Bayesian network mediation analysis with application to the brain functional connectome

Yize Zhao¹² | Tianqi Chen¹ | Jiachen Cai¹ | Sarah Lichenstein³ | Marc N. Potenza³⁴⁶⁷⁸ | Sarah W. Yip³⁴

¹Department of Biostatistics, Yale University School of Public Health, New Haven, Connecticut, USA
²Yale Center for Analytical Sciences, Yale University School of Public Health, New Haven, Connecticut, USA
³Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut, USA
⁴Child Study Center, Yale University School of Medicine, New Haven, Connecticut, USA
⁵Department of Neuroscience, Yale University School of Medicine, New Haven, Connecticut, USA
⁶Connecticut Mental Health Center, New Haven, Connecticut, USA
⁷Connecticut Council on Problem Gambling, Wethersfield, Connecticut, USA
⁸Wu Tsai Institute, Yale University School of Medicine, New Haven, Connecticut, USA

Correspondence
Yize Zhao, Department of Biostatistics, Yale University School of Public Health, 300 George Street, New Haven, CT 06511, USA.
Email: yize.zhao@yale.edu

Funding information
National Center for Advancing Translational Sciences, Grant/Award Number: UL1TR001863; National Institute on Aging, Grant/Award Numbers: 1RF1AG068191, P30AG021342; National Institute on Drug Abuse, Grant/Award Numbers: K01DA039299, P50DA09241, R01DA039136

Statistics in Medicine. 2022;41:3991–4005. wileyonlinelibrary.com/journal/sim © 2022 John Wiley & Sons Ltd.
1 | INTRODUCTION

In the past few decades, the investigation of brain functional organization through a collective set of functional brain connections, known as the functional connectome or connectivity, is a rapidly growing research area that has provided novel insights into large-scale neuronal communication and how individual differences in brain functional networks relate to human behavior and psychiatric disorders. With the latest development of modern brain imaging technology, functional imaging such as positron emission tomography and functional magnetic resonance imaging (fMRI) can provide insight into the foundations of human behavior. These studies provide unique opportunities for integrating complex, whole brain network measures with complex effect pathway analyses, such as those uncovering the underlying effect mechanisms between an exposure and an outcome.

This work is motivated by a recent study on opioid use disorder (OUD) \(^1\) to investigate the impact of a designed treatment on the opioid abstinence. Within the study, 74 OUD patients were recruited from a randomized controlled trial to receive either behavioral therapy plus galantamine or placebo treatment. At the end of the trial, fMRI data were acquired across the whole brain over multiple scans during both resting-state and reward (Monetary Incentive Delay) tasks, and were subsequently used to compute individual functional connectivity matrices (ie, connectomes) for each participant. Our goal here is to quantify the effect mechanisms in the context of a mediation framework by explaining how the therapy impacts the post-treatment abstinence mediated by the interconnected brain circuits along functional networks.

Mediation analysis was initially introduced in social and behavioral research and gradually became a popular analytical tool in other disciplines, including causal inference in statistics. \(^2\), \(^3\) Under certain assumption regulations, a traditional mediation model investigates two effect paths—a direct one straightly from the exposure (eg, treatment) to the outcome (eg, abstinence), and an indirect one intervened by an additional variable, known as a mediator. Among existing mediation analyses, most focus on a single mediator variable and characterize different effects through classical linear regressions. \(^4\) Lately, with the emergence of more complex data structure, mediation analysis has been extended to handle multivariate \(^5\)-\(^10\) and high-dimensional mediators \(^11\)-\(^13\) to tackle specific applications. Particularly, in the field of brain imaging, there is a growing trend to use quantitative neuroimaging measurements as mediators to study the effect mechanism in order to reveal the neuropsychological processes after an intervention and how they further modify human behavior, thus informing future clinical strategies. For example, Lindquist \(^14\) considered regional fMRI time series as an imaging mediator and extended mediation model to functional data; Zhao and Luo \(^15\) further extended the applicable capacity to multiple time series courses. Given the large-scale nature of imaging data, attempts have also been made to handle mediation analysis with a high-dimensional mediator vector. \(^16\) Within all these studies, the considered mediators are concentrated on brain regional measurements which may be correlated but essentially act as distinct units to characterize intermediated variation. However, given the intricate morphology of the human brain, particularly as reflected in neuronal processing interconnectivity, investigating the mediation role neural system plays based on a more state-of-the-art brain connectome traits understanding becomes an exciting and urgent direction.

To this end, we propose here an integrative mediation analysis with networks entailing the mediation effect from the exposure to the outcome. To parameterize network measurements (ie, functional connectomes, in our case) within our unified Bayesian framework, we construct individually specified stochastic block models (SBMs) with the unknown connectivity weights across the connectome modules serving as the latent network mediators. To further remove noninformative components and identify reliable mediating sub-networks over whole brain connectomes, we simultaneously impose a variable selection procedure within our Bayesian model to facilitate identification of distinctions between active and inactive mediating effects. It is worth noting that with a growing interest in studying brain connectivity, multiple studies have been conducted to examine partial or whole brain connectivity to predict behaviors or to integrate with molecular data. Among those, a considerable amount of the analyses simplified the network representation into individual connections, leading to a considerable loss of graphic topological information. In contrast, by retaining the matrix structure of the brain connectome, depending on its role, a number of scalar-to-network and network-to-scalar models have recently been proposed under decomposition, \(^17\) penalization, \(^18\) or graphic models. \(^19\) Despite those studies considered to link brain connectomes with other data constructs, their objectives and modeling schemes are not applicable to our problem with connectomes facilitating mediation.

The major contributions of this article include the following aspects. First, we make the very first attempt to build a mediation analysis with network-variate mediators. Motivated by investigating the mediating effect of brain functional connectomes, we propose a unified Bayesian mediation framework that is generally applicable to study mediated effects within a network format. Second, we simultaneously gain a better understanding of the brain’s intrinsic functional
organization in light of their mediating effects. Utilizing weighted SBMs with unknown block structure embedded within a mediation paradigm, the estimated brain network modules identify neural network components that are likely more clinically relevant than those identified using an unsupervised approach. Finally, we also accomplish selection among network-induced mediators, which removes inactive network features and enhances the analytical power.

The remainder of the article is organized as follows. We present our assumptions and model formulation for the network-variate mediators and different effects in Section 2, followed by the corresponding Bayesian framework on prior specification and posterior inference in Section 3. We assess the model performance by simulations in Section 4, and apply the method to the OUD study in Section 5. We conclude the article with a discussion in Section 6.

2 | METHODS

We start with data structure. For subject \( i, i = 1, \ldots, N \), let \( y_i \) denote the outcome, \( X_i \) a set of \( P \) clinical covariates, and \( z_i \) the binary treatment. For instance, in the OUD study, \( y_i \) summarizes the urine test results, \( X_i \) includes age and gender, and \( z_i \) indicates receiving active galantamine or placebo. For each subject, task-based functional MRI (fMRI) data are acquired \( K_i \) times. Then, each fMRI time-series course is summarized into a brain functional connectivity network \( G_{ik} = (V, E_{ik}) \) with vertex set \( V = \{1, \ldots, V\} \) including \( V \) brain nodes or regions of interest (ROI) defined by a brain atlas, and edge set \( E_{ik} \) collecting functional connections among the nodes. To facilitate analysis, we further represent network \( G_{ik} \) by its corresponding weighted adjacency matrix \( A_{ik} = (a_{ik,jl}) \), which is symmetric with \( a_{ik,jl} \) indicating the connectivity strength between nodes \( j \) and \( l, 0 \leq j, l \leq V \). Of note, one could also dichotomize each element \( a_{ik,jl} \) by a certain threshold to indicate the existence of a connection. In our model illustration and numerical studies, we choose the former way in keeping a continuous scale, but point out that in the case when brain networks are binary, our method can be adjusted by modifying prior specifications and inference algorithms.

In the above context, our goal is to establish effect pathways from the treatment to the outcome mediated by the repeatedly measured brain networks. However, properly modeling network measurements as potential mediators within a mediation analysis is a challenging task and has not yet been investigated. Although one could potentially extract all the unique functional connections throughout the brain and implement existing single- or multivariate-mediator model frameworks, such operations would lead to information loss by not fully considering the topological configuration within brain networks. In addition, given the functional connectivity is non-sparse, the number of unique connections to be considered is \( V(V - 1)/2 \). In brain imaging applications, most of the brain atlases contain at least a moderate number of nodes \( V \) (e.g., \( V = 268 \) in our application). Directly modeling such a large number of edges as mediators would unavoidably bring prohibitive computation complexity along with hurdles in multiple comparisons or model specifications. Motivated by these analytical challenges, we propose here an integrative Bayesian mediation analysis to characterize potential causal effects among the outcome, treatment, and network mediators where the networks are hierarchically modeled by joint SBMs to provide latent mediator surrogates. We demonstrate that our modeling framework permits a desired format to quantify individual path of effect and induce a sub-network level selection with respect to mediation effects.

2.1 | Latent mediators induced by the SBMs

Converging evidence suggests that functional organization of the brain encompasses the cognitive processes through sub-networks or sub-components.\(^{20}\) This supports our anticipation that the brain functional connectome is engaged in the mediation pathways through network modules. Considering prior research, though multiple brain sub-network parcellations under resting or task-based functional connectivity have been constructed,\(^ {21,22} \) they are often formed without the supervision of specific clinical procedures. To enhance analytical power and establish more tractable mediation effects, without pre-specifying the sub-community structure, we assume \( V \) can be divided into \( Q \) unknown latent blocks. For each node \( v \in V \), we introduce a random vector \( G_v = (g_{v1}, \ldots, g_{vQ}) \) to capture community allocation with latent indicator \( g_{vq} = 1 \) if node \( v \) belongs to block \( q \). Thus, \( G_v \) follows a multinomial distribution \( G_v \sim \text{Multi}(1, \pi) \) with \( \pi = (\pi_1, \ldots, \pi_Q) \) the vector of allocation probabilities.

With \( K_i \) functional connectomes collected for subject \( i \), for each of the connectivity matrices \( A_{ik} \), we construct a weighted SBM with individual edges following normal distributions given the community structure

\[
a_{ik,jl}|g_{jq} = 1, g_{lr} = 1 \sim N(m_{ik,jq}, \sigma^2_{qqr}),
\]

(1)
for \( i = 1, \ldots, N; k = 1, \ldots, K; 1 \leq j \neq l \leq V; 1 \leq q, r \leq Q \). Here, \( m_{ikqr} \) is the expected subject/measurement-level connection strength between blocks \( q \) and \( r \); and \( \sigma^2_{qr} \) is the variance. Due to the symmetric structure, we have \( m_{ikqr} = m_{ikrq} \) and \( \sigma^2_{qr} = \sigma^2_{rq} \). Based on model (1), connections within the network become independent given the block membership of the nodes with which they link. Essentially, this approach allows us to transfer the representation of each connectivity matrix to a number of sub-network modules with the topological organization absorbed in the community matrix represented by \( G = (G_1, \ldots, G_V) \), and the modular characteristics quantified by the strength parameter \( m_{ikqr} \) and variance \( \sigma^2_{qr} \) for each block. Particularly, given \( m_{ikqr} \) is not embedded with any topological information, so we can consider it as a regular scalar parameter. It is worth noting that model (1) introduces sub-network encoding for each individual fMRI course from \( \omega^2 \) the vector of latent mediator values when treatment \( z_i \) and (2) pave away to facilitate the characterization of network-variate mediators via these unobserved sub-network connectivity, which we call the latent network mediators. We borrow the word “latent” from Albert et al.\(^\text{24} \) to distinguish our proposed mediating variables \( M_i = (m_{iqr}, 1 \leq q \leq r \leq Q) \) with the commonly used fully observed mediators. With the regulation from brain connectivity topology completely undertaken by the modular allocation, the latent network mediators within \( M_i \) are conditionally independent. As a valuable property, such an independence is anticipated to stabilize our model fitting and improve the parameter estimation,\(^\text{16,25} \) as has been shown under conditions with mediator orthogonality.

### 2.2 Mediation model

Under the assumption that the latent mediator set \( M_i \) directs a partial treatment effect, the relationship among the treatment \( z_i \), outcome \( y_i \), covariates \( X_i \), brain connectivity matrices \( \{A_k\}_{k=1}^K \), and their induced latent mediators \( M_i \) is displayed inside of Figure 1. To formally establish potential causal effects, under a counterfactual representation, we denote \( M_i(Z) \) the vector of latent mediator values when treatment \( z_i = Z \), and \( Y_i(Z, M_i^*) \) is the outcome value when \( z_i = Z \) and latent mediators \( M_i = M_i^* \). Based on these, the natural direct effect (NDE), natural indirect effect (NIE), and total effect (TE) can be defined by

\[
\begin{align*}
\text{NDE} & = E(y_i(Z, M_i(Z^*))) - E(y_i(Z^*, M_i(Z^*_i))), \\
\text{NIE} & = E(y_i(Z, M_i(Z))) - E(y_i(Z, M_i(Z^*_i))), \\
\text{TE} & = E(y_i(Z, M_i(Z))) - E(y_i(Z^*, M_i(Z^*_i))),
\end{align*}
\]

with NDE measuring the expected change on outcome by switching treatment from \( Z^* \) (e.g., control) to \( Z \) (e.g., drug) while maintaining mediators as the original values; NIE quantifying the expected change on outcome when mediators change from \( M_i(Z^*) \) to \( M_i(Z) \) while fixing treatment; and TE characterizing the overall change on outcome by switching treatment with \( \text{TE} = \text{NDE} + \text{NIE} \). In order to identify NDE and NIE from (3), a number of assumptions known as the sequential ignorability assumptions\(^\text{36} \) are required for the latent mediators. Specifically, we assume that (1) \( y_i(Z, M_i) \perp z_i|X_i \), (2) \( M_i(Z) \perp z_i|X_i \), (3) \( y_i(Z, M_i) \perp M_i(Z^*_i)|X_i \), (4) \( y_i(Z, M_i) \perp M_i(Z)|z_i, X_i \). These assumptions describe that no unmeasured confounding for the relationships between outcome and treatment, mediators and treatment, and outcome and mediators with and without additional controlling of treatment. This set of assumptions serves as the standard requirements for a mediation paradigm and has already been extensively discussed previously including those with respect to multivariate and high dimensional mediators.\(^\text{5,13} \) latent mediators,\(^\text{12} \) and neuroimaging applications.\(^\text{14,16} \) Notably, as discussed
in Chen et al.,\textsuperscript{16} though it is extremely hard to rigorously validate the sequential ignorability assumptions in real practice, when one or more of the assumptions fail to hold, we can still use (3) to quantify potential mediating effects in exploratory analyses. Therefore, without making strong causal claims, our analysis can provide insights into exploratory mediation effects. In addition, for our application, though connections within a brain connectome are correlated with each other, our constructed latent mediators are conditionally independent based on the SBM assumption. Therefore, despite our main objective being the uncovering of the average effect pathway bridging the treatment, brain activation along connectomes, and outcome, we do have the capacity to characterize the modular specific indirect effects.

Given the both our outcome and latent mediators are normally distributed, we build regression models for the conditional distributions $y_i|M_i,X_i,z_i$ and $m_{iqr}|X_i,z_i$, $1 \leq q \leq r \leq Q$. Previous empirical studies of individual connections have indicated that a certain proportion of connectivity features within the brain indeed identifies negligible variations between the treatment and placebo groups.\textsuperscript{27} Some recent work on connectome-based prediction also reveals that several key sub-networks play a dominant role in predicting people’s behavior.\textsuperscript{28} The complication of brain neural circuits with respect to their impacts on behavior and correspondence to exposure motivates us to make the following biologically plausible modeling assumptions: first, only a subset of brain network modules and their induced latent network mediators are significantly impacted by the treatment; second, only a subset of brain network modules and their induced latent network mediators significantly influence the outcome. To this end, we have the following sparse regressions

$$y_i = X_i^T \beta_x + M_i^T (\beta_m \circ \tau) + z_i \beta_z + \epsilon_i$$
$$m_{iqr} = X_i^T \alpha_{xqr} + z_i (\alpha_{zqr} \cdot \gamma_{qr}) + \psi_{iqr}, \quad 1 \leq q \leq r \leq Q.$$ (4)

where $X_i$ includes 1 as the first element; $(P+1) \times 1$ vector $\beta_x$, $Q(Q+1)/2 \times 1$ vector $\beta_m = (\beta_{mqr}, 1 \leq q \leq r \leq Q)^T$ and scalar $\beta_z$ represent the effects of covariates, mediators and treatment on the outcome; $(P+1) \times 1$ vector $\alpha_{xqr}$ and
scalar $\alpha_{zqr}$ are the coefficients for covariates and treatment on mediator $m_{iqr}$; and there are random errors $\epsilon_i \sim N(0, \sigma_{\epsilon}^2)$, $\psi_{iqr} \sim N(0, \sigma_{\psi}^2)$. Inside the models (4), we also include latent selection indicator sets $\tau = (\tau_{qr} \in [0, 1], 1 \leq q \leq r \leq Q)$ and $\gamma = (\gamma_{qr} \in (0, 1), 1 \leq q \leq r \leq Q)$, both with a dimension $Q(Q + 1)/2 \times 1$ to impose the desired sparsity, where $\odot$ denotes the entry-wise product. Specifically, we have $\tau_{qr} = 1$ if latent mediator $m_{iqr}$ brings significant impact to outcome, 0 otherwise; and $\gamma_{qr} = 1$ if the treatment effect is significant on mediator $m_{iqr}$, 0 otherwise. Given the fact that brain connectivity modules are influenced by the treatment and those impacting the outcome can be different, to allow flexibility, we use separate selection indicator sets here without any constrain to match between $\tau$ and $\gamma$.

Finally, we impose noninformative IG priors with shape and scale parameters set to be 0.1 for $\alpha_{zqr}$ and $\gamma_{qr}$, to allow flexibility, we use separate selection indicator sets here without any constrain to match between $\tau$ and $\gamma$.

Based on models (3) and (4), under the above sequential ignorability assumptions, the average direct and indirect effects can be identified and expressed as

$$\begin{align*}
NDE &= \beta_\gamma (Z - Z^*), \\
NIE &= (Z - Z^*) \sum_{q \leq r \leq Q} (\alpha_{zqr} \cdot \gamma_{qr})(\beta_{mqr} \cdot \tau_{qr}), \\
TE &= (Z - Z^*) \left\{ \beta_z + \sum_{q \leq r \leq Q} (\alpha_{zqr} \cdot \gamma_{qr})(\beta_{mqr} \cdot \tau_{qr}) \right\}.  \quad (5)
\end{align*}$$

According to the above representation, a latent mediator $m_{iqr}$ will contribute to both NIE and TE only if both $\tau_{qr}$ and $\gamma_{qr}$ are nonzero. In other words, the truly active latent mediators are the ones that are jointly selected in both models in (4). Song et al.\textsuperscript{13} also discussed different patterns of activation within potential mediators. In their paper, they use shrinkage priors on coefficients to regulate small effects to achieve quasi-sparsity and define active mediators as the ones with both related coefficients coming from the larger normal components. Here, we directly impose sparsity using selection indicators, which leads to an explicit definition for each latent network mediator’s activation based on its inclusion or exclusion from the averaged effects.

### 3 A UNIFIED BAYESIAN FRAMEWORK

To jointly model the stochastic block structure of the functional connectome and its associated mediation analysis, we develop a Bayesian estimation and inference framework for models (1), (2), and (4). This enable us to overcome the difficulty in performing inference under a conventional frequentist mediation analysis,\textsuperscript{29} and effectively integrate the construction of mediated brain modules within the effect pathway establishment.

#### 3.1 Prior specification

We assign prior for each unknown parameter within the proposed hierarchical models. For the SBM-related parameters, given that $G_v$ follows a multinomial distribution with probability $\pi$, we have $\pi$ follow a conjugate Dirichlet distribution $\pi \sim \text{Dir}(p_1, \ldots, p_Q)$. Each within block variance is set to follow an inverse gamma (IG) distribution $\sigma_{\epsilon}^2 \sim \text{IG}(a_1, b_1)$, and each inner measurement variance $\sigma_{zm}^2 \sim \text{IG}(a_2, b_2)$. To determine the total number of blocks $Q$, a number of tools have been developed for SBMs using different evaluation metrics such as hypothesis testing,\textsuperscript{30} cross validation,\textsuperscript{31} and Bayes Factor,\textsuperscript{32} among others. Through our analytical experiments, we determine $Q$ based on the Integrated Completed Likelihood (ICL) criterion,\textsuperscript{33} which shows a more robust performance in our numerical studies. In terms of the parameters involved in the regression models, we assign Bernoulli distributions for each side of the mediating effect selection indicators $\gamma_{qr}$ and $\tau_{qr}$, with probabilities $p_\gamma$ and $p_\tau$, respectively. For the regression coefficients, we assume each $\alpha_{zqr} \sim N(0, \sigma_{\alpha}^2)$, $\beta_{mqr} \sim N(0, \sigma_{\beta}^2)$, and the coefficients $\beta_z \sim N(0, \sigma_{\beta}^2)$, $\beta_z \sim N(0, \sigma_{\beta}^2)$, and each $\alpha_{xqr} \sim N(0, \sigma_{\alpha}^2)$, $1 \leq q \leq r \leq Q$. Finally, we impose noninformative IG priors with shape and scale parameters set to be 0.1 for $\sigma_{\alpha}^2, \sigma_{\beta}^2, \sigma_{\alpha}^2, \sigma_{\beta}^2, \sigma_{\gamma}^2$; and set $\sigma_{\epsilon}^2, \sigma_{zm}^2, \sigma_{zm}^2$ to be large values for example, 10. We name our unified Bayesian mediation analysis the Bayesian network mediation model (BNMM) and the core modeling structure along with all the prior specifications is further included in Figure 1.
3.2 | Posterior inference

Given the outcome, exposure, repeatedly measured brain connectomes and covariates, we first write down the joint conditional posterior distribution for the unknown parameters \( \Phi = (G, \beta_x, \beta_m, \tau, \gamma, \beta_c, \alpha_{xqr}, \alpha_{zqr}, \sigma_1^2, \sigma_2^2, M_i, m_{ikqr}, \sigma_q^2, \omega_q^r, k = 1, \ldots, K_i, i = 1, \ldots, N; 1 \leq q \leq r \leq Q) \) given the observed data

\[
f(\Phi|y_i, z_i, X_i, A_{ik}, k = 1, \ldots, K_i; i = 1, \ldots, N) \propto \prod_{i=1}^{N} f(y_i|z_i, X_i, \beta_x, \tau, \beta_c, \sigma_1^2) \prod_{k=1}^{Q} \prod_{1 \leq q \leq Q} f(A_{ik}|m_{ikqr}, \sigma_q^2, G)f(m_{ikqr}|m_{iqr}, \sigma_q^2)f(m_{iqr}|X_i, z_i, \alpha_{xqr}, \alpha_{zqr}, \gamma_{qr}, \sigma_2^2) \prod_{k=1}^{Q} \prod_{1 \leq q \leq Q} f(\sigma_q^2)f(\omega_q^r)f(\gamma_{qr})f(\gamma_{qr})f(\beta_x)f(\beta_c)f(\tau)f(\sigma_1^2)f(\sigma_2^2).
\]

We rely on a Markov Chain Monte Carlo (MCMC) algorithm to conduct the posterior inference. Since we can directly obtain the analytical form of the full conditional distribution of each unknown parameter, the computational cost is not heavy through Gibbs sampler. We briefly describe the sampling steps for each MCMC iteration here and provide the full algorithm in the supporting web materials.

- Given the current value of \( \tau \), split \( \beta_m \) into \( \beta_{m0} \) and \( \beta_{m1} \) corresponding to the unselected \( (\tau = 0) \) and selected \( (\tau = 1) \) brain network blocks for the outcome regression. Update \( \beta_{m0} \sim N(0, \sigma_{m0}^2 I) \); and \( \beta_{m1}|M, \beta_x, \beta_c, \sigma_{m0}^2, \sigma_1^2 \) from its posterior multivariate normal distribution, where \( M = (M_1, \ldots, M_n)^T \).

- Given the current value of \( \gamma_{qr} \), denote \( \alpha_{z0} = (\alpha_{xqr} : \gamma_{qr} = 0) \) and \( \alpha_{z1} = (\alpha_{xqr} : \gamma_{qr} = 1) \) corresponding to brain network blocks receiving or not receiving treatment impact. Update \( \alpha_{z0} \sim N(0, \sigma_{z0}^2 I) \) and \( \alpha_{z1}|\alpha_x, M, \sigma_{z0}^2, \sigma_1^2 \) from its posterior multivariate normal distribution.

- Define \( l_r(\tau_{qr}|M, \tau_{-qr}, \beta_m, \beta_x, \beta_c, \sigma_1^2) := \sum_{y} \Phi[z_i|x, \beta_m, \beta_x, \beta_c, \sigma_1^2] \) with \( \Phi(x) = x^2x/2 \). Update \( \tau_{qr}, 1 \leq q \leq r \leq Q \) from a Bernoulli distribution with probability \( p_r l_r(1)/(p_r l_r(1) + (1 - p_r) l_r(0)) \). In a similar way, we can define \( l_r(\gamma_{qr}|M, \gamma_{-qr}, \alpha_{xqr}, \alpha_{zqr}, \sigma_2^2) \), and update each \( \gamma_{qr} \sim \text{Bern}(p_r l_r(1)/(p_r l_r(1) + (1 - p_r) l_r(0))) \).

- Update nuisance parameters \( \beta_x|M, \beta_m, \tau, \beta_c, \sigma_1^2 \) and \( \alpha_{xqr}|M, \alpha_{zqr}, \gamma_{qr}, \sigma_2^2 \) with \( 1 \leq q \leq r \leq Q \) from their corresponding posterior multivariate normal distributions.

- Update \( \beta_c|M, \beta_x, \beta_m, \tau, \sigma_1^2 \) from its posterior normal distribution.

- Update latent mediators \( M_i|\{m_{ikqr}\}, \beta_x, \beta_m, \tau, \beta_c, \alpha_{xqr}, \alpha_{zqr}, \sigma_1^2, \sigma_2^2, \omega_q^r \) with \( i = 1, \ldots, N \) from the posterior multivariate normal distribution. The conditional subject/measure-level connection strength \( m_{ikqr}|M_i, G, \omega_q^r, \sigma_q^2 \) with \( k = 1, \ldots, K_i, i = 1, \ldots, N, 1 \leq q \leq r \leq Q \) is updated from the posterior normal distribution.

- Update community allocation \( G_v|\pi, \{m_{ikqr}\}, \sigma_q^2 \) for \( v = 1, \ldots, V \) from its posterior multinomial distribution. The hyper-parameter \( \pi|G \) is updated from the posterior Dirichlet distribution.

- Update \( \sigma_q^2|\beta_x, M, \beta_m, \tau, \beta_c, Z; \sigma_2^2|M, \alpha_{xqr}, \gamma; \omega_q^r|\{M_i\}, \{m_{ikqr}\}, \{m_{ikqr}\}, \sigma_{z0}^2, |\alpha_{xqr}|, \sigma_{m0}^2, |\beta_{mqr}| \) from their corresponding posterior IG distributions.

For implementation, we start with random initials for multiple chains and check the posterior convergence after completing the inference by both trace plots and GR method.\(^{34}\) To determine the truly active latent network mediators and the overall NIE and TE, we take the posterior median model\(^{15}\) corresponding to a 0.5 threshold on the marginal posterior inclusion probabilities of \( \tau \) and \( \gamma \) for the effect related parameters. The parcellation of brain sub-networks within the mediation pathways are summarized by the posterior mode of each \( G_v, v = 1, \ldots, V \) to map each brain region to its mediating-related module.

4 | SIMULATION

We next evaluate the finite sample performance of our proposed BNMM compared with existing alternatives by simulations. We consider two data dimensions—(1) 50 subjects with each subject having 6 repeated measures of the connectome over 100 nodes; and (2) 500 subjects with each subject having 6 repeated measures of the connectome over 200 nodes. We also consider both a continuous and a binary exposure, which are primarily distinct on the result interpretation without
a significant modification of our model implementation. However, given the implementation for one of the major competing approaches\textsuperscript{13} is designed for a continuous exposure, we focus on illustrating this case with the exposure generated from a standard normal distribution. To construct latent network mediators, we first set $Q = 10$, and generate the community allocation $G_v$ for each node $v$ from a Dirichlet-multinominal distribution with concentration parameters all set to 3. We include one covariate generated from a standard normal distribution along with an intercept in both exposure and mediator model. Following (4), (2), and (1), we generate the latent network components $\{m_{iqr}\}$, $\{m_{ikqr}\}$, outcome $\{y_i\}$ and brain connectivity matrices $\{A_{ik}\}$ where $\beta_z = 1.5$, $\beta_m = 2$, $\beta_x = (1, 1)$, $\alpha_{zqr} = (0.3, 0.5)$, $\omega_{qr} = 0.1$ and $\alpha_{zqr}, 1 \leq q \leq r \leq Q$ range from 1.5 to 2.5. We set the variance parameters $\sigma_2 = \sigma_{qr} = 0.5$, $\sigma_1 = 1$. For the active latent brain sub-networks along the mediation pathways marked by $\tau$ and $\gamma$, we consider two scenarios. In the first scenario, we set $\tau$ and $\gamma$ to be the same which means the latent mediators that impact the outcome are those and only those influenced by the exposure. This is the prerequisite assumption for any univariate- or multivariate-mediator analysis without imposing mediator selection, but could be unrealistic in clinical studies. In the second scenario, we mimic the real world setting where the non-zeros in $\tau$ and $\gamma$ may be overlapped but contain difference. Hence, we allow certain brain sub-networks to soak impact from the exposure without responding to the outcome or intrinsically alter the outcome without interference by the exposure, and the truly active mediators are the ones that hold effects in both directions. Figure 2 demonstrates the generated sub-network structure and different signal patterns within a connectome to reflect the above considerations under the second scenario.

We generate 100 Monte Carlo datasets for each setting. To implement the BNMM, we set noninformative hyper-priors for the variance parameters as mentioned in the posterior computation in addition to $a_1 = b_1 = a_2 = b_2 = 1$, and assign $p_\tau = p_\gamma = 0.5$. The MCMC is conducted over 5000 iterations with 2000 burn-in. In terms of the alternative methods, given that none of the existing models can be directly applied to handle network-variate mediators, we ought to transfer connectivity networks into a different format in order to implement competing approaches. Specifically, since our proposed BNMM jointly models the latent mediators derived from observed networks and the effect pathways, for the competing methods, we consider (1) a direct use of the true latent sub-network connectivity strengths as mediators, that is, pretending $M_i$ are fully observed with their true values as input; and (2) a direct extraction of unique connections from the averaged brain connectivity matrix over repeated measures as mediators. The mediators are multivariate under both strategies, and we apply an univariate mediation analysis implemented by the R package Mediation on each individual mediator sequentially as well as a recent Bayesian mediation model on high-dimensional mediators (BAMA)\textsuperscript{13} implemented by the R package BAMA. Overall, we implement four additional approaches: univariate mediation analysis under true latent mediator (UMLM), BAMA under true latent mediator (BAMALM), univariate mediation analysis under brain connections (UMBC), and BAMA under brain connections (BAMABC). Of note, the above strategies weaken the influence from the modeling network structure which is infeasible for the competing methods. Particularly, the implementation setting is in favor of the first two competing approaches when the true latent mediators detached from the network configuration are directly imposed, compared with the BNMM with the latent mediators needing to be jointly modeled. Finally, we evaluate both the selection of the truly active mediators measured by sensitivity and specificity, and

![Figure 2](image-url)  
**Figure 2** Generated block structure and the informative sub-networks that are highlighted from a randomly selected brain connectome from the Scenario 2. A set of the induced latent network mediators by the informative sub-networks are (A) associated with the exposure, (B) contributing to the outcome, and (C) the truly active latent mediators.
| Scenario 1 | Sensitivity | Specificity | Bias of NDE | Bias of NIE | Bias of TE |
|------------|-------------|-------------|-------------|-------------|-------------|
| N = 50; V = 100 | BNMM | 0.98 (0.05) | 1.00 (0.00) | 4.22 (0.49) | -4.25 (0.49) | 0.03 (0.08) |
| | UMLM | 0.70 (0.19) | 1.00 (0.00) | 27.58 (0.13) | -8.37 (6.83) | 19.21 (6.31) |
| | BAMALM | 0.51 (0.43) | 1.00 (0.00) | 8.20 (0.43) | 4.99 (34.93) | 13.20 (28.45) |
| | UMBC | 0.61 (0.16) | 0.95 (0.01) | 27.64 (0.11) | 1394.65 (504.07) | 1422.29 (504.00) |
| | BAMABC | 0.05 (0.22) | 1.00 (0.02) | -5.28 (1.31) | 438.07 (2061.96) | 432.79 (2061.83) |
| N = 500; V = 200 | BNMM | 0.98 (0.10) | 1.00 (0.00) | 0.24 (0.59) | -0.24 (0.58) | 0.01 (0.08) |
| | UMLM | 1.00 (0.00) | 1.00 (0.00) | 23.90 (0.14) | -14.99 (6.53) | 8.90 (6.43) |
| | BAMALM | 1.00 (0.00) | 1.00 (0.00) | 3.69 (0.65) | 57.20 (2.20) | 60.89 (1.55) |
| | UMBC | 1.00 (0.00) | 0.94 (0.01) | 27.68 (0.04) | 6419.40 (493.53) | 6447.08 (493.50) |
| | BAMABC | 0.00 (0.00) | 1.00 (0.00) | 4.99 (0.59) | -21.25 (5.93) | -16.26 (6.07) |

| Scenario 2 | Sensitivity | Specificity | Bias of NDE | Bias of NIE | Bias of TE |
|------------|-------------|-------------|-------------|-------------|-------------|
| N = 50; V = 100 | BNMM | 0.99 (0.02) | 1.00 (0.01) | 3.82 (2.06) | -7.80 (1.89) | 3.98 (0.85) |
| | UMLM | 0.63 (0.25) | 1.00 (0.00) | 23.90 (0.14) | -14.99 (6.53) | 8.90 (6.43) |
| | BAMALM | 0.49 (0.45) | 1.00 (0.00) | 8.97 (7.30) | -4.52 (3.24) | 4.45 (24.21) |
| | UMBC | 0.55 (0.20) | 0.96 (0.01) | 24.09 (0.11) | 711.39 (316.80) | 735.48 (316.74) |
| | BAMABC | 0.00 (0.00) | 1.00 (0.00) | 10.74 (3.35) | -42.00 (2.00) | -31.26 (3.35) |
| N = 500; V = 200 | BNMM | 0.97 (0.15) | 1.00 (0.00) | 0.26 (0.62) | -13.57 (1.30) | 13.31 (1.23) |
| | UMLM | 1.00 (0.00) | 1.00 (0.01) | 25.90 (0.05) | -10.78 (1.63) | 15.12 (1.59) |
| | BAMALM | 1.00 (0.00) | 1.00 (0.00) | 3.25 (0.44) | 30.33 (1.24) | 33.58 (0.83) |
| | UMBC | 1.00 (0.00) | 0.96 (0.00) | 24.07 (0.04) | 4171.58 (388.57) | 4195.65 (388.54) |
| | BAMABC | 0.00 (0.00) | 1.00 (0.00) | 4.38 (0.70) | -20.07 (6.86) | -15.68 (6.69) |

Note: The Monte Carlo SD for each evaluation metric is included in the parentheses, and the bolded values indicate the best performance.

The bias of estimating NDE, NIE, and TE. For the selection metrics, to maintain consistent summary, we always map the selected mediators (sub-network or edges) back to the original connectome domain, and all the results are summarized in Table 1.

Based on the results, we can conclude that the proposed BNMM achieves the best or close to the best performance in identifying truly active mediators and estimating different effect components under all the simulation settings. Specifically, our method obtains a satisfactory accuracy in distinguishing mediators from noises within repeatedly measured brain connectomes as shown by the close to one sensitivity and specificity. There is no significant difference in the selection accuracy for all the methods between the two scenarios indicating the overlap between nonzero elements within $\tau$ and $\gamma$ brings limited impact on the model performance. In terms of the effect estimation, we can see from the Table 1 that BNMM obtains the smallest bias in estimating NDE, NIE, and TE under almost all the settings except for the Scenario 2 where BNMM slightly underperforms BAMALM with a smaller network, and UMLM with a larger network in estimating the NIE. However, both BAMALM and UMLM are implemented directly using the true latent mediators, which represents an ideal but unrealistic setup. Overall, we demonstrate the superiority and robustness of BNMM under both low and high dimensional connectome feature spaces with fully and partially overlapping latent mediators. Meanwhile, we also confirm the advantage of joint modeling the network mediator and effect pathways given that our method generally outperforms UMLM and BAMALM which directly use the true latent mediators as inputs, particularly under a small sample size close to our data application. In terms of the comparisons among the competing methods, UMLM and BAMALM outperform their alternative implementations under UMBC and BAMABC, which is anticipated given the former ones directly adopting the true mediator values. Between the univariate mediation analysis and BAMA, BAMA tends to be more
conservative with less mediators selected, resulting in a lower sensitivity. The univariate analysis, although identifying more true positives, suffers from worse performance in effect estimations.

5 | APPLICATION TO OUD DATA

We next apply our proposed approach to the motivated OUD data. Only participants with acceptable neuroimaging data for both pre- and post-treatment fMRI scans are included here, for a total of 41 subjects (18 receiving active drug and 23 receiving placebo). Preprocessing was conducted as previously described and included: discarding the first 6 volumes of each fMRI run, skull stripping, slice-time correction, motion correction, temporal smoothing (Gaussian filter with approximate cutoff frequency=0.12 Hz), normalization and concatenation of functional task runs, and nonlinear registration to structural data. The following covariates were regressed out of the data: linear and quadratic drifts, mean cerebrospinal fluid, mean white-matter signal, and a 24-parameter motion model including six rigid-body motion parameters, six temporal derivatives, and these terms squared. SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) was used for slicetime and motion correction. Additional preprocessing was conducted using BioImage Suite.

For all the subjects, biweekly urine testing was conducted during the trial period and the opioid abstinence outcome was measured by the transferred negative urine specimens test results for non-methadone opioids. We focus on the cognitive task for the connectome mediator where each subject has multiple scans collected and the number of scans are different among subjects ranging from four to six. To construct functional connectivity, we adopt the Shen 268-node brain atlas that includes cortex, subcortex, and cerebellum. The task connectivity is defined on the basis of node-by-node pairwise Pearson’s correlations using the raw task time courses among all the brain regions. These r statistics are further transformed to be normally distributed using Fisher’s z-transformation, and for each participant, we summarize each functional connectome collection into a 268 x 268 connectivity matrix. Finally, we also consider age and gender in both the outcome and the latent mediator models as covariates.

We apply BNMM to jointly dissect the connectivity module topology to assist in mediation analyses and to uncover the effect pathways, in which functional networks mediated relationships between treatment and opioid abstinence. As discussed previously, we determine the block number using the ICL criteria and set Q = 14. The rest of parameter settings and implementation closely follow the procedure in the simulations; and it took us around 2 h (one Intel 2.50 GHz core) to finish the whole posterior inference. Eventually, based on the posterior samples, the estimated NDE, NIE and TE with 95% credible interval are 0.59 (0.12, 1.12), −0.12 (−0.43, 0.66), and 0.48 (0.01, 1.03); and the NIE consists of −0.62 (−1.05, −0.16) for the negative component and 0.51 (0.03, 0.98) for the positive one, aligning with the negative and positive sub-networks typically seen in previous functional connectome analyses. We further investigate how the NIE develops through the brain sub-networks. Based on the estimated allocation matrix G, the 268 nodes are split into 14 sub-networks within the mediation framework. We show this parcellation under the average brain functional connectome in Figure 3. These sub-networks are further compared with the canonical neural networks, and in Table 2, we list the number of nodes within each sub-network, the most overlapping canonical networks, and the number of shared nodes (Nc), with

![Figure 3](image-url)  
**Figure 3** Heat maps depicting the average brain functional connectome from OUD study with the block structure and the highlighted informative sub-networks computed by BNMM. A set of the induced latent network mediators by the informative sub-networks are (A) associated with the treatment, (B) contributing to the outcome, and (C) the truly active latent mediators.
TABLE 2  Established sub-networks within the mediation analyses and their overlaps with canonical brain networks and macroscale regions

| Sub-network | # of nodes | $N_c$ | Canonical network                  | $N_{m}$ | Macroscale regions                  |
|-------------|------------|-------|-----------------------------------|---------|-------------------------------------|
| 1           | 21         | 8     | Frontal-parietal                  | 11      | L-prefrontal                        |
| 2           | 18         | 13    | Default mode                      | 6       | L-prefrontal/L-limbic               |
| 3           | 17         | 5     | Cingular-opercular/ default mode   | 4       | R-temporal                          |
| 4           | 12         | 1     | Default mode                      | 6       | R-cerebellum/L-cerebellum           |
| 5           | 18         | 17    | Visual                            | 8       | L-occipital                         |
| 6           | 17         | 8     | Visual                            | 4       | L-occipital                         |
| 7           | 21         | 10    | Default mode                      | 7       | L-prefrontal                        |
| 8           | 34         | 7     | Default mode                      | 8       | L-temporal                          |
| 9           | 31         | 4     | Default mode                      | 12      | R-cerebellum                        |
| 10          | 23         | 9     | Subcortical                        | 5       | R-subcortical                       |
| 11          | 16         | 7     | Auditory                          | 3       | L-MotorStrip/L-parietal            |
| 12          | 20         | 2     | Default mode                      | 4       | L-limbic                            |
| 13          | 7          | 6     | Somato-motor                      | 2       | L-MotorStrip/L-parietal/R-MotorStrip|
| 14          | 13         | 11    | Somato-motor                      | 3       | L-parietal/R-parietal              |

Note: The sub-networks contribute to the truly active latent mediators are bolded.

FIGURE 4  The identified sub-networks within the truly active latent network mediators involving in the positive and negative contributions to the NIE

the complete results provided in the Web Table 1. As can be seen in the table, our constructed sub-networks in the mediation analysis present a complex pattern of sub-networks that are composed of multiple components of canonical neural networks. This is anticipated given that the canonical networks were constructed in an unsupervised fashion using the resting-state functional connectivity for the default cognitive functions, and our organizations are established specifically in relation to a treatment effect mechanism. Among the latent mediator sets formed by the connectivity weights between and within those sub-networks, we eventually identify six active mediators. Specifically, the latent connection strength between sub-network 13 and sub-networks 1 and 14 positively contribute to the NIE; and the latent connection strength within sub-network 12, between sub-network 4 and sub-networks 7 and 10, and between sub-network 1 and sub-network 14 are negative contributors. We show in Figure 4 the locations of these sub-networks implicate in the positive and negative components of the NIE. To further investigate their network anatomy, we summarize the included macroscale brain regions for all the sub-networks in the supplementary material and highlight the most overlapping regions and the shared node numbers ($N_{m}$) in Table 2. Consistent with previous observations, the abstinence-related network anatomies are complex and occupied connections across different lobes. When further checking the involved
functional systems, we conclude that the three sub-networks with their composed connections positively contribute to the NIE are among corticolimbic (sub-network 1), premotor (sub-network 13) and somatomotor (sub-network 14) systems; and the additional sub-networks with connections negatively contributing to the NIE are concentrated within the cerebellar (sub-network 4), frontoparietal (sub-network 7), prefrontal limbic (sub-network 10), and limbic (sub-network 12) systems. These findings are consistent with recent work in addictions. In particular, recent reviews have highlighted the important role of sensorimotor connectivity in addictions, suggesting that changes in this circuitry may underlie the transition from goal-directed to habitual drug use behavior. Congruently, recent work in this same sample also identified sensorimotor, frontoparietal, cerebellar, and subcortical connections as key predictors of opioid abstinence.

6 | DISCUSSION

In this article, we make the very first attempt to propose a mediation analysis with network-variate mediators. Motivated by an OUD study investigating how brain functional connectomes relate to opiate abstinence, we develop a unified Bayesian mediation analysis named BNMM with latent network mediators that characterize connectivity patterns across sub-networks as mediating effects. Without pre-specifying sub-network structures, we estimate modular allocations along joint modeling to facilitate a more supervised sub-network construction with enhanced mediating effects. Through a simultaneous feature selection based on sparse regressions, we are able to identify truly active mediating network components; and stabilize the model fitting with the conditional independence among latent network mediators. We also demonstrate the superiority of our method in uncovering effect mechanisms relative to other approaches.

Our current method is based on a pre-determination of the number of latent blocks $Q$ in the SBMs based on the ICL criterion. Such a fixed block number is quite common in the general use of SBMs and has been frequently adopted when applying SBMs on brain functional connectome data given that the potential number of blocks within a functional network is typically limited under a certain range in light of the brain functional architecture. To remove such a constraint and simultaneously derive block numbers from the data, we could replace the Dirichlet distribution of $\pi_v$ with a nonparametric Dirichlet process (DP) on the node-specific probability $\pi_v$. The discrete nature of the DP model should facilitate a natural grouping of the nodes, and in the posterior inference, we could resort to an approximate Gibbs sampler under the truncated stick-breaking process representation with an efficient computation.

We assign non-informative IG priors for a number of our variance parameters. An alternative choice that suggested by Gelman for the hierarchical model is the half-t families. There has been a body of literature exploring the optimal prior choice for the variance components that includes comparisons between the half-t family suggested by Gelman and the classical non-informative IG distribution, which is the one used in our model. There is no unique answer and in many cases particularly with small to moderate sample sizes, non-informative IG priors have led to better performance compared with half-t distributions. The major issue with the non-informative IG distribution that Gelman considered is that inferences could be potentially sensitive to the choice of the hyper-parameters when there are low values of the variance parameter. In our application, we consider this as less of a concern given that our model is quite robust to the variance related hyper-parameters as shown by a sensitivity analysis in Appendix A. From a computational perspective, the IG priors provide more straightforward posterior inference under our model setting. The non-informative property may also be valuable for future studies.

While our analysis is motivated by the mediating role of brain functional connectome, the modeling scheme is general to accommodate network-variate mediators and readily applicable to other fields. For instance, it is common in genomics to consider gene pathways or networks in a graph format. Recently, mediation modeling with gene expression mediators has started to be introduced in eQTL analyses and other clinical studies. A natural follow-up is to further investigate the mediation effects under genomic networks. Additionally, within the context of brain imaging, an additional natural extension would be to study mediation effects involving other types of brain networks including structural and multi-modal brain networks. We want to emphasize that the SBM representation for network mediators in our current study is leveraged by the biological insight that a functional connectome tends to employ a modular structure. The modeling assumption for a SBM itself is straightforward and reveals a nice biological interpretation. At the same time, there is a body of literature to extend SBMs for better accommodating more complex network configurations including degree-corrected SBMs. It will be an interesting future direction to assess the performance of different SBM variations under our analytical paradigm for brain functional connectome mediators. Finally, under an alternative application, we could also replace it with other low-rank parametric forms. For instance, we could utilize clique graph representations when modeling brain structural connectomes given that the anatomical neural supports are more concentrated.
Our current model focuses on a single exposure. With recent attempts to incorporate multivariate exposures\textsuperscript{50} and the growing complexity of the collected data, extending BNMM to accommodate multiple or high-dimensional exposure variables is a potential future direction to further generalize our method. For the outcome, we consider a continuous variable here and the regression link function could be replaced to handle other outcome types.\textsuperscript{12} Some more demanding extensions, on which we are working, involve studying the network mediation for survival and longitudinal outcomes, which could offer additional perspectives on effect mechanisms.

**FUNDING INFORMATION**

National Institute on Aging, Grant/Award Numbers: RF1 AG068191, P30 AG021342; National Institute on Drug Abuse, Grant/Award Numbers: K01 DA039299, R01 DA039136, P50 DA09241; National Center for Advancing Translational Science (NCATS), Grant/Award Number: UL1 TR001863.

**DATA AVAILABILITY STATEMENT**

The OUD data used in the case study is subject to third-party restrictions. The software for the proposed method in the form of R code is available at Github: https://github.com/yizekaren/Bayesian_Network_Mediation_Model.

**ORCID**

Yize Zhao \(\text{https://orcid.org/0000-0001-6283-2302}\)

**REFERENCES**

1. Carroll KM, DeVito EE, Yip SW, Nich C, Sofuoglu M. Double-blind placebo-controlled trial of galantamine for methadone-maintained individuals with cocaine use disorder: Secondary analysis of effects on illicit opioid use. *Am J Addict*. 2019;28(4):238-245.
2. Pearl J. *Models, Reasoning and Inference*. Cambridge, UK: Cambridge University Press; 2000.
3. VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology*. 2011;22(4):582-585.
4. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Person Soc Psychol*. 1986;51(6):1173.
5. Wang W, Nelson S, Albert JM. Estimation of causal mediation effects for a dichotomous outcome in multiple-mediator models using the mediation formula. *Stat Med*. 2013;32(24):4211-4228.
6. Imai K, Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Polit Anal*. 2013;21:141-171.
7. Daniel RM, De Stavola BL, Cousens S, Vansteelandt S. Causal mediation analysis with multiple mediators. *Biometrics*. 2015;71(1):1-14.
8. Taguri M, Featherstone J, Cheng J. Causal mediation analysis with multiple causally non-ordered mediators. *Stat Methods Med Res*. 2018;27(1):3-19.
9. Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiology*. 2017;28(2):258.
10. Kim C, Daniels MJ, Hogan JW, Choirat C, Zigler CM. Bayesian methods for multiple mediators: Relating principal stratification and causal mediation in the analysis of power plant emission controls. *Ann Appl Stat*. 2019;13(3):1927-1956.
11. Huang YT, Pan WC. Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics*. 2016;72(2):402-413.
12. Derkach A, Pfeiffer RM, Chen TH, Sampson JN. High dimensional mediation analysis with latent variables. *Biometrics*. 2019;75(3):745-756.
13. Kong D, An B, Zhang J, Zhu H. L2RM: low-rank linear regression models for high-dimensional matrix responses. *J Am Stat Assoc*. 2020;115(529):403-424.
14. Kong D, An B, Zhang J, Zhu H. L2RM: low-rank linear regression models for high-dimensional matrix responses. *J Am Stat Assoc*. 2020;115(529):403-424.
15. Relión JDA, Kessler D, Levina E, Taylor SF. Network classification with applications to brain connectomics. *Ann Appl Stat*. 2019;13(3):121-136.
16. Chén OY, Crainiceanu C, Ogburn EL, Caffo BS, Wager TD, Lindquist MA. High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics*. 2018;19(2):121-136.
17. Kong D, An B, Zhang J, Zhu H. L2RM: low-rank linear regression models for high-dimensional matrix responses. *J Am Stat Assoc*. 2020;115(529):403-424.
18. Relión JDA, Kessler D, Levina E, Taylor SF. Network classification with applications to brain connectomics. *Ann Appl Stat*. 2019;13(3):1648-1677.
19. Xie S, Li X, McColgan P, Scahill RI, Zeng D, Wang Y. Identifying disease-associated biomarker network features through conditional graphical model. *Biometrics*. 2020;76(3):995-1006.
20. Wig GS. Segregated systems of human brain networks. *Trends Cognit Sci*. 2017;21(12):981-996.
21. Yeo BT, Krienen FM, Sepulcre J, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol*. 2011;106(3):1125-1165.
22. Hamdi SM, Wu Y, Angryk R, Krishnamurthy LC, Morris R. Identification of discriminative subnetwork from fmri-based complete functional connectivity networks. *Int J Semant Comput*. 2019;13(01):25-44.
23. Zhang J, Sun WW, Li L. Mixed-effect time-varying network model and application in brain connectivity analysis. *J Am Stat Assoc*. 2020;115(523):2022-2036.
24. Albert JM, Geng C, Nelson S. Causal mediation analysis with a latent mediator. *Biometr J*. 2016;58(3):535-548.
25. Chén OY, Cao H, Phan H, et al. Identifying neural signatures mediating behavioral symptoms and psychosis onset: high-dimensional whole brain functional mediation analysis. *NeuroImage*. 2021;226:117508.
26. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psycholog Methods*. 2010;15(4):309-334.
27. Lichenstein SD, Scheinost D, Potenza MN, Carroll KM, Yip SW. Dissociable neural substrates of opioid and cocaine use identified via connectome-based modelling. *Mol Psych*. 2021;2:4383-4393.
28. Shen X, Finn ES, Scheinost D, et al. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat Protoc*. 2017;12(3):506-518.
29. Yuan Y, MacKinnon DP. Bayesian mediation analysis. *Psychol Methods*. 2009;14(4):301-322.
30. Bickel PJ, Sarkar P. Hypothesis testing for automated community detection in networks. *J Royal Stat Soc Ser B (Stat Methodol)*. 2018;113(521):241-251.
31. Lichenstein SD, Scheinost D, Potenza MN, Carroll KM, Yip SW. Dissociable neural substrates of opioid and cocaine use identified via connectome-based modelling. *NeuroImage*. 2021;226:117508.
32. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psycholog Methods*. 2010;15(4):309-334.
33. Albert JM, Geng C, Nelson S. Causal mediation analysis with a latent mediator. *Biometr J*. 2016;58(3):535-548.
34. Gelman A, Rubin DB, others. Inference from iterative simulation using multiple sequences. *Stat Sci* 1992; 7(4): 457–472.
35. Sheng Y. Investigating a weakly informative prior for item scale hyperparameters in hierarchical 3PNO IRT models. *Front Psychol*. 2017;8:123.
36. Bernardo JM, Smith AF. *Bayesian Theory*. Vol 405. Hoboken: John Wiley & Sons; 2009.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.
APPENDIX A. SENSITIVITY ANALYSIS

We provide here a sensitivity analysis to assess the robustness of BNMM on the prior setting. As it is shown in the right panel of Figure 1, we choose the conjugate priors for all the parameters. In terms of their hyper-parameters which are the potential sensitive part to impact the posterior likelihood and inference results, we have carefully set their values to allow these priors to be non-informative. Non-informative priors themselves have minimal effect on the final inference, which is perfectly suitable if not much is known beyond the data included in the analysis at hand. Specifically, these hyper-parameters either involve in setting the prior probability or relate to the variance parameters in the normal distribution. For all the probability related hyper-parameters, that is, $p_1, \ldots, p_Q, p_y$ and $p_r$, we set them to be 0.5 to induce a neutral preference, which represents the most non-informative setting. For the variance parameters, we either set them to be a large value, that is, 10 or assume them follow a non-informative IG prior including $\sigma_q^2, \sigma_m^2, \sigma_{my}^2, \sigma_{qr}^2, \omega_{qr}^2$. Both these ways have been considered to induce non-informative priors in existing literature to mimic a uniform prior from the normal distributions. To further verify the robustness, we primarily consider $a_1, a_2, b_1, b_2$ which are the shape and scale parameters in the IG prior for $\sigma_q^2$ and $\omega_{qr}^2$. We rerun the real data analysis under different values of $a_1, a_2, b_1, b_2$ but keep them to be small to ensure the induced prior is non-informative. As shown in Table A1, the estimated effects and their corresponding 95% credible intervals almost stay consistent, and we maintain all our conclusions stated in the data application.

| $a_1$ | $a_2$ | $b_1$ | $b_2$ | NIE (CI) | NDE (CI) |
|-------|-------|-------|-------|----------|----------|
| 1     | 1     | 1     | 1     | $-0.12 (-0.43, 0.66)$ | 0.59 (0.12, 1.12) |
| 0.1   | 0.1   | 0.1   | 0.1   | $-0.12 (-0.46, 0.68)$ | 0.58 (0.06, 1.15) |
| 0.1   | 0.1   | 0.01  | 0.01  | $-0.10 (-0.41, 0.73)$ | 0.55 (0.06, 1.11) |
| 0.01  | 0.01  | 0.1   | 0.1   | $-0.11 (-0.43, 0.67)$ | 0.55 (0.06, 1.10) |
| 0.01  | 0.01  | 0.01  | 0.01  | $-0.11 (-0.36, 0.71)$ | 0.55 (0.03, 1.24) |