

COMPOSITIONAL MEDIATION ANALYSIS FOR MICROBIOME STUDIES¹

BY MICHAEL B. SOHN AND HONGZHE LI²

University of Pennsylvania

Motivated by recent advances in causal mediation analysis and problems in the analysis of microbiome data, we consider the setting where the effect of a treatment on an outcome is transmitted through perturbing the microbial communities or compositional mediators. The compositional and high-dimensional nature of such mediators makes the standard mediation analysis not directly applicable to our setting. We propose a sparse compositional mediation model that can be used to estimate the causal direct and indirect (or mediation) effects utilizing the algebra for compositional data in the simplex space. We also propose tests of total and component-wise mediation effects. We conduct extensive simulation studies to assess the performance of the proposed method and apply the method to a real microbiome dataset to investigate an effect of fat intake on body mass index mediated through the gut microbiome.

1. Introduction. It has been shown that fat intake is associated with body mass index (BMI) (Bray and Popkin (1998)) and obesity is associated with the gut microbiome (Ley et al. (2006), Turnbaugh et al. (2006)). From this information, a very natural question to ask is whether fat intake has some effects on BMI mediated through the perturbation of the gut microbiome. The approach to answering this type of questions is known as “mediation analysis.” Mediation analysis is a statistical method of studying the effect of a treatment or exposure on an outcome transmitted through intermediate variables, referred to as “mediators” or “intervening variables.” It has been widely applied in various disciplines, such as sociology, psychology, and epidemiology, since the influential paper by Baron and Kenny (1986) and become increasingly popular due to recent advances in causal inference (Imai, Keele and Yamamoto (2010), Pearl (2000), Rubin (2005)), which clarifies the assumptions needed for causal interpretation. Until recently, mediation analysis has been restricted to a single mediator as depicted in Figure 1, and the effect of a treatment on an outcome transmitted through a mediator is often formulated and implemented within the framework of linear structural equation models (LSEMs).

Received October 2016; revised June 2018.

¹Supported by NIH Grants CA127334 and GM123056.

²Corresponding author.

Key words and phrases. Compositional algebra, 16S sequencing, causal mediation effect, simplex space.

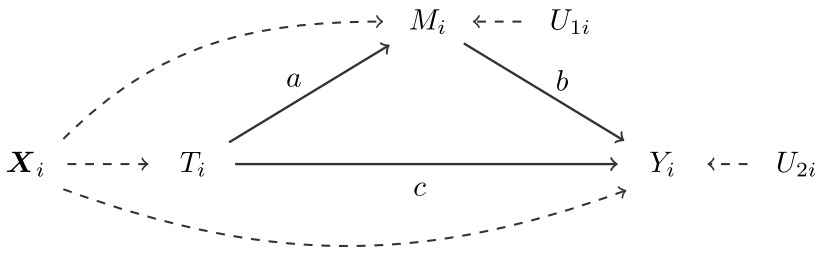


FIG. 1. A single-mediator model: a , b , and c are path coefficients; U_{1i} and U_{2i} are the disturbance variables for a mediator M_i and an outcome Y_i , respectively; T_i is a treatment variable; X_i is a set of pre-treatment variables.

For instance, an LSEM for the path diagram in Figure 1 can be formulated as

$$M_i = a_0 + aT_i + \mathbf{X}_i^\top \mathbf{h} + U_{1i},$$

$$Y_i = c_0 + cT_i + bM_i + \mathbf{X}_i^\top \mathbf{g} + U_{2i},$$

where T_i is a treatment for unit i , M_i a mediator, Y_i an outcome variable, and X_i a set of pre-treatment variables that may affect the treatment, mediator, and outcome; a , b , and c are path coefficients; a_0 and c_0 are intercepts for M_i and Y_i , respectively; \mathbf{h} and \mathbf{g} are nuisance coefficients for X_i ; U_{1i} and U_{2i} are disturbance variables for M_i and Y_i , respectively. Under this model, the effect of T_i on Y_i transmitted through M_i , called the *indirect* or *mediation effect*, is defined by the product of the path coefficients a and b . The effect of T_i on Y_i not transmitted through M_i , the *direct effect*, is defined by the path coefficient c . The model formulation for Figure 1 shows that the total effect of T_i on Y_i is the sum of the direct and indirect effects, $c + ab$.

In recent years, numerous studies have extended the applicability of mediation analysis: incorporating nonlinearity and interaction between a treatment and a mediator (Imai, Keele and Yamamoto (2010), Pearl (2001), VanderWeele and Vansteelandt (2010)); incorporating multiple mediators (Preacher and Hayes (2008), VanderWeele and Vansteelandt (2014)). A few studies have also proposed methods for high-dimensional mediators. Chén et al. (2018) proposed a method to estimate path coefficients of an LSEM by finding the linear combinations of mediators that maximize the likelihood of linear structural equations, which is similar to principal components. Huang and Pan (2016) introduced a transformation model using principal components and included the interaction in their model. Zhao and Luo (2016) proposed a sparse mediation model using a regularized LSEM approach.

In this paper, we contribute to extending the applicability of mediation analysis further by proposing an estimating method for the causal direct and indirect effects when mediators are compositional. Compositional data refer to proportions or percentages of a whole and frequently arise in a wide range of disciplines

such as mineral components of a rock in geology and vote shares of an election in psephology. In microbiome and metagenomic studies, to account for different sizes of sequencing libraries for 16S rRNA or shotgun metagenomic sequencing, the sequencing reads (i.e., count data) are often normalized into proportions. This normalization introduces the unit-sum constraint (i.e., proportions sum to unity), which transforms a k dimensional Euclidian space \mathbb{R}^k into a $k - 1$ dimensional simplex space \mathbb{S}^{k-1} , thus making the statistical models for unconstrained data inappropriate for compositional data. To deal with the nature of compositional data, [Aitchison \(1982\)](#) introduced an axiomatic approach with various operations under logratio transformation, which provides a one-to-one mapping between \mathbb{R}^k and \mathbb{S}^{k-1} , and various researchers including himself have formalized and extended this approach. [Aitchison and Bacon-Shone \(1984\)](#) proposed a linear and quadratic log contrast model for compositional covariates. [Billheimer, Guttorp and Fagan \(2001\)](#) formulated the algebra for compositions in the simplex space. [Lin et al. \(2014\)](#) developed a linear log contrast regression model for compositional covariates in a high-dimensional setting, and [Shi, Zhang and Li \(2016\)](#) generalized the linear log contrast regression model.

We propose a framework for mediation analysis when mediators are high dimensional and compositional. Our framework utilizes two components: (1) an estimation method based on the compositional operators of [Aitchison \(1982\)](#) and the composition algebra of [Billheimer, Guttorp and Fagan \(2001\)](#) and (2) the linear log contrast regression of [Lin et al. \(2014\)](#), [Shi, Zhang and Li \(2016\)](#). We employ the first component (i.e., compositional algebra) to jointly estimate an effect of a randomly assigned treatment on all compositional mediators. To this end, we propose minimizing the difference between observed and estimated compositions with a norm for the composition. We use the second component to quantify an effect of a treatment and compositional mediators on an outcome. Under this compositional mediation framework, we show that the causal direct and indirect effects are identifiable under some assumptions.

Section 2 introduces a compositional mediation model for a continuous outcome and discusses model assumptions and identifiability conditions. Section 3 describes methods of estimating composition and regression parameters and their covariance matrices. We also discuss null hypotheses for the total and component-wise mediation effects and introduce a method for sensitivity analysis. Section 4 compares the performance of our method in extensive simulation studies with two methods that can be applied to compositional mediators. Section 5 presents an application of the proposed method to a gut microbiome dataset. Finally, Section 6 presents a brief discussion of the methods and results.

2. Compositional mediation model and causal interpretation.

2.1. *Notation.* For unit i , we let T_i be a treatment, \mathbf{M}_i a vector of k compositional mediators, Y_i an outcome, and \mathbf{X}_i a set of pre-treatment variables that may

affect the treatment, mediator, and outcome. We denote by \mathcal{T} , \mathcal{M} , and \mathcal{X} , respectively, the support of the distribution of T_i , M_i , and X_i . In the model assumptions and the identification of the causal direct and indirect effects, we adopt the potential outcomes framework. We denote by $M_i(t)$ the potential outcome under $T_i = t$ and by $Y_i(t, \mathbf{m})$ the potential outcome under $T_i = t$ and $M_i = \mathbf{m}$. Thus, we can express an observed variable $M_i = M_i(T_i)$. Similarly, $Y_i = Y_i(T_i, M_i(T_i))$.

2.2. *Compositional mediation model.* Suppose that we have a random sample of size n from a population where we observe Y_i , T_i , X_i , and M_i for each unit i . Note that $M_i \in \mathbb{S}^{k-1}$ for all i , that is, $M_i = \{(M_{i1}, \dots, M_{ik}) : M_{ij} > 0, j = 1, \dots, k, \sum_{j=1}^k M_{ij} = 1\}$. Figure 2 shows the effect of T_i on Y_i mediated through M_i given X_i .

Before introducing a compositional mediation model, we define some compositional operators as in Aitchison, 1982. For two compositions $\eta, \zeta \in \mathbb{S}^{k-1}$, the perturbation operator is defined by

$$\eta \oplus \zeta = \left(\frac{\eta_1 \zeta_1}{\sum_{j=1}^k \eta_j \zeta_j}, \frac{\eta_2 \zeta_2}{\sum_{j=1}^k \eta_j \zeta_j}, \dots, \frac{\eta_k \zeta_k}{\sum_{j=1}^k \eta_j \zeta_j} \right)^\top,$$

and the power transformation for a composition η by a scalar v by

$$\eta^v = \left(\frac{\eta_1^v}{\sum_{j=1}^k \eta_j^v}, \frac{\eta_2^v}{\sum_{j=1}^k \eta_j^v}, \dots, \frac{\eta_k^v}{\sum_{j=1}^k \eta_j^v} \right)^\top.$$

With these operators, we propose the following compositional mediation model:

- (1) $M_i = (\mathbf{m}_0 \oplus \mathbf{a}^{T_i} \oplus \mathbf{h}_1^{X_{i1}} \oplus \dots \oplus \mathbf{h}_q^{X_{iq}}) \oplus U_{1i}$,
- (2) $Y_i = c_0 + cT_i + (\log M_i)^\top \mathbf{b} + X_i^\top \mathbf{g} + U_{2i}$ subject to $\mathbf{b}^\top \mathbf{1}_k = 0$,

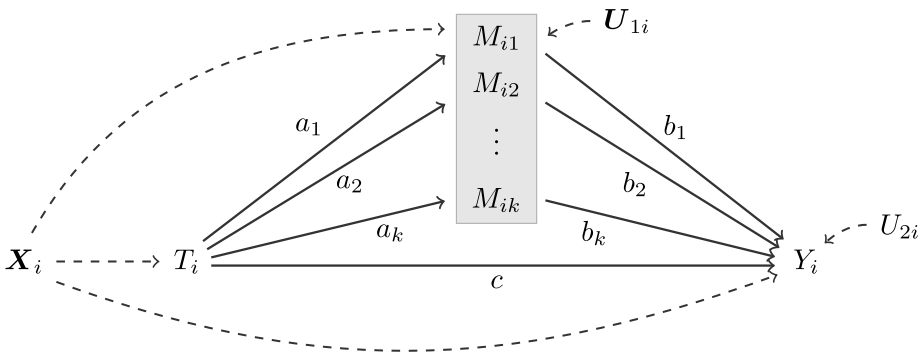


FIG. 2. A compositional mediation model: a_j, b_j and c are path coefficients, $j = 1, \dots, k$; U_{1i} and U_{2i} are disturbance variables for k compositional mediators M_i and an outcome Y_i , respectively; T_i is a treatment variable; X_i is a set of pre-treatment variables.

where \mathbf{m}_0 is the baseline composition (i.e., when $T_i = \mathbb{E}(T_i)$); similarly, c_0 is the baseline for Y_i ; \mathbf{a} , \mathbf{b} , and c are path coefficients; $\mathbf{h}_1, \dots, \mathbf{h}_q$ and \mathbf{g} are nuisance parameters corresponding to \mathbf{X}_i ; $\mathbf{1}_k$ is a vector of k ones. Note that $\mathbf{U}_{1i} \in \mathbb{S}^{k-1}$ since $\mathbf{M}_i \in \mathbb{S}^{k-1}$. We do not specify the distribution of \mathbf{U}_{1i} ; however, we assume that \mathbf{U}_{1i} is perturbed around the identity element \mathcal{J}_{k-1} of \mathbb{S}^{k-1} , that is, $\mathbb{E}(\mathbf{U}_{1i}) = \mathcal{J}_{k-1}$, where $\mathcal{J}_{k-1} = \mathbf{1}_k/k$. We also assume $U_{2i} \sim N(0, \sigma^2)$. To better understand the compositional disturbance term \mathbf{U}_{1i} , we can transform the model (1) using the additive logratio transformation. For a composition vector $\boldsymbol{\eta}$, the additive logratio transformation (alt) of $\boldsymbol{\eta}$ is defined as

$$\text{alt}(\boldsymbol{\eta}) = \left(\log \frac{\eta_1}{\eta_k}, \log \frac{\eta_2}{\eta_k}, \dots, \log \frac{\eta_{k-1}}{\eta_k} \right)^\top.$$

By taking the alt() on both sides of model (1), we have

$$\text{alt}(\mathbf{M}_i) = \text{alt}(\mathbf{m}_0) + T_i \text{alt}(\mathbf{a}) + \sum_{r=1}^q X_{ir} \text{alt}(\mathbf{h}_r) + \text{alt}(\mathbf{U}_{1i}),$$

that is, $\text{alt}(\mathbf{U}_{1i})$ is modeled linearly with other terms.

The model (1) formulates how a treatment perturbs a composition from the baseline composition, which is measured by the composition parameter \mathbf{a} . With the compositional operators, all the calculations are done within the simplex space, leading to intuitive interpretation. The composition parameter \mathbf{a} is directly interpretable as a composition, and a new composition is the baseline composition (\mathbf{m}_0) perturbed by \mathbf{a} for $T_i = 1$ and $\mathbf{h}_1, \dots, \mathbf{h}_q$ for $\mathbf{X}_i = \mathbf{x}$.

The model (2) links a treatment and a composition to an outcome. To account for the compositional nature of \mathbf{M}_i , we impose a linear constraint, $\mathbf{b}^\top \mathbf{1}_k = 0$, which is crucial for an estimator of regression coefficients to have desirable properties for compositional data, such as subcompositional coherence (Aitchison and Bacon-Shone (1984), Lin et al. (2014)) since the regression coefficient \mathbf{b} is independent of an arbitrary scaling of the basis count from which a composition is obtained, that is, $(\log C \mathbf{M}_i)^\top \mathbf{b} = (\log \mathbf{M}_i)^\top \mathbf{b}$ for any constant C . In addition, the regression coefficients remain unaffected by correctly excluding some or all of the zero components Lin et al. (2014). The subcompositional coherence property is one of the principals of compositional data analysis Aitchison (1982). Because of the linear constraint, the interpretation of a given b_j has to be in the context of other nonzero components in \mathbf{b} . Under the linear constraint, model (2) can be rewritten as

$$Y_i = c_0 + cT_i + \text{alt}(\mathbf{M}_i)^\top \mathbf{b}_{-k} + \mathbf{X}_i^\top \mathbf{g} + U_{2i},$$

where $\mathbf{b}_{-k} = (b_1, b_2, \dots, b_{k-1})^\top$.

2.3. *Model assumptions and identification.* Identification of the causal direct and indirect effects requires several assumptions. To define our identifying assumptions, we use the potential outcomes notation described in Section 2.1. Combined with the stable unit treatment value assumption (SUTVA) and the positivity

assumption (i.e., $0 < P(T_i = t | X_i = \mathbf{x})$ and $0 < P(\mathbf{M}_i(t) = \mathbf{m} | T_i = t, X_i = \mathbf{x})$), the assumptions for the compositional mediation model are given by

- (3) $\{Y_i(t', \log(\mathbf{m})), \log \mathbf{M}_i(t)\} \perp\!\!\!\perp T_i | X_i = \mathbf{x}$,
- (4) $Y_i(t', \log(\mathbf{m})) \perp\!\!\!\perp \log \mathbf{M}_i(t) | T_i = t, X_i = \mathbf{x}$,
- (5) No interaction between T_i and \mathbf{M}_i on the response,

for $t, t' \in \mathcal{T}$, $\mathbf{m} \in \mathcal{M}$, and $\mathbf{x} \in \mathcal{X}$. The SUTVA consists of two components: no interference (i.e., no effect of a treatment applied to one unit on an outcome for other units) and no hidden variation of treatment (i.e., consistent treatment levels) (Imbens and Rubin (2015)). The assumptions (3)–(4) with the positivity assumption are an extension of the sequential ignorability assumptions for the single mediator model (Imai, Keele and Tingley (2010)). Basically, they state that there is no unmeasured confounding variable after controlling for X_i . The no interaction assumption (5) can be relaxed; See Supplementary Material VI (Sohn and Li (2019)).

These assumptions appear like those for the multiple mediators model (Imai and Yamamoto (2013), VanderWeele and Vansteelandt (2014)). However, in microbiome studies, the setting we consider is different from the multiple causal mechanisms with multiple mediators considered by Imai and Yamamoto (2013), VanderWeele and Vansteelandt (2014). We assume that a treatment shifts merely the overall microbiome composition from one point in the simplex to another point; therefore, all the components of a composition must be under the same treatment (i.e., $\log \mathbf{M}_i(t)$ or $\log \mathbf{M}_i(t')$). Thus, there are, regardless of k , only four possible cases of conditional independence like the single mediator model. For instance, four possible cases for the assumption (3) are:

- (i) $\{Y_i(t', \log \mathbf{m}), \log \mathbf{M}_i(t)\} \perp\!\!\!\perp T_i = t | X_i = \mathbf{x}$,
- (ii) $\{Y_i(t', \log \mathbf{m}'), \log \mathbf{M}_i(t')\} \perp\!\!\!\perp T_i = t' | X_i = \mathbf{x}$,
- (iii) $\{Y_i(t, \log \mathbf{m}), \log \mathbf{M}_i(t)\} \perp\!\!\!\perp T_i = t | X_i = \mathbf{x}$,
- (iv) $\{Y_i(t, \log \mathbf{m}'), \log \mathbf{M}_i(t')\} \perp\!\!\!\perp T_i = t' | X_i = \mathbf{x}$,

where $\log \mathbf{m}'$ is a vector of the potential values of $\log \mathbf{M}_i(T_i)$ when $T_i = t'$ and $X_i = \mathbf{x}$; t is an observed treatment; t' is a reference value for the treatment. Similarly for the assumption (4).

Under the assumptions of the compositional mediation model, we have the following theorem to show that the models (1) and (2) lead to quantification of the causal direct and total indirect effects.

THEOREM 1 (Identification for the compositional mediation model). *Suppose that the models (1) and (2) are correctly specified, and the assumptions (3)–(5) combined with the SUTVA and the positivity assumptions are satisfied. Then, the*

causal direct effect $\zeta(t)$ and the causal total indirect effect $\delta(t)$ for the compositional mediation model are identifiable and given by

$$(6) \quad \begin{aligned} \zeta(\tau) &\equiv \mathbb{E}[Y_i(t, \log \mathbf{M}_i(\tau)) - Y_i(t_0, \log \mathbf{M}_i(\tau)) | X_i = \mathbf{x}] \\ &= c(t - t_0), \end{aligned}$$

$$(7) \quad \begin{aligned} \delta(\tau) &\equiv \mathbb{E}[Y_i(\tau, \log \mathbf{M}_i(t)) - Y_i(\tau, \log \mathbf{M}_i(t_0)) | X_i = \mathbf{x}] \\ &= (\log \mathbf{a})^\top \mathbf{b}(t - t_0), \end{aligned}$$

where t is an observed treatment for unit i , t_0 a reference value for the treatment, and $\tau = t$ or t_0 .

A proof of Theorem 1 is given in Supplementary Material II. Pearl (2001) calls $\zeta(\tau)$ the average natural direct effect and $\delta(\tau)$ the average natural indirect effect. Pearl (2001) also defines the average controlled direct effect that is defined in terms of a specific value of the mediator, rather than its potential values. Note that $\zeta(t) = \zeta(t_0)$ and $\delta(t) = \delta(t_0)$ under the no interaction assumption. Thus, the controlled and the natural direct effects coincide and are equal to c for a one unit change in t , and similarly, the natural total indirect effect is $(\log \mathbf{a})^\top \mathbf{b}$ for a one unit change in t .

Under the compositional mediation model, the total effect of T_i on Y_i can be decomposed into the direct effect $c(t - t_0)$ and the total indirect effect $(\log \mathbf{a})^\top \mathbf{b}(t - t_0)$, which is the sum of the component-wise indirect effects. Note that under the model (1), \mathbf{a} is interpreted as the expected change in \mathbf{M}_i due to T_i from \mathcal{J}_{k-1} . Thus, in the estimation of the component-wise indirect effects, \mathbf{a} is divided by \mathcal{J}_{k-1} so that $\log(ka_j)$ represents the expected change in $\log(M_{ij})$ from 0 by a one unit change in t . The estimator for a component-wise indirect effect does not possess the scale invariance and subcompositional coherence property. However, it provides indispensable information to understand mediation effect: which components are significant. See Section 3.3 for details.

3. Parameter estimation, variance estimation, and tests of mediation effects.

3.1. Estimation of composition parameters and covariance matrix. The inverse of the perturbation operator is defined by

$$\boldsymbol{\eta} \ominus \boldsymbol{\zeta} = \left(\frac{\eta_1 \zeta_1^{-1}}{\sum_{j=1}^k \eta_j \zeta_j^{-1}}, \frac{\eta_2 \zeta_2^{-1}}{\sum_{j=1}^k \eta_j \zeta_j^{-1}}, \dots, \frac{\eta_k \zeta_k^{-1}}{\sum_{j=1}^k \eta_j \zeta_j^{-1}} \right)^\top.$$

With the composition operators, Billheimer, Guttorp and Fagan (2001) show that \mathbb{S}^{k-1} constitutes a complete inner product space, allowing the definition of a norm for a composition $\boldsymbol{\eta}$:

$$\|\boldsymbol{\eta}\| = (\boldsymbol{\eta}^\top \boldsymbol{\eta})^{1/2} = (\text{alt}(\boldsymbol{\eta})^\top \mathcal{N}^{-1} \text{alt}(\boldsymbol{\eta}))^{1/2},$$

where \mathcal{N}^{-1} is the inverse matrix of a $(k - 1) \times (k - 1)$ matrix \mathcal{N} defined by

$$\mathcal{N} = \mathcal{I}_{k-1} + \mathbf{1}_{k-1}\mathbf{1}_{k-1}^T,$$

where \mathcal{I}_{k-1} is the $(k - 1) \times (k - 1)$ identity matrix.

To estimate the parameters in the model (1), we propose the following objective function, which minimizes the composition norm of the difference between observed and estimated compositions,

$$\begin{aligned} (\hat{\mathbf{a}}, \hat{\mathbf{h}}_r, \hat{\mathbf{m}}_0) &= \operatorname{argmin}_{\mathbf{a}, \mathbf{h}_r, \mathbf{m}_0} \sum_{i=1}^n \|M_i \ominus (\mathbf{m}_0 \oplus \mathbf{a}^{T_i} \oplus \mathbf{h}_1^{X_{i1}} \oplus \dots \oplus \mathbf{h}_q^{X_{iq}})\|^2 \\ &= \operatorname{argmin}_{\mathbf{a}, \mathbf{h}_r, \mathbf{m}_0 \in \mathbb{S}^{k-1}} \sum_{i=1}^n \sum_{j=1}^{k-1} \left\{ (k-1) \left[\log \left(\frac{M_{ij} m_{0k} a_k^{T_i} \prod_{r=1}^q h_{rk}^{X_{ir}}}{M_{ik} m_{0j} a_j^{T_i} \prod_{r=1}^q h_{rj}^{X_{ir}}} \right) \right]^2 \right. \\ (8) \quad &\quad \left. - \log \left(\frac{M_{ij} m_{0k} a_k^{T_i} \prod_{r=1}^q h_{rk}^{X_{ir}}}{M_{ik} m_{0j} a_j^{T_i} \prod_{r=1}^q h_{rj}^{X_{ir}}} \right) \right. \\ &\quad \left. \times \sum_{\ell \neq j}^{k-1} \log \left(\frac{M_{i\ell} m_{0k} a_k^{T_i} \prod_{r=1}^q h_{rk}^{X_{ir}}}{M_{ik} m_{0\ell} a_\ell^{T_i} \prod_{r=1}^q h_{r\ell}^{X_{ir}}} \right) \right\}. \end{aligned}$$

The objective function (8) is not convex in terms of a_j , m_{0j} , and h_{rj} but is in terms of $\operatorname{alt}(\mathbf{a})_j$, $\operatorname{alt}(\mathbf{m}_0)_j$, and $\operatorname{alt}(\mathbf{h}_r)_j$ for $j = 1, \dots, k - 1$; $r = 1, \dots, q$. Thus, the optimal solution can be obtained by solving the following system of linear equations with constraints $\mathbf{m}_0, \mathbf{a}, \mathbf{h}_r \in \mathbb{S}^{k-1}$:

$$\begin{bmatrix} D(1) & D(T) & D(X_1) & \dots & D(X_q) \\ D(T) & D(T^2) & D(TX_1) & \dots & D(TX_q) \\ D(X_1) & D(TX_1) & D(X_1^2) & \dots & D(X_1X_q) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D(X_q) & D(TX_q) & D(X_1X_q) & \dots & D(X_q^2) \end{bmatrix} \begin{bmatrix} \operatorname{alt}(\mathbf{m}_0) \\ \operatorname{alt}(\mathbf{a}) \\ \operatorname{alt}(\mathbf{h}_1) \\ \vdots \\ \operatorname{alt}(\mathbf{h}_q) \end{bmatrix} = \begin{bmatrix} \zeta_0 \\ \zeta_1 \\ \xi_1 \\ \vdots \\ \xi_q \end{bmatrix},$$

where $\zeta_{0j} = k \sum_{i=1}^n \log M_{ij} - \sum_{\ell=1}^k \sum_{i=1}^n \log M_{i\ell}$, $\zeta_{1j} = k \sum_{i=1}^n T_i \log M_{ij} - \sum_{\ell=1}^k \sum_{i=1}^n T_i \log M_{i\ell}$, $\xi_{rj} = k \sum_{i=1}^n X_{ir} \log M_{ij} - \sum_{\ell=1}^k \sum_{i=1}^n X_{ir} \log M_{i\ell}$, and for any v , $D(v)$ is defined as

$$D(v) = \begin{bmatrix} (k-1) \sum_{i=1}^n v_i & - \sum_{i=1}^n v_i & \dots & - \sum_{i=1}^n v_i \\ - \sum_{i=1}^n v_i & (k-1) \sum_{i=1}^n v_i & \dots & - \sum_{i=1}^n v_i \\ \vdots & \vdots & \ddots & \vdots \\ - \sum_{i=1}^n v_i & - \sum_{i=1}^n v_i & \dots & (k-1) \sum_{i=1}^n v_i \end{bmatrix}.$$

To estimate its covariance matrix $\hat{\Sigma}_a$, we use a bootstrap distribution of \hat{a} obtained by the percentile method of Machado and Parente (2005). See Supplementary Material IV for details.

3.2. *Estimation of compositional regression parameters and covariance matrix.* The log-contrast model has been the most general solution to incorporate the unit-sum constraint in the linear regression model for compositional covariates (Aitchison and Bacon-Shone (1984), Lin et al. (2014)). Shi, Zhang and Li (2016) proposed a debias procedure for the ℓ_1 regularized estimates of high dimensional compositional covariates. We use the linear log-contrast model and the debias procedure to estimate regression parameters (i.e., \mathbf{b} and c) and their covariance matrix. Specifically, we first solve the following objective function,

$$(9) \quad \hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2n} \left(\sum_{i=1}^n Y_i - cT_i - (\log \mathbf{M}_i)^\top \mathbf{b} - \mathbf{X}_i^\top \mathbf{g} \right)^2 + \lambda \|\beta\|_1,$$

subject to $\mathbf{b}^\top \mathbf{1}_k = 0$,

where $\hat{\beta} = (\hat{c}, \hat{\mathbf{b}}, \hat{\mathbf{g}})^\top$ and λ is a turning parameter. For simplicity, the intercept is excluded in the model, which can be eliminated by centering all the variables in the model. We then apply the debias procedure of Shi, Zhang and Li (2016) to the solution of the objective function (9) to obtain unbiased estimates and their covariance matrix.

- PROPOSITION 1 (Properties of the proposed estimators of (6)–(7)).
- a. $\zeta(\tau; \hat{c})$ and $\delta(\tau; \hat{\mathbf{a}}, \hat{\mathbf{b}})$ are consistent and unbiased estimators.
 - b. $\delta(\tau; \hat{\mathbf{a}}, \hat{\mathbf{b}})$ possesses essential properties for compositional data analysis: scale invariance; permutation invariance; subcompositional coherence.

A proof of Proposition 1 is given in Supplementary Material III.

3.3. *Hypothesis test of mediation effect.* The null hypothesis of no total compositional mediation effect is given by

$$(10) \quad H_0 : (\log \mathbf{a})^\top \mathbf{b} = 0,$$

and the null hypothesis of no component-wise mediation effect is given by

$$(11) \quad H_0 : \log(ka_j)b_j = 0 \quad \forall j \in \{1, 2, \dots, k\}.$$

The null hypothesis (10) reflects the total mediation effect on an outcome; however, it can disguise the actual mediation effect, which may be captured by the null hypothesis (11). If for instance, the mediation effects of two mediators are equal, but their directions are opposite, then the total mediation effect is zero. In other words, we cannot reject the null hypothesis (10) but might reject the null

hypothesis (11). Therefore, we need to test both hypotheses to avoid a misleading conclusion about the mediation effect.

To test the null hypotheses (10) and (11), we propose two approaches: an extension of the Sobel test (Sobel (1982)) and a bootstrap approach. In testing the null hypothesis (10) with the former, the square root of the first order asymptotic variance of the total indirect effect is computed with the estimated covariance matrices of $\log(k\hat{\mathbf{a}})$ and $\hat{\mathbf{b}}$ by the method described in Bollen (1987). It then is used as a standard error of the total indirect effect in the Z -test. The expressions for the first order asymptotic variances of the total indirect effect and component-wise indirect effects are given in Supplementary Material V.

The distribution of a composition \mathbf{M}_i is not known, but $\text{alt}(\mathbf{M}_i)_j$ is well approximated by a normal distribution (Aitchison (1986)), that is, the distribution of $\log(a_j/a_k)$ is well approximated by a normal distribution. Therefore, $\delta(\tau)$ can also be approximated by a normal distribution assuming the product of two normal variables (i.e., $\log(a_j/a_k)$ and b) follows a normal distribution. Recall that $\sum_{j=1}^{k-1} \log(a_j/a_k) b_j = (\log \mathbf{a})^\top \mathbf{b}$ because $\mathbf{b}^\top \mathbf{1}_k = 0$. In general, the product of two normal variables does not follow a normal distribution. However, a misspecified distribution will just reduce the power but not affect the type I error rate when the null hypothesis of no indirect effect is false (MacKinnon et al. (2002), Shrout and Bolger (2002)).

To avoid the assumption of normality for the indirect effect, we can use a bootstrap approach (Shrout and Bolger (2002), VanderWeele and Vansteelandt (2014)). To this end, we use a nonparametric bootstrap for $\log(k\hat{\mathbf{a}})$ and a parametric bootstrap for $\hat{\mathbf{b}}$ using a multivariate normal distribution to approximate the sampling distribution of $\delta(\tau)$. A p -value for $\delta(\tau)$ is then approximated by utilizing the fact that any bootstrap replicate $\delta(\tau)_b - \delta(\tau)$ should have a distribution close to that of $\delta(\tau)$ when the null hypothesis is true, where $\delta(\tau)_b$ denotes an estimated total indirect effect derived from a resampled dataset (Efron and Tibshirani (1993)). That is, a p -value can be approximated by $2 \sum_{b=1}^{n_b} \mathbf{I}(\delta(\tau)_b - \delta(\tau) \geq \delta(\tau))/n_b$ when $\delta(\tau) \geq 0$ and $2 \sum_{b=1}^{n_b} \mathbf{I}(\delta(\tau)_b - \delta(\tau) < \delta(\tau))/n_b$ when $\delta(\tau) < 0$, where $\mathbf{I}(\cdot)$ is the indicator function and n_b is the number of bootstrap samples. Similarly, the null hypothesis (11) can be tested. In the simulation studies, we observed no significant difference in performance between the two methods regarding power and type I error, concurring with Huang and Pan (2016).

3.4. Sensitivity analysis. In mediation analysis, the assumption (4) of no unmeasured confounding effect may not be satisfied even in a randomized experiment. In this case, the estimated total indirect effect is invalid since the parameter \mathbf{b} in the model (2) is not consistent. To address this problem, we extend the method for the single mediation model of Imai, Keele and Tingley (2010) to assess the sensitivity of the estimated total indirect effect to unmeasured confounding variables by utilizing the relationship among U_{0i} , U_{1i} , and U_{2i} , where

$U_{0i} = Y_i - (c'_0 + c'T_i + \mathbf{X}_i^\top \mathbf{g}')$ is an error in modeling the total treatment effect. Note that the total effect of a treatment is decomposed into the direct and total indirect effects under our model specification. Since we treat k compositional mediators as a single mediator with k components, we assess the sensitivity of the total indirect effect by considering the correlation between the disturbance terms for the mediator and the outcome due to unmeasured confounding variables,

$$(12) \quad \rho \equiv \text{Corr}(\text{alt}(\mathbf{U}_{1i})_j, U_{2i})$$

for all $j = 1, \dots, k - 1$.

Suppose that the assumption (3) is satisfied and our model is correctly specified. Then, for a given correlation ρ , the total indirect effect is identified and given by

$$(13) \quad \delta_\rho(\tau) = (\log \mathbf{a})^\top \mathbf{b}_\rho(t - t_0),$$

where \mathbf{b}_ρ is a solution of the following system of k equations:

$$(14) \quad \sigma_2^2 + 2\rho \mathbf{b}_\rho^\top \text{diag}(\Sigma_\theta^{1/2})\sigma_2 + \mathbf{b}_\rho^\top \Sigma_\theta \mathbf{b}_\rho - \sigma_0^2 = 0,$$

$$(15) \quad \Sigma_\theta \mathbf{b}_\rho = \mathbf{r}(\rho, \sigma_2),$$

where $\sigma_0^2 = \text{Var}(U_{0i})$; $\sigma_2^2 = \text{Var}(U_{2i})$; $\theta_j = \text{alt}(\mathbf{U}_{1i})_j$; $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{k-1})^\top$; $\Sigma_\theta = \text{Var}(\boldsymbol{\theta})$; $\mathbf{r}(\rho, \sigma_2) = (\text{Cov}(U_{0i}, \theta_1), \dots, \text{Cov}(U_{0i}, \theta_{k-1}))^\top - \rho\sigma_2 \text{diag}(\Sigma_\theta^{1/2})^\top$. The derivation of the equations (14) and (15) is given in Supplementary Material VII. Here, σ_0^2 , $\boldsymbol{\theta}$, Σ_θ , and $\text{Cov}(U_{0i}, \theta_j)$ for $j = 1, \dots, k - 1$ can be estimated consistently from the residuals, and σ_2^2 is a part of the solution. We cannot extract any information about ρ from data; therefore, we treat ρ as a sensitivity parameter and obtain corresponding values for the total indirect effect given a plausible range of ρ .

4. Simulation studies. Mediation analysis for multiple or high dimensional mediators typically assumes independence between mediators to establish causal interpretation so principal components of mediators are often used. This approach of using principal components is also applicable to estimate the direct and total indirect effects for compositional mediators. Another naive approach for compositional mediators is to utilize an ℓ_1 regularization, which tends to drop correlated variables. We used these two approaches to evaluate and compare the performance of our compositional mediation model. For the hypothesis test, we used the extension of the Sobel test for fair comparison.

In data generation, we randomly generated the treatment T_i from the standard normal distribution and the baseline composition \mathbf{m}_0 from the standard uniform distribution, $\text{Unif}(0, 1)$, under the unit-sum constraint. The path coefficients \mathbf{a} , \mathbf{b} , and c were selected such that the direct effect is 1.00 and the total indirect effect is approximately 0.00, 0.50, 0.75, or 1.00. For the compositional disturbance, we used a multivariate logistic normal (LN) distribution (Aitchison (1986)) with mean

$\mathbf{0}_{k-1}$ and covariance $2\mathcal{N}$, that is, $\mathbf{U}_{1i} \sim \text{LN}(\mathbf{0}_{k-1}, 2\mathcal{N})$. Note that a multivariate logistic normal distribution is defined by the logistic transformation of a multivariate normal distribution, that is, $\mathbf{R} \sim \text{LN}(\boldsymbol{\mu}, \Sigma) \Leftrightarrow \text{alt}(\mathbf{R}) \sim \text{N}(\boldsymbol{\mu}, \Sigma)$. For the compositional regression disturbance, we used a normal distribution with mean 0 and variance 2, or $U_{2i} \sim \text{N}(0, 2)$. Recall that our model does not specify the distribution of \mathbf{U}_{1i} . The composition \mathbf{M}_i and the outcome Y_i were then generated according to the model (1) and (2), respectively.

4.1. *Comparison of power and type I error.* We first estimated the power and type I error rate in testing the total indirect effect (TIDE) for three methods: the principal component regression (PCR), a two-stage adaptive lasso (TSAL), and our compositional mediation model (CMM). TSAL uses the standard lasso in the first stage to screen irrelevant variables and the adaptive lasso in the second stage to select consistent variables (Bühlmann and van de Geer (2011)). We used 1500 simulations for each k mediators with a sample size $n = 100$ at various probability thresholds, where $k = 5, 49, 99$: 250 simulations with each of TIDE = 1.00, 0.75, 0.50; 250 simulations with no effect of T_i on \mathbf{M}_i (i.e., $a_j = 0, \forall j$); 250 simulations with no effect of \mathbf{M}_i on Y_i (i.e., $b_j = 0, \forall j$); 250 simulations with *inconsistent* TIDEs (i.e., the sum of the component-wise indirect effects is zero). For TIDE $\neq 0$, the first five components of the regression parameter \mathbf{b} are nonzero for $k = 5, 49$, and 99, therefore the effect sizes are the same for $k = 5, 49$ and 99. For PCR, we included only the first k_{pc} principal components that explain 90% of the total variance. As shown in Table 1, while all three methods roughly control the type 1 errors, CMM outperforms both the PCR and TSAL approaches in power, especially when k is large.

To test the bias and variance of estimates of the three methods, we simulated data with $c = 1.00$ and $(\log \mathbf{a})^\top \mathbf{b} = 0.75$ at various sample sizes $n = 100, 200, 500$. The following disturbances were used: $\mathbf{U}_{1i} \sim \text{LN}(\mathbf{0}_{k-1}, 0.2\mathcal{N})$ and $U_{2i} \sim \text{N}(0, 0.2)$. Figure 3 shows the results based on 200 simulations. When the sample size is small, the estimates of all three methods are slightly biased; however, as the sample size increases, they converge to the true values. The results are similar for $k = 5, 49, 99$.

4.2. *Identification of component-wise indirect effect.* PCR is not capable of testing the component-wise indirect effects; therefore, we compared the performance of CMM on the component-wise indirect effects only with TSAL. In data generation, we selected the parameters \mathbf{a} and \mathbf{b} such that the first 7 component-wise indirect effects are at a nonzero constant level but different combinations of a_j and b_j ; the remaining 43 component-wise indirect effects are at zero. The level of nonzero indirect effects was increased by 0.1 from 0, and we used sample sizes $n = 100, 200$. For multiple testing corrections, we used the Benjamini and Yekutieli (2001) false discovery rate (FDR) at 0.05. As a comparison measure, the F_1 score, which is the harmonic mean of precision and recall, was used. Figure 4 shows results: better performance of CMM over TSAL.

TABLE 1

Power and type 1 error rate in testing TIDEs: number of mediators $k = 5, 49, 99$; sample size $n = 100$; significance level $\alpha = 0.001, 0.01, 0.05$. The total of 1500 simulations for each k were used: 750 simulations with nonzero TIDEs and 750 simulations for zero TIDEs. For different $k = 5, 49$, and 99, the nonzero TIDEs take the same values and only the first five components of \mathbf{b} are nonzero. CMM: proposed compositional mediation model; PCR: principal component regression; TSAL: two-stage adaptive lasso

	α	Power			Type 1 error		
		0.001	0.01	0.05	0.001	0.01	0.05
$k = 5$	CMM	0.452	0.623	0.752	0.001	0.011	0.035
	PCR	0.381	0.604	0.732	0.000	0.007	0.040
	TSAL	0.413	0.621	0.740	0.001	0.008	0.037
$k = 49$	CMM	0.476	0.675	0.820	0.001	0.015	0.043
	PCR	0.051	0.207	0.423	0.000	0.005	0.029
	TSAL	0.277	0.487	0.631	0.001	0.012	0.067
$k = 99$	CMM	0.397	0.645	0.791	0.001	0.007	0.040
	PCR	0.023	0.105	0.273	0.000	0.004	0.028
	TSAL	0.304	0.495	0.628	0.003	0.016	0.064

4.3. *Simulation based on real dataset.* We also simulated data using the composition of a real dataset reported in Wu et al. (2011), which is analyzed in Section 5. The composition parameter \mathbf{a} of this dataset was estimated by the Dirichlet regression (Maier (2014)) with randomly generated treatments T_i from $N(0, 1)$. The regression parameter \mathbf{b} under the linear constraint was randomly generated from $\pm 4 \times \text{Unif}(0, 1)$ for k_s components and set to 0 for k_{ns} components, where $k = k_s + k_{ns}$, and k_s was randomly selected between 2 and 5. The direct effect c was set to 1, and we used $U_{2i} \sim N(0, 2)$ to simulate the outcome Y_i . Figure 5 shows that the estimates of the direct effect and TIDE are almost unbiased for all three methods.

Since we randomly generated TIDEs, obtaining particular values of TIDEs was not computationally practical. Thus, we considered only the cases that TIDEs are greater than 0.75 to test the power. Based on 200 simulations, the power was 0.52, 0.38, and 0.48, respectively, for CMM, PCR, and TSAL at the significance level $\alpha = 0.05$, and 0.24, 0.16, and 0.21 at $\alpha = 0.01$. CMM still has the highest power while controlling the type I error.

5. Real data application. We applied CMM to a cross-sectional dataset, referred to as the ‘‘COMBO’’ dataset (Wu et al. (2011)). The dataset consists of 16S rRNA sequences from fecal samples of 98 healthy individuals from the University of Pennsylvania. It also contains demographic and clinical information including fat intake and BMI, where the habitual long-term fat intake was derived from the

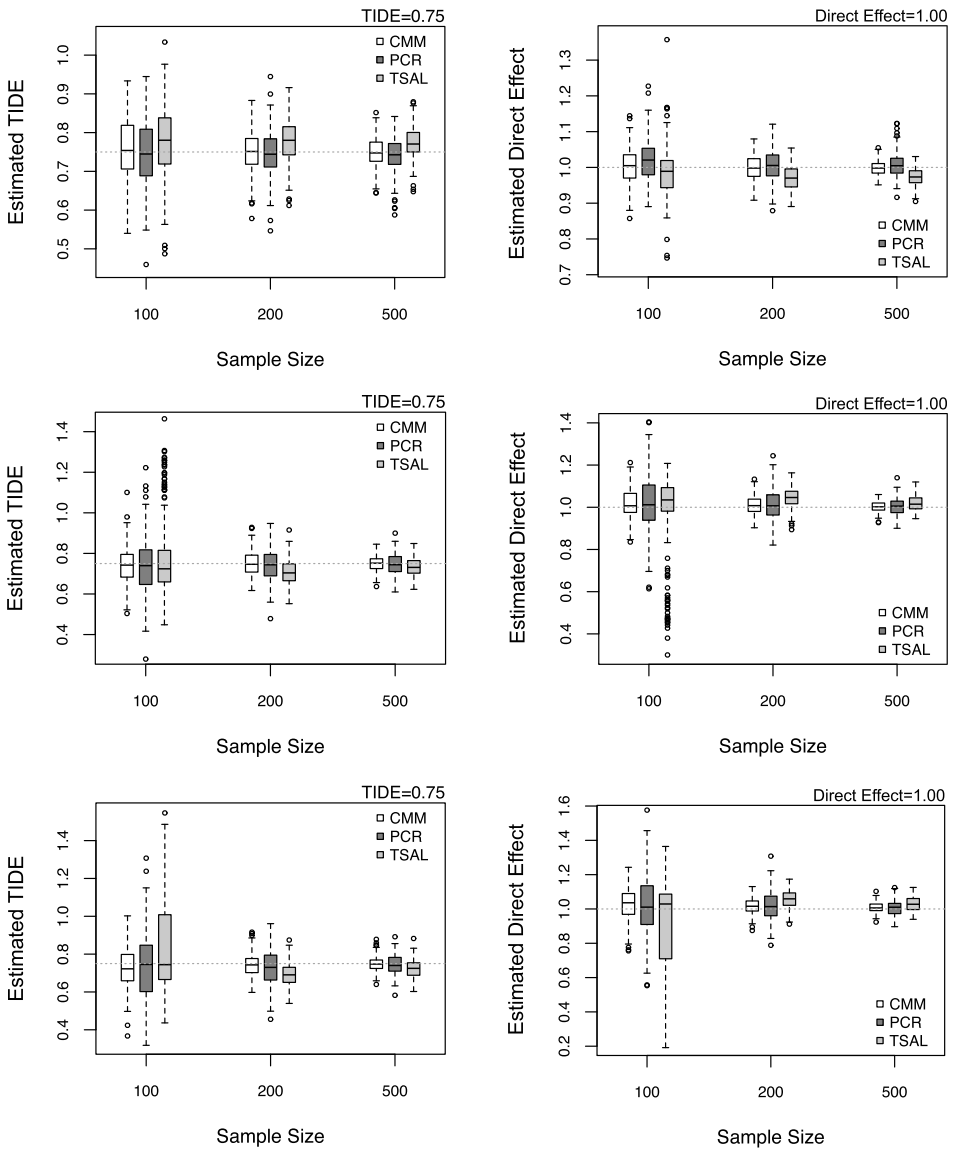


FIG. 3. Estimated direct effects and TIDEs at various sample sizes for the number of mediators $k = 5$ (top), $k = 49$ (middle), and $k = 99$ (bottom) with the true direct effect of 1.00 and the true TIDE of 0.75, which are indicated by dotted lines. The results are based on 200 simulations. CMM: proposed compositional mediation model; PCR: principal component regression; TSAL: two-stage adaptive lasso.

food frequency questionnaire (FFQ). Such measurements are widely applied in nutritional research, and their reproducibility and validity have been validated (Hu et al. (1999)). We summarized operational taxonomic units (OTUs) at the genus

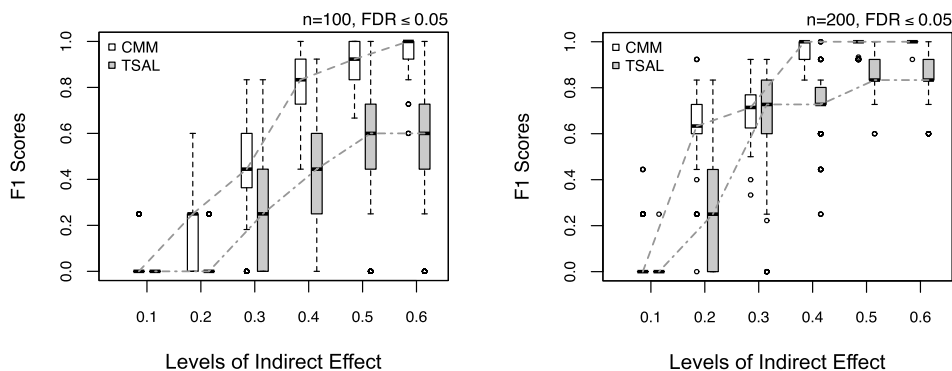


FIG. 4. F_1 scores versus levels of the component-wise indirect effects at $FDR \leq 0.05$. P -values are adjusted by the Benjamini–Yekutieli FDR procedure. CMM: proposed compositional mediation model; TSAL: two-stage adaptive lasso.

level and then filtered out the genera that appear in fewer than 10% of the samples, leaving 45 genera in 98 samples. Because of the difference in the total number of OTUs in each of the sample, the OTU counts assigned to these genera were transformed into proportions after replacing zero counts by the maximum rounding error 0.5, which is commonly used in compositional data analysis (Aitchison (1986)).

5.1. *Estimation of TIDE.* The gut microbiota can influence host adiposity through energy extraction from the diet with variable efficiency depending on community composition; furthermore, the microbiota can affect host adiposity by influencing metabolism throughout the body. It is therefore highly likely that the gut microbiome can potentially mediate the effect of the diet such as fat intake on host adiposity and BMI. Since the 98 samples were roughly randomly sampled,

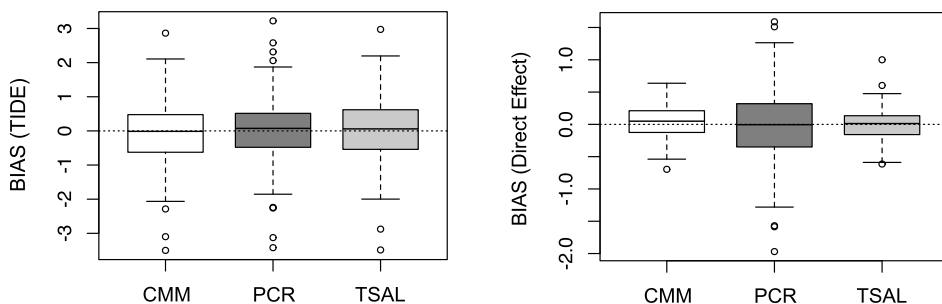


FIG. 5. Biases of the direct effects and TIDEs for data simulated with the composition of the COMBO dataset. The results are based on 200 simulations. CMM: proposed compositional mediation model; PCR: principal component regression; TSAL: two-stage adaptive lasso.

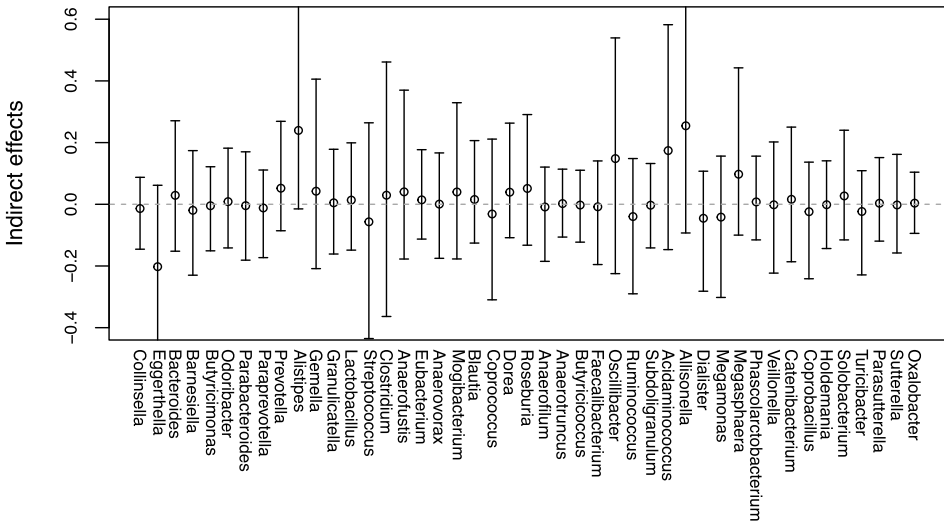


FIG. 6. Estimated component-wise indirect effects of fat intake on BMI through the gut microbiome. Bootstrap CIs for *Eggerthella*, *Alistipes*, and *Allisonella* are $(-0.652, 0.062)$, $(-0.015, 0.668)$, and $(-0.093, 0.786)$, respectively.

it is reasonable to assume that fat intake was randomly assigned. CMM was applied to the COMBO dataset with BMI as the outcome, fat intake as the treatment, and the 45 genera as the compositional mediators. By the bootstrap approach, the estimated direct effect is 0.933 with a 95% bootstrap confidence interval (CI) of (0.005, 1.935) and TIDE is 0.809 with a 95% bootstrap CI of (-0.452, 2.250). The estimated component-wise indirect effects and their 95% bootstrap CIs are shown in Figure 6.

The estimated direct effect and TIDE (i.e., 0.933 and 0.808) correspond to 53.6% and 46.4% of the total effect of fat intake on BMI. Note that under our model specification, the total effect of a treatment can be decomposed into the direct and total indirect effects. The estimated direct effect is statistically significant at $\alpha = 0.05$, but the estimated TIDE, as well as all the component-wise indirect effects, are not. This insignificance is likely due to insufficient sample size: based on the simulation study in which we used the characteristics of parameters estimated from the COMBO dataset, TIDE of 0.75 and component-wise indirect effects at around 0.2 were rarely detected with the sample size of 100.

As shown in Figure 6, the potential genera that could have statistically significant component-wise indirect effects with a sufficient sample size are *Eggerthella*, *Alistipes*, *Oscillibacter*, *Acidaminococcus*, and *Allisonella*. All these genera except *Eggerthella* have positive values for the indirect effects, but their responses to fat intake are entirely different. The abundance of *Alistipes* and *Oscillibacter* is negatively correlated with fat intake and BMI, whereas that of *Acidaminococcus* and *Allisonella* is positively correlated with fat intake and BMI. In other words, we can

hypothesize that the increase in fat intake causes the decrease in the abundance of *Alistipes* and *Oscillibacter* but the increase in the abundance of *Acidaminococcus* and *Allisonella* which in turn cause the increase in BMI. Lam et al. (2012) identified *Oscillibacter*-like organisms as a potentially important gut microbe that mediates high fat-induced gut dysfunction and gut permeability and showed that decrease of *Oscillibacter* led to increased gut permeability, which was shown to be associated with obesity (Teixeira et al. (2012)). This observation is largely consistent with those observed in mice fed with high-fat diet (Daniel et al. (2014)).

As a comparison, we applied TSAL to the COMBO dataset. Its estimates for the direct effect and TIDE are 0.751 and 0.555, respectively, which correspond to 57.5% and 42.5% of the total effect of fat intake on BMI. TSAL selects five genera: *Alistipes* (0.161), *Clostridium* (0.031), *Doria* (0.047), *Acidaminococcus* (0.150), and *Allisonella* (0.166), where the values in parenthesis are the estimated component-wise indirect effects. All except the direct effect are not significant at $\alpha = 0.05$, similar to the results of CMM. Interestingly, *Oscillibacter* is not selected by TSAL.

5.2. *Sensitivity analysis on COMBO.* TIDE of fat intake on BMI is estimated under the assumptions (3) ~ (5). Even though we can assume random assignment of fat intake (i.e., the assumption (3) holds) as mentioned before, we cannot safely assume that the assumption (4) is satisfied. Thus, we estimated TIDE given the correlation ρ between the disturbance terms of the compositional mediators and the outcome using the method proposed in Section 3.4. Figure 7 shows the result. The estimated TIDE ranges from 2.162 to -1.546 for ρ ranging from -0.244

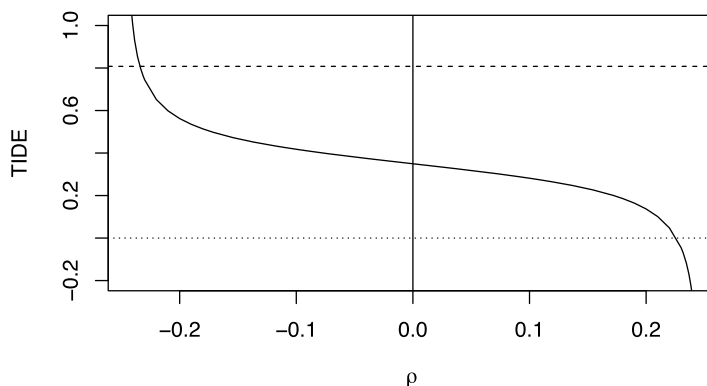


FIG. 7. TIDEs for given correlations ρ between the disturbance terms of the compositional mediators and the outcome. The dashed line is the estimated TIDE (i.e., 0.808) by CMM. The dotted line indicates 0. The solid line denotes TIDEs for given ρ . The vertical line is the 95% bootstrap confidence interval $(-1.735, 2.706)$ of TIDE when $\rho = 0$ (the y-axis is truncated at -0.2 and 1 for better visualization).

to 0.244. These values are within the 95% bootstrap CI of the estimated TIDE assuming $\rho = 0$.

Our simulation results shown in Supplementary Material VIII indicate that when the sample size is small, TIDE can be sensitive to the no confounding assumption (4), that is, the range of ρ where TIDEs are significant is very small. However, as the sample size increases, the estimates become less sensitive.

6. Discussion. In this study, we propose a compositional mediation model for a continuous outcome where a microbial composition is treated as mediators. Our method takes the characteristics of compositional data into account and treats the whole compositional mediators as a unit, that is, it jointly estimates the effect of a treatment on compositional mediators instead of each mediator separately. In our simulation studies, we have shown better performance of our method over the two naive methods for compositional mediators. Our method also provides a clear interpretation of indirect effects. Application to the COMBO dataset indicates potentially mediating effect of the gut microbiome in linking fat intake and BMI. Our results estimated that about 46% of the total effect of fat intake on BMI can be through perturbing the gut microbiome. Although the results are not statistically significant at $\alpha = 0.05$, several potentially significant bacterial genera have been shown to be directly associated with gut permeability and therefore BMI. It would be interesting and important to replicate these results in larger data sets.

Like other causal mediation analysis methods, the identifiability of causal mediation effect under our proposed model that requires certain assumptions. Methods for assessing the validity of these assumptions are important. The assumption (3), which is related to the treatment assignment process, is generally not verifiable with observational data and is usually checked by subject-knowledge matter. The no-confounding assumption between the mediator and the outcome (4) has to be assumed and cannot be easily verified. Instead, we have provided a method for sensitivity analysis. For the no interaction assumption (5), one can empirically evaluate its validity by regression analysis:

$$Y_i = c_0 + cT_i + (\log \mathbf{M}_i)^\top \mathbf{b} + (T_i \log \mathbf{M}_i)^\top \boldsymbol{\gamma} + \mathbf{X}_i^\top \mathbf{g} + U_{2i},$$

subject to $\mathbf{b}^\top \mathbf{1}_k = \boldsymbol{\gamma}^\top \mathbf{1}_k = 0$. If any of $\boldsymbol{\gamma}$ is significant, the no interaction assumption is not satisfied. We included the estimators for the direct and total indirect effects when there are interactions between the treatment and mediators in Supplementary Material VI.

Even though we used a continuous treatment variable in the simulation studies, a binary treatment variable can be used without any modification to our method. In many clinical microbiome studies, the outcome variable is binary such as whether a subject is diseased or not. In this case, the model (2) can be rewritten for logistic or probit regression, assuming the outcome variable is a latent continuous variable indicated by an observed dichotomous variable. Then, the model (1) and the

modified model (2) will provide the identifiability of the direct and indirect effects (Winship and Mare (1983)). Another interesting extension of our method is for longitudinal compositional data, which is also very common in microbial studies.

The Gaussian assumption is made for the error term in modeling the outcomes (i.e., BMI in our application), which often can be approximately achieved by an appropriate transformation. One important future research is to develop bootstrap methods for inference of the direct and total indirect effects allowing heteroscedastic errors and heavy-tailed distributions.

Acknowledgments. We thank the Prof. Kafadar, A.E., and reviewers for many helpful comments.

SUPPLEMENTARY MATERIAL

Supplement to “Compositional mediation analysis for microbiome studies” (DOI: [10.1214/18-AOAS1210SUPP](https://doi.org/10.1214/18-AOAS1210SUPP); .pdf). The online Supplemental Materials include proofs of Theorem 1 and Proposition 1, a detailed computational algorithm for the covariance matrix of composition parameters, variance calculation for the indirect effects, an extension of the model to allow for interactions between a treatment and mediators, and a method for sensitivity analysis.

REFERENCES

- AITCHISON, J. (1982). The statistical analysis of compositional data. *J. Roy. Statist. Soc. Ser. B* **44** 139–177. [MR0676206](#)
- AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data*. CRC Press, London. [MR0865647](#)
- AITCHISON, J. and BACON-SHONE, J. (1984). Log contrast models for experiments with mixtures. *Biometrika* **71** 323–330.
- BARON, R. M. and KENNY, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51** 1173–1182.
- BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. [MR1869245](#)
- BILLHEIMER, D., GUTTORP, P. and FAGAN, W. F. (2001). Statistical interpretation of species composition. *J. Amer. Statist. Assoc.* **96** 1205–1214. [MR1946574](#)
- BOLLEN, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociol. Method.* **17** 37–69.
- BRAY, G. A. and POPKIN, B. M. (1998). Dietary fat intake does affect obesity! *Am. J. Clin. Nutr.* **68** 1157–1173.
- BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. [MR2807761](#)
- CHÉN, O. Y., CRAINICEANU, C., OGBURN, E. L., CAFFO, B. S., WAGER, T. D. and LINDQUIST, M. A. (2018). High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics* **19** 121–136. [MR3799607](#)
- DANIEL, H., GHOLAMI, A. M., BERRY, D., DESMARCHELIER, C., HAHNE, H., LOH, G., MONDOT, S., LEPAGE, P., ROTHBALLER, M., WALKER, A., BÖHM, C., WENNING, M., WAGNER, M., BLAUT, M., SCHMITT-KOPPLIN, P., KUSTER, B., HALLER, D. and CLAVEL, T. (2014). High-fat diet alters gut microbiota physiology in mice. *ISEM J.* **8** 295–308.

- EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. [MR1270903](#)
- HU, F. B., RIMM, E., SMITH-WARNER, S. A., FESKANICH, D., STAMPFER, M. J., ASCHERIO, A., SAMPSON, L. and WILLETT, W. C. (1999). Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am. J. Clin. Nutr.* **69** 243–249.
- HUANG, Y.-T. and PAN, W.-C. (2016). Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics* **72** 401–413. [MR3515767](#)
- IMAI, K., KEELE, L. and TINGLEY, D. (2010). A general approach to causal mediation analysis. *Psychol. Methods* **15** 309–334.
- IMAI, K., KEELE, L. and YAMAMOTO, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statist. Sci.* **25** 51–71. [MR2741814](#)
- IMAI, K. and YAMAMOTO, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Polit. Anal.* **21** 141–171.
- IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—For Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. [MR3309951](#)
- LAM, Y. Y., HA, C. W. Y., CAMPBELL, C. R., MITCHELL, A. J., DINUDOM, A., OSCARSSON, J., COOK, D. I., HUNT, N. H., CATERSON, I. D., HOLMES, A. J. and STORLIEN, L. H. (2012). Increased gut permeability and microbiota change associate with mesenteric fat inflammation and metabolic dysfunction in diet-induced obese mice. *PLoS ONE* **7** e34233.
- LEY, R. E., TURNBAUGH, P. J., KLEIN, S. and GORDON, J. I. (2006). Human gut microbes associated with obesity. *Nature* **444** 1022–1023.
- LIN, W., SHI, P., FENG, R. and LI, H. (2014). Variable selection in regression with compositional covariates. *Biometrika* **101** 785–797. [MR3286917](#)
- MACHADO, J. A. F. and PARENTE, P. (2005). Bootstrap estimation of covariance matrices via the percentile method. *Econom. J.* **8** 70–78. [MR2136930](#)
- MACKINNON, D. P., LOCKWOOD, C. M., HOFFMAN, J. M., WEST, S. G. and SHEETS, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychol. Methods* **7** 83–104.
- MAIER, M. J. (2014). DirichletReg: Dirichlet regression for compositional data in R. Research Report Series, Dept. Statistics and Mathematics, 125. WU Vienna Univ. Economics and Business, Vienna.
- PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge Univ. Press, Cambridge. [MR1744773](#)
- PEARL, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence* 411–420. Morgan Kaufmann, San Francisco, CA.
- PREACHER, K. J. and HAYES, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* **40** 879–891.
- RUBIN, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* **100** 322–331. [MR2166071](#)
- SHI, P., ZHANG, A. and LI, H. (2016). Regression analysis for microbiome compositional data. *Ann. Appl. Stat.* **10** 1019–1040. [MR3528370](#)
- SHROUT, P. E. and BOLGER, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychol. Methods* **7** 422–445.
- SOBEL, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol. Method.* **13** 290–312.
- SOHN, M. B. and LI, H. (2019). Supplement to “Compositional mediation analysis for microbiome studies.” DOI:[10.1214/18-AOAS1210SUPP](#).
- TEIXEIRA, T. F., COLLADO, M. C., FERREIRA, C. L., BRESSAN, J. and PELUZIO, M. C. (2012). Potential mechanisms for the emerging link between obesity and increased intestinal permeability. *Nutr. Res.* **32** 637–47.

- TURNBAUGH, P. J., LEY, R. E., MAHOWALD, M. A., MAGRINI, V., MARDIS, E. R. and GORDON, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444** 1027–1031.
- VANDERWEELE, T. J. and VANSTEELANDT, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *Am. J. Epidemiol.* **172** 1339–1348.
- VANDERWEELE, T. J. and VANSTEELANDT, S. (2014). Mediation analysis with multiple mediators. *Epidemiol. Methods* **2** 95–115.
- WINSHIP, C. and MARE, R. D. (1983). Structural equations and path analysis for discrete data. *Amer. J. Sociol.* **89** 54–110.
- WU, G., CHEN, J., HOFFMANN, C., BITTINGER, K., CHEN, Y. Y., KEILBAUGH, S. A., BEWTRA, M., KNIGHTS, D., WALTERS, W. A., KNIGHT, R., SINHA, R., GILROY, E., GUPTA, K., BALDASSANO, R., NESSEL, L., LI, H., BUSHMAN, F. D. and LEWIS, J. D. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334** 105–108.
- ZHAO, Y. and LUO, X. (2016). Pathway lasso: Estimate and select sparse mediation pathways with high dimensional mediators. [arXiv:1603.07749](https://arxiv.org/abs/1603.07749).

DEPARTMENT OF BIostatISTICS AND EPIDEMIOLOGY
UNIVERSITY OF PENNSYLVANIA SCHOOL OF MEDICINE
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: msohn@mail.med.upenn.edu
hongzhe@upenn.edu