Network-Assisted Mediation Analysis with High-Dimensional Neuroimaging Mediators

Baoyi Shi[†] BS3141@COLUMBIA.EDU
Department of Biostatistics
Columbia University
New York, NY, USA

Ying Liu[†] YING.LIU@NYSPI.COLUMBIA.EDU
Department of Psychiatry
Columbia University
New York, NY, USA

Shanghong Xie SX2@MAILBOX.SC.EDU
Department of Statistics
University of South Carolina
Columbia, SC, USA

Xi Zhu XI.ZHU@NYSPI.COLUMBIA.EDU
Department of Psychiatry
Columbia University
New York, NY , USA

Yuanjia Wang YW2016@CUMC.COLUMBIA.EDU
Department of Biostatistics
Columbia University
New York, NY, USA

[†] *These authors contributed equally to this work*

Abstract

Mediation analysis is a widely used statistical approach to estimate the causal pathways through which an exposure affects an outcome via intermediate variables, i.e., mediators. In many applications, high-dimensional correlated biomarkers are potential mediators, posing challenges to standard mediation analysis approaches. However, some of these biomarkers, such as neuroimaging measures across brain regions, often exhibit hierarchical network structures that can be leveraged to advance mediation analysis. In this paper, we aim to study how brain cortical thickness, characterized by a star-shaped hierarchical network structure, mediates the effect of maternal smoking on children's cognitive abilities within the adolescent brain cognitive development (ABCD) study. We propose a network-assisted mediation analysis approach based on a conditional Gaussian graphical model to account for the star-shaped network structure of neuroimaging mediators. Within our framework, the joint indirect effect of these mediators is decomposed into the indirect effect through hub mediators and the indirect effects solely through each leaf mediator. This decomposition provides mediator-specific insights and informs efficient intervention designs. Additionally, after accounting for hub mediators, the indirect effects solely through each leaf mediator can be identified and evaluated individually, thereby addressing the challenges of high-

dimensional correlated mediators. In our study, our proposed approach identifies a brain region as a significant leaf mediator, a finding that existing approaches cannot discover.

Keywords: pathway analysis; mediation analysis; brain imaging; mental health; RDoC; ABCD study

1. Introduction

Mediation analysis has been widely applied in biomedical research and social sciences to study the causal pathways through which an exposure affects an outcome via intermediate variables, i.e., mediators. Standard mediation analysis with a single mediator or multiple low-dimensional mediators jointly has been well established (Baron and Kenny, 1986; Robins and Greenland, 1992; Pearl, 2014; VanderWeele, 2015). Let A, M, Y, and C denote the exposure, mediator(s), outcome, and confounder(s), respectively. Under the counterfactual framework, let Y_a and M_a represent the counterfactual values of Y and M, respectively, that would have been observed had A been set to a. Let $Y_{aM_{a^*}}$ denote the counterfactual value of Y that would have been observed had A been set to a and M been set to the counterfactual value M_{a^*}. Then, the total effect (TE) of the exposure on the outcome can be defined and decomposed by TE = $\mathrm{E}[Y_a - Y_{a^*}]$ = NIE+NDE, as shown in Figure 1(a), where NIE = $\mathrm{E}[Y_{aM_a} - Y_{aM_{a^*}}]$ represents the natural indirect effect, NDE = $\mathrm{E}[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}]$ is the natural direct effect, and a^* and a are two reference values of A measuring the change in the exposure. Conceptually, the NIE quantifies the effect of A on Y that operates through M, whereas the NDE is the effect of A on Y independent of M. The following assumptions are required to identify the NIE and NDE: (i) no unmeasured exposure-outcome confounding, i.e., $Y_{am} \perp\!\!\!\perp A|C$; (ii) no unmeasured mediator-outcome confounding, i.e., $Y_{am} \perp\!\!\!\perp M|\{A, C\}$; (iii) no unmeasured exposure-mediator confounding, i.e., $M_a \perp\!\!\!\perp A|C$; (iv) cross-world independence between counterfactual outcomes and mediators, i.e., $Y_{am} \perp\!\!\!\perp M_{a^*}|C$. If the assumptions are satisfied, the NIE and NDE can be identified by the following empirical expressions:

$$\begin{aligned}\text{NIE} = \sum_{c,m} & E[Y|C=c, A=a, M=m]\{P(M=m|C=c, A=a) - \\ & P(M=m|C=c, A=a^*)\}P(C=c),\end{aligned}$$

$$\begin{aligned}\text{NDE} = \sum_{c,m} & \{E[Y|C=c, A=a, M=m] - \\ & E[Y|C=c, A=a^*, M=m]\}P(M=m|C=c, A=a^*)P(C=c).\end{aligned}$$

Numerous approaches have been developed to estimate the NIE and NDE. Among these, the regression-based approach (VanderWeele and Vansteelandt, 2014) is commonly used. In this approach, the conditional expectations and probabilities in the above expressions are approximated using corresponding regression models based on the variable's distribution, such as linear regression for continuous variables or logistic regression for binary variables.

In many applications, potential mediators are high-dimensional correlated neuroimaging measures across brain regions (Caffo et al., 2008; Wager et al., 2009; Lindquist, 2012; Chén et al., 2018; Geuter et al., 2020; Zhao et al., 2021). Understanding the role of neuroimaging measures is essential in mental health research to probe the pathology of mental disorders and develop new treatment strategies. This objective is promoted by the Research Domain

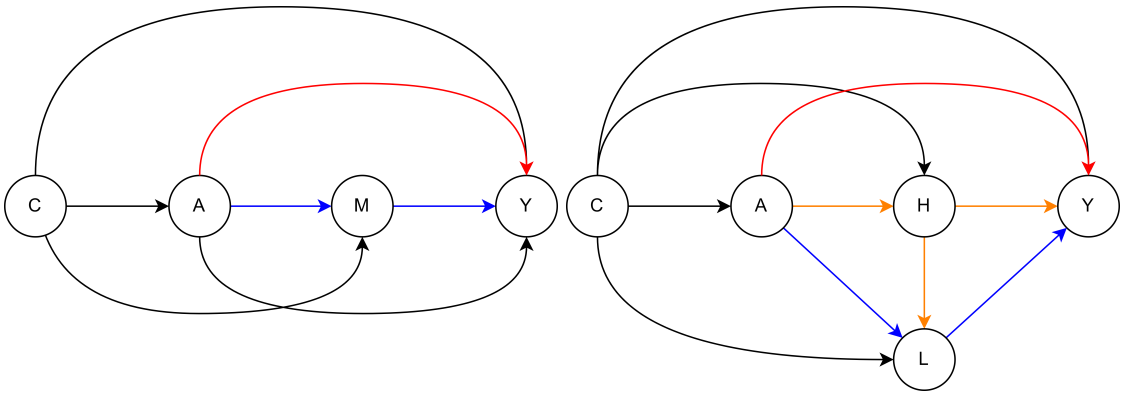

Figure 1: (a): Causal diagram with a single mediator M (NDE in red and NIE in blue). (b): Causal diagram with two causally dependent mediators H and L (NDE in red, NIE_H in orange and NIE_L in blue).

Criteria (RDoC) framework (Cuthbert and Insel, 2013), initiated by the National Institute of Mental Health (NIMH). The RDoC framework advocates for the integration of measures from various behavioral and biological domains to achieve a comprehensive understanding of the constructs and mechanisms underlying mental disorders. For example, intergenerational psychiatry studies the transmission pathways of vulnerability and resilience to mental illness from one generation to the next through potential biomarkers such as the children's brain development (Sawyer et al., 2019). Under the RDoC paradigm, we aim to contribute to the adolescent brain cognitive development (ABCD) study (Karcher and Barch, 2021), the largest population-based mental health study of brain development among U.S. adolescents. Specifically, our study seeks to investigate the effect of maternal smoking on children's cognitive abilities mediated through the development of cortical thickness across brain regions during adolescence, an area that remains unresolved in intergenerational psychiatry.

Several methods have been developed for high-dimensional mediators (Zhang et al., 2016; Huang and Pan, 2016; Chén et al., 2018; Huang, 2019; Zhao et al., 2020; Zhao and Luo, 2022). However, these methods mainly focus on hypothesis testing with adjustment for multiple comparisons or using dimension reduction techniques to combine mediators. Therefore, none of these methods can provide mediator-specific effects except when mediators are causally independent, an uncommon scenario for neuroimaging measures. To better understand how neuroimaging mediators contribute to underlying causal mechanisms and inform more efficient intervention strategies, it is crucial to identify significant mediating pathways involving specific brain regions through mediator-specific NIEs.

To evaluate mediator-specific NIEs, standard mediation analysis approaches can be applied separately for each mediator when mediators are causally independent. However, when mediators are causally dependent, some decomposition is necessary to obtain mediator-specific NIEs. Avin et al. (2005) and Miles et al. (2020) have showed that under several additional identifiability assumptions, the NIE of two causally dependent mediators H and L can be decomposed into the NIE operating solely through H or through both H and L

(NIE_H), and the NIE operating solely through L (NIE_L), as shown in Figure 1(b). With more identifiability assumptions, finer decomposition is possible. Daniel et al. (2015) proposes further decomposing NIE_H into the NIE solely through H and the NIE through both H and L. However, this decomposition requires an inestimable distribution parameter, which must be varied during sensitivity analyses to estimate multiple versions of the NIEs. Consequently, estimating mediator-specific NIEs for more than two causally dependent mediators is impractical, as the number of identifiability assumptions and inestimable distribution parameters required increases exponentially with the number of mediators (Daniel et al., 2015). Thus, the high-dimensionality and dependence among neuroimaging mediators pose substantial analytical challenges to current methodologies, necessitating the development of new approaches.

Fortunately, neuroimaging measures typically exhibit hierarchical network structures, such as star-shaped small-world structures, providing an opportunity to develop new approaches for estimating mediator-specific NIEs of neuroimaging mediators. A small-world topology is characterized by dense local clustering, short distances between nodes, and a limited number of long-range connections. Star-shaped network structures represent a special case of the small-world topology and consist of nodes with high centrality, known as hubs, which are directly connected to leaf clusters that are disconnected from each other, as illustrated in Figure 2(a). With hub regions acting as central communication hubs, these structures facilitate efficient transmission of information across distributed brain regions. These structures have been revealed by many brain imaging techniques, including structural magnetic resonance imaging (sMRI), functional MRI (fMRI), and diffusion tensor imaging (DTI) (He et al., 2007; Bullmore and Sporns, 2009; Harriger et al., 2012; Gollo et al., 2015; Fornito et al., 2016).

Motivated by the star-shaped network structures observed in neuroimaging measures, we propose a hybrid approach that leverages these network structures to obtain mediator-specific NIEs of neuroimaging mediators. In the first step, we estimate the star-shaped network of neuroimaging mediators using a conditional Gaussian graphical model framework. Both the mean of the mediators and their partial correlations, i.e., the precision matrix, are adjusted for the exposure and confounders to facilitate subsequent mediation analysis. To achieve this, our approach integrates domain-specific knowledge, including insights from neuroscience regarding brain network structures, with data-driven techniques. Domain knowledge is essential to determine whether the neuroimaging measures biologically exhibit a star-shaped network structure. During network estimation, we assume sparsity in line with the modularity and the small-world structures of brain networks. Once the network is obtained, hub and leaf mediators can be identified using centrality measures. In the second step, we perform mediation analysis to decompose the joint NIE of the neuroimaging mediators into the NIE through hub mediators, i.e., NIE_H, and the NIEs solely through each leaf mediator, i.e., NIE_Ls. Here, a leaf mediator may consist of a single mediator or a cluster of mediators. One advantage of this decomposition is that it helps identify leaf mediators whose effects cannot be controlled by hub mediators and informs efficient intervention designs. Additionally, after accounting for hub mediators, the NIEs solely through each leaf mediator can be identified and evaluated individually. Thus, our approach addresses the challenges posed by the dependence among these high-dimensional mediators and makes the estimation of mediator-specific NIEs feasible.

The rest of this paper is organized as follows. Section 2 presents our approach for constructing the network structure of neuroimaging mediators and performing mediation analysis. In Section 3, we conduct simulation studies to evaluate the effectiveness of our proposed approach. In Section 4, we illustrate our proposed approach by applying it to the ABCD study. Lastly, Section 5 provides discussions and future directions.

Generalizable Insights about Machine Learning in the Context of Healthcare

This paper makes two significant contributions to machine learning and healthcare, particularly in the context of mediation analysis and neuroimaging studies. First, we propose a network-assisted mediation analysis approach that provides a novel framework for analyzing high-dimensional neuroimaging mediators which exhibit a star-shaped network structure. By incorporating the star-shaped network structure of neuroimaging mediators into mediation analysis, our approach can estimate mediator-specific indirect effects, which current methodologies are unable to achieve. Second, we apply our approach to the adolescent brain cognitive development (ABCD) study (Karcher and Barch, 2021), the largest population-based mental health study of brain development among U.S. adolescents. Our analysis shows that the effect of maternal smoking on children's cognitive abilities is mediated by changes in their cortical thickness across multiple brain regions during adolescence. These findings offer valuable insights into intergenerational psychiatry and demonstrate the practical utility of our approach in discovering significant mediating pathways with neuroimaging mediators.

2. Methods

2.1. Construction of the network

In the first step, we estimate the network structure of neuroimaging mediators across brain regions. To do this, likelihood-based methods are typically used under a Gaussian graphical model framework (Yuan and Lin, 2007). To adjust for the exposure and confounders while estimating the network structure of neuroimaging mediators, we adapt a conditional Gaussian graphical model (Xie et al., 2020). Subsequently, hub and leaf mediators can be identified from the network using Kleinberg's hub centrality scores (Kleinberg, 1999).

Let $\boldsymbol{X}_i = (x_{i1}, ..., x_{iq})^T$ denote a q-dimensional vector of covariates including the exposure A_i and confounders $\boldsymbol{C}_i$, and let $\boldsymbol{M}_i = (M_{i1}, ..., M_{ip})^T$ denote a p-dimensional vector of neuroimaging mediators, for individual $i = 1, ..., n$. The mediators are assumed to follow a multivariate Gaussian distribution with both mean and precision matrix depending on $\boldsymbol{X}_i$ as

$$P(\boldsymbol{M}_i|\boldsymbol{X}_i) \propto \exp(\boldsymbol{\kappa}_i^T \boldsymbol{M}_i - \frac{1}{2}\boldsymbol{M}_i^T \boldsymbol{\Omega}_i \boldsymbol{M}_i), \tag{1}$$

where $\boldsymbol{\kappa}_i = (\boldsymbol{\zeta}_1^T \boldsymbol{X}_i, ..., \boldsymbol{\zeta}_p^T \boldsymbol{X}_i)^T$, and $\boldsymbol{\zeta}_j$ is the q-dimensional coefficient vector of the covariates on the mean of mediator j, $j = 1, ..., p$; $\boldsymbol{\Omega}_i$ is the $p \times p$ precision matrix of $\boldsymbol{M}_i$ with the $(j,k)^{th}$ element $\boldsymbol{\Omega}_i(j,k) = \boldsymbol{\Omega}_i(k,j) = \boldsymbol{\omega}_{jk}^T \boldsymbol{X}_i$ for $j \neq k$ and the j^{th} diagonal element $\boldsymbol{\Omega}_i(j,j) = \frac{1}{\sigma_j^2}$, and $\boldsymbol{\omega}_{jk}$ is the q-dimensional coefficient vector of the covariates on $\boldsymbol{\Omega}_i(j,k)$.

The covariate-adjusted precision matrix $\boldsymbol{\Omega}_i$ characterizes the network of neuroimaging mediators by capturing the partial correlations between mediators, after adjusting for the

exposure and confounders. We assume that individuals share a common network structure. If the Euclidean norm of $\boldsymbol{\omega}_{jk}$, i.e., $\|\boldsymbol{\omega}_{jk}\|_2$, is non-zero, an edge exists between mediators j and k in the network. Conversely, a zero norm indicates no edge, implying conditional independence between mediators j and k, given other mediators, the exposure, and confounders. Additionally, our model allows edge strength, i.e., $\boldsymbol{\Omega}_i(j,k)$, to vary based on individual-specific covariates, and $\boldsymbol{\omega}_{jk}$ represents the effects of each covariate on the strength of the edge between mediators j and k. Thus, our approach enables identifying a common network structure underlying the population, while accommodating individual variations in the magnitude of partial correlations between mediators.

Regularization is crucial to achieve both sparsity and stability in high-dimensional networks. Since neuroimaging measures are highly correlated, where the L_1 regularization tends to be unstable, the L_2 regularization is applied instead. We employ the pseudo-likelihood (Besag, 1975) instead of the joint likelihood to simplify computational complexity while still obtaining consistent parameter estimates. Thus, parameters $\boldsymbol{\zeta} = \{\boldsymbol{\zeta}_j\}_{j=1}^p$, $\boldsymbol{\omega} = \{\boldsymbol{\omega}_{jk}\}_{j,k=1;j\neq k}^p$ and $\boldsymbol{\sigma} = \{\sigma_j\}_{j=1}^p$ are estimated by minimizing the following objective function,

$$-\frac{1}{n}\log L_n(\boldsymbol{\zeta},\boldsymbol{\omega},\boldsymbol{\sigma}) + \lambda\left(\sum_{j=1}^p \boldsymbol{\zeta}_j^T\boldsymbol{\zeta}_j + \sum_{k\neq j}^p \boldsymbol{\omega}_{jk}^T\boldsymbol{\omega}_{jk}\right),$$

where $L_n(\boldsymbol{\zeta},\boldsymbol{\omega},\boldsymbol{\sigma}) = \prod_{i=1}^n\prod_{j=1}^p P(M_{ij}|\boldsymbol{M}_{i,\setminus j},\boldsymbol{X}_i)$ is the pseudo-likelihood, $\boldsymbol{M}_{i,\setminus j}$ is the vector of mediators excluding mediator j for individual i, and λ is a tuning parameter for the L_2 regularization (see Appendix A.1 for details). Based on the modularity and the small-world structures of brain networks, the connections between brain regions are expected to be sparse. To introduce sparsity in the estimated network, we apply hard thresholding by removing edges with the norm $\|\boldsymbol{\omega}_{jk}\|_2$ smaller than a predefined threshold (e.g., $c\log(qp(p+1)/2)/\sqrt{n}$; Bühlmann and Van De Geer, 2011).

The extended Bayesian information criteria (EBIC) (Chen and Chen, 2008; Haslbeck and Waldorp, 2020) is commonly used for model selection in Gaussian graphical models. We select λ and c using an EBIC adapted for our conditional Gaussian graphical model. By maximizing the pseudo-likelihood, our model aligns with neighborhood-selection-based methods for graphical models (Meinshausen and Bühlmann, 2006; Peng et al., 2009). Consequently, the estimation of our model can be regarded as performing ridge regression for each mediator on covariates, as well as on the product of each covariate with each other mediator. Thus, the EBIC is adapted as

$$\text{EBIC} = -2\log L_n(\boldsymbol{\zeta},\boldsymbol{\omega},\boldsymbol{\sigma}) + Eq\log(n) + 2\gamma Eq\log(\sum_{j=1}^p d_j),$$

where E is the number of non-zero edges; $\gamma \in [0,1]$ is a hyperparameter; and $d_j = \text{tr}(\boldsymbol{Z}_j(\boldsymbol{Z}_j^T\boldsymbol{Z}_j + \lambda\boldsymbol{I})^{-1}\boldsymbol{Z}_j^T)$ is the degrees of freedom of the ridge regression for mediator j, where $\boldsymbol{Z}_j$ is the $n\times(q+q(p-1))$ design matrix with columns $\{\boldsymbol{X}_s, \boldsymbol{X}_s\odot\boldsymbol{M}_k\}_{s,k=1;k\neq j}^{q,p}$, $\boldsymbol{X}_s = (x_{1s},...,x_{ns})^T$ is the vector of covariate s for all individuals, $\boldsymbol{X}_s\odot\boldsymbol{M}_k$ is the Hadamard product of $\boldsymbol{X}_s$ and $\boldsymbol{M}_k$, and $\boldsymbol{M}_k = (M_{1k},...,M_{nk})^T$ is the vector of mediator k for all individuals.

To address uncertainty in network estimation, we can apply bootstrap methods to identify edges that appear in more than a threshold percentage (e.g., 80%) of networks estimated from the bootstrapped samples.

2.2. Orientations of edges in an undirected star-shaped network

The methods in Section 2.1 construct an undirected network of neuroimaging mediators. To proceed with mediation analysis, it is necessary to derive a directed acyclic graph (DAG) among mediators. Domain knowledge plays a crucial role in this process. In this section, we show that even without such domain knowledge, we can still assume a DAG with all edges directed from hub mediators to leaf mediators, and there will be at most one mis-oriented edge, i.e., there is at most one leaf mediator preceding hub mediators.

Representing the relationships among variables as a DAG generally requires several conditions (Shalizi, 2013), particularly when domain knowledge of the DAG is lacking: (i) there exists a DAG representing the relationships among variables; (ii) the causal Markov condition: the joint distribution of the variables obeys the Markov property on the DAG; (iii) faithfulness: the joint distribution reflects all and only those conditional independence relations implied by the causal Markov condition. Based on these conditions, we can assume a DAG among mediators from the network obtained by our conditional Gaussian graphical model in Section 2.1. Specifically, this network has the following two properties:

Property 1 *(Pairwise Markov Property) This network is a conditional independence graph* $G = (V, E)$ *with vertices* $V = \{M_j\}_{j=1}^{p}$ *and edges* E *such that*

$$(j,k) \notin E \Leftrightarrow M_j \perp\!\!\!\perp M_k | A, C, M_{V\setminus\{j,k\}}.$$

Property 2 *For this network, a star-shaped structure ensures conditional independence between two leaf mediators* L_j *and* L_k*,*

$$(j,k) \notin E \Leftrightarrow L_j \perp\!\!\!\perp L_k | A, C, H.$$

Based on these two properties, the following lemma can be stated:

Lemma 1 *In a DAG inferred from the star-shaped network obtained by our conditional Gaussian graphical model in Section 2.1, there can be at most one edge directed from leaf mediators to hub mediators.*

Proof *Consider two leaf mediators* L_j *and* L_k *connected by the path* $L_j - H - L_k$*. If the directions on this path are* $L_j \to H \leftarrow L_k$*, then* H *is a collider for* L_j *and* L_k*. Since* H *is a collider, conditioning on* H *prevents* L_j *and* L_k *from being independent, which contradicts Property 2. This contradiction implies that there can be at most one edge directed from leaf mediators to hub mediators. Thus, the directions of all but one of the edges between hub mediators and leaf mediators are identified, as illustrated in Figure 2(b).* ■

Remark: *When there are multiple hub mediators that are pairwise correlated, lemma 1 still hold.*

With conditions (i)-(iii) and Lemma 1, we can assume a DAG among mediators where all edges are directed from hub mediators to leaf mediators, with at most one mis-oriented edge.

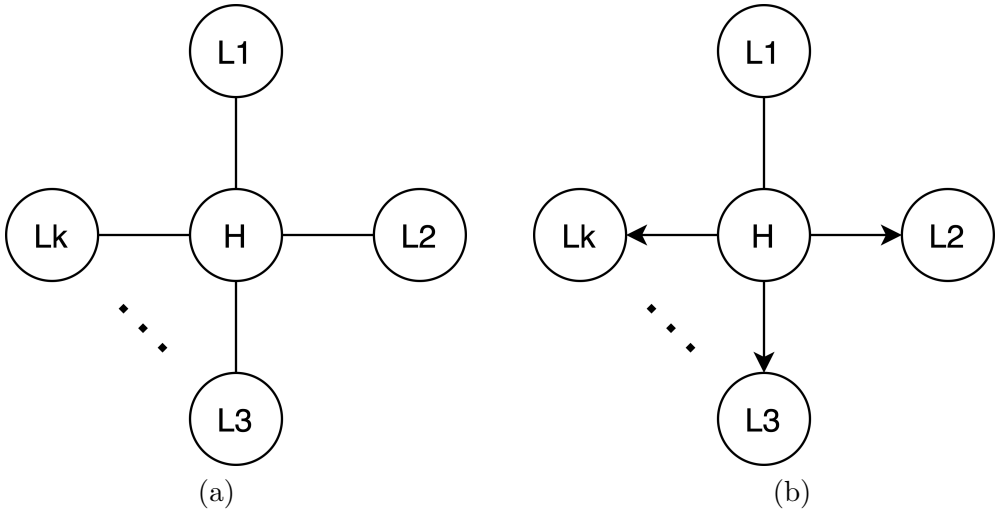

Figure 2: (a): An undirected star-shaped network of mediators.
(b): By Lemma 1, there is at most one edge with an unknown direction between hub mediators and leaf mediators.

2.3. Network-assisted mediation analysis

In the second step, we conduct mediation analysis based on the star-shaped network of the mediators identified in Section 2.1. With the star-shaped network, we differentiate mediators into three categories: (i) independent mediators that are not connected with any other mediators in the network, (ii) hub mediators, and (iii) leaf mediators that are connected only to hub mediators. Here, a leaf mediator may consist of a single mediator or a cluster of mediators. Based on these categories, we will decompose the joint NIE of these mediators into: the NIEs of each independent mediator, the NIE through hub mediators, i.e., NIE_H, and the NIEs solely through each leaf mediator, i.e., NIE_Ls. We choose this decomposition for two reasons. First, this is the finest achievable decomposition without replying on inestimable distribution parameters (Avin et al., 2005; Miles et al., 2020; Daniel et al., 2015). Second, this decomposition is sufficient to inform efficient intervention designs. Conceptually, the NIE_H represents a system-wide effect due to the connections between hub mediators and leaf mediators. Interventions targeting hub regions can effectively control for this effect. In contrast, the NIE_L is the specific effect of a leaf mediator which cannot be controlled by intervening on hub regions. To control for this effect, the intervention must be specifically aimed at this leaf mediator.

Standard mediation analysis approaches for a single mediator or multiple mediators jointly can be applied to estimate NIE_H and the NIEs of each independent mediator. In this section, we focus on the estimation of the NIE_Ls of each leaf mediator. Let H denote the hub mediator(s), and H_a denote the counterfactual value of H that would have been observed had A been set to a. Let L denote a leaf mediator, and $L_{aH_{a^*}}$ denote the counterfactual value of L that would have been observed had A been set to a and H been set to the counterfactual value H_{a^*}. Let $Y_{aH_{a^*}L_{aH_{a^*}}}$ denote the counterfactual value of Y that would

have been observed had A been set to a, H been set to the counterfactual value H_{a^*}, and L been set to the counterfactual value $L_{aH_{a^*}}$. Then, the NIE_L of a leaf mediator can be defined as

$$\text{NIE}_L = E[Y_{aH_{a^*}L_{aH_{a^*}}} - Y_{aH_{a^*}L_{a^*H_{a^*}}}].$$

Under the following assumptions:
(i) No unmeasured exposure-outcome, mediator-outcome, exposure-mediator or hub-leaf confounding:

$$Y_{ahl} \perp\!\!\!\perp A|C, \quad Y_{ahl} \perp\!\!\!\perp (H,L)|\{A,C\}, \quad (H_a, L_{ah}) \perp\!\!\!\perp A|C, \quad L_{ah} \perp\!\!\!\perp (A,H)|C,$$

(ii) Cross-world independence between counterfactual outcomes, hub mediators and leaf mediators:

$$Y_{ahl} \perp\!\!\!\perp (H_{a^*}, L_{a^*})|C, \quad Y_{ahl} \perp\!\!\!\perp (H_{a^*}, L_{ah})|C, \quad L_{ah} \perp\!\!\!\perp H_{a^*}|C,$$

(iii) Conditional independence of each pair of leaf mediators L and L' given the exposure, hub mediators and confounders:

$$L \perp\!\!\!\perp L'|\{A,H,C\},$$

the NIE_L of each leaf mediator can be identified and estimated individually by the following empirical expression:

$$\text{NIE}_L = \sum_{c,h,l} E[Y|c,a,h,l]\{P(l|c,a,h) - P(l|c,a^*,h)\}P(h|c,a^*)P(c). \tag{2}$$

Assumptions (i) and (ii) have been proved by Avin et al. (2005) and are met in the causal diagram shown in Figure 1(b). Assumption (iii) is necessary to enable the separate identification of the NIE_Ls of each leaf mediator, and is satisfied with the star-shaped network structure of the mediators. The conditional expectations and probabilities in equation (2) can be approximated using appropriate regression models. The point estimate of each NIE_L can be obtained through Monte-Carlo simulations with these regression models using equation (2), and bootstrap methods can be used to make inferences.

Continuing from Section 2.2, note that mis-orienting one edge between hub mediators and a leaf mediator will not affect the estimation of the NIE_Ls of other leaf mediators. This can be easily seen from equation (2) and is demonstrated through simulation studies in Section 3. Thus, the NIE_Ls of all but one leaf mediator are guaranteed to be unbiasedly estimated given the identifiability assumptions hold. However, it may bias the estimation of the NIE_H due to unadjusted confounding from the leaf mediator causing hub mediators. Since our primary interest is in NIE_L rather than NIE_H, this issue can be addressed by conducting sensitivity analyses for unmeasured confounding in the NIE_H estimate, such as using the E-Value approach (VanderWeele and Ding, 2017).

3. Simulation Studies

In this section, we conduct simulation studies under two scenarios to demonstrate the unbiasedness of our approach proposed in Section 2 for estimating the NIE_L, and to highlight the advantages and importance of using the network to detect mediator-specific effects of

neuroimaging mediators. In each scenario, 500 datasets are simulated with various sample sizes, including $n = 1000, 5000, 10000$. Each dataset is simulated with a continuous outcome Y, a binary exposure A, a continuous hub mediator H, two continuous leaf mediators L_1 and L_2, and a continuous confounder C, for simplicity. Random errors for each continuous variable are generated from the standard normal distribution $N(0,1)$. The two reference values of A are $a^* = 0$ and $a = 1$.

3.1. Scenario 1

In Scenario 1, our proposed approach is compared with the standard mediation analysis approach for a single mediator (VanderWeele, 2015) in estimating the NIE_L. This scenario highlights the importance of accounting for the hierarchical structure between hub and leaf mediators when estimating the NIE_L. Under this scenario, the relevant counterfactual values of H, L_1, L_2 and Y for each individual i, i.e., $H_{a^*,i}$, $H_{a,i}$, $L_{1a^*H_{a^*},i}$, $L_{1aH_{a^*},i}$, $L_{1aH_a,i}$, $L_{2a^*H_{a^*},i}$, $L_{2aH_{a^*},i}$, $L_{2aH_a,i}$, $Y_{aH_{a^*}L_{a^*H_{a^*}},i}$, $Y_{aH_{a^*}L_{aH_{a^*}},i}$, $Y_{a^*,i}$ and $Y_{a,i}$, are simulated under the following true models:

$$
\begin{gathered}
C \sim N(0,1), \\
logit\{P(A=1|C)\} = -1 + C, \\
E[H|A,C] = 2 + 0.5A + C, \\
E[L_1|A,H,C] = 2 + 2A + \beta_1 H + C, \quad E[L_2|A,H,C] = 2 + 2A + \beta_2 H + C, \\
E[Y|A,H,L_1,L_2,C] = 2 + A + H + 0.5L_1 + 0.5L_2 + C,
\end{gathered}
$$

where β_1 and β_2 vary within the intervals $[0,1]$ and $[-1,0]$, respectively, with an increment of 0.2. The causal diagram under Scenario 1 is illustrated in Figure 3(a). Based on the counterfactual values and the actual exposure for each individual, the corresponding observed values for H, L_1, L_2, and Y are obtained. For example, for individual i with $A_i = a$, the observed values are $H_i = H_{a,i}$, $L_{1i} = L_{1aH_a,i}$, $L_{2i} = L_{2aH_a,i}$, and $Y_i = Y_{a,i}$. According to the true models, the true NIE_L of each L_1 and L_2 is 1.

For each dataset simulated under Scenario 1, both our proposed approach and the standard mediation analysis approach for a single mediator are used to estimate the NIE_Ls of each L_1 and L_2. Since L_1 and L_2 are causally independent, we fit mediation models for each of them individually. When estimating the NIE_Ls of each L_1 and L_2 using the standard mediation analysis approach, H is not included as a mediator-outcome confounder, since this approach incorporates no information from the network. The estimated NIE_Ls are shown in Figure 3(c). As shown in this figure, our proposed approach is consistently unbiased in estimating the NIE_Ls of each L_1 and L_2, while the standard mediation analysis approach is generally biased, except when $\beta_1 = \beta_2 = 0$, i.e., L_1 and L_2 are not affected by H. This is because H is an unadjusted mediator-outcome confounder, violating the identifiability assumption of the standard mediation analysis approach. As expected, the bias increases as the absolute values of β_1 and β_2 increase, indicating a higher dependence between H and each of L_1 and L_2.

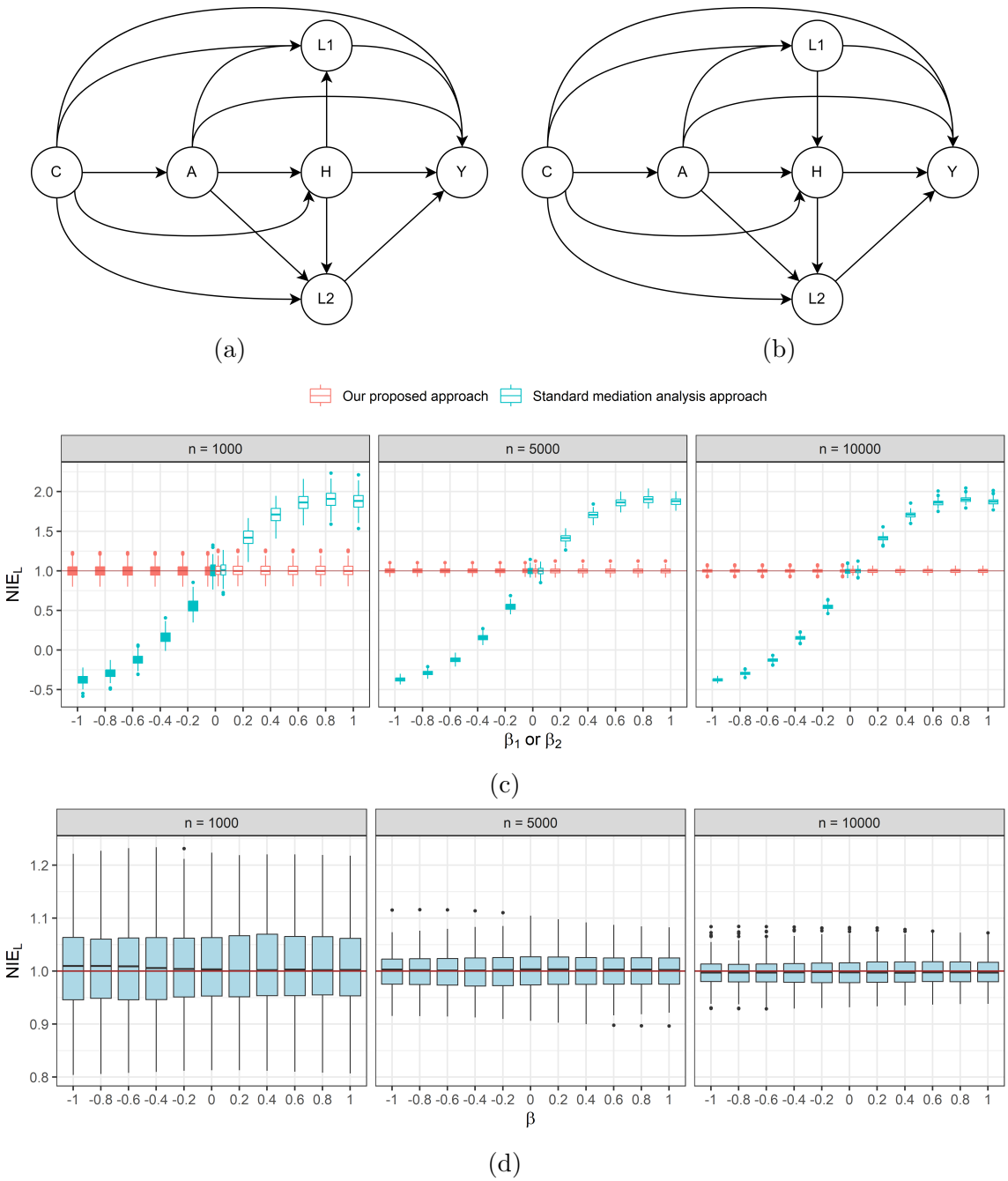

Figure 3: (a): The causal diagram under Scenario 1.
(b): The causal diagram under Scenario 2.
(c): The NIE$_L$s of each L_1 (hollow) and L_2 (solid) estimated from our proposed approach (red) and the standard mediation analysis approach (blue) with varying $\beta_1 \in [0, 1]$ and $\beta_2 \in [-1, 0]$ under Scenario 1.
(d): The NIE$_L$ of L_2 estimated from our proposed approach with varying $\beta \in [-1, 1]$ under Scenario 2.
The true NIE$_L$ is the reference line colored in brown.

3.2. Scenario 2

Scenario 2 shows that our proposed approach remains unbiased in estimating the NIE_L of a leaf mediator even when the edge from hub mediators to another leaf mediator is mis-oriented, i.e., there is another leaf mediator preceding the hub mediators in the causal diagram. We consider the causal diagram shown in Figure 3(b), in which the causal order of H, L_1 and L_2 is $L_1 \to H \to L_2$. Under this scenario, the counterfactual values of H, L_1, L_2 and Y for each individual i are simulated under the following true models:

$$
\begin{aligned}
C &\sim N(0,1),\\
logit\{P(A=1|C)\} &= -1 + C,\\
E[L_1|A,C] &= 2 + 2A + C,\\
E[H|A,L_1,C] &= 2 + 0.5A + \beta L_1 + C,\\
E[L_2|A,H,C] &= 2 + 2A + 0.5H + C,\\
E[Y|A,H,L_1,L_2,C] &= 2 + A + H + 0.5L_1 + 0.5L_2 + C,
\end{aligned}
$$

where β controls the causal dependence between H and L_1. According to the true models, the true NIE_L of L_2 is 1.

The NIE_L of L_2 is estimated using our proposed approach for each simulated dataset, illustrated in Figure 3(d). We see that NIE_L of L_2 estimated from our proposed approach is unbiased regardless of β. Thus, the NIE_L of L_2 estimated from our proposed approach is not affected by the causal order between the hub mediators and other leaf mediators. As expected, the variances of all estimates in both scenarios decrease with increasing sample sizes.

4. Application to the ABCD Study

In the ABCD study, our primary aim is to investigate how the effect of maternal smoking–both before and during pregnancy–on children's cognitive abilities is mediated by the development of brain cortical thickness during adolescence. The dataset for this analysis includes $n = 9,029$ adolescents. The exposure of interest is maternal smoking (whether the mother smoked before or during pregnancy). The outcome is the number of correct trials in the children's emotional n-back test, an emotional regulation task designed to assess the interference effect of emotional processing on working memory and cognitive function (Miller et al., 2009). The candidate mediators are whole-brain structural MRI measures of cortical thickness across 148 brain regions of interest (ROI) in both hemispheres, as defined by the Destriuex atlas (Destrieux et al., 2010). The T1-weighted images were preprocessed using the FreeSurfer 5.1 pipeline (Fischl et al., 1999). In this analysis, confounders include age (the child's age in month), gender (the child's gender, male or female), parental education (whether or not a parent attended college), and race (the child's race, categorized as black, white, or other). Study sites are modeled as random effects.

First, we remove the effects of study sites on the cortical thickness measures by the NeuroCombat package in R (Fortin et al., 2018). Then, we fit the conditional Gaussian graphical model proposed in Section 2.1 to obtain the network of these measures. Using the EBIC criterion with $\gamma = 1$, we identify 101 independent brain regions, while the remaining

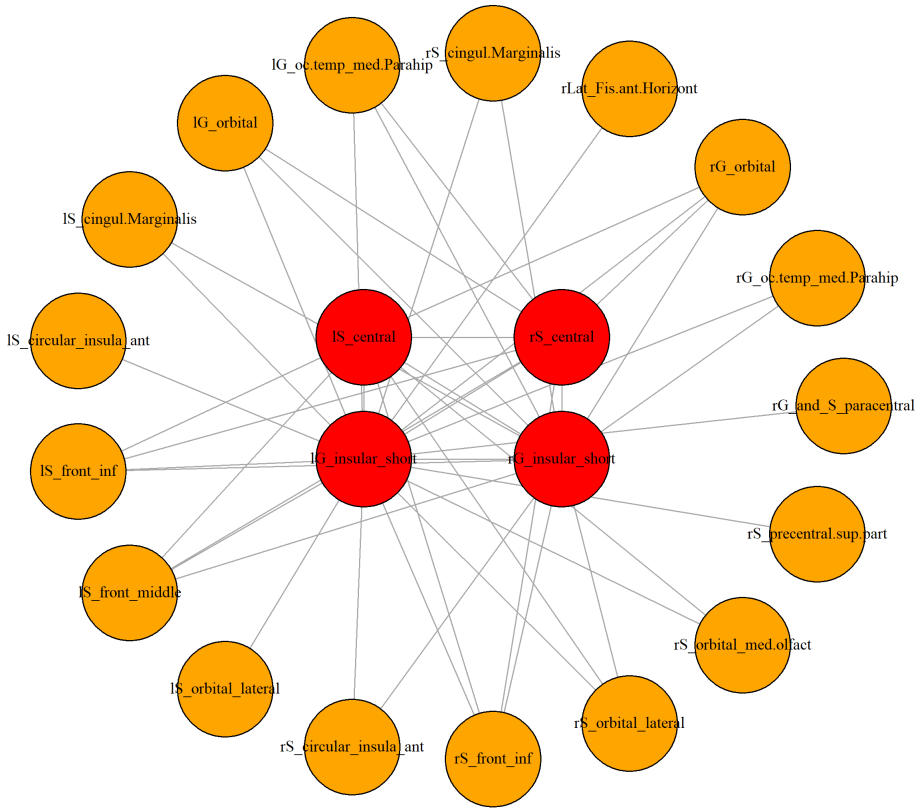

Figure 4: Identified network of brain cortical thickness from the ABCD study.

47 regions are clustered into a network. Based on Kleinberg's hub centrality scores > 0.9, we identify 4 regions as hubs, as shown in Figure 4. The network structure consists of a star-shaped network with 17 leaf regions and a subnetwork with 26 regions. Figure 4 shows the 4 hub regions and connected leaf regions. Our subsequent analysis focuses on this star-shaped network. Abbreviations, full names, and lobes of all brain regions mentioned in this section are provided in Appendix A.2.

The identified hub regions include the left short insular gyri (lG_insular_short), right short insular gyri (lG_insular_short), left central sulcus (lS_central), and right central sulcus (rS_central). Anatomically, the central sulcus is a key ROI that separates the parietal lobe from the frontal lobe and distinguishes the primary motor cortex from the primary somatosensory cortex. The short insular gyri, located in the anterior insula, are part of the salience network and are involved in several functions, including pain perception, emotional regulation, cravings (potentially related to conscious urges to use drugs), and addiction. Importantly, our identified hub regions align with existing literature, which consistently recognizes parietal, frontal, and insular regions as hubs in brain networks across various cortical parcellations (Van den Heuvel and Sporns, 2013); for example, the insular cortex has been identified as a neural source or net emitter, highlighting its role in coordinating complex brain functions.

Before proceeding to mediation analysis, we examine the coefficients of each covariate on the strength of edges, i.e., $\boldsymbol{w}_{jk}$, in the identified network shown in Figure 4. Each coefficient reflects the effect of the corresponding covariate on the partial correlation between the corresponding pair of brain regions, given the covariates and other regions. The magnitude of these coefficients is small, indicating that the edge strength does not vary substantially with these covariates. For example, the mean absolute coefficient of the exposure across edges is 5.59×10^{-4} (SD 5.39×10^{-4}, 1st quartile 1.44×10^{-4}, and 3rd quartile 8.02×10^{-4}). Although the magnitude of the covariate effect is small, the direction of these effects varies across edges. Table 1 presents the coefficients of each covariate on the strength of edges with the strongest exposure effect (> 1 SD from the mean). As shown in Table 1, maternal smoking is negatively associated with the partial correlations between some regions, indicating an attenuated effect, while for others, the effect is positive. These findings suggest a complex pattern of how maternal smoking and other covariates are associated with brain connections.

Table 1: Coefficients ($\times10^{-3}$) of each covariate on the strength of edges with the strongest exposure effect

Region 1	Region 2	Maternal Smoking (Yes)	Race (Black)	Race (White)	Sex (Female)	Age	Parental Education (Yes)
lG_insular_short	rG_oc.temp_med.Parahip	-2.10	-2.13	-0.01	8.96	-0.93	0.11
lG_insular_short	lG_oc.temp_med.Parahip	-2.01	-3.69	1.90	6.49	-1.19	0.64
lG_insular_short	rS_precentral.sup.part	-1.85	3.23	-1.37	0.25	6.27	1.05
rG_insular_short	lG_oc.temp_med.Parahip	-1.65	-4.46	2.99	6.45	0.39	1.21
rG_insular_short	rS_circular_insula_ant	-1.25	-3.68	2.14	-0.56	-2.41	0.99
rG_insular_short	rS_orbital_lateral	1.23	-0.10	0.08	-0.87	0.041	0.67
rS_central	lG_oc.temp_med.Parahip	1.17	3.44	-2.71	3.78	-0.05	-0.58

Following the instructions from Section 2.2, we assume that the directions of edges in the network are from hub mediators to leaf mediators. In the second step, we conducted network-assisted mediation analysis. A generalized linear mixed model with a binomial distribution is used for the outcome, which is modeled as a binary variable where 1 represents a correct response in the emotional n-back test and 0 represents an incorrect response. A linear mixed model is applied to each mediator. In all models, study sites are adjusted for as random intercepts. All effects are reported on the odds ratio (OR) scale. The total effect of maternal smoking on children's emotional n-back accuracy is quantified as an OR of 0.933 (95% CI: [0.892, 0.983]), indicating maternal smoking significantly reduces a child's odds of giving a correct response in the emotional n-back test by 7% (95% CI: [2%, 11%]). The estimated indirect effects are shown in Table 2. The indirect effect though hub regions is 0.996 (95% CI: [0.993, 0.998]), which corresponds to a reduction of a child's odds of giving a correct response by 0.4% (95% CI: [0.2%, 0.7%]). The proportion of the total effect mediated (PM) by hub regions is 4.49% (95% CI: [1.56%, 16.74%]). After accounting for hub regions, the indirect effect solely through the leaf region, left inferior frontal sulcus (lS_front_inf), is

significant at 0.998 (95% CI: [0.996, 0.999]), with a PM of 1.53% (95% CI: [0.85%, 10.04%]). This finding suggests that maternal smoking reduces a child's odds of giving a correct response by 0.2% (95% CI: [0.1%, 0.4%]) solely though changing the thickness of the left inferior frontal sulcus. Additionally, we identify two significant independent regions that are not connected with any other regions in the network: left and right inferior segments of the circular sulcus of the insular (lS_circular_insula_inf and rS_circular_insula_inf). Their indirect effect increases a child's odds of giving a correct response by 0.2% [0.1%, 0.5%]. This effect opposes the direction of the total effect, highlighting the complexity of the mediating role of brain cortical thickness. These indirect effects are expected to be small given the small magnitude of the total effect, the number of brain regions, and other potential factors underlying the pathways between maternal smoking and children's emotional n-back accuracy.

Table 2: Brain regions with a significant indirect effect

Region	Type	OR (95% CI)	PM (95% CI)	$A \to M$	$M \to Y^{\dagger}$
lG_insular_short, rG_insular_short, lS_central, rS_central	NIE_H	0.996 [0.993, 0.998]	4.49% [1.56%, 16.74%]	+ -	- +
lS_front_inf	NIE_L	0.998 [0.996, 0.999]	1.53% [0.85%, 10.04%]	+	-
lS_circular_insula_inf	NIE	1.002 [1.001, 1.005]	N/A	+	+
rS_circular_insula_inf	NIE	1.002 [1.001, 1.005]	N/A	+	+

†: the sign of path effects $A \to M$ and $M \to Y$

By examining the signs of these pathways, we observe that maternal smoking increases cortical thickness in the left and right short insular gyri, left inferior frontal sulcus, and left and right inferior segments of the circular sulcus of the insular. It is important to note that cortical thinning is a typical developmental phenomenon observed in children across all brain regions (Ducharme et al., 2016). Our results suggest that while cortical thinning is a normal part of development, maternal smoking leads to cortical thickening, a deviation from the usual developmental trajectory.

5. Discussions

Mediation analysis with high-dimensional correlated mediators presents significant challenges. In this paper, we introduce a hybrid approach to advance mediation analysis specifically for neuroimaging mediators that exhibit a star-shaped network structure. Although our focus is on neuroimaging mediators, star-shaped network structures are also observed in other domains such as gene regulatory networks (Gerstein et al., 2012) and protein-protein interaction networks (Uetz et al., 2000), where our proposed approach could be applicable. Our approach enables the evaluation of mediator-specific indirect effects by leveraging the network structure. In the first step, we propose a conditional Gaussian graphical model to estimate the network of neuroimaging mediators. This model is general and can be used to learn any network, given the multivariate Gaussian assumption holds. If the network is identified as star-shaped, we can proceed with mediation analysis in the second step. We propose to decompose the joint indirect effect of the mediators into the indirect effect through hub mediators and indirect effects solely through each leaf mediator. After

accounting for hub mediators, the indirect effects solely through each leaf mediator can be estimated individually. In the data application, our findings imply distinct pathways through which maternal smoking affects children's accuracy in the emotional n-back test. These results provide new insights to the RDoC framework, particularly on the mediating role of the seven brain regions in Table 2.

Our proposed approach has several limitations. First, it relies on the structural assumption of a star-shaped network, which simplifies the network and may not accurately represent the truth. However, this simplification facilitates mediation analysis of high-dimensional neuroimaging measures and enhances our understanding of how these measures contribute to underlying causal mechanisms. Second, without domain knowledge, our approach assumes a DAG from the network, specifically that all edges are oriented from hub mediators to leaf mediators. We have shown that we may mis-orient at most one edge, which could result in a biased indirect effect for the corresponding leaf mediator, and we cannot determine which specific edge is mis-oriented without domain knowledge. However, it is important to note that the mis-orientation of a leaf mediator does not impact the identifiability of the indirect effects of other leaf mediators, as demonstrated through our simulation studies. Thus, our approach is primarily designed to help identify potentially significant leaf mediators. Future work is needed to validate the assumed DAG and refine the approach.

Here are some future directions. First, data driven approaches for determining edge directions for our conditional Gaussian graphical model would be highly beneficial. Second, approaches for sensitivity analysis to measure the robustness of the NIE_L to unmeasured confounding are crucial, since identifiability assumptions about no unmeasured confounding are not testable. Third, our proposed approach is applied to analyze brain cortical thickness from structural MRI data at a single time point, which makes it challenging to validate the assumed causal orders between hub mediators and leaf mediators. Future research could explore longitudinal mediators or functional MRI measures. These types of data might make it easier to validate the causal orders between mediators.

Acknowledgments

This research is supported by U.S. NIH grants NS073671, MH123487, and MH124106.

Appendix

A.1 Model estimation

This section provides details on the estimation of our model in Section 2.1. To estimate the parameters in the conditional Gaussian graphical model (1), we employ the pseudo-likelihood instead of the joint likelihood to reduce computational complexity while still

obtaining consistent parameter estimates. The pseudo-likelihood is defined as

$$\begin{aligned} L_n(\boldsymbol{\zeta}, \boldsymbol{\omega}, \boldsymbol{\sigma}) &= \prod_{i=1}^{n} \prod_{j=1}^{p} P(M_{ij} | \boldsymbol{M}_{i,\setminus j}, \boldsymbol{X}_i) \\ &= \prod_{i=1}^{n} \prod_{j=1}^{p} \sqrt{\frac{1}{2\pi\sigma_j^2}} \exp\left[-\frac{1}{2\sigma_j^2} \{ M_{ij} - \sigma_j^2 (\boldsymbol{\zeta}_j^T \boldsymbol{X}_i + \sum_{k \neq j}^{p} \boldsymbol{\omega}_{jk}^T \boldsymbol{X}_i M_{ik}) \}^2 \right], \end{aligned}$$

and the objective function with the L_2 regularization is

$$l(\boldsymbol{\zeta}, \boldsymbol{\omega}, \boldsymbol{\sigma}) = -\frac{1}{n} \log L_n(\boldsymbol{\zeta}, \boldsymbol{\omega}, \boldsymbol{\sigma}) + \lambda \left(\sum_{j=1}^{p} \boldsymbol{\zeta}_j^T \boldsymbol{\zeta}_j + \sum_{k \neq j}^{p} \boldsymbol{\omega}_{jk}^T \boldsymbol{\omega}_{jk} \right).$$

Then, the gradients of the objective function with respect to each parameter are given by

$$\begin{aligned} \frac{\partial l(\boldsymbol{\zeta}, \boldsymbol{\omega}, \boldsymbol{\sigma})}{\partial \zeta_{j'm'}} &= -\frac{1}{n} \sum_{i=1}^{n} X_{im'} M_{ij'} + \frac{1}{n} \sigma_{j'}^2 \sum_{i=1}^{n} X_{im'} (\boldsymbol{\zeta}_{j'}^T \boldsymbol{X}_i + \sum_{k \neq j'}^{p} \boldsymbol{\omega}_{j'k}^T \boldsymbol{X}_i M_{ik}) + 2\lambda \zeta_{j'm'}, \\ \frac{\partial l(\boldsymbol{\zeta}, \boldsymbol{\omega}, \boldsymbol{\sigma})}{\partial \omega_{j'k'm'}} &= \frac{1}{n} \sum_{i=1}^{n} X_{im'} M_{ij'} [-M_{ik'} + \sigma_{k'}^2 (\boldsymbol{\zeta}_{k'}^T \boldsymbol{X}_i + \sum_{j \neq k'}^{p} \boldsymbol{\omega}_{k'j}^T \boldsymbol{X}_i M_{ij})] + \\ &\quad \frac{1}{n} \sum_{i=1}^{n} X_{im'} M_{ik'} [-M_{ij'} + \sigma_{j'}^2 (\boldsymbol{\zeta}_{j'}^T \boldsymbol{X}_i + \sum_{k \neq j'}^{p} \boldsymbol{\omega}_{j'k}^T \boldsymbol{X}_i M_{ik})] + 4\lambda \omega_{j'k'm'}, \\ \frac{\partial l(\boldsymbol{\zeta}, \boldsymbol{\omega}, \boldsymbol{\sigma})}{\partial \sigma_{j'}^2} &= \frac{1}{2\sigma_{j'}^2} - \frac{1}{2n\sigma_{j'}^4} \sum_{i=1}^{n} M_{ij'}^2 + \frac{1}{2n} \sum_{i=1}^{n} (\boldsymbol{\zeta}_{j'}^T \boldsymbol{X}_i + \sum_{k \neq j'}^{p} \boldsymbol{\omega}_{j'k}^T \boldsymbol{X}_i M_{ik})^2, \end{aligned}$$

where $\zeta_{j'm'}$ and $\omega_{j'k'm'}$ are the m'^{th} element in $\boldsymbol{\zeta}_{j'}$ and $\boldsymbol{\omega}_{j'k'}$, $j', k' = 1, ..., p, j' \neq k'; m' = 1, ..., q$.

After solving $\frac{\partial l(\boldsymbol{\zeta}, \boldsymbol{\omega}, \boldsymbol{\sigma})}{\partial \zeta_{j'm'}} = 0$, $\frac{\partial l(\boldsymbol{\zeta}, \boldsymbol{\omega}, \boldsymbol{\sigma})}{\partial \omega_{j'k'm'}} = 0$ and $\frac{\partial l(\boldsymbol{\zeta}, \boldsymbol{\omega}, \boldsymbol{\sigma})}{\partial \sigma_{j'}^2} = 0$, parameters are estimated by

$$\begin{aligned} \hat{\zeta}_{j'm'} &= \frac{\frac{1}{n} \sum_{i=1}^{n} X_{im'} M_{ij'} - \frac{1}{n} \hat{\sigma}_{j'}^2 \sum_{i=1}^{n} X_{im'} (\sum_{m \neq m'}^{q} \hat{\zeta}_{j'm} X_{im} + \sum_{k \neq j'}^{p} \hat{\boldsymbol{\omega}}_{j'k}^T \boldsymbol{X}_i M_{ik})}{\frac{1}{n} \hat{\sigma}_{j'}^2 \sum_{i=1}^{n} X_{im'}^2 + 2\lambda}, \\ \hat{\omega}_{j'k'm'} &= \frac{\frac{1}{n} \sum_{i=1}^{n} X_{im'} (A + B)}{\frac{1}{n} \hat{\sigma}_{\epsilon_{j'}'}^2 \sum_{i=1}^{n} X_{im'}^2 M_{ik'}^2 + \frac{1}{n} \hat{\sigma}_{k'}^2 \sum_{i=1}^{n} X_{im'}^2 M_{ij'}^2 + 4\lambda}, \\ \hat{\sigma}_{j'}^2 &= \frac{-n + \sqrt{n^2 + 4(\sum_{i=1}^{n} M_{ij'}^2)[\sum_{i=1}^{n} (\hat{\boldsymbol{\zeta}}_{j'}^T \boldsymbol{X}_i + \sum_{k \neq j'}^{p} \hat{\omega}_{j'k}^T \boldsymbol{X}_i M_{ik})^2]}}{2 \sum_{i=1}^{n} (\hat{\boldsymbol{\zeta}}_{j'}^T \boldsymbol{X}_i + \sum_{k \neq j'}^{p} \hat{\boldsymbol{\omega}}_{j'k}^T \boldsymbol{X}_i M_{ik})^2}, \end{aligned}$$

where

$$\begin{aligned} A &= M_{ij'} [M_{ik'} - \hat{\sigma}_{k'}^2 (\hat{\boldsymbol{\zeta}}_{k'}^T \boldsymbol{X}_i + \sum_{j \neq k', j'}^{p} \hat{\boldsymbol{\omega}}_{k'j}^T \boldsymbol{X}_i M_{ij} + \sum_{m \neq m'}^{q} \hat{\omega}_{k'j'm} X_{im} M_{ij'})], \\ B &= M_{ik'} [M_{ij'} - \hat{\sigma}_{j'}^2 (\hat{\boldsymbol{\zeta}}_{j'}^T \boldsymbol{X}_i + \sum_{k \neq k', j'}^{p} \hat{\boldsymbol{\omega}}_{j'k}^T \boldsymbol{X}_i M_{ik} + \sum_{m \neq m'}^{q} \hat{\omega}_{j'k'm} X_{im} M_{ik'})]. \end{aligned}$$

A.2 Brain region abbreviations, full names, and their lobes

Table 3: Brain region abbreviations, full names, and their lobes

Abbreviation	Full Name	Lobe
lG_orbital	Left Orbital Gyri	Frontal
rG_orbital	Right Orbital Gyri	Frontal
lS_front_middle	Left Middle Frontal Sulcus	Frontal
lS_orbital_lateral	Left Lateral Orbital Sulcus	Frontal
rS_orbital_lateral	Right Lateral Orbital Sulcus	Frontal
lS_front_inf	Left Inferior Frontal Sulcus	Frontal
rS_front_inf	Right Inferior Frontal Sulcus	Frontal
rLat_Fis.ant.Horizont	Right Horizontal Ramus of the Anterior Segment of the Lateral Sulcus	Frontal
rG_and_S_paracentral	Right Paracentral Lobule and Sulcus	Frontal
rS_precentral.sup.part	Right Superior Part of the Precentral Sulcus	Frontal
rS_orbital_med.olfact	Right Medial Orbital Sulcus	Frontal
lS_central	Left Central Sulcus	Frontal/Parietal
rS_central	Right Central Sulcus	Frontal/Parietal
lS_cingul.Marginalis	Left Marginal Branch of the Cingulate Sulcus	Frontal/Parietal
rS_cingul.Marginalis	Right Marginal Branch of the Cingulate Sulcus	Frontal/Parietal
lG_insular_short	Left Short Insular Gyri	Insular
rG_insular_short	Right Short Insular Gyri	Insular
lS_circular_insula_ant	Left Anterior Segment of the Circular Sulcus of the Insula	Insular
rS_circular_insula_ant	Right Anterior Segment of the Circular Sulcus of the Insula	Insular
lS_circular_insula_inf	Left Inferior Segment of the Circular Sulcus of the Insular	Insular
rS_circular_insula_inf	Right Inferior Segment of the Circular Sulcus of the Insular	Insular
lG_oc.temp_med.Parahip	Left Parahippocampal Gyrus	Temporal
rG_oc.temp_med.Parahip	Right Parahippocampal Gyrus	Temporal

References

Chen Avin, Ilya Shpitser, and Judea Pearl. Identifiability of path-specific effects. *In Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 357–363, 2005.

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

Julian Besag. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society Series D: The Statistician*, 24(3):179–195, 1975.

Peter Bühlmann and Sara Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

Brian Caffo, Sining Chen, Walter Stewart, Karen Bolla, David Yousem, Christos Davatzikos, and Brian S Schwartz. Are brain volumes based on magnetic resonance imaging mediators of the associations of cumulative lead dose with cognitive function? *American Journal of Epidemiology*, 167(4):429–437, 2008.

Jiahua Chen and Zehua Chen. Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–71, 2008.

Oliver Y Chén, Ciprian Crainiceanu, Elizabeth L Ogburn, Brian S Caffo, Tor D Wager, and Martin A Lindquist. High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics*, 19(2):121–136, 2018.

Bruce N Cuthbert and Thomas R Insel. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Medicine*, 11(1):1–8, 2013.

Rhian M Daniel, Bianca L De Stavola, SN Cousens, and Stijn Vansteelandt. Causal mediation analysis with multiple mediators. *Biometrics*, 71(1):1–14, 2015.

Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1–15, 2010.

Simon Ducharme, Matthew D Albaugh, Tuong-Vi Nguyen, James J Hudziak, José María Mateos-Pérez, Aurelie Labbe, Alan C Evans, Sherif Karama, Brain Development Cooperative Group, et al. Trajectories of cortical thickness maturation in normal brain development—the importance of quality control procedures. *Neuroimage*, 125:267–279, 2016.

Bruce Fischl, Martin I Sereno, Roger BH Tootell, and Anders M Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human brain mapping*, 8(4):272–284, 1999.

Alex Fornito, Andrew Zalesky, and Edward Bullmore. *Fundamentals of brain network analysis*. Academic press, 2016.

Jean-Philippe Fortin, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylor, Irem Aselcioglu, Philip A. Cook, Phil Adams, Crystal Cooper, Maurizio Fava, Patrick J. McGrath, Melvin McInnis, Mary L. Phillips, Madhukar H. Trivedi, Myrna M. Weissman, and Russell T. Shinohara. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167:104–120, 2018.

Mark B Gerstein, Anshul Kundaje, Manoj Hariharan, Stephen G Landt, Koon-Kiu Yan, Chao Cheng, Xinmeng Jasmine Mu, Ekta Khurana, Joel Rozowsky, Roger Alexander, et al. Architecture of the human regulatory network derived from encode data. *Nature*, 489(7414):91–100, 2012.

Stephan Geuter, Elizabeth A Reynolds Losin, Mathieu Roy, Lauren Y Atlas, Liane Schmidt, Anjali Krishnan, Leonie Koban, Tor D Wager, and Martin A Lindquist. Multiple brain networks mediating stimulus–pain relationships in humans. *Cerebral Cortex*, 30(7):4204–4219, 2020.

Leonardo L Gollo, Andrew Zalesky, R Matthew Hutchison, Martijn Van Den Heuvel, and Michael Breakspear. Dwelling quietly in the rich club: brain network determinants of slow cortical fluctuations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668):20140165, 2015.

Logan Harriger, Martijn P Van Den Heuvel, and Olaf Sporns. Rich club organization of macaque cerebral cortex and its role in network communication. *PloS One*, 7(9):e46497, 2012.

Jonas M. B. Haslbeck and Lourens J. Waldorp. mgm: Estimating time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software*, 93(8):1–46, 2020.

Yong He, Zhang J Chen, and Alan C Evans. Small-world anatomical networks in the human brain revealed by cortical thickness from mri. *Cerebral Cortex*, 17(10):2407–2419, 2007.

Yen-Tsung Huang. Genome-wide analyses of sparse mediation effects under composite null hypotheses. *Annals of Applied Statistics*, 13(1):60–84, 2019.

Yen-Tsung Huang and Wen-Chi Pan. Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics*, 72(2):402–413, 2016.

Nicole R Karcher and Deanna M Barch. The abcd study: understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology*, 46(1):131–142, 2021.

Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

Martin A Lindquist. Functional causal mediation analysis with an application to brain connectivity. *Journal of the American Statistical Association*, 107(500):1297–1309, 2012.

Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34:1436–1462, 2006.

Caleb H Miles, Ilya Shpitser, Phyllis Kanki, Seema Meloni, and Eric J Tchetgen Tchetgen. On semiparametric estimation of a path-specific effect in the presence of mediator-outcome confounding. *Biometrika*, 107(1):159–172, 2020.

KM Miller, CC Price, MS Okun, H Montijo, and D Bowers. Is the n-back task a valid neuropsychological measure for assessing working memory? *Archives of Clinical Neuropsychology*, 24(7):711–717, 2009.

Judea Pearl. Interpretation and identification of causal mediation. *Psychological Methods*, 19(4):459–481, 2014.

Jie Peng, Pei Wang, Nengfeng Zhou, and Ji Zhu. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104(486):735–746, 2009.

James M Robins and Sander Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2):143–155, 1992.

Kristi M. Sawyer, Patricia A. Zunszain, Paola Dazzan, and Carmine M. Pariante. Intergenerational transmission of depression: clinical observations and molecular mechanisms. *Molecular Psychiatry*, 24(8):1157–1177, 2019.

Cosma Shalizi. Advanced data analysis from an elementary point of view. 2013.

Peter Uetz, Loic Giot, Gerard Cagney, Traci A Mansfield, Richard S Judson, James R Knight, Daniel Lockshon, Vaibhav Narayan, Maithreyan Srinivasan, Pascale Pochart, et al. A comprehensive analysis of protein–protein interactions in saccharomyces cerevisiae. *Nature*, 403(6770):623–627, 2000.

Martijn P Van den Heuvel and Olaf Sporns. Network hubs in the human brain. *Trends in Cognitive Sciences*, 17(12):683–696, 2013.

Tyler VanderWeele and Stijn Vansteelandt. Mediation analysis with multiple mediators. *Epidemiologic methods*, 2(1):95–115, 2014.

Tyler J VanderWeele. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, 2015.

Tyler J VanderWeele and Peng Ding. Sensitivity analysis in observational research: introducing the e-value. *Annals of internal medicine*, 167(4):268–274, 2017.

Tor D Wager, Christian E Waugh, Martin Lindquist, Doug C Noll, Barbara L Fredrickson, and Stephan F Taylor. Brain mediators of cardiovascular responses to social threat: part I: reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. *Neuroimage*, 47(3):821–835, 2009.

Shanghong Xie, Xiang Li, Peter McColgan, Rachael I Scahill, Donglin Zeng, and Yuanjia Wang. Identifying disease-associated biomarker network features through conditional graphical model. *Biometrics*, 76(3):995–1006, 2020.

Ming Yuan and Yi Lin. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

Haixiang Zhang, Yinan Zheng, Zhou Zhang, Tao Gao, Brian Joyce, Grace Yoon, Wei Zhang, Joel Schwartz, Allan Just, Elena Colicino, Pantel Vokonas, Lihui Zhao, Jinchi Lv, Andrea Baccarelli, Lifang Hou, and Lei Liu. Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics*, 32(20):3150–3154, 06 2016.

Yi Zhao and Xi Luo. Pathway lasso: estimate and select sparse mediation pathways with high dimensional mediators. *Statistics and its Interface*, 15:39–50, 2022.

Yi Zhao, Martin A. Lindquist, and Brian S. Caffo. Sparse principal component based high-dimensional mediation analysis. *Computational Statistics & Data Analysis*, 142:106835, 2020.

Yi Zhao, Lexin Li, and Brian S Caffo. Multimodal neuroimaging data integration and pathway analysis. *Biometrics*, 77(3):879–889, 2021.