GENOME-WIDE ANALYSES OF SPARSE MEDIATION EFFECTS UNDER COMPOSITE NULL HYPOTHESES

BY YEN-TSUNG HUANG[1]

Academia Sinica

A genome-wide mediation analysis is conducted to investigate whether epigenetic variations M mediate the effect of socioeconomic adversity S on adiposity Y. The mediation effect can be expressed as a product of two parameters, the S–M association and the M–Y association conditional on S. We show that the joint significance test examining the two parameters separately has smaller p-values than the normality-based or the normal product-based test for the product and is a size α test. However, under multiple tests with sparse signals, the conventional joint significance test has a conservative test size and low power within a study because of the sparsity in signals and not accounting for the composition of different null hypotheses. We develop a novel test assessing the product of two normally distributed test statistics under a composite null hypothesis, where either one parameter is zero or both are zero. We show that the null composition can be adjusted by variances of test statistics without directly estimating proportions of different nulls. Advantages of the new test are illustrated in simulation and the epigenomic study. The new test identifies four methylation loci mediating the socioeconomic effect on adiposity with the false discovery rate less than 20% while existing methods had none surviving this cut-off.

1. Introduction. Mediation analyses first proposed in psychological literature have become a popular approach to interrogate whether there exists an effect of an exposure or intervention on an outcome mediated through an intermediary factor called mediator [Baron and Kenny (1986), MacKinnon (2008)]. Baron and Kenny (1986) studied mediation effects in simple linear models. By employing the counterfactual outcome framework [Rubin (1978)], identifiability issues were discussed [Robins and Greenland (1992)] and no unmeasured confounding assumptions [Pearl (2001), VanderWeele and Vansteelandt (2009)], or equivalently, sequential ignorability assumptions [Imai, Keele and Yamamoto (2010)] have been carefully studied. Causal inference provides a more general framework to extend mediation analyses in linear models to a more complex setting such as generalized linear models [VanderWeele and Vansteelandt (2010)] and survival models for time-to-event data with censoring [Lange and Hansen (2011), Tchetgen Tchetgen

(2011), VanderWeele (2011), Huang and Cai (2016)]. On the practical side mediation analyses have also been applied across various scientific fields to advance scientific discoveries [Bullock, Green and Ha (2010), Pan et al. (2015), Song et al. (2015)]. Mediation analyses for a large number of mediators have also been proposed for transcriptomic markers [Huang and Pan (2016)] and epigenetic markers [Zhang et al. (2016)].

This paper is motivated by an epigenome-wide mediation analysis for the effect of socioeconomic adversity on obesity. The association of socioeconomic disadvantage with obesity and epigeneic DNA methylation and the association between obesity and DNA methylation have been reported [Agha et al. (2015), Borghol et al. (2012), Senese et al. (2009)]. Building on the existing evidence, we hypothesized that socioeconomic disadvantage during the childhood altered epigenetic DNA methylation which in turn affected adiposity at adulthood. We conducted analyses using the New England Family Study (NEFS) where the childhood socioeconomic disadvantage at age seven using a socioeconomic index and body mass index (kg/m^2) at the mean age of 47 years were assessed in 74 women. DNA methylation of adipose tissue samples collected in the 74 women at adulthood was measured using the Infinium HumanMethylation450K BeadChip (Illumina, San Diego, CA, United States). We conducted epigenome-wide mediation analyses to investigate the mechanism link between the childhood socioeconomic disadvantage and obesity. The epigenome-wide mediation analyses for 285,163 epigenetic DNA methylation loci were conducted one at a time to investigate whether the effect of socioeconomic disadvantage S on adiposity measured by body mass index (BMI) Y was mediated by the DNA methylation level of a locus M.

It has been well established that the mediation effect is proportional to the product $\alpha_S \beta_M$, which is comprised of the association of S and Y, α_S and the association of M and Y conditional on S, β_M [MacKinnon (2008)], provided that confounding is fully adjusted. A normality-based test was performed on $\hat{\alpha}_S \hat{\beta}_M$, where $\hat{\alpha}_S$ and $\hat{\beta}_M$ were maximum likelihood estimators of α_S and β_M respectively. Suppose that $\sigma_{\alpha n}^2$ and $\sigma_{\beta n}^2$ respectively are variances of $\hat{\alpha}_S$ and $\hat{\beta}_M$, the variance of $\hat{\alpha}_S \hat{\beta}_M$ can be approximated using the delta method, $\hat{\beta}_M^2 \sigma_{\alpha n}^2 + \hat{\alpha}_S^2 \sigma_{\beta n}^2$, also known as Sobel's test [MacKinnon (2008), MacKinnon et al. (2002), Sobel (1982)]. The histogram of the p-values from Sobel's test denoted as p_N obtained from normal approximation for $\hat{\alpha}_S \hat{\beta}_M$ is shown in Figure 1. Even for a collection of hypothesis tests conducted entirely under the null, we expect the histogram of p-values to be uniformly distributed. Figure 1 suggests that the test is overly conservative, that is, the signals are much weaker than the expected under the null, only 0.12% of tests with $p < 0.05$ compared with the expected 5%.

In addition to the normality-based test, a variety of hypothesis tests for mediation have been available [MacKinnon et al. (2002)] and will be formally defined in Section 3. Unfortunately, none of these existing tests provides a valid study-wise test size. MacKinnon et al. (2002) and Barfield et al. (2017) concluded via numerical simulation that the joint significance test of $\hat{\alpha}_S$ and $\hat{\beta}_M$ examining $H_0 : \alpha_S = 0$
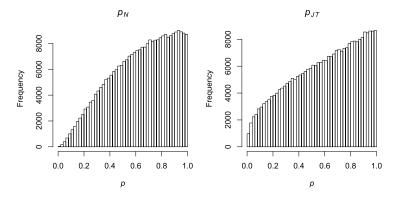
FIG. 1. *Histograms of* 285,163 *p-values in epigenome-wide mediation analyses of CpG loci using the normality-based test* (p_N) *and the joint significance test* (p_{JT}).

and $H_0 : \beta_M = 0$ separately dominates other tests by providing the best balance of Type I error rate and statistical power. However, joint significance tests still have a considerably conservative departure from the theoretical null in the motivating study (Figure 1). These tests were proposed without taking into account the composition of null hypotheses and have very conservative study-wise performance when such information is available within a study. We define $H_0^{(1)} : \alpha_S = \beta_M = 0$, $H_0^{(2)} : \alpha_S \neq 0, \beta_M = 0$ and $H_0^{(3)} : \alpha_S = 0, \beta_M \neq 0$, and the test of mediation effects is conducted under a composite null hypothesis,

$$(1) \qquad H_0 : H_0^{(1)} \cup H_0^{(2)} \cup H_0^{(3)},$$

against an alternative hypothesis,

$$H_A : \alpha_S \neq 0, \beta_M \neq 0.$$

The previous empirical findings are based on numerical studies and have a lack of theoretical justification. Furthermore, with a rapid growth of research data, there is an imperative need for an efficient hypothesis test to facilitate studies that exploit high-dimensional data to examine the effects in multiple mediation analyses. Similar issues of power losses in multiple mediation analyses have also been reported [Barfield et al. (2017)]. To address these, this paper aims to: (1) establish theoretical properties of various hypothesis tests for mediation effects, and (2) develop a new testing procedure for multiple mediation tests that takes into consideration the nature of the composite null hypothesis. The proposed method is expected to provide uniformly distributed p-values under the composite null.

2. Causal mediation model. We are interested in the effect of the socioeconomic status S on the body mass index Y mediated by a mediator M, the methylation level of a locus. The natural indirect [VanderWeele and Vansteelandt (2009)]

or mediation effect of $S = s_1$ vs. s_0 is defined as

$$(2) \qquad \mathrm{E}[Y(s_1, M(s_1))] - \mathrm{E}[Y(s_1, M(s_0))],$$

where $Y(s, m)$ denotes the BMI Y, that is, the potential outcome that would have been observed had S and M been set to s and m respectively, and $M(s)$ denotes the measured value of DNA methylation M, that is, the mediator that would have been observed had S been set to s. We assume two models:

$$(3) \qquad M_i = \boldsymbol{\alpha}_X^T \mathbf{X}_i + \alpha_S S_i + \varepsilon_{Mi},$$

$$(4) \qquad h(Y_i) = \boldsymbol{\beta}_X^T \mathbf{X}_i + \beta_S S_i + \beta_M M_i + \varepsilon_{Yi},$$

where $\mathbf{X}$ is a vector of potential confounders with the first element being 1 for the intercept, ε_{Mi} and ε_{Yi} follow two independent arbitrary zero mean distributions, $i = 1, \ldots, n$, and $h(\cdot)$ is a known or unknown strictly increasing smooth transformation function. The independence of ε_{Mi} and ε_{Yi} holds by the following assumptions. By denoting the independence of A and B conditional on C as $A \perp B | C$, we list four no unmeasured confounding assumptions that are sufficient to identify the effect in (2): conditional on known confounders $\mathbf{X}$, (1) $Y(s) \perp S | \mathbf{X}$, no confounding for the relation of S and Y; (2) $Y(s, m) \perp M | (S, \mathbf{X})$, no confounding for the relation of M and Y, conditional on S; (3) $M(s) \perp S | \mathbf{X}$, no confounding for the relation of S and M; and (4) $Y(s, m) \perp M(s^*) | \mathbf{X}$, no confounder for the M–Y relation that is affected by S. Under the above assumptions, $\mathrm{E}[h\{Y(s_a, M(s_b))\}|\mathbf{X}] = \int \mathrm{E}[h(Y)|s_a, m, \mathbf{X}] \, d F_M(m|s_b, \mathbf{X}) = \boldsymbol{\beta}_X^T \mathbf{X} + \boldsymbol{\alpha}_X^T \mathbf{X} \beta_M + \beta_S s_a + \alpha_S \beta_M s_b$, and mediation effect in (2) can be expressed as

$$(5) \qquad \mathrm{E}[h\{Y(s_1, M(s_1))\}] - \mathrm{E}[h\{Y(s_1, M(s_0))\}] = \alpha_S \beta_M (s_1 - s_0).$$

Since $h(\cdot)$ is strictly increasing, $\mathrm{E}[Y(s_1, M(s_1))] - \mathrm{E}[Y(s_1, M(s_0))] = 0 \leftrightarrow \mathrm{E}[h\{Y(s_1, M(s_1))\}] - \mathrm{E}[h\{Y(s_1, M(s_0))\}] = 0$. It follows that a test for the mediation effect is to test the null hypothesis: $H_0 : \alpha_S \beta_M = 0$ which is equivalent to (1).

We stress that the results in this paper are applicable to pure indirect effect [VanderWeele (2013)], $\mathrm{E}[Y(s_0, M(s_1))] - \mathrm{E}[Y(s_0, M(s_0))]$ and not limited to the model specification in (3) and (4); the product expression of the mediation effect is shared by various models. Under logistic regression models [VanderWeele and Vansteelandt (2010)] or accelerated failure time models [VanderWeele (2011)] assuming a rare outcome or probit models for a survival time [Huang and Cai (2016)], mediation effects all have the same product expression, $\alpha_S \beta_M$, if a monotone function of the outcome Y is determined linearly by $\beta_S S + \beta_M M$. Furthermore, even for cases that have no explicit product expression for mediation effects, its null hypothesis in general can still be shown to be the composite null (1) [Huang (2015)]. Therefore, the results to be introduced are readily generalizable to various model assumptions provided that α_S and β_M respectively characterize the unconfounded association of S and M and that of M and Y given S.

We let $\hat{\alpha}_S$ and $\hat{\beta}_M$ be $\sqrt{n}$-consistent estimators for α_S and β_M respectively, where $\sqrt{n}(\hat{\alpha}_S - \alpha_S) \xrightarrow{d} N(0, \sigma_\alpha^2)$ and $\sqrt{n}(\hat{\beta}_M - \beta_M) \xrightarrow{d} N(0, \sigma_\beta^2)$ as the sample size $n \to \infty$. We construct test statistics $a(\alpha_S) = \sqrt{n}(\hat{\alpha}_S - \alpha_S)/(\sqrt{n}\sigma_{\alpha n})$ and $b(\beta_M) = \sqrt{n}(\hat{\beta}_M - \beta_M)/(\sqrt{n}\sigma_{\beta n})$, where $n\sigma_{\alpha n}^2 \xrightarrow{p} \sigma_\alpha^2$ and $n\sigma_{\beta n}^2 \xrightarrow{p} \sigma_\beta^2$. $a \equiv a(0) \xrightarrow{d} N(0, 1)$ under $\alpha_S = 0$, and $b \equiv b(0) \xrightarrow{d} N(0, 1)$ under $\beta_M = 0$, where a and b are independent under the above no unmeasured confounding assumptions.

3. Hypothesis tests of mediation effect.

3.1. *Single hypothesis test.* We define and study the following five existing tests for mediation effects that were commonly used and compared in the previous literature [MacKinnon et al. (2002)]:

• Normality-based test [Sobel (1982)]: perform a z-test on the test statistic $|ab|/\sqrt{a^2 + b^2}$ to obtain the p-value, $p_N \equiv 2 \times \Phi(-|ab|/\sqrt{a^2 + b^2})$, where $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$.

• Normal product-based test under the alternative: calculate two times of the smaller tail probability of the normal product distribution $z_1 z_2$ from zero to $+\infty$ or $-\infty$, where $z_1 \sim N(a, 1)$, $z_2 \sim N(b, 1)$, and z_1 is independent of z_2. Denote the p-value as p_{NP1}

• Normal product-based test centered at 0: calculate two times the smaller tail probability of the normal product distribution $z_1 z_2$ at $|ab|$ or more extreme, where z_1 and z_2 are independent standard normals. Denote the p-value as $p_{\text{NP0}} \equiv F(ab)$ where $F(z) = 2 \int_{-\infty}^{\infty} \int_{|\frac{z}{x}|}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \, dy \, dx$.

• Joint significance test: conduct z-tests on a and b and obtain the larger p-value of the two tests, that is, $p_{\text{JT}} \equiv 2 \times \max\{\Phi(-|a|), \Phi(-|b|)\}$.

• Bootstrap test: collect bootstrap samples of $\hat{\alpha}_S$ and $\hat{\beta}_M$: $\{\hat{\alpha}_S^{(b)}\}$ and $\{\hat{\beta}_M^{(b)}\}$, where $b = 1, \ldots, B$, B is the number of bootstrap and calculate two times of the smaller tail probability of $\{\hat{\alpha}_S^{(b)} \hat{\beta}_M^{(b)}\}$ at 0 or more extreme as the p-value, p_B.

Let $\sqrt{n}(\hat{\alpha}_S - \alpha_S) \xrightarrow{d} U - \sqrt{n}\alpha_S \sim N(0, \sigma_\alpha^2)$ and $\sqrt{n}(\hat{\beta}_M - \beta_M) \xrightarrow{d} V - \sqrt{n}\beta_M \sim N(0, \sigma_\beta^2)$. The product UV follows a normal product distribution [Aroian (1947)] with moment generating function (m.g.f.):

$$(6) \qquad M_{UV}(t) = \frac{1}{\sqrt{1 - t^2 \sigma_\alpha^2 \sigma_\beta^2}} e^{\frac{1}{1 - t^2 \sigma_\alpha^2 \sigma_\beta^2} \{n\alpha_S \beta_M t + (n\alpha_S^2 \sigma_\beta^2 + n\beta_M^2 \sigma_\alpha^2)t^2/2\}}.$$

Suppose $\zeta = (UV - n\alpha_S \beta_M)/\sqrt{n\alpha_S^2 \sigma_\beta^2 + n\beta_M^2 \sigma_\alpha^2}$ and $n\alpha_S^2 \sigma_\beta^2 + n\beta_M^2 \sigma_\alpha^2 \to \infty$, the m.g.f. of ζ, $M_\zeta(t)$ converges to $e^{t^2/2}$, the m.g.f. of standard normal which justifies the normality-based test with the test statistic $n\hat{\alpha}_S \hat{\beta}_M / \sqrt{n^2 \hat{\alpha}_S^2 \sigma_{\beta n}^2 + n^2 \hat{\beta}_M^2 \sigma_\alpha^2} =$

$ab/\sqrt{a^2 + b^2}$ and the p-value p_N. We note that the normal approximation depends on the condition $n\alpha_S^2\sigma_\beta^2 + n\beta_M^2\sigma_\alpha^2 \to \infty$, which only holds if at least one of α_S and β_M is not zero, that is, under $H_0^{(2)}$ and $H_0^{(3)}$, and is not satisfied under $H_0^{(1)}$: $\alpha_S = \beta_M = 0$. However, in the real application such as genome-wide analyses, the majority of the tests are conducted under $H_0^{(1)}$ due to the sparse signals. Therefore, the normality-based test performs poorly in applications with weak signals.

THEOREM 3.1. *The joint significance test is an intersection-union test* (IUT) *with size α, and its p-value of joint significance tests p_{JT} is smaller than that of normality-based tests, p_N and normal product-based tests under the alternative, p_{NP1}.*

It is straightforward to show the inequality $|ab|/\sqrt{a^2 + b^2} \leq \min(|a|, |b|)$, where the equality is attained if either a or b is zero. It follows that $p_N = \Phi(-|ab|/\sqrt{a^2 + b^2}) \geq \max\{\Phi(-|a|), \Phi(-|b|)\} = p_{JT}$. The proof that p_{JT} is smaller than p_{NP1} is provided in the Appendix. Because bootstrap distributions of $\sqrt{n}\hat{\alpha}_S$ and $\sqrt{n}\hat{\beta}_M$ converge to normal distributions with a reasonably large sample size, the bootstrap distribution of $n\hat{\alpha}_S\hat{\beta}_M$ should converge to a normal distribution, if the condition $n\alpha_S^2\sigma_\beta^2 + n\beta_M^2\sigma_\alpha^2 \to \infty$ is satisfied or a normal product distribution. We therefore expect that as $n \to \infty$, p_B converges to p_{NP1} or p_N and thus is also larger than p_{JT}.

The results shown in this section hold regardless of the composition for the null mixtures. Because of its better power and proper test size shown in Theorem 3.1, we applied the joint significance test in the epigenome-wide mediation analysis. Unfortunately, the joint significance test still suffers from severe power losses with Type I error rate within the study much less than the nominal level [Figure 1(b)]. Although we show that the joint significance test is a size α test, that is, $\sup_{\theta\in\Theta_0} \pi(\theta) = \alpha$ where $\pi(\theta)$ is the power function of parameter θ and Θ_0 is the parameter space under the null, it should be noted that the supremum only occurs when either α_S or β_M is extremely large. Again, in the genome-wide setting with sparse signals the majority of the tests are conducted under small α_S and β_M.

3.2. *Multiple hypothesis tests.* We emphasize that this study is not specifically for developing multiplicity methodology but more for producing uniformly distributed nominal p-values under the null for subsequent multiplicity adjustment using existing methods. Under multiple tests the composition of $H_0^{(1)}$, $H_0^{(2)}$, and $H_0^{(3)}$ may affect the performance of different tests. To address this, we propose a testing procedure under the composite null hypothesis and provide an analytic explanation about its advantage over the joint significance test.

One test on the list in Section 3.1 that we have not discussed is the normal product-based test centered at zero. It examines a very restricted null hypothesis, that is, $H^{(1)} : \alpha_S = \beta_M = 0$, and therefore may mistakenly reject other possible null hypotheses such as $\alpha_S = 0, \beta_M \neq 0$ if one is interested in testing (1).

Nevertheless, we use it as a basis to construct the proposed test. We provide an overview of the following theoretical development. We first derive a technical result in Lemma 3.2 for approximating the weighted sum of multiple $F(\cdot)$ functions into a single $F(\cdot)$. With that we are able to develop in Theorem 3.3 a closed form p-value formula for our proposed composite test, if the nonzero mean of test statistics under $H_0^{(2)}$ and $H_0^{(3)}$ follows normal distribution. In Theorem 3.4 we further generalize the result to the condition that the nonzero mean of test statistics under $H_0^{(2)}$ and $H_0^{(3)}$ follows a mixture of normal distributions.

LEMMA 3.2. *With arbitrary weights $w_1, \ldots, w_J$, $\sum_{j=1}^{J} w_j F(ab/\sqrt{1+\sigma_j^2})$ can be expressed as $(\sum_{j=1}^{J} w_j) F(ab/\sqrt{1+\sigma^2}) + \delta$, where $\sigma^2 = \sum_{j=1}^{J} w_j \sigma_j^2 / \sum_{j=1}^{J} w_j$. And, $|\delta|$ is bounded by $2\pi^{-1} \sum_j w_j (|ab|/\sqrt{1+\sigma^2} - |ab|/\sqrt{1+\sigma^{*2}}) \times K_0(|ab|/\sqrt{1+\sigma^{*2}})$, where $K_0(z)$ is the modified Bessel function of the second kind with order zero, and σ^{*2} satisfies $\sum_{j=1}^{J} w_j F(ab/\sqrt{1+\sigma_j^2}) = (\sum_{j=1}^{J} w_j) \times F(ab/\sqrt{1+\sigma^{*2}})$.*

We note that $\delta \to 0$ if σ_j^2 are similar or close to 0. In the setting where $w_1 = w_2 = 0.5$, $\sigma_1^2 = 0$, $\sigma_2^2 = 1$, and the ratio of c' to c [both defined in the Supplementary Material, Huang (2019)] is set to 1.3, we show in Figure S1a that the bound of $|\delta|$ approaches 0. We suppose that under the null the three hypotheses $H_0^{(1)}$, $H_0^{(2)}$ and $H_0^{(3)}$ occur with unknown probabilities π_0, π_a and π_b respectively, and $\pi_0 + \pi_a + \pi_b = 1$. Using Lemma 3.2, we prove in the Appendix the following main result.

THEOREM 3.3. *Suppose that with probability π_0, the two independent test statistics a and b follow standard normal distribution; with probability π_a, b follows the standard normal and a follows $N(\mu_a, 1)$; with probability π_b, a follows the standard normal and b follows $N(\mu_b, 1)$, where $\mu_a \sim N(0, \sigma_a^2)$ and $\mu_b \sim N(0, \sigma_b^2)$. The p-value for testing the composite null hypothesis* (1) *is $p_{\mathrm{comp}} = \pi_a F(ab/\sqrt{1+\sigma_a^2}) + \pi_b F(ab/\sqrt{1+\sigma_b^2}) + \pi_0 F(ab) = \hat{p}_{\mathrm{comp}}^{N} + \delta_N$, where*

$$
(7) \qquad \hat{p}_{\mathrm{comp}}^{N} = F\left(\frac{ab}{\sqrt{\mathrm{Var}(a)}}\right) + F\left(\frac{ab}{\sqrt{\mathrm{Var}(b)}}\right) - F(ab),
$$

and the bound for $|\delta_N|$ is provided in the Appendix.

The result enables us to calculate the p-value using (7) without estimating the unknown π_0, π_a and π_b. As shown in the Supplementary Material [Huang (2019)], $|\delta_N|$ tends to zero rapidly and its magnitude relative to p_{comp} is so small that in practice, $\hat{p}_{\mathrm{comp}}^{N}$ is sufficient to approximate p_{comp} with decent accuracy. $F(z) = 2 \int_{|z|}^{\infty} f(z) \, dz$ is calculated by numerical integration of probability density function

of normal product distribution: $f(z) = K_0(z)/\pi$. The function $K_0(z)$ is available in R, besselK(, nu=0). We establish a similar result in Theorem 3.4 for a more general setting where the distributions of μ_a and μ_b are not Gaussian but can be represented as a mixture of Gaussians.

THEOREM 3.4.   *Suppose that with probability π_0, two independent test statistics a and b follow standard normal distribution; with probability π_a, b follows the standard normal and a follows $N(\mu_a, 1)$; with probability π_b, a follows the standard normal and b follows $N(\mu_b, 1)$, where $\mu_a \sim Q_a(\mu_a)$ and $\mu_b \sim Q_b(\mu_b)$; and both $Q_a(\mu_a)$ and $Q_b(\mu_b)$ are zero-mean symmetric distributions with respective variances σ_{Qa}^2 and σ_{Qb}^2 and can be represented by a mixture of Gaussians. Under the composite null hypothesis (1), p-value is $p_{\mathrm{comp}} = \pi_a F(ab/\sqrt{1+\sigma_{Qa}^2}) + \pi_b F(ab/\sqrt{1+\sigma_{Qb}^2}) + \pi_0 F(ab) + \delta_{Q1}$ and $p_{\mathrm{comp}} = \hat{p}_{\mathrm{comp}} + \delta_{Q1} + \delta_{Q2}$, where*

$$(8) \qquad \hat{p}_{\mathrm{comp}} = F\left(\frac{ab}{\sqrt{\mathrm{Var}(a)}}\right) + F\left(\frac{ab}{\sqrt{\mathrm{Var}(b)}}\right) - F(ab),$$

$\delta_{Q1} = \pi_a \delta_{Q1a} + \pi_b \delta_{Q1b}$ with bounds for $|\delta_{Q1a}|$, $|\delta_{Q1b}|$ and $|\delta_{Q2}|$ provided in the Appendix.

Note that $\hat{p}_{\mathrm{comp}}^N$ and $\hat{p}_{\mathrm{comp}}$ share an identical expression. Because $\hat{p}_{\mathrm{comp}}$ of the composite tests considers both $\hat{\alpha}_S$ and $\hat{\beta}_M$, its value decreases as the absolute value of either one increases. For $|b| > |a|$, p_{JT} given the same $|a|$ is identical no matter how large the value of $|b|$ is. Although p_{JT} is robust under any arbitrary null by not assuming the mixture composition of null hypotheses, it loses considerable power, that is, a strong signal in $\hat{\beta}_M$ does not help the test for $\alpha_S \beta_M = 0$ if the signal in $\hat{\alpha}_S$ is weak. The advantage of the composite test over the conventional joint significance test is formally characterized as follows.

PROPOSITION 3.5.   *Suppose $|b| > |a|$, the p-value of the conventional joint significance test p_{JT} is larger than p_{comp} if $|b| > \sqrt{1+\sigma^{*2}}\Phi^{-1}((1+\Phi(|a|))/2)$ where $p_{\mathrm{comp}} = F(ab/\sqrt{1+\sigma^{*2}})$.*

$|b| > \sqrt{1+\sigma^{*2}}\Phi^{-1}((1+\Phi(|a|))/2)$ is a sufficient condition but not necessary. By Lemma 3.2 the condition for $\hat{p}_{\mathrm{comp}} < p_{\mathrm{JT}}$ is approximately $|b| > \sqrt{\mathrm{Var}(a) + \mathrm{Var}(b) - 1}\Phi^{-1}((1+\Phi(|a|))/2)$. The higher variability in a or b requires even larger $|b|$ to ensure a smaller p-value in the composite tests than p_{JT}. As long as the variances of a and b are finite, $\hat{p}_{\mathrm{comp}}$ would eventually be smaller than p_{JT} if $|b|$ is large enough. Because a and b are independent and symmetric, the above result also implies the following: $p_{\mathrm{JT}} > p_{\mathrm{comp}}$ if $|a| > |b|$ and $|a| > \sqrt{1+\sigma^{*2}}\Phi^{-1}((1+\Phi(|b|))/2)$.

4. Simulation. We conducted numerical studies to evaluate the performance of the proposed testing procedures in Theorems 3.3 and 3.4 in the setting of multiple tests. We simulated 10^7 pairs of test statistics (a_j, b_j), $j = 1, \ldots, 10^7$, assuming all a's and b's are independent and repeated the experiment for 20 times. Among the 10^7 pairs of test statistics, there were m_0 pairs of a_j and b_j both following $N(0, 1)$, m_1 pairs with a_j following $N(\mu_a, 1)$ and b_j following $N(0, 1)$, m_2 pairs with a_j following $N(0, 1)$ and b_j following $N(\mu_b, 1)$, m_3 pairs with a_j following $N(\mu_a, 1)$ and b_j following $N(\mu_b, 1)$, and $m = m_0 + m_1 + m_2 + m_3 = 10^7$. We explored four different distributions for μ_a and μ_b: (1) both μ_a and μ_b followed independent zero-mean normal, (2) μ_a followed $N(0, 0.5^2)$ with probability of 0.3 and $N(0, 1.5^2)$ with probability of 0.7, and μ_b followed $N(0, 0.5^2)$ with probability of 0.4 and $N(0, 1.5^2)$ with probability of 0.6, (3) μ_a and μ_b followed two independent t-distributions with five degrees of freedom, and (4) μ_a and μ_b followed two independent uniform distributions both ranging from -1 to 1. The four distributions were standardized to have standard deviation σ_a for μ_a and σ_b for μ_b.

The joint significance test and three composite tests were conducted. The first composite test calculated the p-value p_{comp} by plugging in the true $\pi_a = m_1/(m_0 + m_1 + m_2)$, $\pi_b = m_2/(m_0 + m_1 + m_2)$, σ_a^2 and σ_b^2. The second method estimated the proportion of null [Langaas, Lindqvist and Ferkingstad (2005)] in $\{a_j, j = 1, \ldots, 10{,}000\}$ and $\{b_j\}$, denoted by $\hat{\pi}_a$ and $\hat{\pi}_b$ respectively, and subsequently, the proportion of $H_0^{(3)}$, $H_0^{(2)}$ and $H_0^{(1)}$ as $w_3 = \hat{\pi}_a(1 - \hat{\pi}_b)/\hat{\pi}_{ab}$, $w_2 = \hat{\pi}_b(1 - \hat{\pi}_a)/\hat{\pi}_{ab}$ and $w_1 = \hat{\pi}_a\hat{\pi}_b/\hat{\pi}_{ab}$ respectively, where $\hat{\pi}_{ab} = \hat{\pi}_a + \hat{\pi}_b - \hat{\pi}_a\hat{\pi}_b$. With the proportion of three types of null, we calculated the p-value under each type of null and weighted by the proportion. Specifically, we calculated p-value denoted as $\tilde{p}_{\text{comp}}$ as $w_3\Phi(-|a|) + w_2\Phi(-|b|) + w_1 F(ab)$. The third was our proposed method calculating p-value $\hat{p}_{\text{comp}} = F(ab/\sqrt{\text{Var}(a)}) + F(ab/\sqrt{\text{Var}(b)}) - F(ab)$.

Under the null where $m_0 = 5 \times 10^6$, $m_1 = 3 \times 10^6$, $m_2 = 2 \times 10^6$, $m_3 = 0$ and $\sigma_a = \sigma_b = 0.5$ (Table 1), our proposed method $\hat{p}_{\text{comp}}$ had a comparable performance to p_{comp}. Their Type I error rates were close to the expected under the setting where the distributions of μ_a and μ_b are normal, mixture of normals or uniform; at the significance level of 10^{-6} or lower, there was an inflation of Type I error rate when μ_a and μ_b followed t distributions. The joint significance test had a much more conservative Type I error rate even at the level of 0.1, and $\tilde{p}_{\text{comp}}$ had a conservative Type I error rate at the level of 0.001 or lower. Under the alternative where $m_0 = 1 \times 10^6$, $m_1 = 2 \times 10^6$, $m_2 = 2 \times 10^6$, $m_3 = 5 \times 10^6$ and $\sigma_a = \sigma_b = 1.5$ (Table 2), the proposed test outperformed other tests at the level of 0.001 or lower, but $\tilde{p}_{\text{comp}}$ had better power at the level of 0.1 and 0.01. Similar patterns emerged when μ_a and μ_b followed t distributions, mixtures of normals or uniforms.

We further conducted three sets of simulation studies to investigate the impact of $(\pi_0, \pi_a, \pi_b, \sigma_a, \sigma_b)$ on the performance of $\hat{p}_{\text{comp}}$. In the first simulation study

TABLE 1
Proportions (with Monte Carlo standard deviations) of p-values at various cut-offs using different tests under the null where $m_0 = 5 \times 10^6$, $m_1 = 3 \times 10^6$, $m_2 = 2 \times 10^6$ and $m_3 = 0$

Cut-offs	p_{comp}	$\hat{p}_{\text{comp}}$	$\tilde{p}_{\text{comp}}$	p_{JT}
$(\mu_a, \mu_b) \sim$ normal distribution				
$p < 10^{-1}$	0.99 (0.0008) $\times 10^{-1}$	0.99 (0.0005) $\times 10^{-1}$	1.06 (0.0007) $\times 10^{-1}$	0.12 (0.0004) $\times 10^{-1}$
$p < 10^{-2}$	0.99 (0.003) $\times 10^{-2}$	1.01 (0.003) $\times 10^{-2}$	1.01 (0.003) $\times 10^{-2}$	0.02 (0.0004) $\times 10^{-2}$
$p < 10^{-3}$	1.00 (0.01) $\times 10^{-3}$	1.04 (0.01) $\times 10^{-3}$	0.65 (0.008) $\times 10^{-3}$	0.002 (0.0004) $\times 10^{-3}$
$p < 10^{-4}$	1.00 (0.03) $\times 10^{-4}$	1.11 (0.03) $\times 10^{-4}$	0.22 (0.01) $\times 10^{-4}$	0.0003 (0.0006) $\times 10^{-4}$
$p < 10^{-5}$	0.99 (0.11) $\times 10^{-5}$	1.18 (0.12) $\times 10^{-5}$	0.04 (0.02) $\times 10^{-5}$	0.0 (0.0)
$p < 10^{-6}$	1.06 (0.27) $\times 10^{-6}$	1.36 (0.31) $\times 10^{-6}$	0.02 (0.04) $\times 10^{-6}$	0.0 (0.0)
$p < 10^{-7}$	0.85 (0.99) $\times 10^{-7}$	1.25 (1.25) $\times 10^{-7}$	0.0 (0.0)	0.0 (0.0)
$(\mu_a, \mu_b) \sim$ mixture of normal distribution				
$p < 10^{-1}$	0.99 (0.001) $\times 10^{-1}$	0.99 (0.0007) $\times 10^{-1}$	1.06 (0.0008) $\times 10^{-1}$	0.12 (0.0003) $\times 10^{-1}$
$p < 10^{-2}$	1.00 (0.004) $\times 10^{-2}$	1.01 (0.003) $\times 10^{-2}$	1.02 (0.002) $\times 10^{-2}$	0.02 (0.0004) $\times 10^{-2}$
$p < 10^{-3}$	1.02 (0.01) $\times 10^{-3}$	1.06 (0.01) $\times 10^{-3}$	0.68 (0.009) $\times 10^{-3}$	0.002 (0.0005) $\times 10^{-3}$
$p < 10^{-4}$	1.05 (0.03) $\times 10^{-4}$	1.16 (0.03) $\times 10^{-4}$	0.24 (0.01) $\times 10^{-4}$	0.0002 (0.0004) $\times 10^{-4}$
$p < 10^{-5}$	1.09 (0.11) $\times 10^{-5}$	1.29 (0.10) $\times 10^{-5}$	0.04 (0.02) $\times 10^{-5}$	0.0 (0.0)
$p < 10^{-6}$	1.04 (0.28) $\times 10^{-6}$	1.37 (0.32) $\times 10^{-6}$	0.0 (0.0)	0.0 (0.0)
$p < 10^{-7}$	1.15 (0.67) $\times 10^{-7}$	1.50 (1.00) $\times 10^{-7}$	0.0 (0.0)	0.0 (0.0)
$(\mu_a, \mu_b) \sim t$ distribution				
$p < 10^{-1}$	0.99 (0.001) $\times 10^{-1}$	0.99 (0.0007) $\times 10^{-1}$	1.06 (0.0008) $\times 10^{-1}$	0.12 (0.0003) $\times 10^{-1}$
$p < 10^{-2}$	1.00 (0.002) $\times 10^{-2}$	1.01 (0.003) $\times 10^{-2}$	1.03 (0.002) $\times 10^{-2}$	0.02 (0.0005) $\times 10^{-2}$
$p < 10^{-3}$	1.06 (0.01) $\times 10^{-3}$	1.11 (0.01) $\times 10^{-3}$	0.71 (0.009) $\times 10^{-3}$	0.002 (0.0006) $\times 10^{-3}$
$p < 10^{-4}$	1.22 (0.04) $\times 10^{-4}$	1.34 (0.03) $\times 10^{-4}$	0.27 (0.02) $\times 10^{-4}$	0.0005 (0.0008) $\times 10^{-4}$
$p < 10^{-5}$	1.69 (0.13) $\times 10^{-5}$	1.92 (0.16) $\times 10^{-5}$	0.06 (0.02) $\times 10^{-5}$	0.0 (0.0)
$p < 10^{-6}$	3.17 (0.61) $\times 10^{-6}$	3.75 (0.62) $\times 10^{-6}$	0.02 (0.04) $\times 10^{-6}$	0.0 (0.0)
$p < 10^{-7}$	9.65 (3.15) $\times 10^{-7}$	11.4 (3.41) $\times 10^{-7}$	0.0 (0.0)	0.0 (0.0)
$(\mu_a, \mu_b) \sim$ uniform distribution				
$p < 10^{-1}$	1.00 (0.0007) $\times 10^{-1}$	0.99 (0.0006) $\times 10^{-1}$	1.06 (0.0007) $\times 10^{-1}$	0.12 (0.0004) $\times 10^{-1}$
$p < 10^{-2}$	0.99 (0.004) $\times 10^{-2}$	1.00 (0.003) $\times 10^{-2}$	0.99 (0.004) $\times 10^{-2}$	0.02 (0.0002) $\times 10^{-2}$
$p < 10^{-3}$	0.97 (0.009) $\times 10^{-3}$	1.02 (0.01) $\times 10^{-3}$	0.62 (0.008) $\times 10^{-3}$	0.002 (0.0005) $\times 10^{-3}$
$p < 10^{-4}$	0.95 (0.03) $\times 10^{-4}$	1.04 (0.03) $\times 10^{-4}$	0.19 (0.01) $\times 10^{-4}$	0.0003 (0.0005) $\times 10^{-4}$
$p < 10^{-5}$	0.88 (0.09) $\times 10^{-5}$	1.04 (0.08) $\times 10^{-5}$	0.04 (0.02) $\times 10^{-5}$	0.0 (0.0)
$p < 10^{-6}$	0.90 (0.42) $\times 10^{-6}$	1.13 (0.50) $\times 10^{-6}$	0.01 (0.02) $\times 10^{-6}$	0.0 (0.0)
$p < 10^{-7}$	1.05 (1.47) $\times 10^{-7}$	1.55 (1.67) $\times 10^{-7}$	0.0 (0.0)	0.0 (0.0)

we set up a very sparse effect: 90% $H_0^{(1)}$, 5% $H_0^{(2)}$ and 5% $H_0^{(3)}$; in the second study we made π_a and π_b very different: 50% $H_0^{(2)}$ and 0% $H_0^{(3)}$ (and 50% $H_0^{(1)}$); and finally, we made $\sigma_a = 0.2$ and $\sigma_b = 0.8$. Because the proposed composite test accommodates the proportion of null with the variance of a and b, $\hat{p}_{\text{comp}}$ still well protected Type I error rate up to the significance level of 10^{-7} provided that μ_a and μ_b followed normal, mixture of normals or uniform distributions (Tables S1–S3).

For approximation the proposed method requires small σ_a^2 and σ_b^2, a function of sample size, and thus the assumption may be violated with a very large sample

TABLE 2
Proportions (with Monte Carlo standard deviations) of p-values at various cut-offs using different tests under the alternative where $m_0 = 1 \times 10^6$, $m_1 = 2 \times 10^6$, $m_2 = 2 \times 10^6$ and $m_3 = 5 \times 10^6$

Cut-offs	p_{comp}	$\hat{p}_{\text{comp}}$	$\tilde{p}_{\text{comp}}$	p_{JT}
	$(\mu_a, \mu_b) \sim$ normal distribution			
$p < 10^{-1}$	2.99 (0.002) $\times 10^{-1}$	1.32 (0.0009) $\times 10^{-1}$	2.23 (0.001) $\times 10^{-1}$	0.81 (0.0009) $\times 10^{-1}$
$p < 10^{-2}$	6.22 (0.008) $\times 10^{-2}$	3.17 (0.005) $\times 10^{-2}$	3.56 (0.005) $\times 10^{-2}$	1.23 (0.003) $\times 10^{-2}$
$p < 10^{-3}$	14.7 (0.03) $\times 10^{-3}$	8.80 (0.02) $\times 10^{-3}$	6.39 (0.03) $\times 10^{-3}$	2.33 (0.02) $\times 10^{-3}$
$p < 10^{-4}$	37.2 (0.16) $\times 10^{-4}$	26.0 (0.16) $\times 10^{-4}$	12.7 (0.13) $\times 10^{-4}$	4.78 (0.08) $\times 10^{-4}$
$p < 10^{-5}$	97.6 (1.13) $\times 10^{-5}$	78.6 (1.03) $\times 10^{-5}$	26.5 (0.55) $\times 10^{-5}$	10.1 (0.29) $\times 10^{-5}$
$p < 10^{-6}$	260 (6.14) $\times 10^{-6}$	242 (5.60) $\times 10^{-6}$	56.8 (2.87) $\times 10^{-6}$	22.6 (1.87) $\times 10^{-6}$
$p < 10^{-7}$	701 (31.4) $\times 10^{-7}$	749 (29.2) $\times 10^{-7}$	130 (10.9) $\times 10^{-7}$	49.8 (5.23) $\times 10^{-7}$
	$(\mu_a, \mu_b) \sim$ mixture of normal distribution			
$p < 10^{-1}$	2.75 (0.001) $\times 10^{-1}$	1.20 (0.001) $\times 10^{-1}$	2.10 (0.0009) $\times 10^{-1}$	0.70 (0.001) $\times 10^{-1}$
$p < 10^{-2}$	5.82 (0.008) $\times 10^{-2}$	3.11 (0.005) $\times 10^{-2}$	3.36 (0.006) $\times 10^{-2}$	1.14 (0.003) $\times 10^{-2}$
$p < 10^{-3}$	15.6 (0.03) $\times 10^{-3}$	10.0 (0.02) $\times 10^{-3}$	6.99 (0.03) $\times 10^{-3}$	2.75 (0.02) $\times 10^{-3}$
$p < 10^{-4}$	48.5 (0.25) $\times 10^{-4}$	36.0 (0.25) $\times 10^{-4}$	18.0 (0.16) $\times 10^{-4}$	7.65 (0.09) $\times 10^{-4}$
$p < 10^{-5}$	162 (1.46) $\times 10^{-5}$	137 (1.25) $\times 10^{-5}$	51.5 (0.75) $\times 10^{-5}$	22.5 (0.64) $\times 10^{-5}$
$p < 10^{-6}$	573 (8.08) $\times 10^{-6}$	540 (7.57) $\times 10^{-6}$	153 (4.77) $\times 10^{-6}$	68.0 (3.05) $\times 10^{-6}$
$p < 10^{-7}$	2069 (59.5) $\times 10^{-7}$	2182 (63.2) $\times 10^{-7}$	465 (22.8) $\times 10^{-7}$	208 (12.7) $\times 10^{-7}$
	$(\mu_a, \mu_b) \sim t$ distribution			
$p < 10^{-1}$	2.70 (0.002) $\times 10^{-1}$	1.15 (0.001) $\times 10^{-1}$	2.07 (0.001) $\times 10^{-1}$	0.67 (0.0009) $\times 10^{-1}$
$p < 10^{-2}$	5.41 (0.008) $\times 10^{-2}$	2.86 (0.005) $\times 10^{-2}$	3.07 (0.005) $\times 10^{-2}$	0.97 (0.003) $\times 10^{-2}$
$p < 10^{-3}$	14.4 (0.05) $\times 10^{-3}$	9.39 (0.03) $\times 10^{-3}$	5.67 (0.03) $\times 10^{-3}$	2.13 (0.01) $\times 10^{-3}$
$p < 10^{-4}$	48.0 (0.30) $\times 10^{-4}$	36.8 (0.22) $\times 10^{-4}$	14.0 (0.12) $\times 10^{-4}$	6.08 (0.09) $\times 10^{-4}$
$p < 10^{-5}$	189 (2.09) $\times 10^{-5}$	165 (1.89) $\times 10^{-5}$	43.3 (0.86) $\times 10^{-5}$	21.3 (0.67) $\times 10^{-5}$
$p < 10^{-6}$	850 (12.9) $\times 10^{-6}$	815 (12.1) $\times 10^{-6}$	159 (5.48) $\times 10^{-6}$	85.2 (3.30) $\times 10^{-6}$
$p < 10^{-7}$	4218 (103) $\times 10^{-7}$	4362 (102) $\times 10^{-7}$	667 (31.3) $\times 10^{-7}$	388 (27.2) $\times 10^{-7}$
	$(\mu_a, \mu_b) \sim$ uniform distribution			
$p < 10^{-1}$	3.31 (0.002) $\times 10^{-1}$	1.49 (0.001) $\times 10^{-1}$	2.41 (0.001) $\times 10^{-1}$	0.98 (0.001) $\times 10^{-1}$
$p < 10^{-2}$	6.73 (0.008) $\times 10^{-2}$	3.14 (0.004) $\times 10^{-2}$	3.82 (0.005) $\times 10^{-2}$	1.31 (0.004) $\times 10^{-2}$
$p < 10^{-3}$	12.3 (0.03) $\times 10^{-3}$	6.34 (0.02) $\times 10^{-3}$	5.09 (0.02) $\times 10^{-3}$	1.57 (0.009) $\times 10^{-3}$
$p < 10^{-4}$	19.3 (0.11) $\times 10^{-4}$	11.5 (0.08) $\times 10^{-4}$	5.65 (0.08) $\times 10^{-4}$	1.61 (0.04) $\times 10^{-4}$
$p < 10^{-5}$	26.2 (0.52) $\times 10^{-5}$	18.7 (0.45) $\times 10^{-5}$	5.24 (0.19) $\times 10^{-5}$	1.38 (0.11) $\times 10^{-5}$
$p < 10^{-6}$	30.6 (1.60) $\times 10^{-6}$	27.0 (1.48) $\times 10^{-6}$	4.24 (0.71) $\times 10^{-6}$	1.02 (0.32) $\times 10^{-6}$
$p < 10^{-7}$	31.1 (6.68) $\times 10^{-7}$	35.4 (6.90) $\times 10^{-7}$	3.00 (2.18) $\times 10^{-7}$	0.70 (0.66) $\times 10^{-7}$

size. We conducted another set of simulation studies to investigate the influence of sample size (n) provided that the nonzero effects of exposure-mediator (20% of tests) and mediator-outcome (30%) were drawn from the indicated distributions (Figure 2). Specifically, we simulated the effects α_S and β_M from the aforementioned normal distribution with mean following normal, mixture of normals, t or uniform distribution and then used models (3) and (4) to generate mediator and outcome, where $S_i \sim N(0, 1)$, $\boldsymbol{\alpha}_X^T \mathbf{X}_i = \boldsymbol{\beta}_X^T \mathbf{X}_i = \beta_S = 1$.
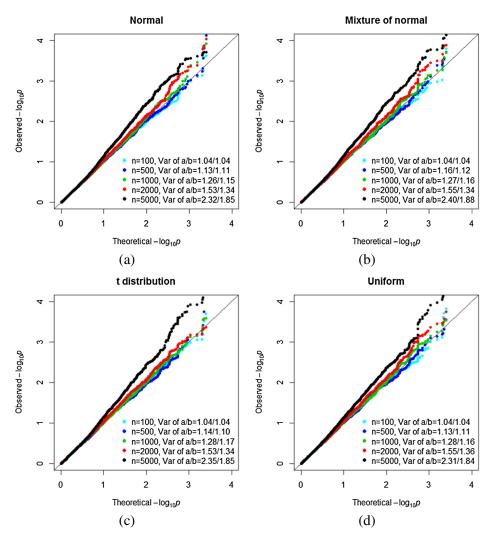
FIG. 2. *Quantile–quantile plots of $\hat{p}_{comp}$ by different sample sizes n:* (a) *the noncentrality parameters for α_S and β_M (μ_a, μ_b) ~ normal distribution;* (b) *(μ_a, μ_b) ~ mixture of normal distribution;* (c) *(μ_a, μ_b) ~ t distribution; and* (d) *(μ_a, μ_b) ~ uniform distribution.*

In Figure 2 the QQ plot suggests that the Type I error rate was well protected if the sample size was less than 500, and would be inflated if the sample size was larger than 2000. The impact of sample size may be better reflected by variances of the test statistics a and b. A practical suggestion is to apply our method when Var(a) and Var(b) are both less than 1.5. When the signal is not sparse, for example, due to the large sample size, or large Var(a) and Var(b), the normal approximation for the majority of the tests is more likely to hold as shown in the

preceding Section. Therefore, under such scenarios one may resort to Sobel's test or joint significance test.

In order to investigate the effect of number of mediators on our proposed method, we conducted a simulation study with the number of mediators m being 100, 1000, 10,000 and 100,000 under the condition that μ_a and μ_b follow various types of distributions with 50%, 30% and 20% of m being m_0, m_1 and m_2 respectively. Type I error at the cut-offs = 0.1, 0.01, 0.001, and 0.0001 was evaluated (Table S3). The results suggest that the number of mediators did not seem to largely affect the performance of our proposed composite tests.

Taken together, our proposed testing procedure $\hat{p}_{comp}$ well protects the Type I error rate across various scenarios with sparse effects and dominates the joint significance test and the test estimating the proportion of null mixtures $\tilde{p}_{comp}$. The method has been implemented as an R software JT_Comp available at http://www.stat.sinica.edu.tw/ythuang/JT-Comp.zip.

5. Epigenome-wide mediation analyses. The DNA methylation data measured by the Infinium HumanMethylation450K BeadChip were preprocessed and normalized using the R methylumi package [Davis et al. (2015)] and the Beta-Mixture Quantile Dilation approach [Teschendorff et al. (2013)]. The beta values of methylation levels (ranging from 0 to 1) were logit transformed prior to statistical analyses. After filtering by variation and genomic locations, there were 285,163 methylation markers [Huang et al. (2016)]. Socioeconomic adversity of the 74 women was measured at age seven using a socioeconomic index assessing prospectively by averaging the percentile of both parents' educational attainment, occupation and income relative to the general US population [Loucks et al. (2016)]. Potential confounders including sex, age, race and cigarette smoking (current number of cigarettes smoked per day) were adjusted as covariates throughout the analyses.

We performed an epigenome-wide mediation analysis examining whether the 285,163 DNA methylation loci one at a time mediate the effect of socioeconomic disadvantage on adiposity. Four testing methods were conducted: the conventional joint significance test (p_{JT}), the normality-based test (p_N), the composite test estimating the proportion of null ($\tilde{p}_{comp}$) and our proposed composite tests ($\hat{p}_{comp}$). The quantile–quantile plots and histograms of p-values in Figures 1, 3 and 4 show that $\hat{p}_{comp}$ performed well as opposed to power losses in p_{JT} and p_N and potential inflation of Type I error rate in $\tilde{p}_{comp}$.

After adjusting for multiple comparisons with the false discovery rate (FDR) [Benjamini and Hochberg (1995)], only $\hat{p}_{comp}$ had four methylation loci, cg04145890 (at the gene *FGFRL1*), cg21120176 (at *LOC285780*), cg26750548 (at *LDB3*) and cg05157970 (at *FASN*), with FDR < 0.2, and the other methods had none surviving this cut-off (Table 3). As shown in Figure 5, the four CpG loci are also among top candidates using the other tests. The overall magnitude of p-values in p_N and p_{JT} is larger than $\hat{p}_{comp}$.
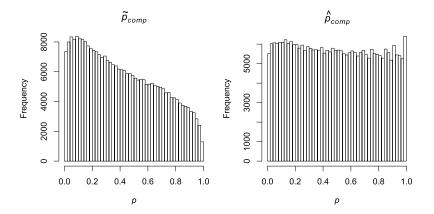
FIG. 3. *Histograms of* 285,163 *p-values in epigenome-wide mediation analyses of CpG loci using the $\tilde{p}_{\text{comp}}$ defined in Section* 4 *and $\hat{p}_{\text{comp}}$ defined in* (8).

Mutations in the gene *LDB3*, LIM domain binding 3 have been shown associated with several muscular diseases [Lopez-Ayala et al. (2014), Selcen and Engel (2005)], and the genes *FGFRL1*, fibroblast growth factor receptor-like 1 and *FASN*, fatty acid synthase respectively have been reported for the association with bone formation [Niu et al. (2015)] and body weight [Loftus et al. (2000)]. While it has been documented that individuals with socioeconomic adversity may have different epigenetic profile by consuming less methyl donor foods such as fruits and vegetables [Darmon and Drewnowski (2008), Giskes et al. (2010)], our analyses support the hypothesis that socioeconomic status may alter epigenetics of these genes via the dietary habit and the *FGFRL1/LDB3/FASN* expression regulated by epigenetics has a subsequent influence on body weight or skeletomuscular growth.

We also used the bootstrap method to analyze the 16,394 CpG loci on chromosome 17. The bootstrap test was carried out with 1000 replicates, and the smallest p_B was set to 5×10^{-4}. There were 95 (0.58%), 96 (0.59%), 14 (0.09%), 27 (0.16%) and 0 (0%) methylation loci with $p < 0.005$ for $\hat{p}_{\text{comp}}$, $\tilde{p}_{\text{comp}}$, p_{JT}, p_B and p_N respectively; the respective computation time was 122.83 seconds, 115.55 seconds, 108.71 seconds, 29.23 hours and 108.68 seconds using a laptop with Intel® Core™ i7-6500 CPU at 2.50 GHz/2.50 GHz and 8.00 GB RAM. We note that the computation time was based on 1000 bootstrap resampling. To reach the genome-wide significance, 10^7 replicates or more are usually required which makes it not feasible for the epigenome-wide analysis. Similar to the epigenome-wide analysis, $\hat{p}_{\text{comp}}$ has a better performance in both Type I error rate and power than $\tilde{p}_{\text{comp}}$, p_{JT}, p_B and p_N [Supplementary Material, Huang (2019)].

6. Discussion. The model (4) can be extended to incorporate an exposure-by-mediator cross-product interaction $\beta_{SM} S_i M_i$:

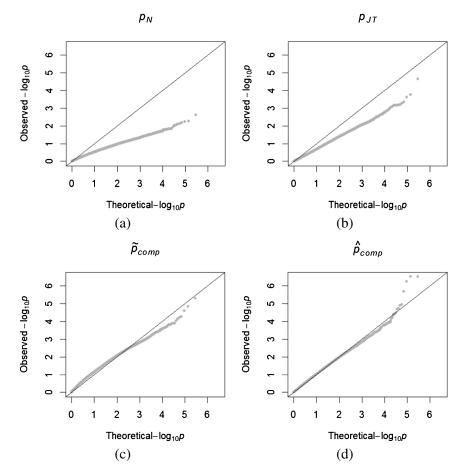$$h(Y_i) = \boldsymbol{\beta}_X^T \mathbf{X}_i + \beta_S S_i + \beta_M M_i + \beta_{SM} S_i M_i + \varepsilon_{Yi}.$$

FIG. 4.   *Quantile–quantile plots of mediation p-values of* 285,163 *CpG loci for the socioeconomic index-BMI association.* (a) p_N, *p-value of the normality-based test*; (b) p_{JT}, *p-value of the joint significance test*; (c) $\tilde{p}_{comp}$ *defined in Section* 4; *and* (d) $\hat{p}_{comp}$ *defined in* (8).

TABLE 3
The four CpG loci with FDR < 0.2 *based on* $\hat{p}_{comp}$ *for the mediation effect of socioeconomic index on BMI*

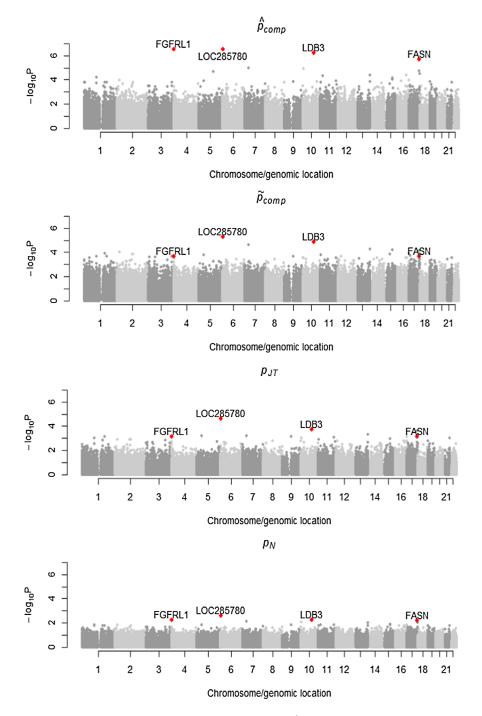| Probe name	Gene	$\hat{p}_{comp}$	FDR	$\tilde{p}_{comp}$	p_{JT}	p_N
cg04145890	*FGFRL1*	3.1×10^{-7}	0.04	2.0×10^{-4}	6.9×10^{-4}	0.0059
cg21120176	*LOC285780*	3.1×10^{-7}	0.04	5.0×10^{-6}	2.3×10^{-5}	0.0024
cg26750548	*LDB3*	5.8×10^{-7}	0.05	1.4×10^{-5}	1.8×10^{-4}	0.0053
cg05157970	*FASN*	1.1×10^{-6}	0.14	2.1×10^{-4}	7.1×10^{-4}	0.0066

FIG. 5. *p-values of epigenome-wide mediation analyses.* $\hat{p}_{comp}$ *defined in* (8); $\tilde{p}_{comp}$ *defined in Section* 4; p_{JT}, *p-value of joint significance tests*; *and* p_N, *p-value of normality-based tests. Red dots: FDR* < 0.2 *based on* $\hat{p}_{comp}$.

It follows that

$$E[h\{Y(s_a, M(s_b))\}] = \boldsymbol{\beta}_X^T \mathbf{X} + \boldsymbol{\alpha}_X^T \mathbf{X}(\beta_M + \beta_{SM}s_a) + \beta_S s_a + \alpha_S(\beta_M + \beta_{SM}s_a)s_b,$$

and mediation effect (2) is proportional to

$$\alpha_S(\beta_M + \beta_{SM}s_1).$$

By letting $\hat{\beta}_{SM}$ be the estimator for β_{SM}, $b(\beta_M^*) = \frac{\sqrt{n}(\hat{\beta}_M^* - \beta_M^*)}{\sqrt{n}\sigma_{\beta n}}$ where $\beta_M^* = \beta_M + s_1\beta_{SM}$ and $\hat{\beta}_M^* = \hat{\beta}_M + s_1\hat{\beta}_{SM}$. The results in Section 3 are readily applicable to such a generalization.

The proposed method depends on the independence within a and b and between each other. The dependence between a and b may exist if the assumptions of no unmeasured confounding in Section 2 do not hold. Sensitivity analyses may be conducted to evaluate the influence by the potential confounding [Imai, Keele and Yamamoto (2010), VanderWeele (2010)]. The independence within a is equivalent to the independence of $\hat{\alpha}_{Sj}$, $j = 1, \ldots, m$. One sufficient condition is that M_j is independent conditional on S and $\mathbf{X}$. Similarly, we require the independence of $\hat{\beta}_{Mj}$. These independence assumptions may not necessarily hold in applications. How to account for the dependence of the test statistics warrants further research.

The proof of Theorems 3.3 and 3.4 utilizes the equalities $\mathrm{Var}(a) = 1 + \pi_a \sigma_{Qa}^2$ and $\mathrm{Var}(b) = 1 + \pi_b \sigma_{Qb}^2$, which only hold under the null, $\pi_0 + \pi_a + \pi_b = 1$. Suppose that with probability π_c, $a \sim N(\mu_a, 1)$ and $b \sim N(\mu_b, 1)$ where $\mu_a \sim N(0, \sigma_a^2)$ and $\mu_b \sim N(0, \sigma_b^2)$, then variances of a and b would be $1 + (\pi_a + \pi_c)\sigma_a^2$ and $1 + (\pi_b + \pi_c)\sigma_b^2$ respectively. Under the alternative, $\pi_0 + \pi_a + \pi_b + \pi_c = 1$, $\hat{p}_{\mathrm{comp}}$ is larger than the theoretical p-value, p_{comp} because

$$F(ab/\sqrt{\mathrm{Var}(a)}) + F(ab/\sqrt{\mathrm{Var}(b)}) - F(ab)$$
$$> F\left(ab/\sqrt{1 + \pi_a \sigma_{Qa}^2}\right) + F\left(ab/\sqrt{1 + \pi_b \sigma_{Qb}^2}\right) - F(ab).$$

If π_c, σ_a^2, and σ_b^2 can be consistently estimated, one may gain power by a revised p-value, $F(ab/\sqrt{\mathrm{Var}(a) - \pi_c \sigma_a^2}) + F(ab/\sqrt{\mathrm{Var}(b) - \pi_c \sigma_b^2}) - F(ab)$.

We express the general test statistics a or b for the jth mediator as $\sqrt{n}(\hat{\theta}_j - \theta_j)/\sqrt{n\sigma_{\theta n}^2}$ where $\hat{\theta}_j$ is a maximum likelihood estimator for the true parameter θ_j, $n\sigma_{\theta n}^2 = O_p(1)$, and n is the sample size. It follows that the noncentrality of the test statistics, μ_a and μ_b, is $\theta_j/\sqrt{\sigma_{\theta n}^2}$, which we assume is a random draw from a normal distribution, $\theta_j/\sqrt{\sigma_{\theta n}^2} \sim N(0, \sigma_\theta^2)$. We note that $N(0, \sigma_\theta^2)$ is a working distribution where σ_θ^2 represents the dispersion of the noncentrality, and that $\sigma_\theta^2 = \sigma_a^2$ if $\theta_j = \alpha_S$ and $\hat{\theta}_j = \hat{\alpha}_S$, and $\sigma_\theta^2 = \sigma_b^2$ if $\theta_j = \beta_M$ and $\hat{\theta}_j = \hat{\beta}_M$. If the nonzero effect decreases with square root of the sample size, $\theta_j = O_p(n^{-1/2})$, then σ_θ^2 is

bounded. We actually require the nonzero effect to be small enough to ensure a small σ_θ^2 and thus good approximation by (7) and (8). This setting corresponds to the simulation studies in Tables 1 and 2. Alternatively, if θ_j is fixed and does not vary by the sample size, then σ_θ^2 grows with n, $\sigma_\theta^2 = O_p(n)$ which, depending on the magnitude of the fixed θ_j, will eventually violate the assumption required for our proposed method as n gets large. This is illustrated in simulation studies in Figure 2. For application we suggest checking the sample variance of the test statistics and to apply the proposed method if they are less than 1.5.

As shown in Section 3.1, the conventional test such as the normality-based test is not valid under $H_0^{(1)}$, and the joint significance test has a size α only under $H_0^{(2)}$ and $H_0^{(3)}$ with a very large nonzero α_S or β_M. With a single test we have no information about the proportion of $H_0^{(1)}$, $H_0^{(2)}$ and $H_0^{(3)}$. We are able to estimate the mixture proportion of null or apply our proposed method only if a collection of hypothesis tests for the similar purpose, for example, epigenetic mediation for the socioeconomic effect on adiposity, is available to us.

APPENDIX

Proof of Theorem 3.1. One can show that

$$
\begin{aligned}
p_{\text{NP1}} &= \int_{-\infty}^{0} \int_{0}^{\infty} \frac{1}{2\pi} e^{-\frac{(x-|a|)^2+(y-|b|)^2}{2}} \, dx \, dy \\
&\quad + \int_{0}^{\infty} \int_{-\infty}^{0} \frac{1}{2\pi} e^{-\frac{(x-|a|)^2+(y-|b|)^2}{2}} \, dx \, dy \\
&= \int_{-\infty}^{-|b|} \int_{-|a|}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \, dx \, dy + \int_{-|b|}^{\infty} \int_{-\infty}^{-|a|} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \, dx \, dy \\
&= \Phi(-|b|)\{1 - \Phi(-|a|)\} + \{1 - \Phi(-|b|)\}\Phi(-|a|) \\
&= \Phi(-|b|)\Phi(-|a|) + \Phi(-|b|)\{\Phi(|a|) - \Phi(-|a|)\} \\
&\quad + \{1 - \Phi(-|b|)\}\Phi(-|a|) \\
&= \Phi(-|a|) + \Phi(-|b|)\{\Phi(|a|) - \Phi(-|a|)\}.
\end{aligned}
$$

Similarly, once can express p_{NP1} as

$$
\Phi(-|b|) + \Phi(-|a|)\{\Phi(|b|) - \Phi(-|b|)\}.
$$

Therefore,

$$
p_{\text{NP1}} > \max\{\Phi(-|a|, \Phi(-|b|))\} = p_{\text{JT}}.
$$

The proof is based on one-sided tests, and the result for two-sided tests are identical by multiplying the above expressions by two.

Null hypothesis (1) can be re-expressed as

$$
H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_A : \theta \in \Theta_0^c,
$$

where $\theta = (\alpha_S, \beta_M)^T$, $\Theta_0 = \Theta_1 \cup \Theta_2$, and $\Theta_1 = \{\theta : \alpha_S = 0, \beta_M \in \Re\}$, $\Theta_2 = \{\theta : \alpha_S \in \Re, \beta_M = 0\}$. Intersection-union test (IUT) is a test that has a rejection region of the form $R = \bigcap_{k=1}^{K} R_k$ [Berger and Hsu (1996)]. The rejection region of the joint significance test on a and b can be expressed as $R = R_1 \cap R_2$, where R_1 and R_2 are the rejection regions for $H_{01} : \theta \in \Theta_1$ and $H_{02} : \theta \in \Theta_2$ respectively. Consequently, JT is an IUT, which has been shown the level α test, that is, $\sup_{\theta \in \Theta_0} \pi(\theta) \leq \alpha$ [Berger and Hsu (1996)], where $\pi(\theta)$ is the power function. We next apply Theorem 2 in Berger and Hsu (1996) to prove that the joint significance test is a size α test, that is, $\sup_{\theta \in \Theta_0} \pi(\theta) = \alpha$. To apply the theorem, one needs to show that there exists a sequence of parameters $\theta_l \in \Theta_i$ $(l = 1, \ldots, \infty)$ such that:

1. $\lim_{l \to \infty} P_{\theta_l}(\mathbf{Z} \in R_i) = \alpha$,
2. for $i' \neq i$, $\lim_{l \to \infty} P_{\theta_l}(\mathbf{Z} \in R_{i'}) = 1$,

where $\mathbf{Z}$ is the data. Suppose that $\theta_l = (\alpha_{Sl}, \beta_{Ml})^T = (0, \beta_{Ml})^T \in \Theta_1$ and $\lim_{l \to \infty} \beta_{Ml} = \infty$. It is straightforward that $\lim_{l \to \infty} P_{\theta_l}(\mathbf{Z} \in R_1) = \alpha$ and $\lim_{l \to \infty} P_{\theta_l}(\mathbf{Z} \in R_2) = 1$. Therefore, by Theorem 2 in Berger and Hsu (1996), the joint significance test is a size α test.

Proof of Theorem 3.3. The two-sided tail probability at $|ab|$ or more extreme is

$$F(ab) = 2 \int_{-\infty}^{\infty} \int_{|\frac{ab}{x}|}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \, dy \, dx = 2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left\{ 1 - \Phi\left(\left| \frac{ab}{x} \right| \right) \right\} dx.$$

By letting x be centered at 0 and y be centered at μ_b, we express this new normal product distribution $F(ab)$ as

$$F(ab; \mu_b) = 2 \int_{-\infty}^{\infty} \int_{|\frac{ab}{x}|}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+(y-\mu_b)^2}{2}} \, dy \, dx$$

$$= 2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left\{ 1 - \Phi\left(\left| \frac{ab}{x} \right| - \mu_b \right) \right\} dx.$$

Similarly, we let x be centered at μ_a and y be centered at 0 and obtain $F(ab; \mu_a) = 2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \{ 1 - \Phi(|\frac{ab}{y}| - \mu_a) \} \, dy$. p-value under the composite null hypothesis equals

$$(A.1) \qquad p_{\text{comp}} \equiv \pi_b \int_{-\infty}^{\infty} \frac{F(ab; \mu_b)}{\sqrt{2\pi} \sigma_b} e^{-\frac{\mu_b^2}{2\sigma_b^2}} \, d\mu_b + \pi_a \int_{-\infty}^{\infty} \frac{F(ab; \mu_a)}{\sqrt{2\pi} \sigma_a} e^{-\frac{\mu_a^2}{2\sigma_a^2}} \, d\mu_a$$

$$+ \pi_0 F(ab).$$

We express the integral of the first term in (A.1) $\int_{-\infty}^{\infty} \frac{F(ab; \mu_b)}{\sqrt{2\pi} \sigma_b} e^{-\frac{\mu_b^2}{2\sigma_b^2}} \, d\mu_b$ as

$\int_{-\infty}^{\infty} [2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \{ 1 - \Phi(|\frac{ab}{x}| - \mu_b) \} \, dx] \frac{1}{\sqrt{2\pi} \sigma_b} e^{-\frac{\mu_b^2}{2\sigma_b^2}} \, d\mu_b$ and it can be further

re-expressed as

$$2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \Phi\left(\mu_b - \left|\frac{ab}{x}\right|\right) \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{\mu_b^2}{2\sigma_b^2}} d\mu_b \right] \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$= 2 \int_{-\infty}^{\infty} \left[\Phi\left(-\frac{|ab/x|}{\sqrt{1+\sigma_b^2}}\right) \right] \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$= 2 \int_{-\infty}^{\infty} \left[\int_{\frac{|ab/x|}{\sqrt{1+\sigma_b^2}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \right] \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$= 2 \int_{-\infty}^{\infty} \int_{\frac{|ab/x|}{\sqrt{1+\sigma_b^2}}}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dy\, dx.$$

The last expression is $F(\frac{ab}{\sqrt{1+\sigma_b^2}})$. Similarly, one can show

$$\int_{-\infty}^{\infty} \frac{F(ab; \mu_a)}{\sqrt{2\pi}\sigma_a} e^{-\frac{\mu_a^2}{2\sigma_a^2}} d\mu_a = F\left(\frac{ab}{\sqrt{1+\sigma_a^2}}\right).$$

Plugging them into (A.1), we obtain

(A.2) $$p_{\text{comp}} = \pi_b F\left(\frac{ab}{\sqrt{1+\sigma_b^2}}\right) + \pi_a F\left(\frac{ab}{\sqrt{1+\sigma_a^2}}\right) + \pi_0 F(ab).$$

Note that unless we know π_0, π_a and π_b, we are not able to calculate a p-value using (A.2). We re-express (A.2) as

$$p_{\text{comp}} = \pi_b F\left(\frac{ab}{\sqrt{1+\sigma_b^2}}\right) + (1-\pi_b)F(ab)$$

$$+ \pi_a F\left(\frac{ab}{\sqrt{1+\sigma_a^2}}\right) + (1-\pi_a)F(ab) - F(ab)$$

$$= F\left(\frac{ab}{\sqrt{1+\pi_b\sigma_b^2}}\right) + F\left(\frac{ab}{\sqrt{1+\pi_a\sigma_a^2}}\right) - F(ab) + \delta_N,$$

where by Lemma 3.2, $|\delta_N| < \sum_{k=a,b} \frac{2}{\pi}(\frac{|ab|}{\sqrt{1+\pi_k\sigma_k^2}} - \frac{|ab|}{\sqrt{1+\sigma_k^{*2}}})K_0(\frac{|ab|}{\sqrt{1+\sigma_k^{*2}}})$, σ_a^{*2} and σ_b^{*2} satisfy the following two equalities: $F(\frac{ab}{\sqrt{1+\sigma_b^{*2}}}) = \pi_b F(\frac{ab}{\sqrt{1+\sigma_b^2}}) + (1-\pi_b)F(ab)$ and $F(\frac{ab}{\sqrt{1+\sigma_a^{*2}}}) = \pi_a F(\frac{ab}{\sqrt{1+\sigma_a^2}}) + (1-\pi_a)F(ab)$. Under the null, a is distributed as $N(0,1)$ with probability $1-\pi_a$ and as $N(\mu_a, 1)$ with probability π_a where $\mu_a \sim N(0, \sigma_a^2)$. The marginal variance of a is $(1-\pi_a) \times 1 +$

$\pi_a(\text{Var}(\text{E}(a|\mu_a)) + \text{E}(\text{Var}(a|\mu_a))) = 1 + \pi_a\sigma_a^2$, and the marginal variance of b is $1 + \pi_b\sigma_b^2$. Therefore, we obtain the following equality:

$$F\left(\frac{ab}{\sqrt{1+\pi_a\sigma_a^2}}\right) + F\left(\frac{ab}{\sqrt{1+\pi_b\sigma_b^2}}\right) - F(ab)$$

$$= F\left(\frac{ab}{\sqrt{\text{Var}(a)}}\right) + F\left(\frac{ab}{\sqrt{\text{Var}(b)}}\right) - F(ab).$$

Proof of Theorem 3.4. The p-value under the composite null hypothesis equals

(A.3)
$$p_{\text{comp}} \equiv \pi_b \int_{-\infty}^{\infty} F(ab; \mu_b)\, dQ_b(\mu_b) + \pi_a \int_{-\infty}^{\infty} F(ab; \mu_a)\, dQ_a(\mu_a)$$
$$+ \pi_0 F(ab),$$

where

$$\frac{dQ_a(\mu_a)}{d\mu_a} = \sum_{j=1}^{J_a} w_{aj} \frac{1}{\sqrt{2\pi}\sigma_{aj}} e^{-\mu_a^2/2\sigma_{aj}^2},$$

$$\frac{dQ_b(\mu_b)}{d\mu_b} = \sum_{j=1}^{J_b} w_{bj} \frac{1}{\sqrt{2\pi}\sigma_{bj}} e^{-\mu_b^2/2\sigma_{bj}^2}$$

and

$$\sum_{j=1}^{J_a} w_{aj} = \sum_{j=1}^{J_b} w_{bj} = 1.$$

We express the integral of the first term as

$$\int_{-\infty}^{\infty} F(ab; \mu_b) \sum_{j=1}^{J_b} w_{bj} \frac{1}{\sqrt{2\pi}\sigma_{bj}} e^{-\frac{\mu_b^2}{2\sigma_{bj}^2}}\, d\mu_b$$

$$= \sum_{j=1}^{J_b} w_{bj} \left[\int_{-\infty}^{\infty} F(ab; \mu_b) \frac{1}{\sqrt{2\pi}\sigma_{bj}} e^{-\frac{\mu_b^2}{2\sigma_{bj}^2}}\, d\mu_b\right]$$

$$= \sum_{j=1}^{J_b} w_{bj} F\left(\frac{ab}{\sqrt{1+\sigma_{bj}^2}}\right)$$

$$= F\left(\frac{ab}{\sqrt{1+\sigma_{Qb}^2}}\right) + \delta_{Q1b},$$

where $\sigma_{Qb}^2 = \sum_{j=1}^{J_b} w_{bj}\sigma_{bj}^2$, $|\delta_{Q1b}| < \frac{2}{\pi}(\frac{|ab|}{\sqrt{1+\sigma_{Qb}^2}} - \frac{|ab|}{\sqrt{1+\sigma_{Qb}^{*2}}})K_0(\frac{|ab|}{\sqrt{1+\sigma_{Qb}^{*2}}})$, and σ_{Qb}^{*2}

satisfies $\sum_{j=1}^{J_b} w_{bj} F(\frac{ab}{\sqrt{1+\sigma_{bj}^2}}) = F(\frac{ab}{\sqrt{1+\sigma_{Qb}^{*2}}})$. The last equality is by Lemma 3.2.

Similarly, we show

$$\int_{-\infty}^{\infty} F(ab; \mu_a)\, dQ_a(\mu_a) = F\left(\frac{ab}{\sqrt{1+\sigma_{Qa}^2}}\right) + \delta_{Q1a},$$

where $\sigma_{Qa}^2 = \sum_{j=1}^{J_a} w_{aj}\sigma_{aj}^2$ and $|\delta_{Q1a}| < \frac{2}{\pi}(\frac{|ab|}{\sqrt{1+\sigma_{Qa}^2}} - \frac{|ab|}{\sqrt{1+\sigma_{Qa}^{*2}}})K_0(\frac{|ab|}{\sqrt{1+\sigma_{Qa}^{*2}}})$, and

σ_{Qa}^{*2} satisfies the equality, $\sum_{j=1}^{J_a} w_{aj} F(\frac{ab}{\sqrt{1+\sigma_{aj}^2}}) = F(\frac{ab}{\sqrt{1+\sigma_{Qa}^{*2}}})$. It follows that

$p_{\text{comp}} - \pi_a\delta_{Q1a} - \pi_b\delta_{Q1b} = p_{\text{comp}} - \delta_{Q1}$ equals

$$\pi_a F\left(\frac{ab}{\sqrt{1+\sigma_{Qa}^2}}\right) + \pi_b F\left(\frac{ab}{\sqrt{1+\sigma_{Qb}^2}}\right) + \pi_0 F(ab)$$

$$= F\left(\frac{ab}{\sqrt{1+\pi_a\sigma_{Qa}^2}}\right) + F\left(\frac{ab}{\sqrt{1+\pi_b\sigma_{Qb}^2}}\right) - F(ab) + \delta_{Q2},$$

where by Lemma 3.2, $|\delta_{Q2}| < \sum_{k=a,b} \frac{2}{\pi}(\frac{|ab|}{\sqrt{1+\pi_k\sigma_{Qk}^2}} - \frac{|ab|}{\sqrt{1+\sigma_{Qk}^{*2}}})K_0(\frac{|ab|}{\sqrt{1+\sigma_{Qk}^{*2}}})$,

and σ_{Qa}^{*2} and σ_{Qb}^{*2} satisfy $F(\frac{ab}{\sqrt{1+\sigma_{Qa}^{*2}}}) = \pi_a F(\frac{ab}{\sqrt{1+\sigma_{Qa}^{*2}}}) + (1 - \pi_a) F(ab)$ and

$F(\frac{ab}{\sqrt{1+\sigma_{Qb}^{*2}}}) = \pi_b F(\frac{ab}{\sqrt{1+\sigma_{Qb}^{*2}}}) + (1 - \pi_b) F(ab)$. Under the null, variances of a and

b respectively are $1 + \pi_a\sigma_{Qa}^2$ and $1 + \pi_b\sigma_{Qb}^2$, which do not depend on the entire distribution of μ_a and μ_b, but only their first two moments. Therefore, we obtain

$$F\left(\frac{ab}{\sqrt{1+\pi_a\sigma_{Qa}^2}}\right) + F\left(\frac{ab}{\sqrt{1+\pi_b\sigma_{Qb}^2}}\right) - F(ab)$$

$$= F\left(\frac{ab}{\sqrt{\text{Var}(a)}}\right) + F\left(\frac{ab}{\sqrt{\text{Var}(b)}}\right) - F(ab).$$

Proof of Proposition 3.5.   We express the difference of p_{JT} and p_{comp}:

$$\frac{1}{4}(p_{\text{JT}} - p_{\text{comp}}) = \frac{1}{2}\Phi(-|a|) - \int_0^\infty \int_{\frac{|ab/x|}{\sqrt{1+\sigma^{*2}}}}^\infty \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}\, dy\, dx$$

$$= \int_{|a|}^\infty \int_0^{\frac{|ab/x|}{\sqrt{1+\sigma^{*2}}}} \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}\, dy\, dx$$

$$- \int_{\frac{|b|}{\sqrt{1+\sigma^{*2}}}}^\infty \int_{\frac{|ab/x|}{\sqrt{1+\sigma^{*2}}}}^{|a|} \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}\, dy\, dx$$

$$> \int_{|a|}^{\frac{|b|}{\sqrt{1+\sigma^{*2}}}} \int_{0}^{|a|} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \, dy \, dx$$

$$- \int_{\frac{|b|}{\sqrt{1+\sigma^{*2}}}}^{\infty} \int_{0}^{|a|} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \, dy \, dx$$

$$= \{\Phi(|a|) - \Phi(|0|)\}\left[\Phi\left(\frac{|b|}{\sqrt{1+\sigma^{*2}}}\right)\right.$$

$$\left. - \Phi(|a|) - \left\{1 - \Phi\left(\frac{|b|}{\sqrt{1+\sigma^{*2}}}\right)\right\}\right]$$

$$= \{\Phi(|a|) - \Phi(|0|)\}\left[2\Phi\left(\frac{|b|}{\sqrt{1+\sigma^{*2}}}\right) - \Phi(|a|) - 1\right].$$

The last expression is greater than 0 because of $|b| > \sqrt{1+\sigma^{*2}}\Phi^{-1}(\frac{1+\Phi(|a|)}{2})$.

SUPPLEMENTARY MATERIAL

Supplement to "Genome-wide analyses of sparse mediation effects under composite null hypotheses" (DOI: 10.1214/18-AOAS1181SUPP; .pdf). The supplementary material includes the proof of Lemma 3.2, additional simulation studies for very sparse signals, unbalanced $H_0^{(2)}$ and $H_0^{(3)}$, different number of hypothesis tests within a study and $\tilde{p}_{\text{comp}}$, and analysis results of chromosome 17 in the epigenomic study.

REFERENCES

AGHA, G., HOUSEMAN, E. A., KELSEY, K. T., EATON, C. B., BUKA, S. L. and LOUCKS, E. B. (2015). Adiposity is associated with DNA methylation profile in adipose tissue. *Int. J. Epidemiol.* **44** 1277–1287.

AROIAN, L. A. (1947). The probability function of the product of two normally distributed variables. *Ann. Math. Stat.* **18** 265–271. MR0021284

BARFIELD, R., SHEN, J., JUST, A. C., VOKONAS, P. S., SCHWARTZ, J., BACCARELLI, A. A., VANDERWEELE, T. J. and LIN, X. (2017). Testing for the indirect effect under the null for genome-wide mediation analyses. *Genet. Epidemiol.* **41** 824–833.

BARON, R. M. and KENNY, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical consideration. *J. Pers. Soc. Psychol.* **51** 1173–1182.

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

BERGER, R. L. and HSU, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statist. Sci.* **11** 283–319. MR1445984

BORGHOL, N., SUDERMAN, M., MCARDLE, W., RACINE, A., HALLETT, M., PEMBREY, M., HERTZMAN, C., POWER, C. and SZYF, M. (2012). Associations with early-life socioeconomic position in adult DNA methylation. *Int. J. Epidemiol.* **41** 62–74.

BULLOCK, J. G., GREEN, D. P. and HA, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *J. Pers. Soc. Psychol.* **98** 550–558.

DARMON, N. and DREWNOWSKI, A. (2008). Does social class predict diet quality? *Am. J. Clin. Nutr.* **87** 1107–1117.

DAVIS, S., DU, P., BILKE, S., TRICHE, T. J. and BOOTWALLA, M. (2015). methylumi: Handle Illumina methylation data. R package version 2.14.0.

GISKES, K., AVENDANO, M., BRUG, J. and KUNST, A. E. (2010). A systematic review of studies on socioeconomic inequalities in dietary intakes associated with weight gain and overweight/obesity conducted among European adults. *Obes. Rev.* **11** 413–429.

HUANG, Y.-T. (2015). Integrative modeling of multi-platform genomic data under the framework of mediation analysis. *Stat. Med.* **34** 162–178. MR3286246

HUANG, Y.-T (2019). Supplement to "Genome-wide analyses of sparse mediation effects under composite null hypotheses." DOI:10.1214/18-AOAS1181SUPP.

HUANG, Y.-T. and CAI, T. (2016). Mediation analysis for survival data using semiparametric probit models. *Biometrics* **72** 563–574. MR3515783

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

HUANG, Y. T., CHU, S., LOUCKS, E. B., LIN, C. L., EATON, C. B., BUKA, S. L. and KELSEY, K. T. (2016). Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood. *Epigenetics* **11** 227–236.

IMAI, K., KEELE, L. and YAMAMOTO, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statist. Sci.* **25** 51–71. MR2741814

LANGAAS, M., LINDQVIST, B. H. and FERKINGSTAD, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *J. Roy. Statist. Soc. Ser. B* **67** 555–572. MR2168204

LANGE, T. and HANSEN, J. V. (2011). Direct and indirect effects in a survival context. *Epidemiology* **22** 575–581.

LOFTUS, T. M., JAWORSKY, D. E., FREHYWOT, G. L., TOWNSEND, C. A., RONNETT, G. V., LANE, M. D. and KUHAJDA, F. P. (2000). Reduced food intake and body weight in mice treated with fatty acid synthase inhibitor. *Science* **288** 2379–2381.

LOPEZ-AYALA, J. M., ORTIZ-GENGA, M., GOMEZ-MILANES, I., LOPEZ-CUENCA, D., RUIZ-ESPEJO, F., SANCHEZ-MUNOZ, J. J., OLIVA-SANDOVAL, M. J., MONSERRAT, L. and GIMENO, J. R. (2014). A mutation in the Z-line Cypher/ZASP protein is associated with arrhythmogenic right ventricular cardiomyopathy. *Clin. Genet.* **88** 172–176.

LOUCKS, E. B., HUANG, Y. T., AGHA, G., CHU, S., EATON, C. B., GILMAN, S. E., BUKA, S. L. and KELSEY, K. T. (2016). Epigenetic mediators between childhood socioeconomic disadvantage and mid-life body mass index: The New England Family Study. *Psychosom. Med.* **78** 1053–1065.

MACKINNON, D. P. (2008). *Introduction to Statistical Mediation Analysis*. Taylor & Francis, New York.

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.

NIU, T., LIU, N., ZHAO, M., XIE, G., ZHANG, L., LI, J., PEI, Y. F., SHEN, H., FU, X., HE, H., LU, S., CHEN, X. D., TAN, L. J., YANG, T. L., GUO, Y., LEO, P. J., DUNCAN, E. L., SHEN, J., GUO, Y. F., NICHOLSON, G. C., PRINCE, R. L., EISMAN, J. A., JONES, G., SAMBROOK, P. N., HU, X., DAS, P. M., TIAN, Q., ZHU, X. Z., PAPASIAN, C. J., BROWN, M. A., UITTERLINDEN, A. G., WANG, Y. P., XIANG, S. and DENG, H. W. (2015). Identification of a

novel FGFRL1 microRNA target site polymorphism for bone mineral density in meta-analyses of genome-wide association studies. *Hum. Mol. Genet.* **24** 4710–4727.

PAN, W. C., WU, C. D., CHEN, M. J., HUANG, Y. T., CHEN, C. J., SU, H. J. and YANG, H. I. (2015). Fine particle pollution, alanine transaminase, and liver cancer: A Taiwanese prospective cohort study (REVEAL-HBV). *J. Natl. Cancer Inst.* **108** Art. ID djv341.

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.

ROBINS, J. M. and GREENLAND, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3** 143–155.

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152

SELCEN, D. and ENGEL, A. G. (2005). Mutations in ZASP define a novel form of muscular dystrophy in humans. *Ann. Neurol.* **57** 269–276.

SENESE, L. C., ALMEIDA, N. D., FATH, A. K., SMITH, B. T. and LOUCKS, E. B. (2009). Associations between childhood socioeconomic position and adulthood obesity. *Epidemiol. Rev.* **31** 21–51.

SOBEL, M. E. (1982). *Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models*. American Sociological Association, Washington, DC.

SONG, Y., HUANG, Y. T., SONG, Y., HEVENER, A. L., RYCKMAN, K. K., QI, L., LEBLANC, E. S., KAZLAUSKAITE, R., BRENNAN, K. M. and LIU, S. (2015). Birthweight, mediating biomarkers and the development of type 2 diabetes later in life: A prospective study of multi-ethnic women. *Diabetologia* **58** 1220–1230.

TCHETGEN TCHETGEN, E. J. (2011). On causal mediation analysis with a survival outcome. *Int. J. Biostat.* **7** Art. ID 33. MR2843528

TESCHENDORFF, A. E., MARABITA, F., LECHNER, M., BARTLETT, T., TEGNER, J., GOMEZ-CABRERO, D. and BECK, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29** 189–196.

VANDERWEELE, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology* **21** 540–551.

VANDERWEELE, T. J. (2011). Causal mediation analysis with survival data. *Epidemiology* **22** 582–585.

VANDERWEELE, T. J. (2013). A three-way decomposition of a total effect into direct, indirect, and interactive effects. *Epidemiology* **24** 224–232.

VANDERWEELE, T. J. and VANSTEELANDT, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Stat. Interface* **2** 457–468. MR2576399

VANDERWEELE, T. J. and VANSTEELANDT, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *Am. J. Epidemiol.* **172** 1339–1348.

ZHANG, H., ZHENG, Y., ZHANG, Z., GAO, T., JOYCE, B., YOON, G., ZHANG, W., SCHWARTZ, J., JUST, A., COLICINO, E., VOKONAS, P., ZHAO, L., LV, J., BACCARELLI, A., HOU, L. and LIU, L. (2016). Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics* **32** 3150–3154.

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
128 ACADEMIA ROAD, SECTION 2
TAIPEI 11529
TAIWAN
E-MAIL: ythuang@stat.sinica.edu.tw