Journal of Data Science 23 (3), 521–541
July 2025
DOI: 10.6339/25-JDS1169
Statistical Data Science

High-dimensional Confounding in Causal Mediation: A Comparison Study of Double Machine Learning and Regularized Partial Correlation Network

Ming Chen[1], Tanya T. Nguyen[2], and Jinyuan Liu[3,*]

[1]*University of Wisconsin-Madison, USA*
[2]*Sam and Rose Stein Institute for Research on Aging, University of California, San Diego, USA*
[3]*Department of Biostatistics, Vanderbilt University, USA*

Abstract

In causal mediation analyses, of interest are the direct or indirect pathways from exposure to an outcome variable. For observation studies, massive baseline characteristics are collected as potential confounders to mitigate selection bias, possibly approaching or exceeding the sample size. Accordingly, flexible machine learning approaches are promising in filtering a subset of relevant confounders, along with estimation using the efficient influence function to avoid overfitting. Among various confounding selection strategies, two attract growing attention. One is the popular debiased, or double machine learning (DML), and another is the penalized partial correlation via fitting a Gaussian graphical network model between the confounders and the response variable. Nonetheless, for causal mediation analyses when encountering high-dimensional confounders, there is a gap in determining the best strategy for confounding selection. Therefore, we exemplify a motivating study on the human microbiome, where the dimensions of mediator and confounders approach or exceed the sample size to compare possible combinations of confounding selection methods. By deriving the multiply robust causal direct and indirect effects across various hypotheses, our comprehensive illustrations offer methodological implications on how the confounding selection impacts the final causal target parameter estimation while generating causality insights in demystifying the "gut-brain axis". Our results highlighted the practicality and necessity of the discussed methods, which not only guide real-world applications for practitioners but also motivate future advancements for this crucial topic in the era of big data.

Keywords *efficient influence functions; gut-brain axis; multiply robust; Neyman orthogonality; regularization bias*

1 Introduction

Increasingly, causal mediation analyses are recognized by various domains. Upon evaluating the total causal effect of some treatment or exposure, investigators are further interested in its direct or indirect pathways through some mediator variable (Tchetgen and Shpitser, 2012; Imai et al., 2010). For ubiquitous observational studies, the utmost challenge is to mitigate the selection bias by collecting as many baseline characteristics as possible. However, when the number of confounders approaches or exceeds the sample size, the model fitting becomes unstable without some filtering (Chernozhukov et al., 2018, 2022). In quantifying the causal mediation effect under

*Corresponding author. Email: talliechen108@gmail.com or jinyuan.liu@vumc.org.

Received October 1, 2024; Accepted January 9, 2025

high-dimensional confounders, selecting a subset of relevant confounders is essential to avoiding overfitting and drawing consistent inferential conclusions (Xue and Qu, 2022).

However, there is a gap in determining the best strategy for confounding selection in this growing problem of high-dimensional causal mediation analysis. In this paper, we focus on two promising approaches. One is the mediation extension of the double machine learning (DML) that incorporates penalization in modeling the nuisance functions that appeared in the efficient influence functions to handle the massive confounders (Farbmacher et al., 2022). Another is motivated by the emerging partial correlation network in deciphering the complex interplay among psychosocial variables (Epskamp et al., 2018). Unlike the conventional Lasso (Tibshirani, 1996; Zou, 2006), which penalizes the coefficients in the regression model, the penalization is applied at the outset when fitting a Gaussian graphical model between the confounders and the response variable (Lauritzen, 1996). The deduced penalized partial correlation coefficients encourage selecting confounders with direct effects on the response while discouraging the selection of irrelevant ones by partial them out (Xue and Qu, 2022).

To address the gap and guide real-world applications, we illustrated different combinations of confounding selection methods to demystify the causality in the "gut-brain axis" in a motivating observational study (Liu et al., 2022; Nguyen et al., 2021). Our comprehensive comparison results offer methodological implications on how the confounding selection impacts the final causal target parameter estimation, above and beyond the valuable real-world scientific insights.

Essentially, our results are precious for the ubiquitous observational data, where the unknown confounding mechanisms are modeled as a nuisance using flexible machine learning approaches, especially when encountering massive confounders. The rest of the paper is organized as follows. In Section 2, we provide an overview of the motivating study on the gut-brain axis. In Section 3, we first detail the efficient influence functions (EIF) under the potential outcome framework in causal and mediation settings and then discuss the two confounding selection strategies. In Section 4, we offer comparisons among various combinations to evaluate their impacts on the final estimation of target parameters, which are applied to the motivating real-world data. We conclude the paper in Section 5.

2 Motivation

2.1 Background

The human microbiome consists of the microbes, their genetic elements, and their interactions with surrounding environments throughout the human body (Cho and Blaser, 2012). Numerous studies have suggested that the microbiome is the missing link between genetics, environment, and disease (Virgin and Todd, 2011; Nguyen et al., 2021; Liu et al., 2023), incentivizing statistical advancements to decipher their inherent mechanisms, especially for the causal pathways, direct or indirect.

Fueled by the technological breakthrough of next-generation sequencing, the human microbiome composition can be interrogated using high-throughput sequencing. Marker genes can be amplified and sequenced and then clustered into Operational Taxonomic Units (OTUs) or amplicon sequence variants (ASVs). By comparing them with reference databases, one can establish taxonomic abundance profiles for each subject as the basis for statistical analyses (Liu et al., 2022). Such abundance data are challenging to analyze. First, the number of taxa features exceeds the number of subjects in many studies. Second, those counts are quite sparse with a preponderance of zeros. Third, to overcome the heterogeneity or artifact in sampling

depth, the absolute abundance is frequently normalized into relative abundance using centered log-ratio transformation, creating the compositional data that needs special statistical consideration. Lastly, the human microbiome is highly dynamic and varies on a day-to-day basis. Longitudinal studies with repeated measures will help depict more comprehensive insights but appropriate associational and causal inference tools are still lacking.

2.2 Gut-Brain-Axis

The gut-brain axis refers to the complex communications between the gut microbiota and the brain. For almost a century the human microbiota has been linked to neuropsychiatric disorders associated with neurodevelopment (e.g., autism spectrum disorder and schizophrenia), neurodegeneration (e.g., Parkinson's disease, Alzheimer's disease, and multiple sclerosis), and mood (e.g., depression and anxiety) (Morais et al., 2021). However, we are still at the beginning of deciphering their innate mechanisms and causal pathways. A growing body of literature has suggested a strong connection between the gut microbiome and the central nervous system, evidencing the key role of the gut-brain axis. For example, in a cross-sectional study by Meyer et al. (2022), the beta-diversity, a measure of gut microbial community composition, was significantly associated with all measures of cognitive function. Also, major depressive disorder, a maladaptive response to chronic stress (or stress during early life), has been hypothesized to be mediated by the gut microbiome since stress is a disruptor of gut microbiota composition in animals and humans (Foster et al., 2021). Further, mechanisms underlying microbial-mediated changes in social behavior in mouse models of autism spectrum disorder have been confirmed (Sgritta et al., 2019). Still, more causality evidence in human studies is needed to move beyond correlation to the validation of causal mechanisms.

2.3 Motivating Study

Given most human gut-brain-axis research is limited by its observational and cross-sectional nature, showing association but not causation, a recent longitudinal study was conducted for a group of aging residents to overcome this challenge. In this study, demographics, physical, cognitive, and psychosocial instruments were assessed at baseline. Their fecal microbiome was sequenced at baseline and the follow-up visit after six months. Albeit observational, various potential confounders were collected ($p = 81$), which approaches the sample size ($n = 92$). Along with the time lag, it allows for repeated measures from the same individual to strengthen the cause-and-effect insights.

We hypothesize that exposure to loneliness and cognitively stimulating activities alters cognitive functioning over six months via the microbiome composition, which serves as the mediator (see Figure 1). The implication is pivotal in supporting the potential for gut microbiota targets in preventing or treating cognitive decline.

Denoted by $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ih})$ the microbial taxa counts at six months, compounding the issue of $n \ll h = 363$, $\mathbf{Y}_i$ are highly-skewed and zero-inflated as we mentioned earlier. An exploratory screening using the Lasso (Zou, 2006) indicated that 71 out of the 363 features had non-zero coefficients regarding cognitive decline. Investigating causal mediation impacts on the individual taxa based on this information will not only violate the rule of post-selection inference (Liu et al., 2024) but is unlikely to survive the multiple comparisons, and hence, harm the study reproducibility. Whereas the feature aggregation provides a promising solution, specifically, we aggregated the taxonomic abundances at the within-subject level, yielding Faith's phylogenetic

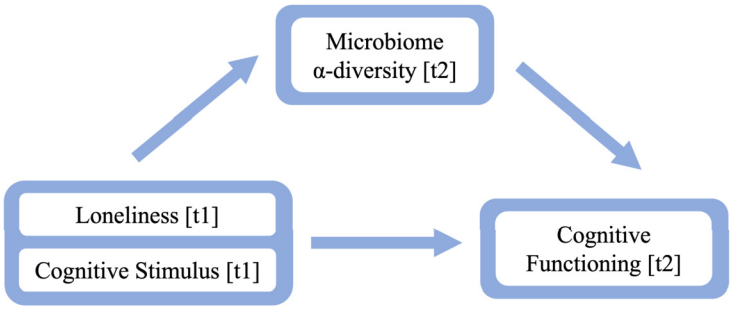

Figure 1: Hypothesized causal effect between gut microbiota and cognitive functioning.

alpha-diversity $f(\mathbf{Y}_i)$, which also incorporates a biologically relevant tree structure across microbial taxa units (Nguyen et al., 2021). Our goal is to investigate the modulation effect of the microbiome alpha-diversity on cognitive decline. Accordingly, the causal mediation pathway is examined by leveraging the exposure at baseline (loneliness, cognitive stimulating activities), the microbiome alpha-diversity as the mediator, and the cognitive functioning at six months as the outcome. We consider baseline measures of demographics, physical, and psychosocial instruments as potential confounders. To overcome the inferential challenge of massive confounders $n \ll p$, we focus on data-adaptive approaches such as double machine learning (DML) and regularized partial correlation networks to select a subset for further causality analyses.

3 Method

We review Neyman's potential outcome framework in causal and mediation analysis.

3.1 Causal Effect

Consider a binary exposure $E_i = e \in \{0, 1\}$, where each subject is equipped with a pair of potential, or counterfactual, outcomes $\{Y_i^1, Y_i^0\}$ but only one of them is observed. The mean-difference type of *average causal effect* (ACE) aims to quantify the deviation of the outcome $Y_i \in \mathbb{R}$ from the exposure arm to the control, averaging over the confounders $\mathbf{X}_i \in \chi$ across the population (Rubin, 1990):

$$\Psi^{\text{CE}}(P_{\theta_0}) = \Delta_{ACE} = \text{E}_{\theta_0}\big(Y_i^1\big) - \text{E}_{\theta_0}\big(Y_i^0\big) = \xi^1 - \xi^0, \tag{1}$$

where P_{θ_0} is the underlying data generating process with θ_0 its true parameter.

In (1), the exposure-specific mean potential outcome $\xi^e = \text{E}_{\theta_0}(Y_i^e)$ is termed the *causal functional*, and $\Psi^{\text{CE}}(P_{\theta_0})$ is identifiable with three assumptions:

C1. *SUTVA*: the observed outcome satisfies $Y_i = E_i Y_i^1 + (1 - E_i) Y_i^0$.
C2. *Strong ignorability*: $E_i \perp \{Y_i^1, Y_i^0\} \mid \mathbf{X}_i = \mathbf{x}_i$.
C3. *Positivity*: $\Pr(E_i = e \mid \mathbf{X}_i) > 0$, w.p.1. for each $e \in \{0, 1\}$.

To estimate and make inference about $\Psi^{\text{CE}}(P_{\theta_0})$, the *efficient influence function* (EIF) for ξ^e has been proposed (Chernozhukov et al., 2022; Tsiatis, 2006):

$$\varphi_{\xi^e}^{\text{EIF}}\big(\xi_0^e, \eta_0\big) = \frac{I(E_i = e)}{\pi_0(\mathbf{X}_i)}\big\{Y_i - \mu(e, \mathbf{X}_i; P_{\theta_0})\big\} + \mu(e, \mathbf{X}_i; P_{\theta_0}) - \xi^e(P_{\theta_0}), \tag{2}$$

which is verified to be *Neyman orthogonal* (Chernozhukov et al., 2018; Neyman, 1979), and hence, the estimators of ξ^e solved from the estimating equations constructed from (2) are deemed doubly robust, namely, the final estimator is consistent, provide one of the two nuisance functions in (2) are specified correctly. The nuisance functions are the mean function $\mu_i(e, \mathbf{X}_i) := \mathrm{E}_{\theta_0}(Y_i \mid E_i = e, \mathbf{X}_i)$ and propensity score (PS) $\pi_0(\mathbf{X}_i) := \Pr_{\theta_0}(E_i = e \mid \mathbf{X}_i)$. As a special case of coarsened or missing data, the corresponding estimators of ξ^e are shown to reach the *semiparametric efficiency bound* when both nuisance functions are specified correctly (Tsiatis, 2006; Liu et al., 2022).

Many machine learning (ML) approaches perform well by employing regularization to select among vast confounders. However, the induced *regularization bias* may propagate to invalidate the naive estimators of causal effects. The Neyman orthogonality motivated advancements such as debiased ML (DML) and target learning (Zheng and Van Der Laan MJ, 2012; Wang et al., 2023) to remove the regularization bias, where one key component is deploying this Neyman-orthogonal score or EIF in (2) that is robust to the unknown patterns from confounders.

3.2 Causal Mediation Effect

Increasingly, investigators are interested in causal mediation analysis. For instance, upon evaluating the *total causal effect* of the exposure, they would like to further examine the *direct* or *indirect* pathways of the exposure, possibly through a *mediator* variable (Tchetgen and Shpitser, 2012). Following the literature (Robins and Greenland, 1992; Pearl, 2014), we consider natural (pure) direct effects where the mediator is viewed as random in what follows.

For notation, a three-component (exposure E_i, mediator M_i, and outcome Y_i) causal study design, where $M_i \in \Omega$ is the mediator from the exposure to an outcome. For instance, being exposed to cognitively stimulating activities E_i has been hypothesized to alter the cognitive functioning Y_i through the microbiome composition M_i as in our motivating example.

With M_i and Y_i the respective *observed* mediator and outcome, we define two counterfactual quantities under $e \in \{0, 1\}$:

- M_i^e – *counterfactual mediator* for the exposure arm $E_i = e$;
- $Y_i^{e,m}$ – *counterfactual outcome* for the *exposure-mediator combination* $(E_i, M_i^{e'}) = (e, m)$.

In our example, the microbiome as a mediator can be affected by the cognitive stimulating activities and admits M_i^1 and M_i^0, but for each subject i, only one of them can be observed, and hence, is counterfactual. The potential outcomes depend on both the mediator and exposure, which differ from the pure causal setting in (1) where the potential outcome only depends on exposure. The individual causal *indirect (mediation) effect* has been defined as $Y_i^{e,M_i^1} - Y_i^{e,M_i^0}$, $e = 0, 1$, which answers a counterfactual question: what is the difference in the outcome when the value of mediator switches from M_i^0 (under the control) to M_i^1 (under the exposure) while holding the exposure status at e (Pearl, 2014).

In contrast, the individual causal *direct effect* is defined as $Y_i^{1,M_i^e} - Y_i^{0,M_i^e}$, $e = 0, 1$. In our example, the direct stimulation effect on the subject i's cognitive functioning while holding his or her microbiome composition constant at the level under no stimulation is $Y_i^{1,M_i^0} - Y_i^{0,M_i^0}$. Further, if one assumes no interaction between E_i and M_i, then the indirect and direct effects will no longer vary with e. Akin to the classical setting, those individual effects are unidentifiable due to their counterfactual nature.

In the presence of a mediator, we decompose the average causal effect (ACE) of E on Y in (1) as (drop the subscript i):

$$\Psi^{\mathrm{CE}}(P_{\theta_0}) = \mathrm{E}_{\theta_0}\big(Y^{1,M^1} - Y^{0,M^0}\big) = \mathrm{E}_{\theta_0}\big[\big(Y^{1,M^1} - Y^{1,M^0}\big) + \big(Y^{1,M^0} - Y^{0,M^0}\big)\big], \tag{3}$$

which respectively defines the *average mediation effect* (AME) and *average direct effect* (ADE):

$$\Psi^{\mathrm{ME}}(P_{\theta_0}) = \mathrm{E}_{\theta_0}\big(Y^{1,M^1} - Y^{1,M^0}\big), \quad \Psi^{\mathrm{DE}}(P_{\theta_0}) = \mathrm{E}_{\theta_0}\big(Y^{1,M^0} - Y^{0,M^0}\big). \tag{4}$$

In particular, $\Psi^{\mathrm{ME}}(P_{\theta_0})$ is the comparison of the potential outcome from the exposure arm when the mediator is switched on (M^1) and off (M^0). And $\Psi^{\mathrm{DE}}(P_{\theta_0})$ directly contrasts the potential outcomes between the two arms when the mediator was switched off (M^0) (Tchetgen and Shpitser, 2012). Both averaged across the entire population.

Not tied to any specific model, the definitions in (4) generalize previous discussions by Gunzler et al. (2014) using linear structural equation model (SEM), where one posits two models, one for the outcome Y with $E(Y \mid \mathbf{X}, E, M) = \alpha_0 + \alpha_E E + \alpha_M M + \boldsymbol{\alpha}_X^\top \mathbf{X}$, the other for the mediator M with $E(M \mid \mathbf{X}, E) = \beta_0 + \beta_E E + \boldsymbol{\beta}_X^\top \mathbf{X}$. Hence, it is easily deduced from (4) that $\Psi^{\mathrm{ME}}(P_{\theta_0}) = \alpha_M \beta_E$, $\Psi^{\mathrm{DE}}(P_{\theta_0}) = \alpha_E$. Often, it is of interest to test (1) the existence of mediation effect $H_{01} : \Psi^{\mathrm{ME}}(P_{\theta_0}) = 0$, and (2) full vs. partial mediation by testing the direct effect $H_{02} : \Psi^{\mathrm{DE}}(P_{\theta_0}) = 0$.

3.2.1 Assumptions

Denote by $\mathbf{X} \in \chi$ a set of pre-exposure variables sufficient to account for selection bias, then under three stronger assumptions, $\Psi^{\mathrm{ME}}(P_{\theta_0})$ and $\Psi^{\mathrm{DE}}(P_{\theta_0})$ become identifiable.

D1. *Counterfactual Consistency*:

(i) The observed $M = EM^1 + (1-E)M^0$: if $E = e$, then $M^e = M$, w.p.1.

(ii) The observed $Y = EY^{1,m} + (1-E)Y^{0,m}$: If $E_i = e$ and the observed $M = m$, then $Y^{e,m} = Y$, w.p.1.

D2. *Strong sequential ignorability*:

(i) $E \perp \{Y^{e',m}, M^e\} \mid \mathbf{x}$: given the observed confounders, the exposure assignment is statistically independent of the potential outcomes and potential mediators;

(ii) $Y^{e',m} \perp M^e \mid E = e, \mathbf{x}$: given the observed exposure and confounders, the mediator is ignorable.

D3. *Positivity*: C3 holds and $\Pr(M \mid e, \mathbf{x}) > 0$, w.p.1. for each $m \in \Omega$.

In D2, the two ignorability assumptions are sequential: first, given $\mathbf{x}$, the exposure is assumed to be ignorable, which can sometimes be enforced by randomization (Imai et al., 2010); second, the mediator is ignorable given the observed value of the ignorable exposure and $\mathbf{x}$. It implies that among those subjects under the same exposure status and pre-exposure characteristics, the mediator can be regarded as if being randomized, which can be strong. Hence, we use sensitivity analysis to further validate the ignorability assumptions (see Section 4).

3.2.2 Identification

Akin to the causal effect $\Psi^{\mathrm{CE}}(P_{\theta_0})$, the causal direct and indirect effects can be constructed to be *multiply robust* and *locally efficient*, even with a large number of pre-exposure confounders. Under assumptions D1–D3, the *causal mediation functional* $\mathrm{E}_{\theta_0}(Y^{1,M^0}) := \nu$ is identifiable (Imai et al., 2010; Pearl, 2014), which appeared in (4), we can further deduce that

$$\Psi^{\mathrm{CE}}(P_{\theta_0}) = \xi^1 - \xi^0, \quad \Psi^{\mathrm{ME}}(P_{\theta_0}) = \xi^1 - \nu, \quad \Psi^{\mathrm{DE}}(P_{\theta_0}) = \nu - \xi^0.$$

By the linearity, the EIF for the causal functional ξ^e in (2) can be used, and we only need to estimate the mediation functional ν for the estimation of AME and ADE. Using the Gautaex derivative and the strategy of a point mass (Chernozhukov et al., 2018), the efficient influence function (EIF) for ν is found to be (Tchetgen and Shpitser, 2012)

$$\begin{aligned}\varphi_{\nu}^{\mathrm{EIF}}(\upsilon_0, \eta_0) = &\frac{I(E=1)f_0(M \mid E=0, \mathbf{X})}{\pi_0(\mathbf{X})f_0(M \mid E=1, \mathbf{X})}\big\{Y - \mu(1, M, \mathbf{X}; P_{\theta_0})\big\} \\ &+ \frac{I(E=0)}{1-\pi_0(\mathbf{X})}\big\{\mu(1, M, \mathbf{X}; P_{\theta_0}) - \mathrm{E}_{\theta_0}\big[\mu(1, M, \mathbf{X}) \mid E=0, \mathbf{X}\big]\big\} \\ &+ \mathrm{E}_{\theta_0}\big[\mu(1, M, \mathbf{X}) \mid E=0, \mathbf{X}\big] - \nu(P_{\theta_0}), \end{aligned} \tag{5}$$

where $\pi_0(\mathbf{X}_i)$ is the true propensity score (PS) as in (2); also, the true outcome mean given exposure, mediator, and confounders is denoted as $\mu(e, M, \mathbf{X}; P_{\theta_0}) := \mathrm{E}_{\theta_0}(Y \mid E=e, M, \mathbf{X})$. The additional part $f_0(M \mid E=e, \mathbf{X})$ denotes the true conditional density of mediators; for discrete mediators, the probability mass function $\Pr_{\theta_0}(M=m \mid E=e, \mathbf{X})$ can be substituted.

To bypass estimating the density $f_0(M \mid E=0, \mathbf{X})$ when M is continuous, or even multi-dimensional, an alternative form has been proposed based on Bayes law (Farbmacher et al., 2022), which substitutes the first component in (5) by

$$\frac{I(E=1)[1-p_0(M, \mathbf{X})]}{[1-\pi_0(\mathbf{X})]p_0(M, \mathbf{X})}\big\{Y - \mu(1, M, \mathbf{X})\big\}, \quad \text{where } p_0(M, \mathbf{X}) = \Pr(E=1 \mid M, \mathbf{X}; P_{\theta_0}).$$

3.3 Confounding Selection

For the ubiquitous observational studies where randomization is not feasible or ethical, the utmost challenge is that the confounding mechanisms are unknown. To mitigate the potential bias, the investigators tend to collect as many baseline characteristics as possible, hoping to avoid unmeasured confounding. However, when the number of confounders approaches or exceeds the sample size, the model fitting becomes unstable without some filtering, either by adding a penalized term or using other strategies.

The previously introduced efficient influence functions (EIF) follow Neyman's orthogonality condition (Liu et al., 2022). Therefore, the resulting estimators should not be strongly affected by the quality or precision in estimating the confounding patterns (Neyman, 1979). In practice, different selection strategies may lead to distinct sets of confounders, hence propagating to yield different causal estimators. We briefly review some recent developments for selecting among the vast number of confounders in the context of causal mediation analyses.

3.3.1 Double Machine Learning (medDML)

The debiased machine learning (DML) (Chernozhukov et al., 2018) is a popular approach deploying the orthogonal score (or EIF) and flexible nonparametric methods to handle the massive confounders. Conventional DML only estimates the causal effect $\Psi^{\mathrm{CE}}(P_{\theta_0})$. A recent development in Farbmacher et al. (2022) extends the scope into a causal mediation framework, which yields asymptotically normal and $\sqrt{n}$-consistent average direct effect (ADE) and mediation effect (AME) estimators while automatically selecting among high-dimensional confounders.

Akin to the DML, two components are essential for obtaining well-behaved final estimators. First, to overcome the regularization bias in selecting the confounders, the Neyman orthogonal

scores, or EIFs, are adopted, specifically, (5) for ν and (2) for ξ^e. This procedure ensures the resulting estimators of ADE and AME are robust to misspecifications of the outcome and mediator models, also referred to as the "multiply robust" property.

Second, *cross-fitting* is another essential piece to prevent overfitting. By randomly splitting the sample and estimating the nuisance parameters in one half (*auxiliary* part) while estimating the ADE and AME in the other half (*main* part), one avoids potential overfitting in estimating the nuisance, which can occur, say, involving irrelevant confounders. Then, by swapping the role of main and auxiliary parts to obtain a second estimator, the full efficiency is recovered by taking their averages. Notably, it only requires weaker assumptions to shrink an empirical process term to zero, where the Donsker conditions do not apply in this high-dimensional setting (Chernozhukov et al., 2018).

The cross-fitting can be generalized to a K-fold version. To deploy the K-fold cross-fitting procedure for estimating ν and ξ^e in the mediation analyses, we denote the observed data by $\mathbf{O}_i = (Y_i, M_i, D_i, X_i)^\top$ as an *i.i.d.* sample of size n. One first randomly partition the observation indices $I = \{1, \ldots, n\}$ into K disjoint subsets with equal size, denoted as I_k for $k = 1, \ldots, K$, and denote by I_k^c the complement set of I_k.

Next, within each fold, some ML algorithm is applied to the observations in the complement set $\mathbf{O}_{i \in I_k^c}$ to estimate the nuisance parameters, which include the conditional outcome means, mediator densities, and propensity scores appeared in (5) and (2), collectively denoted as $\widehat{\eta}(\mathbf{O}_{i \in I_k^c})$. This process is repeated for all K folds, yielding K estimators by solving the empirical analog of EIFs (5) and (2) using estimating equations but with η_0 replaced by $\widehat{\eta}(\mathbf{O}_{i \in I_k^c})$ for each fold. The final estimators for ν and ξ^e are averaged across the K estimators.

Essentially, these two key components ensure the $\sqrt{n}$-convergence of the estimated ADE and AME, where the regularity conditions are attained by various ML algorithms, including Lasso.

3.3.2 Regularized Partial Correlation Network

The emergence of the *partial correlation network* modeling of psychosocial variables overcame the historical challenges of deciphering their complex interplay (Epskamp et al., 2018). Further, regularization has been incorporated for a more interpretable and parsimonious network structure. Inherently differing from social networks that represent the connections among subjects, such networks reflect the connection (*edges*) between psychological variables (*nodes*) based on partial correlations (McNally et al., 2015; Borsboom and Cramer, 2013), also termed the *Gaussian graphical models* (Lauritzen, 1996), which belong to the general class of *pairwise Markov random fields* (Koller and Friedman, 2009; Murphy, 2012).

In the context of variable selection, consider a $p \times 1$ vector of potential confounders $\mathbf{X}$ (we suppressed the subscript i for each subject in what follows), denoted by $\rho_j = \mathrm{Corr}(Y, X_j \mid \mathbf{X}_{-j})$ $(j = 1, \ldots, p)$ the *partial correlation coefficient* between the response Y and a confounder X_j, after conditioning on other variables $\mathbf{X}_{-j} = \{X_{j'} : j' = 1, \ldots, p; j' \neq j\}$. Under the joint normal assumption, ρ_j captures the linear relationship between Y and X_j, partial out other confounders. To estimate ρ_j from the sample, denote the *precision matrix* by $\mathbf{K} = \mathbf{\Sigma}^{-1} = (\kappa_{jj'})$ $(j, j' = 1, \ldots, p+1)$ as the inverse of the variance-covariance matrix from $(Y, \mathbf{X})^\top \sim N(\mathbf{0}, \mathbf{\Sigma})$. It is readily shown that by standardizing the elements $\kappa_{jj'}$, one obtains a *partial correlation matrix* $\Omega = (\rho_{jj'})$, whose elements $\rho_{jj'} = -\kappa_{jj'}/(\sqrt{\kappa_{jj}}\sqrt{\kappa_{j'j'}})$. The first column of Ω $(j\prime = 1)$ readily recovers each ρ_j, the partial correlations between Y and X_j.

However, during the estimation of $\mathbf{K}$, one often encounters *spurious* edges that are not exactly zero due to the sampling variation (Costantini et al., 2015). To avoid overfitting, a *regularized* partial correlation network (Foygel and Drton, 2010), or *partial correlation graphical Lasso* (glasso), has become prevalent. Accordingly, the log-likelihood adds a L_1 penalty to the sum of absolute covariance values to yield the MLE of $\mathbf{K}$:

$$\mathbf{K}^{\lambda}(\mathbf{S}) = \arg\min\left\{\text{tr}(\mathbf{SK}) - \log\det(\mathbf{K}) + \lambda \sum_{j=1}^{p+1}\sum_{j'=1}^{p+1} |\kappa_{jj'}|\right\},$$

where $\mathbf{S}$ is the *sample* covariance matrix of $(Y, \mathbf{X})^{\top}$ and λ a tuning parameter to control the sparsity level. $\mathbf{K}^{\lambda}(\mathbf{S})$ is well-suited for high-dimensional settings where $p+1 > n$, under which $\det(\mathbf{K}) = 0$ and regularization is essential as in our context (Williams and Rast, 2020).

To choose λ, one can start by grid search and select the optimal one with minimal EBIC (Extended Bayesian Information Criterion) (Chen and Chen, 2008), which works well in retrieving the true network structure, especially when they are sparse. Let l denote the penalized log-likelihood, m the number of non-zero edges, and n the sample size, the EBIC is defined as $-2l + m\log(n) + 4\gamma m \log(p+1)$, which contains a hyperparameter γ to control the preference over sparsity (i.e., fewer edges), suggested to be set between 0 and 0.5 (Epskamp et al., 2018), with higher values indicating more parsimonious network models are preferred.

Accordingly, the resulting penalized partial correlation coefficients ρ_j^{γ} (range: $[-1, 1]$) encode the conditional association between two variables after controlling for all others and removing spurious connections. This is important for variable selection since Y and X_j being conditionally dependent is equivalent to ρ_j^{γ} being non-zero, implying a stronger signal of X_j, and hence a more *relevant* confounder after penalization. Thus, ρ_j^{γ} captures the relationship between the response and a relevant confounder, conditional on, or partial out, all other covariate effects, whereas $\rho_j^{\gamma} \neq 0$ corresponds to irrelevant confounders, where Y and X_j are conditionally independent.

On the other hand, by directly penalizing elements in the variance-covariance matrix among the vast confounders, this variable selection strategy differs from conventional Lasso, which penalizes the coefficients in the regression model as in the DML approach. As the regularized partial correlations have become standard when estimating psychopathology networks, they provide a promising alternative for selecting confounders in the causal mediation analysis.

4 Real Data Analyses

4.1 Setup

With the motivating dataset described in Section 2, we now investigate whether exposure to loneliness and cognitively stimulating activities will alter cognitive functioning over time (after six months) via modulating the microbiome composition. This observational collected a large number of baseline characteristic variables, including the psychiatric and mental health instruments such as the PHQ9 severity score of depression, positive psychological traits such as Connor Davidson Resilience score, physical health such as sleep, and medications. They help mitigate unmeasured confounding but challenge our statistical inference since the number of confounders ($p = 81$) approaches the sample size ($n = 92$).

Hence, we need to select relevant confounders carefully. We first applied the regularized partial correlation network to screen the 81 potential confounders. Several hyperparameters γ

Table 1: Non-zero connections to loneliness (uclalst) and cognitive stimulus (csascore) with varying γ values for sparsity.

Treatment	Confounder	$\gamma = 0$	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma \geqslant 0.24$
Loneliness	mencomp	−0.011	–	–	No connection
	ldr2	0.080	0.065	0.049	
	lotrt	−0.067	−0.065	−0.060	
	mlq_ps	−0.030	−0.034	−0.035	
	cse_sff	−0.056	−0.062	−0.063	
	nefftot	−0.129	−0.127	−0.123	
	wsdm_cd	0.080	0.062	0.042	
	sdw_sa	−0.013	−0.009	−0.001	
	sdw_psb	−0.030	−0.020	−0.007	
	phycomp	−0.081	−0.072	−0.060	
	prsd8a_ss	0.099	0.072	0.045	
	prsi_ss	0.105	0.084	0.059	
	prdsa_ss	−0.090	−0.069	−0.048	
Cognitive stimulus	socposc	−0.036	–	–	No connection
	lotrt	−0.078	−0.046	−0.006	
	nefftot	−0.028	–	–	
	wsdm_cd	0.054	0.012	–	
	nsictot	−0.014	–	–	

in the EBIC were explored, from dense ($\gamma = 0$) to very sparse ($\gamma = 0.5$), each visualized by using the qgraph package in R (Epskamp et al., 2018). Since γ influences the sparsity of the network connection, the chosen range covers a spectrum of complexities of the model.

The resulting network was estimated using the glasso algorithm, where we examined the number of edges and their strengths (i.e., the magnitude of partial correlations) across different γ values. We present the final models under $\gamma = 0$ and $\gamma = 0.1$ to compare the impact of the confounder selection. In the partial correlation network, the exposure of loneliness and cognitively stimulating activities yielded distinct sets of variables carrying non-zero coefficients, which were used as respective confounders for further causal mediation analyses.

We compared three procedures for deriving the causal mediation effects as follows:

(1) Traditional mediation analysis with network-selected confounders (*Non-EIF + Network-selection*) was conducted with the mediate() function from the mediation package. For the binary outcome, we deployed logistic regression for the outcome model and the linear model for the continuous mediator. For continuous outcome, we used the linear model for the mediator, and a generalized additive model (GAM) with smooth terms for the outcome to allow for a treatment-mediator interaction, which did not impose the same direct and indirect effect under treatment and control and hence is more flexible. Bootstrap was implemented with robust standard errors to account for potential heteroscedasticity.

(2) Double machine learning, or DML with network-selected confounders (*EIF + DML-selection + Network-selection*): the medDML() function from the causalweight package was applied using the selected confounders from the regularized partial correlation network. Cross-validation was employed to tune the hyperparameters.

(3) DML with full confounder set (*EIF + DML-selection*): the DML analysis was repeated using all 81 potential confounders, with the same configuration as in (2) but with an expanded set of confounders.

The three combinations were compared regarding their parameter estimates, standard error, and p-value for the average total, direct, and indirect effects under treatment and control.

Finally, we conducted sensitivity analyses to assess if the assumption of no unmeasured pre-treatment confounders holds using the function medsens() in the mediation package. This function yields a sensitivity parameter at which the causal indirect effect equals zero. More particular, it measures the correlation between the residuals from the mediator and outcome models, namely, let $\varepsilon_{mi} = M_i - \mathrm{E}_{\theta_0}(M_i \mid E_i = e, \mathbf{X}_i)$, $\varepsilon_{yi} = Y_i - \mathrm{E}_{\theta_0}(Y_i \mid E_i = e, M_i^e, \mathbf{X}_i)$, we have

$$r := corr(\varepsilon_{mi}, \varepsilon_{yi}).$$

The interpretation is that when the assumption of no unmeasured pre-treatment confounders holds, the correlation between two model residuals should be close to zero. Thus, larger r values suggest less robust causal and mediation estimators, and caution is warranted (Chi et al., 2022).

We conducted the mediation analyses to investigate the two main scientific questions: first, the path from loneliness to microbiome to cognitive impairment, and second, the path from cognitively stimulating activities to microbiome to cognitive impairment. For each, we considered the continuous and binary outcome of cognitive impairment, measured by the MoCA instrument as a continuous score but also dichotomized by the clinical cutoff (Nasreddine et al., 2005). Sensitivities analyses for each analysis were presented by reporting the corresponding r values. Since this sensitivity analysis is implemented only for linear mediator and outcome models and linear mediator and binary probit outcome models, we assessed the sensitivity for binary MoCA outcome using the probit link even though the final results were based on the logit link.

4.2 Scientific Insights

4.2.1 Loneliness → Microbiome → Cognitive Impairment (MoCA)

With a prevalence rate of 20% to 35% among U.S. adults over the past decade (McGinty et al., 2020), loneliness is considered the latest global health epidemic with serious health implications, including depression, cognitive impairment, hypertension, and frailty (Holt-Lunstad, 2017). However, whether the impact of loneliness on cognitive impairment via the human microbiome, namely, the gut-brain axis, has not been previously investigated, our approach provided some insights as follows.

Binary MoCA outcome (cognitive impairment vs. healthy control): When $\gamma = 0$ as shown in Table 2, the approach (2) deploying DML with network-selected confounders identified marginal indirect control effect ($\Psi^{\mathrm{ME}} = 0.106$, $se = 0.060$, $p\text{-}val = 0.079$); while the DML with a full set of confounders in approach (3) identified an intensified effect ($\Psi^{\mathrm{ME}} = 0.141$, $se = 0.072$, $p\text{-}val = 0.050$), suggesting being exposed to loneliness will increase the odds of being cognitively impaired, mediated through the microbiome alpha-diversity, in particular, the odds ratio of cognitive impairment and healthy control is $exp(0.141) = 1.15$ times when the values of alpha-diversity switches from M_i^0 (under control) to M_i^1 (under the exposure) while holding the actual exposure status at not exposing to loneliness. No significant effect appeared in approach (1).

Table 2: Mediation effects of loneliness/cognitive stimulus on cognitive functioning via gut microbiome ($\gamma = 0$).

Causal path	Outcome type	Effect type	TMA $\gamma = 0$ (1)			DML $\gamma = 0$ (2)			DML-FCS (3)			r_1	r_0
			Effect	SE	p	Effect	SE	p	Effect	SE	p		
L→M→COG	Binary	ACE	0.035	0.192	0.680	0.227	0.280	0.417	0.162	0.303	0.593	−0.1	−0.1
		ADE_1	0.054	0.205	0.610	0.121	0.270	0.654	0.021	0.341	0.951		
		ADE_0	0.055	0.213	0.610	0.167	0.314	0.596	0.078	0.332	0.813		
		AME_1	−0.020	0.089	0.610	0.061	0.118	0.608	0.084	0.083	0.313		
		AME_0	−0.019	0.077	0.620	0.106	0.060	0.079	0.141	0.072	0.050		
	Continuous	ACE	0.711	1.960	0.620	1.068	1.067	0.317	0.393	0.861	0.648	0.2	0.2
		ADE_1	0.399	1.938	0.840	0.763	1.099	0.488	0.273	0.921	0.767		
		ADE_0	0.456	2.056	0.730	0.870	1.065	0.414	0.273	0.954	0.775		
		AME_1	0.254	0.620	0.620	0.198	0.488	0.685	0.120	0.280	0.669		
		AME_0	0.312	0.801	0.430	0.305	0.260	0.241	0.119	0.234	0.610		
S→M→COG	Binary	ACE	−0.005	0.162	0.960	−0.088	0.169	0.603	−0.355	0.315	0.259	−0.2	−0.2
		ADE_1	−0.001	0.164	1.000	−0.091	0.168	0.587	−0.274	0.313	0.382		
		ADE_0	−0.001	0.163	1.000	−0.093	0.170	0.586	−1.677	2.555	0.512		
		AME_1	−0.005	0.035	0.820	0.005	0.009	0.572	1.322	2.610	0.612		
		AME_0	−0.005	0.036	0.820	0.004	0.006	0.539	−0.082	0.050	0.100		
	Continuous	ACE	1.218	0.971	0.330	−0.082	1.041	0.937	1.997	0.971	0.040	0	0.1
		ADE_1	1.187	0.994	0.340	−0.082	1.041	0.937	1.648	0.999	0.099		
		ADE_0	1.138	0.954	0.350	0.831	1.399	0.553	11.476	15.604	0.462		
		AME_1	0.080	0.533	0.840	−0.913	1.086	0.401	−9.479	15.748	0.547		
		AME_0	0.032	0.273	0.910	0.000	0.000	1.000	0.350	0.244	0.152		

Notes: TMA = Traditional Mediation Analysis; DML = Double Machine Learning; DML-FCS = DML with full confounder set; ACE = Average total causal effect; ADE_1 = Average direct treatment effect; ADE_0 = Average direct control effect; AME_1 = Average indirect treatment effect; AME_0 = Average indirect control effect. L = Loneliness (*uclalst*); S = Cognitive Stimulus (*csascore*); M = Microbiome α-diversity (faith_pd), COG = Cognitive Functioning (MoCA); r_1 = sensitivity parameter under exposure; r_0 = sensitivity parameter under control.

Continuous MoCA outcome: No significant effects were identified for the continuous outcome.

4.2.2 Cognitively Stimulating Activities → Microbiome → Cognitive Impairment (MoCA)

The composite score of cognitive activity participation ranges from 1 to 5, with higher scores indicating more frequent participation in cognitively stimulating activities, which include education and training courses, reading, crossword puzzles, and playing chess or card games. Such activities have been shown to impact the cognitive functioning of older adults. For example, a study found that a 1-point increase in cognitive activity score was associated with a 33% reduction in the risk of Alzheimer's disease (AD) (Wilson et al., 2002). Here, we validate the gut-brain axis to assess whether this path is partly through the microbiome using causal mediation analyses.

Binary MoCA outcome: When $\gamma = 0.1$ as shown in Table 3, the approach (2) using DML with network-selected confounders identified a significant indirect exposure effect ($\Psi^{\text{ME}} = 0.014$, $se = 0.005$, $p\text{-}val = 0.013$), supporting the mediation effect through the microbiome alpha-diversity, in particular, the odds ratio of cognitive impairment and healthy control is $exp(0.014) = 1.014$ times when the values of alpha-diversity are switch on (i.e., M_i^1 under the exposure) while holding the actual exposure to cognitive stimulates.

Continuous MoCA outcome (the higher, the less cognitively impaired): With $\gamma = 0.1$, the partial network selected only LOT-R Total Score (Optimism) and 3D Wisdom Scale – Cognitive dimension as confounders. The mediation package with these two confounders in approach (1) identified a significant total effect ($\Psi^{\text{TE}} = 1.941$, $se = 0.853$, $p\text{-}val = 0.024$), as well as direct effect under the exposure and control ($\Psi^{\text{DE}} = 1.909$, exposure $p\text{-}val = 0.024$, control $p\text{-}val = 0.018$); they suggest that being exposed to cognitive activities can improve the cognitive functioning.

Interestingly, the network selected three more confounders with less stringent EBIC where $\gamma = 0$, including Hollingshead Index of Social Position (ISP) – Current Status, Neff Self-Compassion Scale score, and Nutrition Screening Checklist Total Score. The previous two effects were no longer significant after adding them.

Also, the DML approach with network-selected confounders in (2) did not find any significant effect for $\gamma = 0$ or 0.1.

The DML approach in (3) with a full set of confounders identified a significant total effect ($\Psi^{\text{TE}} = 1.997$, $se = 0.971$, $p\text{-}val = 0.040$), yet the direct effects were no longer significant.

4.3 Methodological Implications

By comparing the analyses across the three combinations in selecting confounders and mediation estimation, we present some implications in statistical methods as follows.

In network-based confounder selection, the confounding sets were stable for the exposure of loneliness across different γ values, which induces more comparable final causal estimators. For example, when $\gamma = 0$, this loose EBIC criteria only yields one additional confounder of the SF-36 Mental Component Scale, compared with more stringent $\gamma = 0.1$ or 0.2, which both selected 12 confounders. It partly explained the similar final results of loneliness to microbiome to binary cognitive impairment (MoCA) when comparing $\gamma = 0$ and 0.1. However, this was not the case

Table 3: Mediation effects of loneliness/cognitive stimulus on cognitive functioning via gut microbiome ($\gamma = 0.1$).

Causal path	Outcome type	Effect type	TMA $\gamma = 0.1$ (1)			DML $\gamma = 0.1$ (2)			DML-FCS (3)			r_1	r_0
			Effect	SE	p	Effect	SE	p	Effect	SE	p		
L→M→COG	Binary	ACE	0.057	0.180	0.710	0.196	0.277	0.479		/		−0.1	−0.1
		ADE_1	0.068	0.193	0.610	0.137	0.269	0.610					
		ADE_0	0.068	0.206	0.610	0.137	0.309	0.657					
		AME_1	−0.011	0.083	0.650	0.059	0.116	0.612					
		AME_0	−0.010	0.074	0.660	0.059	0.066	0.368					
	Continuous	ACE	0.774	1.628	0.610	1.395	1.044	0.181		/		0.1	0
		ADE_1	0.404	1.765	0.860	1.168	1.075	0.277					
		ADE_0	0.668	1.876	0.670	1.040	1.072	0.332					
		AME_1	0.106	0.804	0.900	0.356	0.451	0.431					
		AME_0	0.370	0.730	0.430	0.227	0.259	0.380					
S→M→COG	Binary	ACE	−0.004	0.151	0.900	−0.116	0.176	0.510		/		−0.1	−0.1
		ADE_1	0.001	0.152	0.930	−0.120	0.175	0.496					
		ADE_0	0.001	0.154	0.930	−0.129	0.177	0.465					
		AME_1	−0.005	0.028	0.780	0.014	0.005	0.013					
		AME_0	−0.005	0.028	0.780	0.004	0.006	0.506					
	Continuous	ACE	1.941	0.853	0.024	−1.428	1.283	0.266		/		0.3	0
		ADE_1	1.909	0.872	0.024	−1.428	1.283	0.266					
		ADE_0	1.909	0.864	0.018	−1.269	1.195	0.288					
		AME_1	0.032	0.270	0.822	−0.159	0.188	0.397					
		AME_0	0.032	0.236	0.884	0.000	0.000	1.000					

Notes: TMA = Traditional Mediation Analysis; DML = Double Machine Learning; DML-FCS = DML with full confounder set; ACE = Average total causal effect; ADE_1 = Average direct treatment effect; ADE_0 = Average direct control effect; AME_1 = Average indirect treatment effect; AME_0 = Average indirect control effect. L = Loneliness (*uclalst*); S = Cognitive Stimulus (*csascore*); M = Microbiome α-diversity (faith_pd), COG = Cognitive Functioning (MoCA); r_1 = sensitivity parameter under exposure; r_0 = sensitivity parameter under control.

for cognitively stimulating activities, where the selected sets were much smaller. For example, when $\gamma = 0$, five confounders were selected, which reduced to two when $\gamma = 0.1$ and one when $\gamma = 0.2$, as shown in Table 1.

DML with distinct confounder sets can lead to different mediation estimations, regardless of continuous or binary outcomes. For example, when $\gamma = 0$, albeit the same sign for the indirect control effect of microbiome under approaches (2) and (3), DML with the full sets yielded a larger effect (0.106 *vs.* 0.141) and a smaller p-value (0.079 *vs.* 0.050) from loneliness to the binary MoCA. This is especially the case if more distinct confounder sets are selected. In particular, for the path of cognitively stimulating activities to microbiome to binary MoCA, the indirect exposure effect was significantly positive when $\gamma = 0.1$ but almost negligible when $\gamma = 0$ (0.014 *vs.* 0.005).

DML shows the potential for being more sensitive than traditional mediation analyses in identifying total, direct, or indirect effects. In the eight evaluated hypotheses, the DML-based approach identified significant or marginally significant effects in six cases, four of which used the complete set of confounders. These results indicate that the confounder screening process in DML may help reduce overfitting and support causal inference, with cross-fitting and Neyman orthogonality contributing to its performance.

In this dataset, DML performed better in detecting effects for binary outcomes than continuous ones. Four of the six significant effects identified by the DML-based approach were related to binary outcomes, while the corresponding continuous outcomes did not show significant results. For example, in the pathway from cognitive stimulating activities to the microbiome and then to binary MoCA, DML with network-based confounders identified a significant indirect effect ($\Psi^{\mathrm{ME}} = 0.014$, $se = 0.005$, $p\text{-}val = 0.013$). In contrast, no strong effect was observed for the continuous MoCA outcome ($p\text{-}val = 0.397$).

The network-based confounder selection can help add to the DML approach, especially in improving the robustness and stability of the model fit. For instance, in the path of cognitively stimulating activities to microbiome to continuous MoCA, DML with the full sets in approach (3) yielded inflated direct and indirect effects compared with approach (2) that deploys the network-selected sets (*e.g.*, 0.83 *vs.* 11.48), which could be an artifact of the sample-splitting over a small sample size.

5 Discussion

In this paper, we addressed a timely issue of quantifying the causal mediation effect encountering high-dimensional confounders. Under the counterfactual framework, we first showed that the average causal effect (ACE) is decomposed into the average indirect (or mediation, AME) and direct effects (ADE), which facilitated constructing the nonparametric target causal functionals without attaching to any specific model. Later, two confounding selection strategies were carefully studied, including double machine learning (DML) and regularized partial correlation network. To our knowledge, these two promising approaches have not been compared in the growing causal mediation setting under high dimensionality.

We, hence, offered thorough comparisons among various combinations to evaluate their impacts on the final estimation of target parameters, which not only guides real-world applications for practitioners but also incentivizes future advancements for this important topic.

In our motivating data from a longitudinal observational study on the human microbiome, we encountered high dimensionality in both the mediator and confounders, coupled with a small

to moderate sample size. For the mediator of microbiome taxa counts, we leveraged the feature aggregation to enrich signals and domain-specific structures using diversity metrics (Liu et al., 2024), which have been well-recognized in the field (Cho and Blaser, 2012; Meyer et al., 2022). For the massive confounders, we considered the two confounding selection strategies. Along with the efficient influence functions for the causal mediation effects, three combinations were carefully studied to demystify causality in the "gut-brain axis."Our results are consistent with the scientific literature but offer nuanced methodological implications on how the confounding selection impacts the final causal target parameter estimation, above and beyond the real-world scientific insights.

However, the study results are still limited by the relatively small sample size and the imputation of the missing covariates in the follow-up visit. Even with the Neyman orthogonality, the performance of DML still depends on the accuracy of the nuisance function estimators, including the propensity score, outcome, and mediation models. This emphasizes the importance of selecting appropriate ML algorithms; in some settings, expert knowledge can guide the choice based on the characteristics of the data. In our application, for instance, the default LASSO method in the function *medDML()* was used due to the sparsity. As highlighted in Hünermund et al. (2023), while DML is a powerful tool for variable selection in high-dimensional data, it is crucial to use it cautiously within the empirical context, as inappropriate choices may compromise the conclusion's validity. Accordingly, we have also discussed the possibility of using the regularized partial correlation network in conjunction with the model fit to improve the stability of the model. Nonetheless, the sensitivity analyses suggested that the assumption of no unmeasured pre-treatment confounders was not strongly violated in the various hypotheses we tested.

Finally, our causal mediation analyses showcased the exposure impact of loneliness and cognitively stimulative activities on cognitive functioning for the aging population, which is mediated by their microbiome composition. Albeit in the early stage, some clinical trials have been administered to examine probiotics as a treatment option for mental disorders including cognitive impairment (Northumbria University, 2019; Cohen-Kadosh, 2020). The term psycho-biotic was coined to describe live bacteria or prebiotics that confer mental health benefits, such as improved mood, reduced anxiety, and enhanced cognitive function (Sasso et al., 2023). Our derived insights contribute to support that augmenting psychosocial and behavioral modulations (e.g., strengthen social support and reduce loneliness, expand cognitive stimulus activities) may improve the therapeutical effect of psycho-biotics, especially for the aging population (Meyer et al., 2022). Our thorough comparison studies are valuable in many other growing fields encountering high dimensionality, such as the metabolomics or functional connectivity in neuroimage that are commonly hypothesized as the mediator (Booth et al., 2013; Lindquist, 2012).

In summary, our results highlighted the practicality and necessity of the discussed methods in mitigating selection bias in causal mediation analysis, especially when the dimension of mediator and confounders exceed the sample size.

Supplementary Material

Contains Figures 2, 3, 4, and 5.

A Appendix

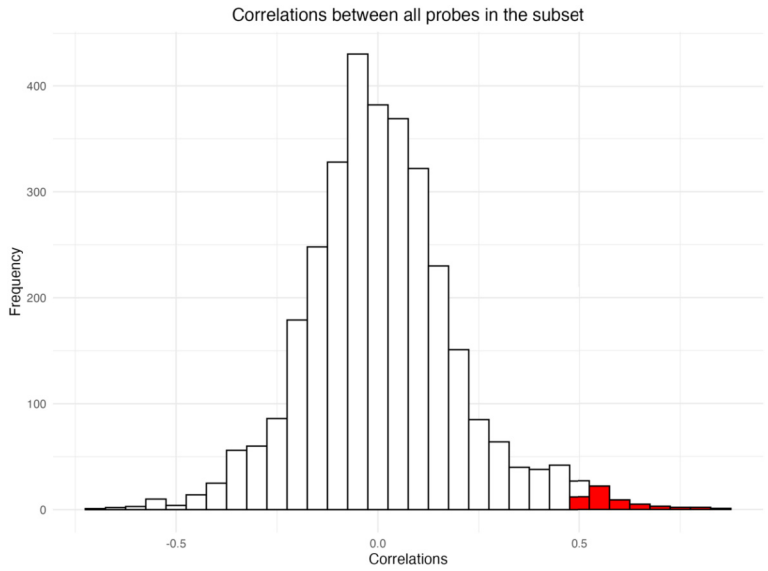

Figure 2: Pairwise correlation among confounders.

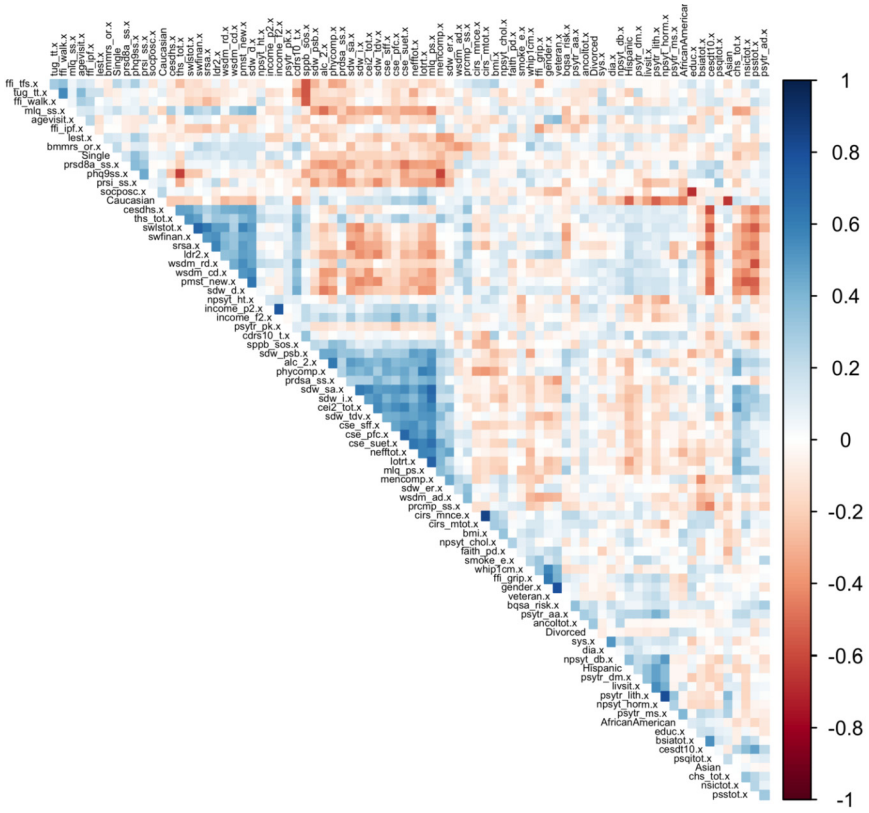

Figure 3: Correlation heatmap between confounders.

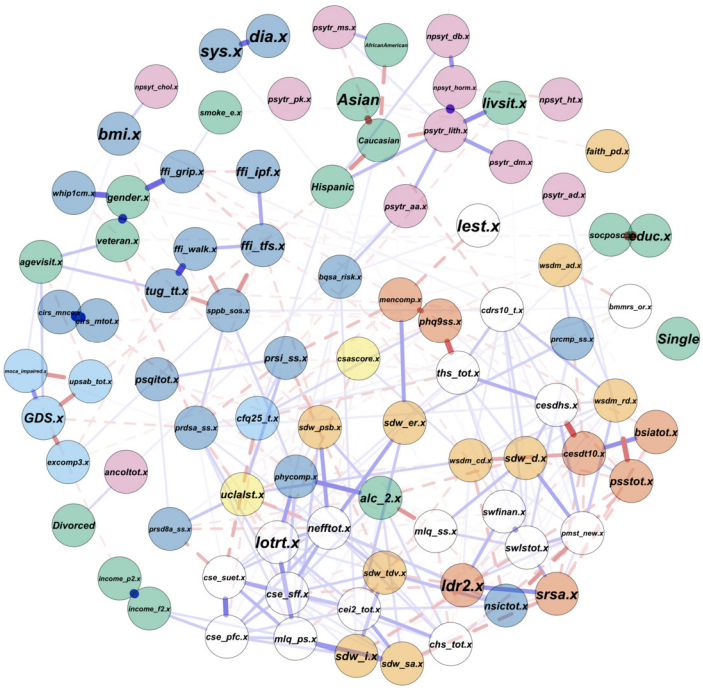

Figure 4: Network at $\gamma = 0$.

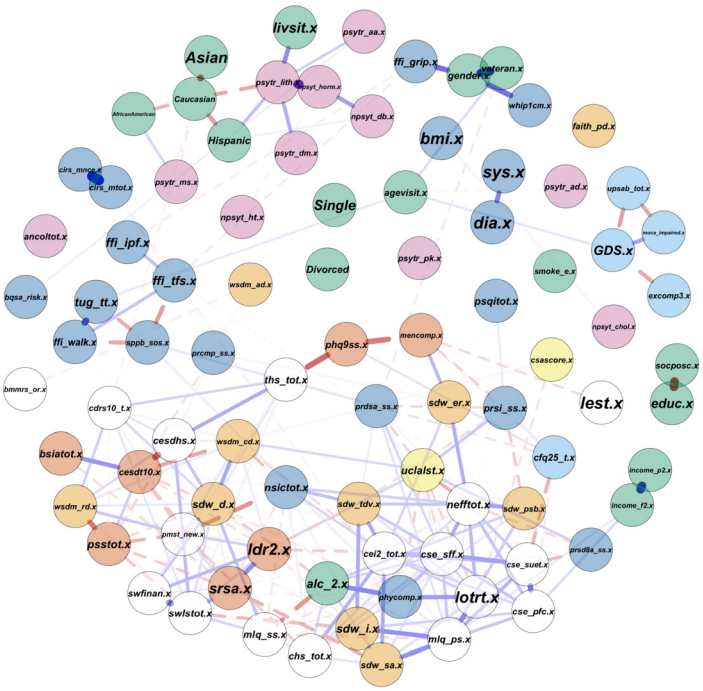

Figure 5: Network at $\gamma = 0.1$.

References

Booth SL, Centi A, Smith SR, Gundberg C (2013). The role of osteocalcin in human glucose metabolism: marker or mediator? *Nature Reviews Endocrinology*, 9(1): 43–55. https://doi.org/10.1038/nrendo.2012.201

Borsboom D, Cramer AO (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9(1): 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Chen J, Chen Z (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3): 759–771. https://doi.org/10.1093/biomet/asn034

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, et al. (2018). Double/debiased machine learning for treatment and structural parameters.

Chernozhukov V, Escanciano JC, Ichimura H, Newey WK, Robins JM (2022). Locally robust semiparametric estimation. *Econometrica*, 90(4): 1501–1535. https://doi.org/10.3982/ECTA16294

Chi WE, Huang S, Jeon M, Park ES, Melguizo T, Kezar A (2022). A practical guide to causal mediation analysis: illustration with a comprehensive college transition program and non-program peer and faculty interactions. *Frontiers in Education*, 7: 886722. https://doi.org/10.3389/feduc.2022.886722

Cho I, Blaser MJ (2012). The human microbiome: at the interface of health and disease. *Nature Reviews. Genetics*, 13(4): 260–270. https://doi.org/10.1038/nrg3182

Cohen-Kadosh K (2020). The role of the microbiota-gut-brain axis in brain development and mental health: Behavioural. ClinicalTrials.gov Identifier: NCT04616937. Updated November 15, 2020. Accessed November 28, 2022. Available at: https://clinicaltrials.gov/ct2/show/NCT04616937.

Costantini G, et al. (2015). Network analysis: A new perspective on personality psychology.

Epskamp S, Maris G, Waldorp LJ, Borsboom D (2018). Network psychometrics. In: Irwing P, Booth T, Hughes DJ (eds.), *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development*, 953–986.

Farbmacher H, Huber M, Lafférs L, Langen H, Spindler M (2022). Causal mediation analysis with double machine learning. *Econometrics Journal*, 25(2): 277–300. https://doi.org/10.1093/ectj/utac003

Foster JA, Baker GB, Dursun SM (2021). The relationship between the gut microbiome-immune system-brain axis and major depressive disorder. *Frontiers in Neurology*, 12: 721126. https://doi.org/10.3389/fneur.2021.721126

Foygel R, Drton M (2010). Extended bayesian information criteria for gaussian graphical models. *Advances in Neural Information Processing Systems*, 23.

Gunzler D, Tang W, Lu N, Wu P, Tu X (2014). A class of distribution-free models for longitudinal mediation analysis. *Psychometrika*, 79: 543–568. https://doi.org/10.1007/s11336-013-9355-z

Holt-Lunstad J (2017). The potential public health relevance of social isolation and loneliness: prevalence, epidemiology, and risk factors. *Public Policy & Aging Report*, 27(4): 127–130. https://doi.org/10.1093/ppar/prx030

Hünermund P, Louw B, Caspi I (2023). Double machine learning and automated confounder selection: a cautionary tale. *Journal of Causal Inference*, 11(1): 20220078. https://doi.org/10.1515/jci-2022-0078

Imai K, Keele L, Tingley D (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15(4): 309. https://doi.org/10.1037/a0020761

Koller D, Friedman N (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Lauritzen S (1996). *Graphical Models*. Clarendon Press.

Lindquist MA (2012). Functional causal mediation analysis with an application to brain connectivity. *Journal of the American Statistical Association*, 107(500): 1297–1309. https://doi.org/10.1080/01621459.2012.695640

Liu J, Lin T, Chen T, Zhang X, Tu XM (2022). On semiparametric efficiency of an emerging class of regression models for between-subject attributes. arXiv preprint: https://arxiv.org/abs/2205.08036.

Liu J, Xu K, Wu T, Yao L, Nguyen TT, Jeste D, et al. (2023). Deciphering the 'gut–brain axis' through microbiome diversity. *General Psychiatry*, 36(5): e101090. https://doi.org/10.1136/gpsych-2023-101090

Liu J, Zhang X, Chen T, Wu T, Lin T, Jiang L, et al. (2022). A semiparametric model for between-subject attributes: applications to beta-diversity of microbiome data. *Biometrics*, 78(3): 950–962. https://doi.org/10.1111/biom.13487

Liu J, Zhang X, Lin T, Chen R, Zhong Y, Chen T, et al. (2024). A new paradigm for high-dimensional data: distance-based semiparametric feature aggregation framework via between-subject attributes. *Scandinavian Journal of Statistics*, 51(2): 672–696. https://doi.org/10.1111/sjos.12695

McGinty EE, Presskreischer R, Han H, Barry CL (2020). Psychological distress and loneliness reported by us adults in 2018 and April 2020. *JAMA*, 324(1): 93–94. https://doi.org/10.1001/jama.2020.9740

McNally RJ, Robinaugh DJ, Wu GW, Wang L, Deserno MK, Borsboom D (2015). Mental disorders as causal systems: a network approach to posttraumatic stress disorder. *Clinical Psychological Science*, 3(6): 836–849. https://doi.org/10.1177/2167702614553230

Meyer K, Lulla A, Debroy K, Shikany JM, Yaffe K, Meirelles O, et al. (2022). Association of the gut microbiota with cognitive function in midlife. *JAMA Network Open*, 5(2): e2143941. https://doi.org/10.1001/jamanetworkopen.2021.43941

Morais LH, Schreiber HL IV, Mazmanian SK (2021). The gut microbiota–brain axis in behaviour and brain disorders. *Nature Review, Microbiology*, 19(4): 241–255. https://doi.org/10.1038/s41579-020-00460-0

Murphy KP (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. (2005). The Montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4): 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Neyman J (1979). $C(\alpha)$ tests and their use. *Sankhya. Series A*, 41(1/2): 1–21.

Nguyen TT, Zhang X, Wu TC, Liu J, Le C, Tu XM, et al. (2021). Association of loneliness and wisdom with gut microbial diversity and composition: an exploratory study. *Frontiers in Psychiatry*, 12: 648475. https://doi.org/10.3389/fpsyt.2021.648475

Northumbria University (2019). The cognitive effects of 6 weeks administration with a probiotic: a randomized, placebo controlled proof-of-concept study in healthy elderly humans. ClinicalTrials.gov Identifier: NCT03601559. Updated June 18, 2019. Accessed November 28, 2022. Available at: https://clinicaltrials.gov/ct2/show/NCT03601559.

Pearl J (2014). Interpretation and identification of causal mediation. *Psychological Methods*, 19(4): 459. https://doi.org/10.1037/a0036434

Robins JM, Greenland S (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2): 143–155. https://doi.org/10.1097/00001648-199203000-00013

Rubin DB (1990). Formal mode of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, 25(3): 279–292. https://doi.org/10.1016/0378-3758(90)90077-8

Sasso J, Ammar R, Tenchov R, Lemmel S, Kelber O, Grieswelle M, et al. (2023). Gut microbiome–brain alliance: a landscape view into mental and gastrointestinal health and disorders. *ACS Chemical Neuroscience*, 14(10): 1717–1763. https://doi.org/10.1021/acschemneuro.3c00127

Sgritta M, Dooling SW, Buffington SA, Momin EN, Francis MB, Britton RA, et al. (2019). Mechanisms underlying microbial-mediated changes in social behavior in mouse models of autism spectrum disorder. *Neuron*, 101(2): 246–259. https://doi.org/10.1016/j.neuron.2018.11.018

Tchetgen EJT, Shpitser I (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *The Annals of Statistics*, 40(3): 1816.

Tibshirani R (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 58(1): 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tsiatis AA (2006). *Semiparametric Theory and Missing Data*, volume 4. Springer.

Virgin HW, Todd JA (2011). Metagenomics and personalized medicine. *Cell*, 147(1): 44–56. https://doi.org/10.1016/j.cell.2011.09.009

Wang Z, van der Laan L, Petersen M, Gerds T, Kvist K, van der Laan M (2023). Targeted maximum likelihood based estimation for longitudinal mediation analysis. arXiv preprint.

Williams DR, Rast P (2020). Back to the basics: rethinking partial correlation network methodology. *British Journal of Mathematical & Statistical Psychology*, 73(2): 187–212. https://doi.org/10.1111/bmsp.12173

Wilson RS, De Leon CFM, Barnes LL, Schneider JA, Bienias JL, Evans DA, et al. (2002). Participation in cognitively stimulating activities and risk of incident Alzheimer disease. *JAMA*, 287(6): 742–748. https://doi.org/10.1001/jama.287.6.742

Xue F, Qu A (2022). Semi-standard partial covariance variable selection when irrepresentable conditions fail. *Statistica Sinica*, 32(4): 1881–1909.

Zheng W, Van Der Laan MJ (2012). Targeted maximum likelihood estimation of natural direct effects. *The International Journal of Biostatistics*, 8(1): 1–40. https://doi.org/10.2202/1557-4679.1361

Zou H (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476): 1418–1429. https://doi.org/10.1198/016214506000000735