# HT-MMIOW：一种用于微生物组中介作用的假设检验方法

# 采用逆优势加权

Yuka Moroishi[1,2], Zhigang Li[3], Juliette C. Madan[2,4], Hongzhe Li[5], Margaret R. Karagas[2], Jiang Gui[1*]

作者单位

[1]达特茅斯盖塞尔医学院生物医学数据科学系，美国新罕布什尔州汉诺威，邮编 03755

[2]达特茅斯盖塞尔医学院流行病学系，美国新罕布什尔州汉诺威，邮编 03755

[3]D佛罗里达大学生物统计学系，佛罗里达州盖恩斯维尔，邮编 32603，美国 A

[4]达特茅斯儿童医院儿科系，黎巴嫩，新罕布什尔州 03766，

USA

[5]宾夕法尼亚大学佩雷尔曼医学院生物统计学与流行病学系，宾夕法尼亚州费城，邮编 19104-6021，美国

* 致谁 通讯事宜请联系：Jiang Gui <Jiang.Gui@dartmouth.edu >.

摘要

人类微生物组在决定健康方面发挥着重要作用。中介分析可量化微生物组在暴露与疾病之间因果路径中的贡献；然而，现有中介模型无法充分捕捉微生物组数据的高维、相关性及组成型特征，通常也无法处理二分类结局。我们提出一种新方法，利用逆优势加权检验微生物组的中介效应。我们通过模拟证明，该方法在处理高维中介变量时具有更高的检验效能，且无需考虑暴露与中介变量之间交互作用的影响。我们将该方法应用于新罕布什尔州出生队列研究的婴儿肠道微生物组数据，发现6周龄婴儿肠道微生物组在母亲妊娠期间抗生素使用与儿童至5岁时过敏发病率之间发挥中介作用。

# 1. 引言

人类肠道微生物组与免疫系统相互作用，形成双向关系（Zheng 等，2020）；在婴儿期，发育中的肠道微生物组对免疫系统的成熟起着至关重要的作用（Gensollen 等，2016）。因此，肠道微生物组的扰动与婴儿期、儿童期乃至成年期的临床结局相关。分娩方式、饮食和抗生素使用等外部因素会影响肠道微生物组（Hasan 和 Yang，2019）。了解外部因素、肠道微生物组与临床结局三者之间的相互作用，有助于寻找干预机会，例如使用治疗药物和益生菌调节肠道微生物组，从而改善健康结局。

中介分析用于量化某个变量在暴露或治疗与结局之间因果路径中的贡献。这类分析在健康研究中尤其有用，可用于确定可能的干预措施，以预防疾病或改善健康结局。随着中介分析方法的发展，包括传统方法（Baron 和 Kenny，1986；MacKinnon 等，2002）和因果推断方法（Robins 和 Greenland，1992；Pearl，2001；Rubin，2005；VanderWeele 和 Vansteelandt，2010；Imai 和 Yamamoto，2013），近年来也发表了许多中介分析的扩展方法。其中包括针对多个中介变量的模型（Imai 和 Yamamoto，2013；VanderWeele 和 Vansteelandt，2014），以及考虑暴露与中介变量之间交互作用的模型（Valeri 和 Vanderweele，2013；VanderWeele，2014）。其他方法还包括逆优势比加权（IORW）（Tchetgen Tchetgen，2013）或逆优势加权（IOW）（Nguyen 等，2015），用于量化总效应、直接效应和间接效应。检验

中介效应可以采用常见方法进行检验，例如 Sobel检验（Sobel，1982）和联合显著性检验（MacKinnon 等，2002）。

针对高维中介变量的中介分析在生物标志物建模中日益受到关注。这类方法包括利用主成分分析和正则化进行降维的方法（Huang and Pan, 2016; Chén 等, 2018; Zhang, 2021; Zhao 等, 2020）、多重检验程序（Boca 等, 2014; Millstein 等, 2016; Sampson 等, 2018; Djordjilović 等, 2019; Liu 等, 2022; Dai 等, 2022）以及其他方法（Gao 等, 2019; Song 等, 2020）。中介分析方法也已扩展到微生物组数据：许多方法检验连续型结局的中介效应（Sohn and Li, 2019; Zhang 等, 2018; Wang 等, 2020; Zhang, Chen, Feng, 等, 2021; Zhang, Chen, Li, 等, 2021; Wu 等, 2021）；一些方法考虑暴露与中介变量之间的交互作用（Wang 等, 2020; Wu 等, 2021）；还有一种方法针对二分类结局开展中介分析（Sohn 等, 2022）。然而，这些方法大多需要指定微生物组模型；由于微生物组数据具有稀疏、高维和组成型特征，这一要求难以满足。此外，据我们所知，这些方法均无法同时处理二分类结局和高维微生物组中介变量。

我们提出一种新方法 HT-MMIOW，即基于 IOW 的微生物组中介假设检验方法。该方法利用 IOW 检验高维微生物组数据对连续结局和二分类结局的中介效应。它采用等距对数比变换（ilr）处理微生物组数据的组成型特征，并

利用均匀流形近似与投影（UMAP）进行降维。UMAP 生成的分量作为中介变量。我们通过置换检验确定整体微生物中介效应的统计显著性。模拟结果表明，当真实中介变量数量较多或间接效应较大时，这种新方法具有较高的检验效能。我们将该方法应用于婴儿肠道微生物组数据，以检测抗生素暴露与5岁前过敏诊断之间的中介效应。

# 2. 方法

## 2.1 可识别性

假设暴露、中介变量和结局分别表示为一个 $n \times 1$ 向量 $E$、一个$n \times p$ 矩阵 $\boldsymbol{M}$和一个 $n \times 1$ 向量 $Y$ 。已知 $Y$ 先于 $\boldsymbol{M}$发生，且 $\boldsymbol{M}$ 先于 $E$发生。微生物组矩阵 $\boldsymbol{M}$ 由 $n$中每个受试者的 $p$ 个分类单元的丰度构成，且 $\boldsymbol{M}$中有 $t$ 个真实中介变量。为简便起见，令 $E$ 和 $Y$ 均为二分变量。令 $X$ 为协变量矩阵。在 Baron 和 Kenny 针对单个中介变量 $M$提出的经典中介框架下，中介分析可由以下方程表示:

$$logit(\Pr(Y = 1 \mid e)) = a_0 + a_1 E$$

$$E[M|e] = b_0 + b_1 E$$

$$logit(\Pr(Y = 1 \mid e, m)) = c_0 + c_1 E + c_2 M$$

$E$ 对 $Y$ 的直接效应为 $c_1$，$E$ 通过 $M$ 对 $Y$ 产生的间接效应为 $b_1c_2$，总效应为 $a_1 = c_1 + b_1c_2$。Sobel检验通过假设 $H_0$： $b_1c_2 = 0$ 对比 $H_A$： $b_1c_2 \neq 0$，判断是否存在间接效应。

## 2.2 微生物组数据的降维

标准中介模型需要知道真实中介变量。虽然已有针对多个中介变量的中介模型，但其中许多模型假设中介变量彼此独立，或已知中介变量之间的因果关系。然而，微生物组数据具有高维、组成型和相关性等特点。我们不知道哪些微生物分类群是真实中介变量，也不知道分类单元之间的因果关系。我们提出将微生物组数据从 Aitchison单纯形映射到欧几里得空间，并进行降维，以获得规模更小的一组中介特征。

ilr是针对组成数据的对数比变换的扩展，拓展了传统的加性（alr）和中心化（clr）对数比变换。简而言之，alr是组成中某个值与参考值的对数比，而clr是某个值与组成中各值的几何平均数之比的对数。不过，这些方法存在局限：alr产生的变换无法保留距离，而clr产生的变换具有奇异协方差矩阵。采用ilr时，数据从$p$维Aitchison空间变换到$p-1$维欧几里得空间后，距离得以保留，所得向量彼此正交，且在分析中具有可解释性（Egozcue 等，2003年）。

ilr变换的每个值都是$\boldsymbol{M}$两个子集之间的“平衡”：位于平衡左侧的子集记作$R$，位于平衡右侧的子集记作$S$：

$$ilr(R,S) = \sqrt{\frac{rs}{r+s}}\log\left(\frac{g(m_R)}{g(m_S)}\right)$$

其中，$r$ 是$R$的基数，$s$ 是$S$的基数，$m_R$ 是$R$中的值，$m_S$ 是$S$中的值，而$g(\cdot)$是几何平均函数。

我们为零值填补伪计数 0.5，并对微生物组数据应用等距对数比变换，以处理组成性。完成等距对数比变换后，我们使用 UMAP 进行降维。简单来说，UMAP 是一种快速的降维程序，它通过寻找具有尽可能相近的模糊拓扑结构的低维投影，利用模糊拓扑结构对流形进行建模（McInnes 等，2020）。UMAP 支持非线性降维，并能有效分离各个簇。对$ilr(\boldsymbol{M})$应用 UMAP 后，微生物组数据会转换为一个$n \times c$ 矩阵$\boldsymbol{U}$，其中$c$ 是用户指定的成分数量。

2.3 计算间接效应并进行中介置换检验 IORW 和 IOW 方法用于因果中介分析，可将总效应分解为直接效应和间接效应。该方法通过加权程序消除了对多个中介变量与暴露之间关系进行回归建模的需求，因此能够处理多个中介变量。这种方法适用于微生物组数据，因为高维、

组成型且彼此相关的微生物组特征的联合条件密度难以建模。使用 IOW 方法，在给定协变量 $\boldsymbol{X}$ 的情况下，暴露 $E$ 对结局 $Y$ 的总效应可通过以下模型估计：

$$h(Y|E,\boldsymbol{X}) = \beta_0 + E\beta_1 + \boldsymbol{X}\beta_2 \qquad (1)$$

其中，$\beta_1$ 表示在给定$\boldsymbol{X}$的条件下，$E$ 对$Y$的总效应。h(·)是用户指定的连接函数，$\varepsilon$ 表示误差。我们采用 IOW 来概括$E$ 与中介变量$\boldsymbol{U}$之间的关联，并以$\boldsymbol{X}$为条件。权重可通过采用用户指定连接函数$j$(·)的模型进行估计：

$$j(E|\boldsymbol{U},\boldsymbol{X}) = \alpha_0 + \boldsymbol{U}\alpha_1 + \boldsymbol{X}\alpha_2 \qquad (2)$$

对于每个观测值，暴露组的权重为预测优势的倒数，未暴露组的权重为 1。与 IORW 相比，使用 IOW 可稳定权重，但可能会给估计值带来轻微偏倚（Nguyen 等，2015）。然后，我们使用加权回归模型估计以 . 为条件时$E$ 对$Y$的直接效应。

$$h(WY|WE,W\boldsymbol{X}) = \gamma_0 W + WE\gamma_1 + W\boldsymbol{X}\gamma_2 \qquad (3)$$

其中，$\gamma_1$ 表示在给定$\boldsymbol{X}$的情况下，$E$ 对$Y$的直接效应，$W$ 是一个$n \times 1$ 权重向量。由于采用了这一加权程序，我们无需指定$E$ 与$\boldsymbol{U}$之间的交互作用。

现在，我们可以利用（1）和（3）中的参数估计间接效应，并将其作为置换检验的观测检验统计量：$T_{obs} = \beta_1 - \gamma_1$。微生物组不存在中介效应的零假设表示为：

$$H_0 : \beta_1 - \gamma_1 = 0,$$

以及备择　　假设微生物组存在中介效应可表示为　　：

$$H_A : \beta_1 - \gamma_1 \neq 0.$$

我们使用置换检验，并根据公式 j∑　$I(\mathrm{m}_{j=1}^{B}$　$T^{(i)}$ >m|$T_{obs}$|s)/$B$ 计算 P 值，其中 B 是置换总次数，$I(\cdot)$ 是示性函数，$T^{(j)}$ 是第 $j = 1$ 次置换下零假设对应的检验统计量，...，$B$。

# 3. 模拟

## 3.1 模拟概述

我们通过模拟评估了所提出方法在连续结局和二分类结局上的检验效能。我们使用 SparseDOSSA（Ma 等，2021）模拟了 $E$ 和 $\boldsymbol{M}$。简而言之，该工具使用零膨胀对数正态分布，模拟真实的人类粪便微生物群落结构，以及与模拟的微生物组数据相关的协变量。我们对 $\boldsymbol{M}$ 应用 UMAP，将维度降至 2 个分量。使用 SparseDOSSA 进行模拟数据生成时，我们将与暴露相关的微生物特征比例设为 50%，并将暴露对中介变量的效应量设为 3。我们使用标准逻辑回归模拟结局 $Y$，模型纳入暴露以及经缩放的真实微生物分类群中介变量相对丰度；这些中介变量从与暴露相关的分类群中随机选取，其中

暴露对结局的效应量设为 5，随机噪声为 $N(0,1)$。暴露变量为二分类变量，结局变量为连续变量和二分类变量。对于连续结局和二分类结局，我们还在零假设下评估了不同样本量（$n$ = 50、70、100、150、300、500）和分类单元数量（$p$ = n、2n）时的 I 类错误。

为进一步优化我们的方法，我们评估了 HT-MMIOW 在采用各种降维技术时的表现，包括：1）前文所述的 UMAP；2）先用 PCA 降至 n 个主成分，再进行 UMAP；3）使用能够解释 100% 方差的 PCA 主成分；以及 4）使用能够解释 80% 方差的 PCA 主成分。我们考察了真实中介变量对结局的不同效应量（效应量为 = 0.5、1、5，分别代表较小到较大的间接效应），以及微生物组样本中真实中介变量数量的变化（$t$ = 1、5、10）。我们还改变了样本量（$n$ = 50、70、100、150、300、500）和微生物组数据中的分类单元数量（$p$ = n、2n）。我们将 HT-MMIOW 的表现与采用 Bray-Curtis 和 Jaccard距离的基于距离的总体检验进行了比较（Zhang 等人，2018）。尽管总体检验是为连续型结局设计的，但由于缺乏针对高维微生物组中介变量和二分类结局的方法，我们仍考察了该检验在二分类结局中的表现。所有分析均使用 R 4.0.2 版以及“SparseDOSSA2”、“magrittr”、“dplyr”、“ggplot2”、“compositions”、“umap”、“foreach”、“doParallel”和“bda”程序包完成。

## 3.2 模拟结果

图1和图2展示了在不同条件下，HT-MMIOW针对连续型结局的模拟结果，并将其与总体检验进行了比较。当分类单元数量是样本量的两倍时，采用 UMAP降维的 HT-MMIOW 优于采用 PCA 和 UMAP 的 HT-MMIOW。对于所有条件，HT-MMIOW 的表现也优于基于距离的总体检验，尤其是在真实中介变量数量较少时。仅采用 PCA 作为降维技术的 HT-MMIOW 表现不佳。采用能够解释 80% 方差的 PCA 主成分的 HT-MMIOW 通常优于采用能够解释 100% 方差的 PCA 主成分的 HT-MMIOW；这可能是因为在逆优势比中介分析方法中，使用较少的主成分可提高检验效能。这些功效计算基于 0.05 的经验阈值。尽管根据 0.05 的经验阈值，连续型结局的 I类错误约为 0.05，但根据 0.01 经验阈值得到的 I类错误更小（补充图1）。使用 0.01 经验阈值的 HT-MMIOW，其检验效能仍高于总体距离检验（补充图2）。

对于二分类结局，当中介变量对 Y 的效应更强，且 M 中真实中介变量的数量更多时，HT-MMIOW 表现良好（图3、图4）。使用 UMAP 的表现与使用 PCA 和 UMAP 一样好。在所有条件下，HT-MMIOW 的表现也优于基于距离的总体检验。同样，仅使用 PCA 作为降维技术时，HT-MMIOW 的表现较差。正如预期，随着样本量增加，连续型和二分类结局的表现均有所提升。同样，以上功效计算基于 0.05 的经验阈值。对于

二分类结局，使用 0.05 的经验阈值时，I类错误膨胀；然而，基于 0.01 经验阈值的 I类错误接近于零（补充图1）。对于二分类结局，使用 0.01 经验阈值的 HT-MMIOW 检验效能高于总体距离检验（补充图2）。

# 4. NHBCS数据的应用

产前抗生素使用与婴儿期和儿童期过敏的发生有关（Baron等, 2020），也与婴儿肠道微生物组的组成差异相关（Dierikx 等, 2020; Coker 等, 2020）。我们利用新罕布什尔州出生队列研究（NHBCS）的数据，应用 HT-MMIOW 检验婴儿肠道微生物组作为产前抗生素使用与过敏之间因果路径中潜在中介变量的作用。NHBCS 是一项前瞻性出生队列研究，纳入在新罕布什尔州诊所接受产前护理的母婴配对。该研究招募了年龄在 18 至 45 岁之间、单胎妊娠且使用私人供水系统的孕妇，具体情况此前已有描述（Gilbert-Diamond 等, 2011）。产前抗生素使用情况记录在产前记录中。婴儿约 4、8、12 和 18 月龄时，以及此后每隔一年进行的电话访谈中，儿童的照护者报告儿童是否曾被医生诊断患有任何过敏。婴儿约 6 周龄时采集粪便样本，并在伍兹霍尔的海洋生物学实验室使用 Illumina MiSeq（Illumina，美国加利福尼亚州圣迭戈）对细菌 16S rRNA基因测序的 V4-V5高变区进行测序。我们推断出扩增子序列变体

使用 DADA2（Callahan 等, 2016）进行推断，并使用 SILVA 数据库进行分类学注释（Pruesse等, 2007）。

本分析共纳入 412 对母婴配对。其中，72 人（17.5%）自我报告在妊娠期间使用过抗生素，39 名儿童（9.5%）在 5 岁前被诊断出患有过敏。根据 6 周龄粪便样本的 16S 测序读段，我们鉴定出 705 个不同的细菌属。在中介检验中，HT-MMIOW 得出的 P 值为 0.008；这为我们提供了拒绝零假设的证据，并表明婴儿肠道微生物组在妊娠期间抗生素使用与过敏发病率之间起中介作用。由于随访时长各异，我们在模型中调整了首次确诊过敏时的年龄或末次随访时的年龄。

# 5. 讨论

我们提出了一种用于检验微生物组中介效应的新型假设检验方法，该方法结合了微生物组数据的降维程序和基于 IOW 的中介分析程序，以检验是否存在间接效应。我们在不同条件下设置了模拟情景，以评估该方法的检验效能。当中介变量对微生物组的效应较大，或微生物组数据中的真实中介变量数量较多时，HT-MMIOW 的检验效能很高。与传统假设检验相比，HT-MMIOW 对连续型和二分类结局的表现通常更好。我们的方法是少数能够针对高维微生物组中介变量和二分类结局检验间接效应的假设检验方法之一。

我们提出的假设检验方法在应用上具有灵活性。HT-MMIOW 可用于多种类型的暴露数据和结局数据；用户可以在回归模型中指定自己的链接函数。通过将中心化对数比变换步骤替换为用户选择的变换方法，我们的方法还可以调整以适用于其他类型的高维中介分析，包括基因组学。总效应、直接效应和间接效应均可通过 IOW 方法计算，置信区间则可通过自助法估计。

我们方法的主要优势在于，HT-MMIOW 将高维微生物组数据简化为少数能够代表微生物群落结构的独立成分。IOW程序可在中介模型中纳入多个中介变量。此外，IOW框架无需指定一个模型来对暴露与多个组成型中介变量的联合条件密度进行回归。另一个优势是，即使暴露与中介变量之间存在交互作用，IOW也不要求我们对其进行指定。

尽管 HT-MMIOW 具有诸多优势，但也存在一些局限性。我们的方法有一项局限：真实微生物中介变量的中介效应必须足够大，才能检测到间接效应。我们还必须假设不存在未测量混杂。此外，当中介效应较小且样本量较小时，真实中介变量的数量必须足够多。我们的方法也无法检测特定微生物或微生物簇的中介效应

。尽管如此，这种方法仍是判断是否存在中介效应的重要第一步。未来仍需进一步研究，以识别驱动这一效应的关键微生物特征和交互作用。


## 致谢

我们感谢新罕布什尔州出生队列研究的所有参与者和工作人员。我们还感谢 Hilary G. Morrison 处理婴儿粪便样本。

## 经费

本研究获得美国国立卫生研究院的资助，资助编号为 UH3OD023275、美国国家环境健康科学研究所 P01ES022832 和美国国家普通医学科学研究所 P20GM104416；此外，还获得美国环境保护署的资助，资助编号为 RD83544201。


## 作者贡献

Y.M.、M.R.K. 和 J.G. 构思了本研究。Y.M. 开发了统计方法、进行了模拟，并实现了 R 软件包。Z.L.、H.L. 和 J.G. 为方法的开发作出了贡献。M.R.K. 和 J.C.M. 参与了数据采集和处理。Y.M. 撰写了稿件初稿。所有作者均审阅并编辑了稿件。

所有作者均声明不存在利益冲突。

## 数据和代码可用性

本研究使用的微生物组数据可通过登录号 PRJNA296814 在在线数据库 http://www.ncbi.nlm.nih.gov/sra中获取。代码可按需提供。HT-MMIOW 的 R 软件包可在 GitHub 上获取：~~https://github.com/yukamoro~~/HTMMIOW

# 参考文献


Baron,R. 等（2020）产前抗生素暴露和婴儿抗生素用药与儿童过敏之间的关系：系统综述。BMC 儿科，20，312。Baron,R.M. 和 Kenny,D.A.（1986）社会心理学研究中调节变量与中介变量的区分：概念、策略和统计学考量。《人格与社会心理学杂志》，51，1173–1182。Boca,S.M. 等（2014）同时检验多个生物学中介变量。《生物信息学》，30，214–220。Callahan,B.J. 等（2016）DADA2：基于 Illumina 扩增子数据进行高分辨率样本推断。《自然·方法》，13，581–583。Chén,O.Y. 等（2018）高维多变量中介分析及其在神经影像数据中的应用。《生物统计学》，19，121–136。Coker,M. 等（2020）特定类别的产时抗生素与婴儿肠道微生物群成熟相关：一项前瞻性队列研究。BJOG，127，217–227。Dai,J.Y. 等（2022）针对高维中介假设的多重检验程序。Journal of the American Statistical Association，117，198–213。Dierikx,T.H. 等（2020）产前和产时抗生素对婴儿肠道微生物群定植的影响：系统综述。《感染杂志》，81，190–204。Djordjilović,V. 等（2019）高维中介分析的全局检验：检验潜在中介变量组。《医学统计学》，38，3346–3360。Egozcue,J.J. 等（2003）用于组成数据分析的等距对数比变换。《数学地质学》，35，279–300。Gao,Y. 等（2019）检验高维表观遗传学研究中的中介效应。《遗传学前沿》，10。Gensollen,T. 等（2016）生命早期微生物群定植如何塑造免疫系统。Science，352，539–544。Gilbert-Diamond,D. 等（2011）大米摄入导致美国女性砷暴露。Proc Natl Acad Sci U S A，108，20656–20660。Hasan,N. 和 Yang,H.（2019）影响肠道微生物群组成及其调节的因素。PeerJ，7，e7502。Huang,Y.-T. 和 Pan,W.-C.（2016）高维连续中介变量因果中介模型中的中介效应假设检验。Biometrics，72，402–413。

Imai,K. 和 Yamamoto,T.（2013）多重因果机制的可识别性与敏感性分析：重新审视框架实验的证据。Political Analysis，21，141–171。Liu,Z. 等（2022）用于因果中介效应的大规模假设检验及其在全基因组表观遗传学研究中的应用。Journal of the American Statistical Association，117，67–81。Ma,S. 等（2021）用于描述和模拟微生物群落谱的统计模型。PLOS Computational Biology，17，e1008913。MacKinnon,D.P. 等（2002）检验中介效应及其他中介变量效应的方法比较。Psychol Methods，7，83–104。McInnes,L. 等（2020）UMAP：用于降维的均匀流形近似与投影。arXiv:1802.*03426 [*cs, stat*]*。Millstein,J. 等（2016）cit：用于基因组应用中介分析的假设检验软件。《生物信息学》，32，2364–2365。Nguyen,Q.C. 等（2015）使用逆优势比加权进行多中介变量中介分析的实用指南。American Journal of Epidemiology，181，349–356。Pearl,J.（2001）直接效应和间接效应。载于《第十七届人工智能不确定性会议论文集》，UAI’01。Morgan Kaufmann Publishers Inc.，旧金山，加利福尼亚州，美国，第411–420页。Pruesse,E. 等（2007）SILVA：与 ARB 兼容、经过质量检查并完成比对的核糖体RNA序列数据综合在线资源。Nucleic Acids Res，35，7188–7196。Robins,J.M. 和 Greenland,S.（1992）直接效应和间接效应的可识别性与可交换性。Epidemiology，3，143–155。Rubin,D.B.（2005）使用潜在结果进行因果推断。Journal of the American Statistical Association，100，322–331。Sampson,J.N. 等（2018）检验多个中介变量时的家族错误率和假发现率控制。《生物信息学》，34，2418–2424。Sobel,M.E.（1982）结构方程模型中间接效应的渐近置信区间。Sociological Methodology，13，290–312。Sohn,M.B. 等（2022）用于二元结局的组成型中介模型：在微生物组研究中的应用。《生物信息学》，38，16–21。Sohn,M.B. 和 Li,H.（2019）微生物组研究中的组成数据中介分析。The Annals of Applied Statistics，13，661–681。Song,Y. 等（2020）组学研究中高维因果中介效应的贝叶斯收缩估计。Biometrics，76，700–710。Tchetgen Tchetgen,E.J.（2013）用于因果中介分析的逆优势比加权估计。Stat Med，32，4567–4580。Valeri,L. 和 Vanderweele,T.J.（2013）允许存在暴露-中介变量交互作用并进行因果解释的中介分析：理论假设及使用 SAS 和 SPSS宏的实现。Psychol Methods，18，137–150。VanderWeele,T. 和 Vansteelandt,S.（2014）多个中介变量的中介分析。Epidemiologic Methods，2，95–115。

VanderWeele,T.J. (2014) 中介与交互作用的统一：四向分解。*Epidemiology*, **25**, 749–761. VanderWeele,T.J. and Vansteelandt,S. (2010) 二分类结局中介分析的优势比。*American Journal of Epidemiology*, **172**, 1339–1348. Wang,C. *et al.* (2020) 利用高维组成型微生物组数据估计和检验微生物因果中介效应。《生物信息学》, **36**, 347–355. Wu,Q. *et al.* (2021) MedZIM：针对零膨胀中介变量的中介分析及其在微生物组数据中的应用。*arXiv:1906.09175 [stat]*. Zhang,H., Chen,J., Feng,Y., *et al.* (2021) 高维组成型微生物组数据中的中介效应筛选。《医学统计学》, **40**, 885–896. Zhang,H., Chen,J., Li,Z., *et al.* (2021) 中介效应检验及其在人类微生物组数据中的应用。《统计生物科学》, **13**, 313–328. Zhang,J. *et al.* (2018) 一种检验人类微生物组中介效应的基于距离的方法。《生物信息学》, **34**, 1875–1883. Zhang,Q. (2021) 高维中介分析及其在因果基因识别中的应用。《统计生物科学》. Zhao,Y. *et al.* (2020) 基于稀疏主成分的高维中介分析。《计算统计学与数据分析》, **142**, 106835. Zheng,D. *et al.* (2020) 健康与疾病中的微生物群与免疫交互作用。《细胞研究》, **30**, 492–506.

图

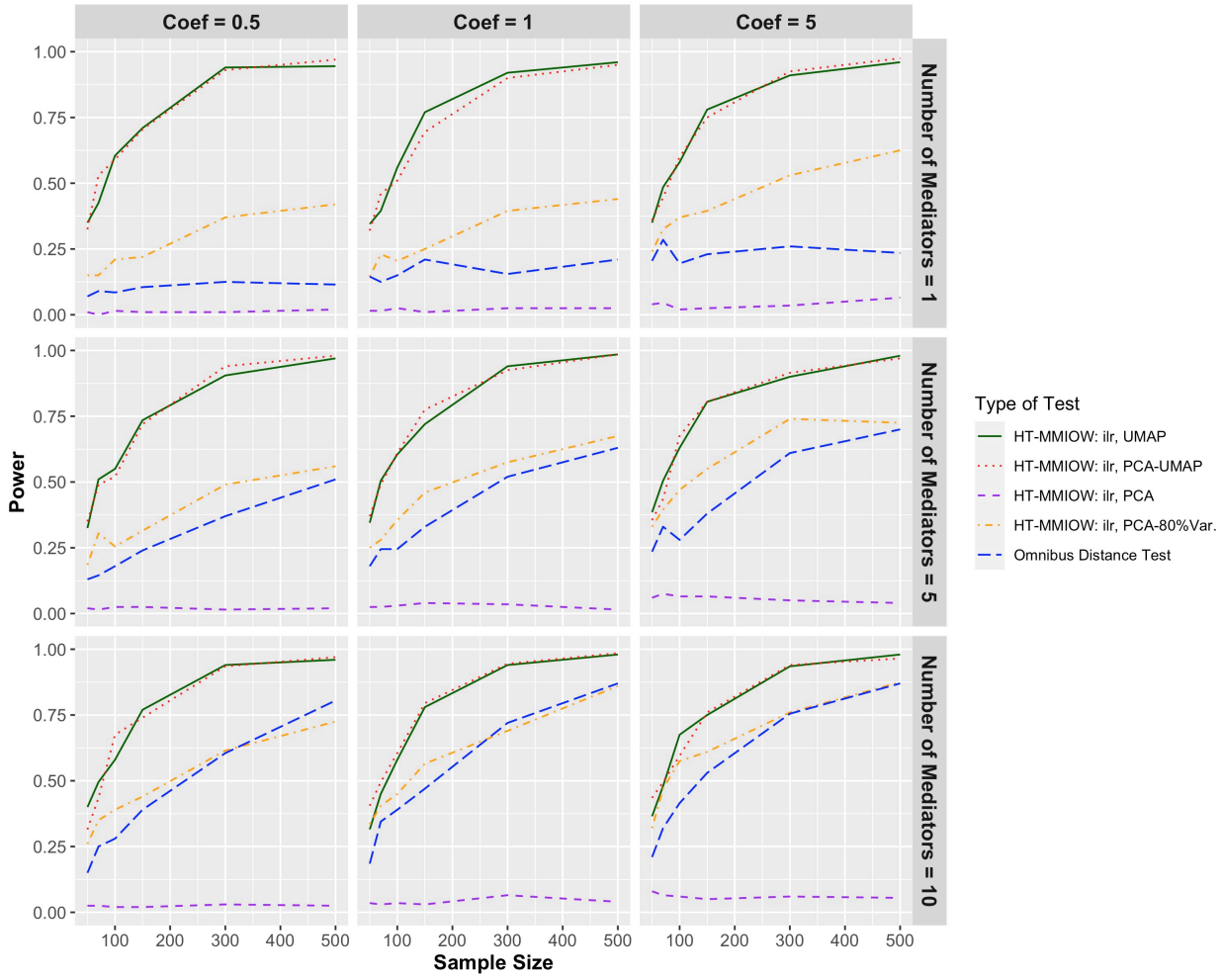



图 1. 不同样本量、效应量和真实中介变量数量下，针对连续型结局和 $p = n$ 的功效计算。结果基于 200次模拟。每条线代表所评估的假设检验程序。

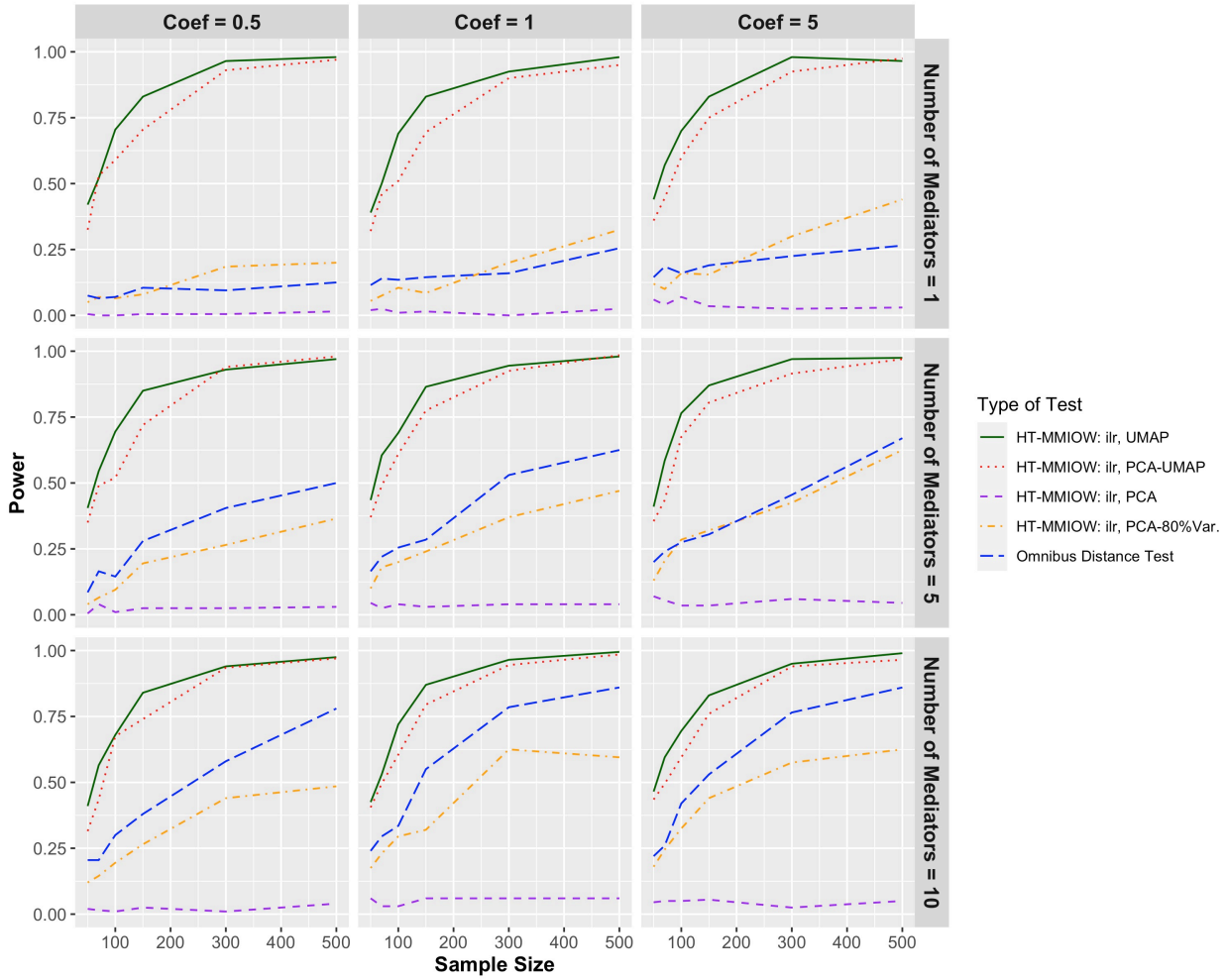



图 2. 连续型结局和 $p = 2n$ 下，不同样本量、效应量及真实中介变量数量的功效计算。结果基于 200次模拟。每条线代表所评估的假设检验程序。

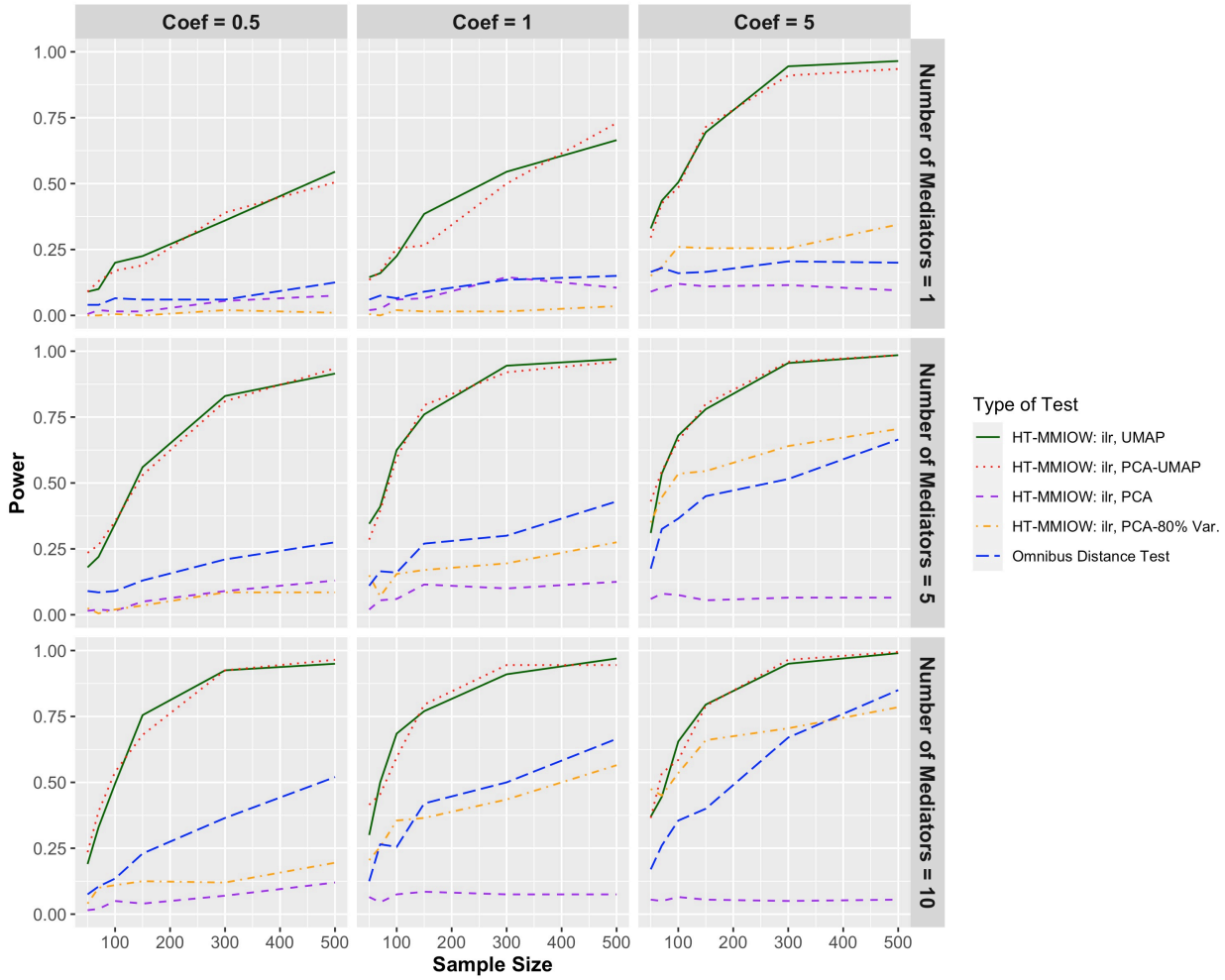



图3. 二分类结局和$p = n$在不同样本量、效应量及真实中介变量数量下的功效计算。结果基于200次模拟。每条线代表所评估的假设检验程序。

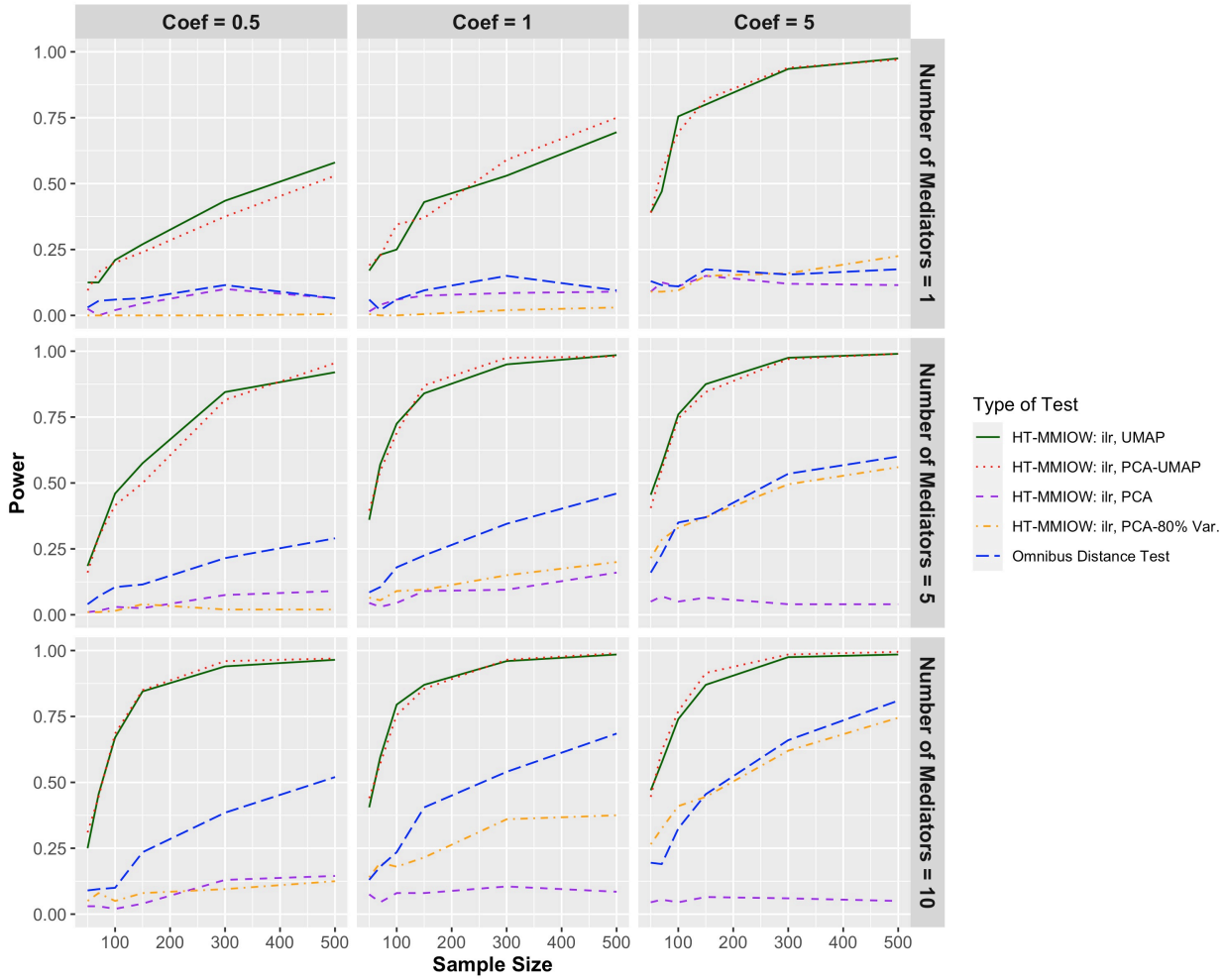
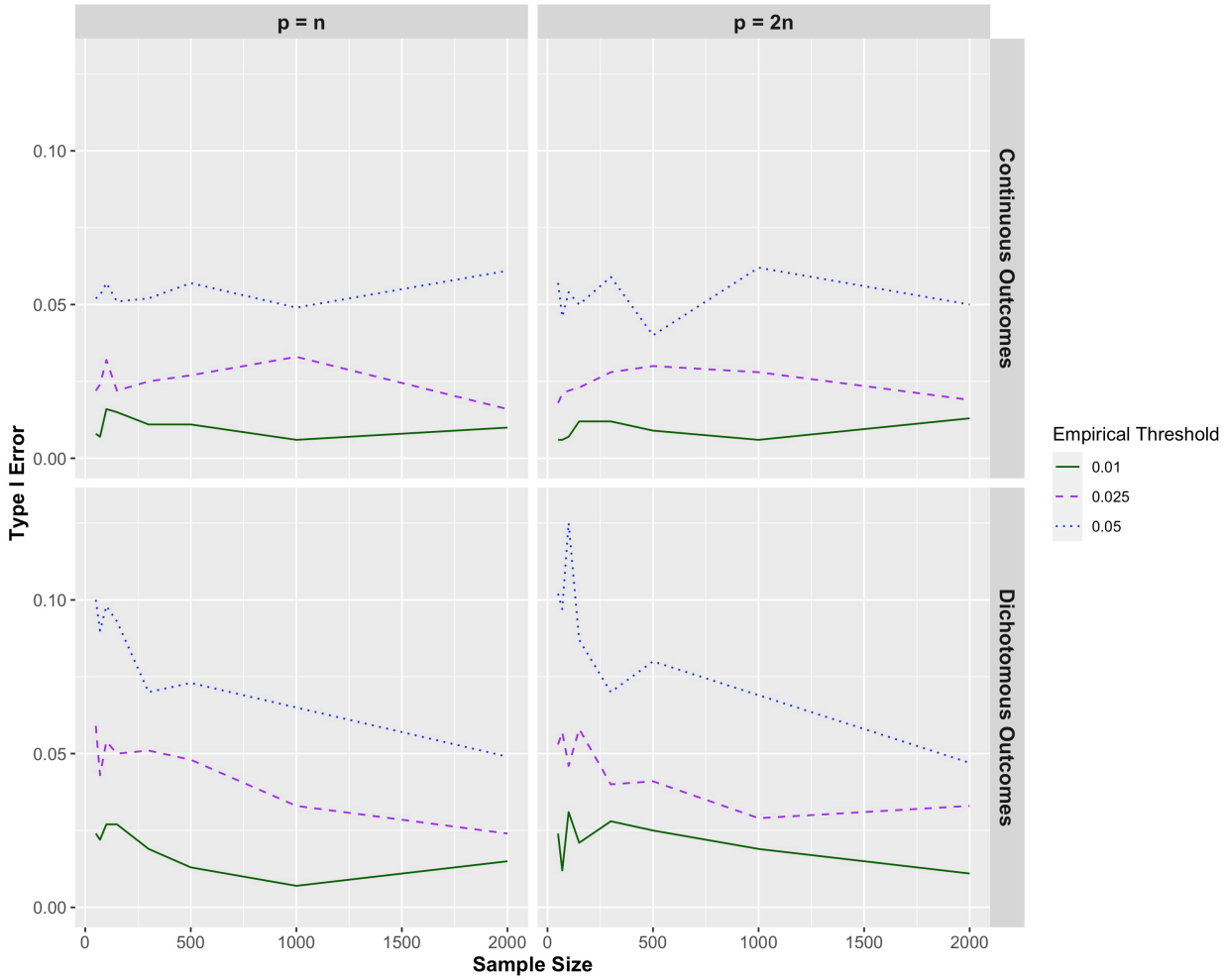
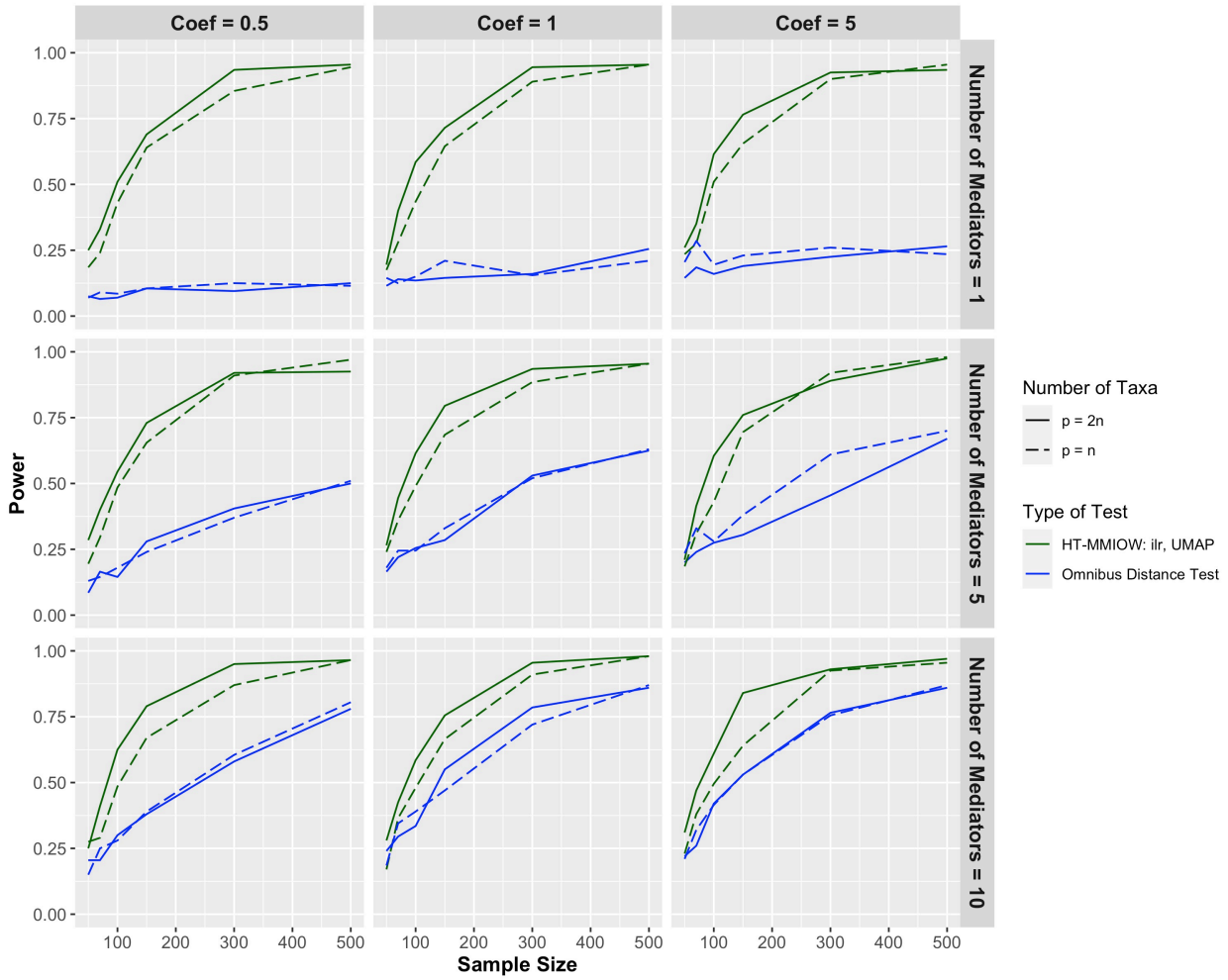



图4. 二分类结局和 $p = 2n$ 下不同样本量、效应量和真实中介变量数量的功效计算。结果基于200次模拟。每条线代表所评估的假设检验程序。

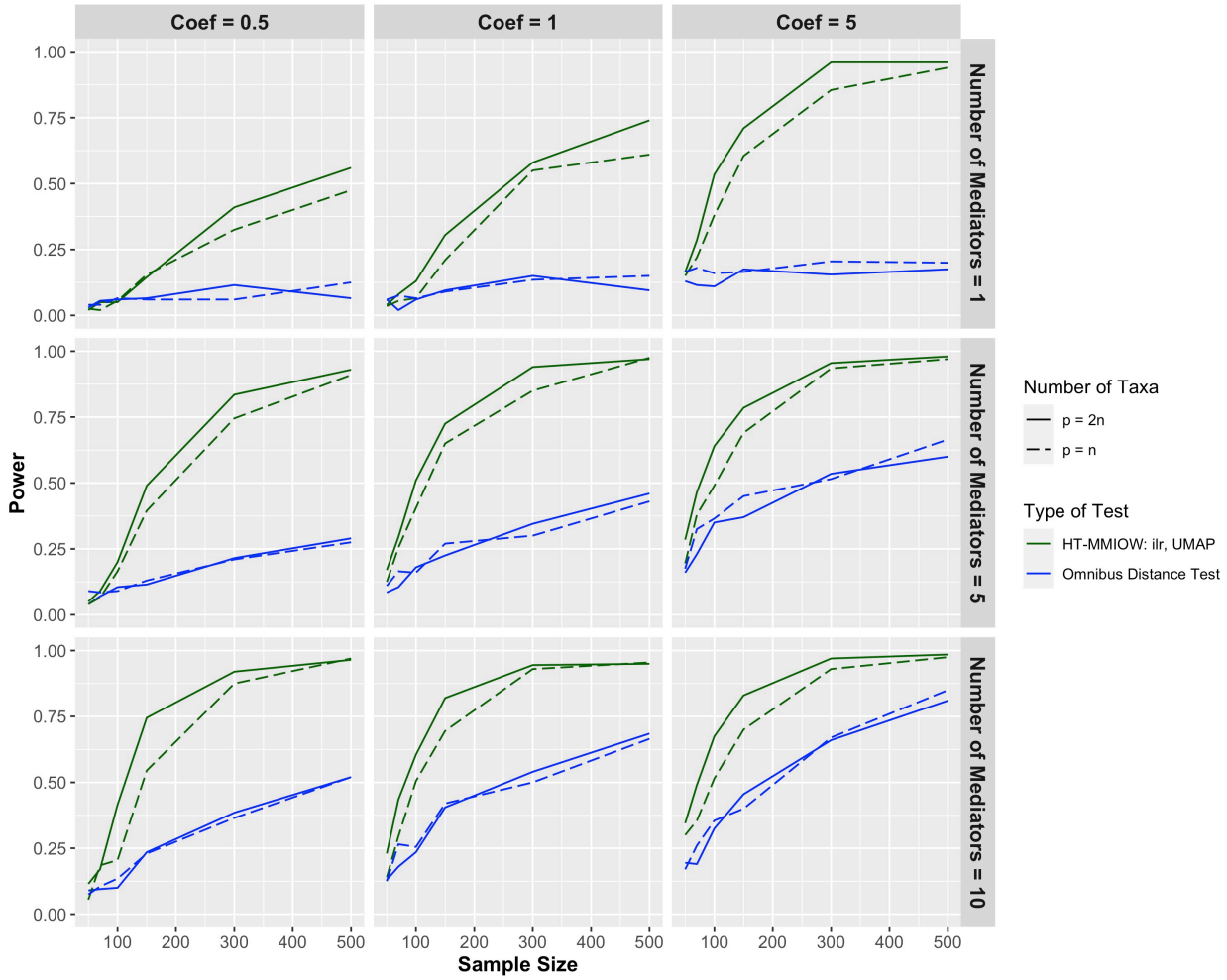
补充材料




补充图1. 不同样本量、分类单元数量和结局类型下的I类错误计算。结果基于1000次模拟。每条线表示不同经验阈值下的I类错误。




补充图2. 在连续型结局下，采用0.01阈值和总体距离检验的 HT-MMIOW 功效计算，展示样本量、效应量、真实中介变量数量和分类单元数量变化时的结果。结果基于200次模拟。实线表示$p = 2n$，虚线表示$p = n$。绿色线表示采用0.01经验阈值的 HT-MMIOW，蓝色线表示总体距离检验。




补充图3。针对二分类结局，采用0.01阈值和总体距离检验的HT-MMIOW功效计算，展示样本量、效应量、真实中介变量数量和分类单元数量变化时的结果。结果基于200次模拟。实线表示$p = 2n$，虚线表示$p = n$。绿色线表示采用0.01经验阈值的HT-MMIOW，蓝色线表示总体距离检验。