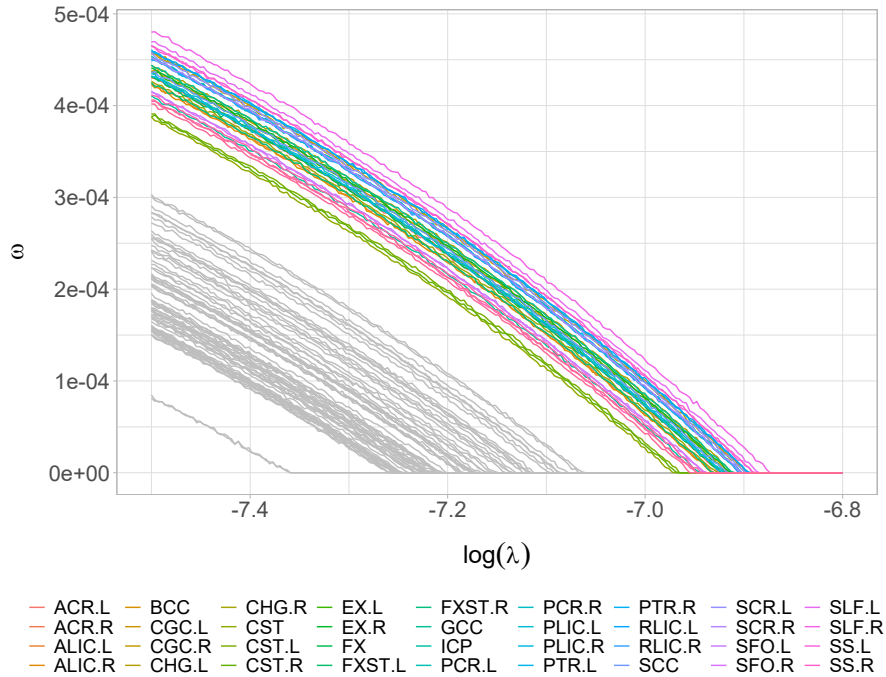


Fig 2: The regularization paths for weight vector of neuroimaging measures: weights shrink toward zero when λ increases. Gray matter cortical thickness measurements are depicted in gray, while white matter integrity measurements are represented using various colors. The weights associated with white matter integrity consistently surpass those of cortical thickness. This implies that white matter tracts are more likely to be preserved across different tuning parameter choices, which indicates that white matter plays a more significant role than cortical thickness in explaining age-induced cognitive decline.

the precise role of the corpus callosum is unclear, our findings may support the widely hypothesized role of the corpus callosum in cognitive function, especially in integrating and recalling verbal and visual information.

3.2.1. Reproducibility and robustness analysis. To evaluate the robustness of the above results, we carried out an extensive validation analysis. Specifically, we resampled the entire cohort 200 times. For each resample, we randomly selected 10,000 participants as the training set. The remaining data were used as a blind testing dataset. We set the tuning parameters and calculated the MMP parameters using only the training data. Then, we locked these parameters and applied them to the testing dataset. We evaluated the MP by MMP using the testing dataset. As shown in Figure 3, the median of the MP for the independent testing dataset was 0.9218 (the first and third quantiles were 0.8934, and 0.9527, respectively). Therefore, our MMP approach can robustly explain a high MP for the independent testing datasets. These results confirm that age-related cognitive decline is mediated by the white-matter integrity measures.

To assess the consistency of the proposed model in selecting mediators from resampled data, selection probabilities for each brain region are presented in the Supplementary Material (Lee et al., 2024). The selection probabilities of FAs are generally high, with 18 FAs having a probability of 1, and another 19 FAs having probabilities between 0.5 and 1. All 18 brain regions, which have a selection probability of 1, are a subset of the previously selected 30 regions in the analysis of the entire set of observations. We display the 18 brain regions in

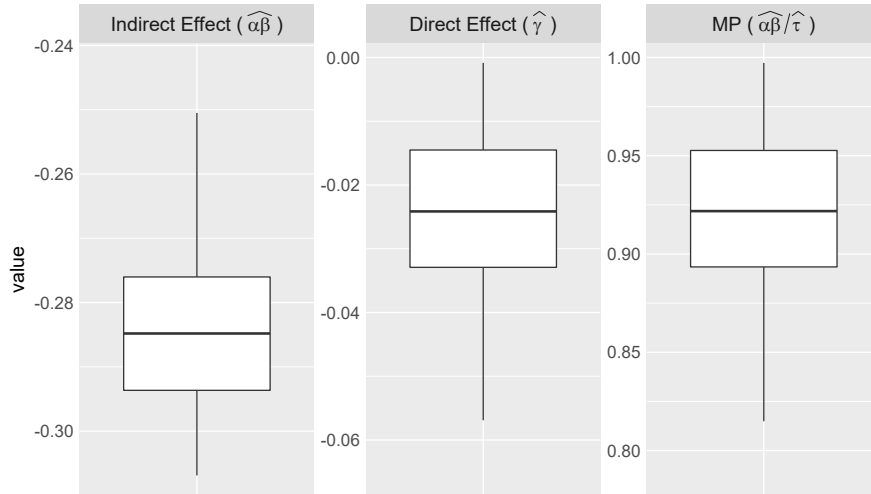


Fig 3: Reproducibility analysis results. We demonstrate the performance of the MMP-based mediation analysis based on 200 testing datasets. The three boxplots represent the indirect effect (left), the direct effect (middle), and the mediation proportion (right). The aggregate mediator is estimated based on the training datasets. The median of MPs from the testing datasets remains above 92%.

Figure 4. In contrast, the selection probabilities of all CTs are close to zero with the highest probability equal to 2%. Therefore, the selected mediators are highly consistent across all resamples.

We performed additional analyses to evaluate i) whether FA measures were preferentially selected over CT measures due to varying signal-to-noise ratios (SNR), and ii) the estimated MP of 96.93% is over-inflated by our model. First, we calculated SNRs for both FA and CT measures, and compared their mean differences, finding no significant difference between them (see the Supplementary Material (Lee et al., 2024) for further details). We next conducted separate mediation analyses using only FA measures and only CT measures, respectively. The analyses revealed that the MP is only 9.45% when using CTs alone as mediators, in contrast to a significant increase to 94.40% with FA mediators alone. The notably low MP value obtained from CT mediation indicates that our method does not artificially inflate the MP. Furthermore, the high MP of 94.40% obtained solely from FA mediators closely aligns with the 96.93% reported in our main findings. Moreover, the identified white matter tracts from the FA-only mediation analysis, detailed in the Supplementary Material (Lee et al., 2024), show high consistency with those listed in the main findings. Finally, we conducted an additional mediation analysis employing the method proposed by Zhang et al. (2016). The results showed that i) 25 neuroimaging mediators, all derived from FA measures, were selected, and ii) the MP-value is 92.96%. These results closely align with our main findings.

3.2.2. Remarks. Both cortical thickness and white-matter integrity have a downward trend during the aging process (Kochunov et al., 2011). Decreases in both cortical thickness and white-matter integrity are associated with the progressive loss of cognitive performance. Therefore, performance on cognitive tests deteriorates with age for older people.

Our findings suggest that white matter integrity appears to play a more substantial role in mediating age-related cognitive decline compared to cortical thickness. Our results are well aligned with previous neurobiological findings that cognitive function is more attributable

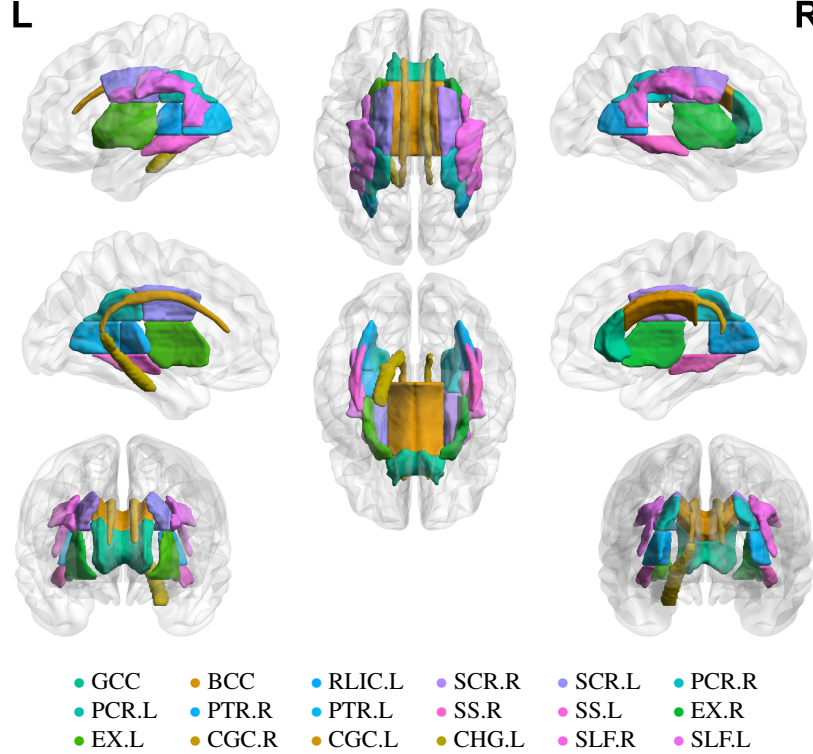


Fig 4: The 18 brain regions that are consistently selected from every resampling in [Section 3.2.1](#) are displayed. The abbreviations of the white matter tracts (regions) are also given. The color code for each region corresponds to that used in [Figure 2](#). The complete names of the region abbreviations can be found in the Supplementary Material ([Lee et al., 2024](#)).

to the degeneration of white matter than cortical thinning in the elderly population “*Cognition in healthy aging is related to regional white matter integrity, but not cortical thickness*” ([Ziegler et al., 2010](#)). Additionally, in the study conducted by [Hedden et al. \(2014\)](#), when simultaneously considering different neuroimaging measures, including FA and CT, obtained from various imaging modalities as mediators, FA was found to be the most significant mediator explaining age-related cognition.

4. Simulation. In this section, using simulated data, we evaluate the performance of our proposed mediation model and benchmark it against existing methods. We consider two settings. In Setting 1, we investigate how well our method identifies the active mediators and maximally uncovers the mediation effect, when the mediation pathway dominates the relation. In Setting 2, we simulate data with none and medium mediation effects, and evaluate if our method yields an overinflated estimate of the mediation proportion.

4.1. *Uncovering the mediation pathway: identifying active mediators and MP estimation.* We focused on the setting where the effect of exposure on the outcome is completely mediated by multiple mediators. Specifically, we generated the univariate exposure variable $\mathbf{X} \sim \text{Normal}(0, \sigma^2)$ and the multiple mediator $\mathbf{M} = \mathbf{X}\boldsymbol{\alpha}^\top + \mathbf{E}$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q, 0, \dots, 0)^\top \in \mathbb{R}^p$ and $\text{vec}(\mathbf{E}) \sim \text{Normal}(\mathbf{0}_{np}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$. Then, we generated the outcome variable as $\mathbf{Y} = \mathbf{M}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r, 0, \dots, 0)^\top \in \mathbb{R}^p$ and $\boldsymbol{\varepsilon} \sim \text{Normal}(0, \sigma^2)$.

We considered sample sizes of $n \in \{100, 200\}$, and varied the dimension of the mediators $p \in \{20, 200\}$. We set $q = 5$ and $r = 6$ to reflect the number of active mediators. The magnitude of the non-zero coefficients was set to $\{\alpha_j\}_{j=1}^q = \{\beta_j\}_{j=1}^r = -0.5$. We incorporated a correlation structure between the mediators because brain-imaging mediators in our motivating data example are correlated. Specifically, we set $\Sigma = \mathbf{I}_p + \mathbf{R}$, where the (i, j) th element of \mathbf{R} is given by $\mathbf{R}_{ij} = \rho \cdot \mathbb{I}(i \neq j)$. We used $\rho = \{0, 0.25, 0.5\}$. Finally, we fixed $\sigma^2 = 1$ across all settings. Each setting was replicated 100 times.

We applied the MMP and other comparable methods to the simulated datasets. These methods include the screening based two-stage method (SIS+MCP: Zhang et al., 2016) as implemented in the HIMA package, Directions of Mediation (DMs: Chén et al., 2018), the Sparse PCA-based method (SPCMA: Zhao, Lindquist and Caffo, 2020), and Pathway Lasso (Zhao and Luo, 2022). Like MMP, DMs use a linear combination of the neuroimaging mediators to reduce the number of high-dimensional mediators. However, the weights of DMs are determined by maximizing the LSEM likelihood rather than directly maximizing the MP. In addition, instead of relying on a single score, DMs makes use of k orthogonally transformed mediators. In this simulation study, we consider the DMs for $k = 1$, and 2 as suggested by the algorithm. SPCMA also bears resemblance to MMP in its use of sparse linear combinations of original mediators for reducing dimensionality. However, while MMP seeks a weight vector ω that maximally recovers of the mediation effect, SPCMA aims to identify multiple mediation pathways using orthogonal mediators. In contrast to the linear combination-based methods (MMP, DMs, and SPCMA), pathway lasso directly imposes penalty terms on the regression coefficients in (1), which has the following form: $\lambda \left[\sum_{j=1}^p (|\alpha_j \beta_j| + \phi(\alpha_j^2 + \beta_j^2)) + \gamma \right] + \psi \left[\sum_{j=1}^p (|\alpha_j| + |\beta_j|) \right]$. Here $\lambda, \phi, \psi \geq 0$ play a crucial role in tuning the model. Note that the implementation of pathway lasso used in this simulation was not fully optimized. We tuned only λ whereas the other two parameters were constant: $\phi = 2$ and $\psi = 0$. We also considered the special case $\lambda = 0$, known as the two-stage (TS) lasso penalty, which shrinks mediation pathways separately for each α_j and β_j .

For evaluation measures, we first considered the MP to assess whether our proposed method was well-optimized. Note that in these experiments, we emulated the full mediation mechanism by generating the outcome from exposure only through multiple mediators. Nevertheless, the true MP cannot be directly prespecified, since the exposure and mediators were correlated. However, since intuitively a higher value is more likely to reflect the full mediation mechanism, it was also considered as a measure when assessing the performance of a model. In addition, to evaluate the performance of active mediator selection, we use the comparison criteria of accuracy = $\frac{TP+TN}{TP+FP+TN+FN}$, precision = $\frac{TP}{TP+FP}$, recall = $\frac{TP}{TP+FN}$, and $F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$ where TP, TN, FP , and FN denote the true positive, true negative, false positive, and false negative count, respectively.

The results for sample size $n = 100$, and $n = 200$ are summarized in Table 1, and Table 2, respectively. First, our proposed MMP approach outperformed the comparable methods in terms of the MP, achieving values almost reaching 1. This indicates that the exposure and outcome relation can be maximally explained by the mediated pathway. In addition, our proposed method can accurately select informative mediators, attaining the highest performance for roughly 3 out of the 4 variable selection measures. In particular, we focused on the F_1 score, which is based on both precision and recall, for a small proportion of the informative mediators. The results show that our method outperformed the other methods across all settings. When the number of mediators was increased to $p = 200$, there were false positive mediators in the models for pathway lasso and TS lasso, as evidenced by the low precision values. As suggested by Chén et al. (2018), we only apply DMs to the settings with $p < n$

TABLE 1

Results ($n = 100$) of the experiment described in Section 4.1 which compare the performance of active mediator selection and the mediation proportion. All evaluation measures reported in the table are averaged over the 100 replications, providing an overall estimate of the method's performance across various scenarios (e.g., different p - dimension of the multiple mediators, and ρ - correlation among the multiple mediators).

		$n = 100$						
	p	Method	Accuracy	Precision	Recall	F_1	M.P.	
$\rho = 0$	20	MMP	0.9475	0.8370	1.0000	0.9085	0.9650	
		SIS+MCP	0.9095	0.7536	0.9880	0.8504	0.9197	
		DMs1	0.2500	0.2500	1.0000	0.4000	0.8830	
		DMs2	0.2500	0.2500	1.0000	0.4000	0.9476	
		SPCMA	0.7485	0.5012	0.7480	0.5902	0.4393	
		Pathway	0.8995	0.9072	0.7060	0.7776	0.0097	
		TS	0.7595	0.5357	0.4060	0.4372	0.0146	
	200	MMP	0.9964	0.9052	0.9980	0.9427	0.9501	
		SIS+MCP	0.9532	0.4019	0.9780	0.5513	0.8314	
		SPCMA	0.8558	0.1188	0.7380	0.2037	0.5274	
		Pathway	0.8344	0.0018	0.0120	0.0031	0.0005	
		TS	0.9776	0.5539	1.0000	0.7055	0.0130	
		$\rho = 0.25$	MMP	0.9925	0.9821	0.9920	0.9857	0.9938
			SIS+MCP	0.9125	0.7633	0.9820	0.8543	0.9436
DMs1	0.2500		0.2500	1.0000	0.4000	0.8680		
DMs2	0.2500		0.2500	1.0000	0.4000	0.9589		
SPCMA	0.7505		0.5006	1.0000	0.6671	0.7657		
Pathway	0.6860		0.4833	0.7220	0.5538	0.0167		
TS	0.5845		0.2781	0.4100	0.3239	0.0168		
200	MMP	0.9966	0.9105	0.9920	0.9436	0.9979		
	SIS+MCP	0.9794	0.5871	0.9560	0.7168	0.8650		
	SPCMA	0.8268	0.1195	0.9340	0.2118	0.6714		
	Pathway	0.3088	0.0006	0.0140	0.0011	0.0026		
	TS	0.5644	0.0629	1.0000	0.1172	0.0182		
	$\rho = 0.5$	MMP	0.9990	1.0000	0.9960	0.9978	0.9990	
		SIS+MCP	0.8945	0.7396	0.9200	0.8149	0.9314	
DMs1		0.2500	0.2500	1.0000	0.4000	0.7620		
DMs2		0.2500	0.2500	1.0000	0.4000	0.9586		
SPCMA		0.7465	0.4964	0.9920	0.6617	0.7643		
Pathway		0.5835	0.3749	0.7820	0.4980	0.0222		
TS		0.5835	0.2619	0.3180	0.2722	0.0171		
200	MMP	0.9994	0.9805	1.0000	0.9889	0.9985		
	SIS+MCP	0.9776	0.5562	0.7920	0.6454	0.7247		
	SPCMA	0.8305	0.1125	0.8440	0.1982	0.5763		
	Pathway	0.1018	0.0014	0.0460	0.0026	0.0059		
	TS	0.0900	0.0268	1.0000	0.0522	0.0215		

to ensure fair comparisons. In addition, since the loadings of factors by DMs may involve all mediators, we did not prioritize comparing the mediator selection criteria of DMs. Overall, our method exhibited robust performance across various settings. Therefore, our proposed method may be an effective tool for selecting active mediators for mediation analysis with a complete mediation effect. In summary, MMP provides a robust way for maximizing the mediation effect and active mediator selection when dealing with multiple mediators.

4.2. *Assessing MP inflation under settings of none and medium mediation effects*. Since our objective function seeks to maximize the mediation proportion, we perform additional

TABLE 2

Results ($n = 200$) of the experiment described in Section 4.1 which compare the performance of active mediator selection and the mediation proportion. All evaluation measures reported in the table are averaged over the 100 replications, providing an overall estimate of the method's performance across various scenarios (e.g., different p - dimension of the multiple mediators, and ρ - correlation among the multiple mediators).

$n = 200$							
p	Method	Accuracy	Precision	Recall	F_1	M.P.	
$\rho = 0$	20	MMP	0.9875	0.9600	0.9980	0.9771	0.9530
		SIS+MCP	0.9345	0.8017	0.9980	0.8871	0.9650
		DMs1	0.2500	0.2500	1.0000	0.4000	0.9340
		DMs2	0.2500	0.2500	1.0000	0.4000	0.9718
		SPCMA	0.7465	0.4809	0.5620	0.4919	0.4489
		Pathway	0.9705	0.9950	0.8880	0.9255	0.0059
		TS	0.9320	0.9095	0.8900	0.8770	0.0035
	200	MMP	0.9946	0.8427	1.0000	0.9094	0.9736
		SIS+MCP	0.9867	0.7024	1.0000	0.8141	0.9389
		DMs1	0.0250	0.0250	1.0000	0.0488	0.9196
		DMs2	0.0250	0.0250	1.0000	0.0488	0.9358
		SPCMA	0.8458	0.1224	0.8460	0.2134	0.5070
		Pathway	0.9981	0.9983	0.9260	0.9551	0.0043
		TS	0.9882	0.8666	0.9040	0.8489	0.0034
$\rho = 0.25$	20	MMP	0.9830	0.9448	1.0000	0.9695	0.9998
		SIS+MCP	0.9400	0.8107	1.0000	0.8944	0.9673
		DMs1	0.2500	0.2500	1.0000	0.4000	0.8999
		DMs2	0.2500	0.2500	1.0000	0.4000	0.9715
		SPCMA	0.7470	0.4924	0.8540	0.6191	0.7182
		Pathway	0.9670	0.9206	0.9880	0.9461	0.0053
		TS	0.9585	0.9334	0.9840	0.9467	0.0032
	200	MMP	0.9990	0.9723	1.0000	0.9836	0.9988
		SIS+MCP	0.9906	0.7510	1.0000	0.8522	0.9536
		DMs1	0.0250	0.0250	1.0000	0.0488	0.8450
		DMs2	0.0250	0.0250	1.0000	0.0488	0.9371
		SPCMA	0.8250	0.1250	1.0000	0.2222	0.6400
		Pathway	0.9896	0.8419	0.9520	0.8732	0.0049
		TS	0.9948	0.9134	0.9860	0.9349	0.0028
$\rho = 0.5$	20	MMP	0.9975	0.9917	1.0000	0.9955	1.0000
		SIS+MCP	0.9360	0.8034	0.9980	0.8886	0.9678
		DMs1	0.2500	0.2500	1.0000	0.4000	0.8843
		DMs2	0.2500	0.2500	1.0000	0.4000	0.9739
		SPCMA	0.7650	0.5262	0.8600	0.6486	0.7715
		Pathway	0.8585	0.7139	1.0000	0.8127	0.0082
		TS	0.6945	0.5271	0.9980	0.6646	0.0043
	200	MMP	1.0000	0.9983	1.0000	0.9991	1.0000
		SIS+MCP	0.9892	0.7279	0.9840	0.8305	0.9283
		DMs1	0.0250	0.0250	1.0000	0.0488	0.7615
		DMs2	0.0250	0.0250	1.0000	0.0488	0.9556
		SPCMA	0.8250	0.1248	0.9980	0.2219	0.6194
		Pathway	0.9005	0.4256	1.0000	0.5395	0.0110
		TS	0.8099	0.2861	0.9960	0.3867	0.0035

TABLE 3

The table presents the comparison of absolute bias in mediation proportion (MP) estimates obtained from different methods. The absolute bias values are reported as averages over 100 replications. The values indicate the degree of deviation of the estimates from the true values, with lower values indicating less bias. The standard deviation of the MP estimates is shown in parentheses.

Scenario 1 (Noninformative Mediator : MP=0)								
n	p	MMP	SIS+MCP	DMS1	DMS2	SPCA	Pathway	TS
100	10	0.017 (0.02)	0.018 (0.02)	0.046 (0.05)	0.046 (0.05)	0.043 (0.04)	0.018 (0.02)	0.015 (0.01)
	50	0.024 (0.03)	0.027 (0.03)	0.140 (0.11)	0.139 (0.11)	0.157 (0.15)	0.018 (0.02)	0.024 (0.03)
200	10	0.010 (0.01)	0.010 (0.01)	0.023 (0.02)	0.023 (0.02)	0.019 (0.02)	0.011 (0.01)	0.009 (0.00)
	50	0.013 (0.01)	0.011 (0.01)	0.069 (0.05)	0.069 (0.05)	0.057 (0.04)	0.010 (0.01)	0.010 (0.01)
Scenario 2 (Informative Mediator: MP=0.5)								
n	p	MMP	SIS+MCP	DMS1	DMS2	SPCA	Pathway	TS
100	10	0.006 (0.01)	0.005 (0.01)	0.020 (0.02)	0.020 (0.02)	0.019 (0.02)	0.055 (0.03)	0.052 (0.02)
	50	0.015 (0.07)	0.020 (0.03)	0.076 (0.06)	0.072 (0.06)	0.069 (0.05)	0.232 (0.21)	0.080 (0.02)
200	10	0.002 (0.00)	0.001 (0.00)	0.011 (0.01)	0.011 (0.01)	0.010 (0.01)	0.075 (0.02)	0.038 (0.01)
	50	0.010 (0.05)	0.003 (0.01)	0.026 (0.02)	0.026 (0.02)	0.028 (0.02)	0.076 (0.04)	0.069 (0.02)

simulations to evaluate whether MMP artificially amplifies the underlying mediation proportion. We investigated two scenarios: (i) the absence of any mediation effect, where none of the mediators play a role in the relationship between the exposure and outcome variables (i.e., the MP is zero), (ii) the case of partial mediation, where the mediator has a partial effect on the relationship between the exposure and outcome variables (i.e., the MP is around 0.5). We generated the exposure, outcome, and mediator simultaneously using a precision matrix. Specifically, let Θ_{ij} represent the (i, j) th element of the precision matrix $\Theta = \Sigma^{-1}$ of $(\mathbf{X}, \mathbf{Y}, \mathbf{M}) \in \mathbb{R}^{p+2}$, where Σ is the covariance matrix. It is worth noting that the partial correlation between the i th and j th variables can be obtained by $r_{ij} = -\Theta_{ij} / \sqrt{\Theta_{ii}\Theta_{jj}}$. In both scenarios, we specified the direct effect by setting $r_{\mathbf{X}\mathbf{Y}} = 0.5$. In the first scenario, we assumed that all mediators are not correlated with either the exposure or outcome variables ($r_{\mathbf{X}\mathbf{M}_j} = r_{\mathbf{Y}\mathbf{M}_j} = 0$ for all $j = 1, \dots, p$), resulting in a zero indirect effect. In the second scenario, we assumed that the mediators are equally and partially correlated with both the exposure and outcome variables (e.g., $r_{\mathbf{X}\mathbf{M}_j} = r_{\mathbf{Y}\mathbf{M}_j} = 0.5$). Note that in this case, it is challenging to directly specify the mediation proportion due to the limitation of controlling the marginal correlations between the mediators and the exposure variable using the precision matrix. Under our model parameter specification, the mediation proportions vary around 0.5 (see Table 3).

The comparison results of estimating MP are provided in Table 3, where the absolute bias is averaged over 100 replications. In both scenarios, the estimated mediation proportions by MMP are accurate with no inflation, and our method demonstrates comparable performance to the screening-based method and outperforms the shrinkage methods. In the scenario of partial mediation (Scenario 2), the methods based on direct shrinkage (TS, and pathway lasso) exhibit an underestimation of the mediation proportion (see Figure in the Supplementary Material (Lee et al., 2024)). More detailed results and inference, including the estimation of each parameter and corresponding p -values, are provided in the Supplementary Material (Lee et al., 2024).

4.3. *Assessing mediator selection under different Signal-to-Noise Ratios (SNR).* This simulation study stems from our real data application involving two sets of multiple neuroimaging measures from distinct imaging modalities, where we observed that the nonzero elements of ω were predominantly linked to white matter integrity. Given the potential hetero-

geneity in the Signal-to-Noise Ratio (SNR) characteristics across different imaging modalities, this variability could impact the reliability and robustness of active mediator selection. For example, there is a possibility that methods could introduce a bias in the selection of mediators, favoring those from imaging measures associated with a modality that has a higher SNR. To investigate whether this issue occurs within our model, we conducted an additional simulation study based on the following scenario.

We consider two sets of multiple mediators, \mathbf{M}_1 and \mathbf{M}_2 , each characterized by distinct SNR levels. Specifically, these mediators are sampled from multivariate normal distributions with mean vectors set to zero and differing covariance matrices: Σ_1 for \mathbf{M}_1 and Σ_2 for \mathbf{M}_2 , where $\Sigma_1 \in \mathbb{R}^{p_1 \times p_1}$ and $\Sigma_2 \in \mathbb{R}^{p_2 \times p_2}$. For the sake of the simplicity, we employ diagonal covariance matrices of the same dimensions ($p_1 = p_2$), with distinct variances σ_1^2 and σ_2^2 for Σ_1 and Σ_2 , respectively. The $p = p_1 + p_2$ dimensional combined multiple mediator \mathbf{M} is then constructed by column-wise concatenation of \mathbf{M}_1 and \mathbf{M}_2 , denoted as $\mathbf{M} = [\mathbf{M}_1 | \mathbf{M}_2]$. Within each set of mediators, we evaluate the performance of MMP in selecting active mediators and compare it with four other methods specifically designed for sparse mediator selection: SIS+MCP, Pathway, TS, and SPCMA.

The results based on 100 replications are summarized in Table 4. The MMP method generally outperforms competing methods in both sets of mediators. This indicates a robust ability to identify active mediators without introducing selection bias between the two sets, i.e., \mathbf{M}_1 and \mathbf{M}_2 . Additionally, this further confirms the validity of the preferred selection of FA measures over CT measures in the real data analysis.

5. Discussion. In this paper, we introduce a novel mediation model with multiple mediators within the LSEM framework, aiming to fully elucidate the effect of exposure on the outcome through mediation effects. Our model development was driven by the practical challenge in neuroscience to understand the neural mechanisms explaining age-related cognitive decline. While existing multivariate mediation methods aim to maximize the likelihood or identify sparse mediation pathways, MMP focuses on fully uncovering the aggregate mediation effect with parsimoniously selected mediators. We developed a new objective function with a quadratic form so that the optimization does not involve non-identifiable parameters. We also devised new algorithms to facilitate efficient implementation of the optimization. The proposed method achieved excellent performance in our data example and simulation analysis.

We applied our method to analyze a large cohort of multimodal neuroimaging data from UKBB. The results revealed the importance of white-matter integrity for the cognitive decline during the aging process. We hope that the findings will provide new directions for neuroscience research, as the identified white-matter tracts may serve as potential target areas for developing new treatments and interventions aimed at mitigating accelerated cognitive aging. The estimated aggregate mediator as a linear projection of neuroimaging variables, reliably explains a remarkably high proportion of the age effect on cognitive performance. Furthermore, the estimated aggregate mediator on its own offers a function-specific (e.g., cognitive performance and sensorimotor control) measure of brain aging, rather than an overall brain aging. In addition, as an objectively measured (neuroimaging-based) metric, it may become a more reliable alternative to subjective cognitive test scores that are currently widely used. The objective measure may reduce the measurement errors due to subjective evaluations of cognitive performance. It may also be useful in clinical investigations of new treatments for neurodegenerative disorders as a more accurate primary endpoint.

The proposed method is limited to identifying unknown causal orderings (pathways) among mediators. The extension of our method to address this aspect presents an intriguing and promising avenue for future research.

TABLE 4

Result ($n = 100$) of active mediator selection in the presence of two blocks of mediators with different SNRs, as described in Section 4.3. The table presents averaged metrics over 100 replications, including Accuracy, Precision, Recall, and F_1 Score, to evaluate the performance of active mediator selection under scenarios with different dimensions (p) and Signal-to-Noise Ratios (SNR) for each mediator block.

p	Method	First block ($\sigma_1^2 = 1$)				Second block ($\sigma_2^2 = 2$)			
		Accuracy	Precision	Recall	F_1	Accuracy	Precision	Recall	F_1
40	MMP	0.9110	0.7608	0.9920	0.8552	0.9060	0.7434	1.0000	0.8484
	SIS+MCP	0.8960	0.7306	0.9820	0.8318	0.9090	0.7491	1.0000	0.8525
	SPCMA	0.7455	0.5008	0.7540	0.5928	0.7420	0.4985	0.9640	0.6544
	Pathway	0.7420	0.5062	1.0000	0.6678	0.6970	0.4667	1.0000	0.6317
	TS	0.8665	0.6868	1.0000	0.8051	0.7945	0.5894	1.0000	0.7295
80		First block ($\sigma_1^2 = 1$)				Second block ($\sigma_2^2 = 2$)			
	MMP	0.9488	0.7489	0.9800	0.8382	0.9455	0.7197	0.9880	0.8269
	SIS+MCP	0.8580	0.5496	0.9620	0.6772	0.8582	0.5451	0.9980	0.6844
	SPCMA	0.6630	0.2516	0.8480	0.3865	0.6350	0.2532	0.9720	0.4011
	Pathway	0.6567	0.2762	0.9960	0.4297	0.6132	0.2499	1.0000	0.3983
TS	0.6120	0.2562	0.9960	0.4037	0.5420	0.2203	1.0000	0.3596	
40		First block ($\sigma_1^2 = 1$)				Second block ($\sigma_2^2 = 4$)			
	MMP	0.9200	0.7846	0.9880	0.8682	0.9170	0.7621	1.0000	0.8621
	SIS+MCP	0.9010	0.7426	0.9800	0.8381	0.9145	0.7580	1.0000	0.8590
	SPCMA	0.7495	0.5120	0.6182	0.5375	0.7420	0.5000	0.9820	0.6594
	Pathway	0.7405	0.5072	1.0000	0.6679	0.6820	0.4535	0.9980	0.6193
TS	0.6740	0.4469	1.0000	0.6138	0.4965	0.3372	0.9980	0.5023	
80		First block ($\sigma_1^2 = 1$)				Second block ($\sigma_2^2 = 4$)			
	MMP	0.9535	0.7662	0.9860	0.8519	0.9440	0.7175	0.9980	0.8273
	SIS+MCP	0.8588	0.5579	0.9740	0.6849	0.8605	0.5557	1.0000	0.6925
	SPCMA	0.6958	0.2607	0.7340	0.3756	0.6362	0.2552	0.9840	0.4049
	Pathway	0.6370	0.2644	0.9960	0.4155	0.5860	0.2354	0.9960	0.3798
TS	0.4940	0.2053	0.9960	0.3384	0.3440	0.1619	0.9980	0.2781	

To summarize, the newly introduced MMP allows for prioritizing the mediation pathway while selecting active mediators. While our method is developed with a specific focus on the brain’s role in cognitive aging, it offers advantages in studying complex systems where high-throughput mediators play a key role in mediating the effect of exposure on the outcome (e.g., gut microbiome mediating the antibiotic treatment on nutrient absorbing, Wang et al. (2019)). In addition, the efficient optimization algorithms utilized in MMP make it highly scalable. We provide a user-friendly software package at <https://github.com/hwiyoungstat/MMP>, and also included in the Supplementary Material (Lee et al., 2024).

Funding. This research was partly funded by National Institutes of Health (NIH) grant 1DP1DA04896801, EB008432, and EB008281.

Data Availability Statement. The raw UKBB data used in this study can be accessed via <https://www.ukbiobank.ac.uk>.

SUPPLEMENTARY MATERIAL

Supplement to “A New Multiple-mediator Model Maximally Uncovering the Mediation Pathway: Evaluating the Role of Neuroimaging Measures in Age-Related Cognitive

Decline”

This document contains a detailed derivation of [Algorithm 1](#), the proof of [Theorem 2.1](#), and additional results of real data applications and simulation studies.

R package

This file contains the source code required to implement the proposed method. Future updates will be hosted on GitHub at <https://github.com/hwiyoungstat/MMP>.

REFERENCES

- ANDREWS, R. M. and DIDELEZ, V. (2021). Insights into the Cross-world Independence Assumption of Causal Mediation Analysis. *Epidemiology* **32** 209–219. <https://doi.org/10.1097/EDE.0000000000001313>
- BARON, R. M. and 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** 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- BECK, A. (2017). *First-Order Methods in Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA. <https://doi.org/10.1137/1.9781611974997>
- BETHLEHEM, R. A. I., SEIDLITZ, J., WHITE, S. R. and ET AL. (2022). Brain charts for the human lifespan. *Nature* **604** 525–533. <https://doi.org/10.1038/s41586-022-04554-y>
- BETTIO, L. E. B., RAJENDRAN, L. and GIL-MOHAPEL, J. (2017). The effects of aging in the hippocampus and cognitive decline. *Neuroscience and Biobehavioral Reviews* **79** 66–86. <https://doi.org/10.1016/j.neubiorev.2017.04.030>
- BOYD, S., PARIKH, N., CHU, E., PELEATO, B. and ECKSTEIN, J. (2011). Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.* **3** 1122. <https://doi.org/10.1561/22000000016>
- BUTLER, E. R., CHEN, A., RAMADAN, R., LE, T. T., RUPAREL, K., MOORE, T. M., SATTERTHWAITTE, T. D., ZHANG, F., SHOU, H., GUR, R. C. et al. (2021). Pitfalls in brain age analyses. *Human Brain Mapping* **42** 4092.
- CHÉN, O. Y., CRAINICEANU, C., OGBURN, E. L., CAFFO, B. S., WAGER, T. D. and LINDQUIST, M. A. (2018). High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics* **19** 121–136. <https://doi.org/10.1093/biostatistics/kxx027>
- COLE, J. H. and FRANKE, K. (2017). Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends in Neurosciences* **40** 681–690. <https://doi.org/10.1016/j.tins.2017.10.001>
- DESIKAN, R. S., SÉGONNE, F., FISCHL, B., QUINN, B. T., DICKERSON, B. C., BLACKER, D., BUCKNER, R. L., DALE, A. M., MAGUIRE, R. P., HYMAN, B. T., ALBERT, M. S. and 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** 968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>
- FRANKE, K. and GASER, C. (2019). Ten Years of BrainAGE as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained? *Frontiers in Neurology* **10**. <https://doi.org/10.3389/fneur.2019.00789>
- GRADY, C. L. (1998). Brain imaging and age-related changes in cognition. *Experimental Gerontology* **33** 661–673. [https://doi.org/10.1016/S0531-5565\(98\)00022-9](https://doi.org/10.1016/S0531-5565(98)00022-9)
- GRADY, C. L. (2000). Functional brain imaging and age-related changes in cognition. *Biological Psychology* **54** 259–281. [https://doi.org/10.1016/S0301-0511\(00\)00059-4](https://doi.org/10.1016/S0301-0511(00)00059-4)
- HEDDEN, T., SCHULTZ, A. P., RIECKMANN, A., MORMINO, E. C., JOHNSON, K. A., SPERLING, R. A. and BUCKNER, R. L. (2014). Multiple Brain Markers are Linked to Age-Related Variation in Cognition. *Cerebral Cortex* **26** 1388–1400. <https://doi.org/10.1093/cercor/bhu238>
- HOOGENDAM, Y. Y., HOFMAN, A., VAN DER GEEST, J. N., VAN DER LUGT, A. and IKRAM, M. A. (2014). Patterns of cognitive function in aging: the Rotterdam Study. *European journal of epidemiology* **29** 133–140.
- IMAI, K., KEELE, L. and TINGLEY, D. (2010). A general approach to causal mediation analysis. *Psychological Methods* **15** 309–334. <https://doi.org/10.1037/a0020761>
- IMAI, K., KEELE, L. and YAMAMOTO, T. (2010). Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science* **25** 51 – 71. <https://doi.org/10.1214/10-STS321>
- KAUP, A. R., MIRZAKHANIAN, H., JESTE, D. V. and EYLER, L. T. (2011). A Review of the Brain Structure Correlates of Successful Cognitive Aging. *The Journal of Neuropsychiatry and Clinical Neurosciences* **23** 6–15. PMID: 21304134. <https://doi.org/10.1176/jnp.23.1.jnp6>
- KOCHUNOV, P., GLAHLN, D. C., LANCASTER, J., THOMPSON, P. M., KOCHUNOV, V., ROGERS, B., FOX, P., BLANGERO, J. and WILLIAMSON, D. E. (2011). Fractional anisotropy of cerebral white matter and thickness of cortical gray matter across the lifespan. *NeuroImage* **58** 41–49. <https://doi.org/10.1016/j.neuroimage.2011.05.050>

- KOCHUNOV, P., JAHANSHAD, N., MARCUS, D., WINKLER, A., SPROOTEN, E., NICHOLS, T. E., WRIGHT, S. N., HONG, L. E., PATEL, B., BEHRENS, T. et al. (2015). Heritability of fractional anisotropy in human white matter: a comparison of Human Connectome Project and ENIGMA-DTI data. *Neuroimage* **111** 300–311. <https://doi.org/10.1016/j.neuroimage.2015.02.050>
- LEE, H., CHEN, C., KOCHUNOV, P., HONG, L. E. and CHEN, S. (2024). Supplement to “A New Multiple-mediator Model Maximally Uncovering the Mediation Pathway: Evaluating the Role of Neuroimaging Measures in Age-Related Cognitive Decline”. <https://doi.org/10.1214/24-AOAS190>[willbeaddedbythetypesetter].
- LIEM, F., VAROQUAUX, G., KYNAST, J., BEYER, F., KHARABIAN MASOULEH, S., HUNTENBURG, J. M., LAMPE, L., RAHIM, M., ABRAHAM, A., CRADDOCK, R. C., RIEDEL-HELLER, S., LUCK, T., LOEFLER, M., SCHROETER, M. L., WITTE, A. V., VILLRINGER, A. and MARGULIES, D. S. (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage* **148** 179–188. <https://doi.org/10.1016/j.neuroimage.2016.11.005>
- LIN, Y.-C., SHIH, Y.-C., TSENG, W.-Y. I., CHU, Y.-H., WU, M.-T., CHEN, T.-F., TANG, P.-F. and CHIU, M.-J. (2014). Cingulum Correlates of Cognitive Functions in Patients with Mild Cognitive Impairment and Early Alzheimers Disease: A Diffusion Spectrum Imaging Study. *Brain Topography* volume **27** 393–402. <https://doi.org/10.1007/s10548-013-0346-2>
- LUO, Z.-Q., MA, W.-K., SO, A. M.-C., YE, Y. and ZHANG, S. (2010). Semidefinite Relaxation of Quadratic Optimization Problems. *IEEE Signal Processing Magazine* **27** 20–34. <https://doi.org/10.1109/MSP.2010.936019>
- MACKINNON, D. P., KRULL, J. L. and LOCKWOOD, C. M. (2000). Equivalence of the Mediation, Confounding and Suppression Effect. *Prevention Science* **1** 173–181. <https://doi.org/10.1023/A:1026595011371>
- MORRISON, J. H. and BAXTER, M. G. (2012). The ageing cortical synapse: hallmarks and implications for cognitive decline. *Nature Reviews Neuroscience* **13** 240–250.
- NIU, X., ZHANG, F., KOUNIOS, J. and LIANG, H. (2020). Improved prediction of brain age using multimodal neuroimaging data. *Human brain mapping* **41** 1626–1643.
- NÄSLUND, J., HAROUTUNIAN, V., MOHS, R., DAVIS, K. L., DAVIES, P., GREENGARD, P. and BUXBAUM, J. D. (2000). Correlation Between Elevated Levels of Amyloid β -Peptide in the Brain and Cognitive Decline. *JAMA* **283** 1571–1577. <https://doi.org/10.1001/jama.283.12.1571>
- PARIKH, N. and BOYD, S. (2014). Proximal Algorithms. *Found. Trends Optim.* **1** 127239. <https://doi.org/10.1561/24000000003>
- PARK, D. C. and REUTER-LORENZ, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annual review of psychology* **60** 173.
- RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66** 688–701. <https://doi.org/10.1037/h0037350>
- SALTHOUSE, T. A. (2011). Neuroanatomical substrates of age-related cognitive decline. *Psychological Bulletin* **137** 753–784. <https://doi.org/10.1037/a0023262>
- SERANG, S., JACOBUCCI, R., BRIMHALL, K. C. and GRIMM, K. J. (2017). Exploratory Mediation Analysis via Regularization. *Structural Equation Modeling: A Multidisciplinary Journal* **24** 733–744. PMID: 29225454. <https://doi.org/10.1080/10705511.2017.1311775>
- TIBSHIRANI, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58** 267–288.
- TIBSHIRANI, R. J. and TAYLOR, J. (2011). The solution path of the generalized lasso. *The Annals of Statistics* **39** 1335 – 1371. <https://doi.org/10.1214/11-AOS878>
- VANDENBERGHE, L. and BOYD, S. (1996). Semidefinite Programming. *SIAM Review* **38** 49–95.
- VANDERWEELE, T. and VANSTEELENDT, S. (2014). Mediation Analysis with Multiple Mediators. *Epidemiologic Methods* **2** 95–115. <https://doi.org/doi:10.1515/em-2012-0010>
- WANG, C., HU, J., BLASER, M. J. and LI, H. (2019). Estimating and testing the microbial causal mediation effect with high-dimensional and compositional microbiome data. *Bioinformatics* **36** 347–355. <https://doi.org/10.1093/bioinformatics/btz565>
- ZHANG, H., ZHENG, Y., ZHANG, Z., GAO, T., JOYCE, B., YOON, G., ZHANG, W., SCHWARTZ, J., JUST, A., COLICINO, E., VOKONAS, P., ZHAO, L., LV, J., BACCARELLI, A., HOU, L. and LIU, L. (2016). Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics* **32** 3150–3154. <https://doi.org/10.1093/bioinformatics/btw351>
- ZHAO, Y., LI, L. and CAFFO, B. S. (2021). Multimodal neuroimaging data integration and pathway analysis. *Biometrics* **77** 879–889. <https://doi.org/10.1111/biom.13351>
- ZHAO, Y., LI, L. and ALZHEIMER’S DISEASE NEUROIMAGING INITIATIVE (2022). Multimodal data integration via mediation analysis with high-dimensional exposures and mediators. *Human Brain Mapping* **43** 2519–2533. <https://doi.org/10.1002/hbm.25800>
- ZHAO, Y., LINDQUIST, M. A. and CAFFO, B. S. (2020). Sparse principal component based high-dimensional mediation analysis. *Computational Statistics & Data Analysis* **142** 106835. <https://doi.org/10.1016/j.csda.2019.106835>

- ZHAO, Y. and LUO, X. (2022). Pathway Lasso: pathway estimation and selection with high-dimensional mediators. *Statistics and Its Interface* **15** 39 – 50. <https://dx.doi.org/10.4310/21-SII673>
- ZHAO, B., LI, T., YANG, Y., WANG, X., LUO, T., SHAN, Y., ZHU, Z., XIONG, D., HAUBERG, M. E., BENDL, J., FULLARD, J. F., ROUSSOS, P., LI, Y., STEIN, J. L. and ZHU, H. (2021). Common genetic variation influencing human white matter microstructure. *Science* **372** 1304. <https://doi.org/10.1126/science.abf3736>
- ZIEGLER, D. A., PIGUET, O., SALAT, D. H., PRINCE, K., CONNALLY, E. and CORKIN, S. (2010). Cognition in healthy aging is related to regional white matter integrity, but not cortical thickness. *Neurobiology of Aging* **31** 1912-1926. <https://doi.org/10.1016/j.neurobiolaging.2008.10.015>
- ZOU, H. and HASTIE, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **67** 301–320.