Causal mediation analysis in presence of multiple mediators uncausally related

Allan Jérolon, Laura Baglietto, Etienne Birmelé, Flora Alarcon, Vittorio Perduca

Allan Jérolon*, Laura Baglietto, Etienne Birmelé, Flora Alarcon and Vittorio Perduca

Causal mediation analysis in presence of multiple mediators uncausally related

https://doi.org/10.1515/ijb-2019-0088
Received September 5, 2019; accepted August 6, 2020; published online September 30, 2020

Abstract: Mediation analysis aims at disentangling the effects of a treatment on an outcome through alternative causal mechanisms and has become a popular practice in biomedical and social science applications. The causal framework based on counterfactuals is currently the standard approach to mediation, with important methodological advances introduced in the literature in the last decade, especially for simple mediation, that is with one mediator at the time. Among a variety of alternative approaches, Imai et al. showed theoretical results and developed an R package to deal with simple mediation as well as with multiple mediation involving multiple mediators conditionally independent given the treatment and baseline covariates. This approach does not allow to consider the often encountered situation in which an unobserved common cause induces a spurious correlation between the mediators. In this context, which we refer to as mediation with uncausally related mediators, we show that, under appropriate hypothesis, the natural direct and joint indirect effects are non-parametrically identifiable. Moreover, we adopt the quasi-Bayesian algorithm developed by Imai et al. and propose a procedure based on the simulation of counterfactual distributions to estimate not only the direct and joint indirect effects but also the indirect effects through individual mediators. We study the properties of the proposed estimators through simulations. As an illustration, we apply our method on a real data set from a large cohort to assess the effect of hormone replacement treatment on breast cancer risk through three mediators, namely dense mammographic area, nondense area and body mass index.

Keywords: correlated mediators; direct and indirect effects; independent mediators; multiple mediators; simulation of counterfactuals.

1 Introduction

Causal mediation analysis comprises statistical methods to study the mechanisms underlying the relationships between a cause, an outcome and a set of intermediate variables. This approach has become increasingly popular in various domains such as biostatistics, epidemiology and social sciences. Mediation analysis applies to the situation depicted by the causal directed acyclic graph of Figure 1, where an exposure (or treatment) T affects an outcome Y either directly or through one or more intermediate variables referred to as *mediators*. The aim of the analysis is to assess the total causal effect of T on Y by decomposing it into a *direct* effect and an *indirect* effect through the mediator(s).

Mediation analysis originally developed within the setting of linear structural equation modelling (LSEM) [1–3]. Following the seminal works by Robins and Greenland [4] and Pearl [5], a formal framework based on counterfactual variables established itself as the standard approach to mediation analysis, with a growing methodological literature, see for instance [6–9] and the comprehensive book [10].

The authors Flora Alarcon and Vittorio Perduca shared last autorship.

***Corresponding author: Allan Jérolon**, Laboratoire MAP5 (UMR CNRS 8145), Université de Paris, Paris, Île-de-France, France, E-mail: allan.jerolon@parisdescartes.fr
Laura Baglietto, Department of Clinical and Experimental Medicine, Università di Pisa, Pisa, Italy
Etienne Birmelé, Flora Alarcon and Vittorio Perduca, Laboratoire MAP5 (UMR CNRS 8145), Université de Paris, Paris, Île-de-France, France

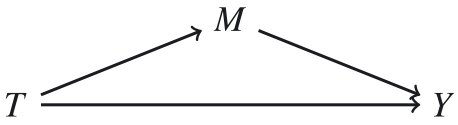


Figure 1: Simple mediation model with one mediator M and no confounding covariates.

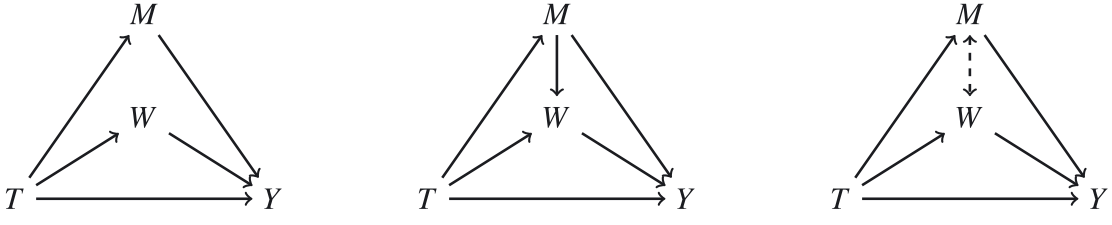
In this work, we adopt the point of view and formalism of [11] and [12], who put forward a general approach based on counterfactuals to define, identify and estimate causal mediation effects without assuming any specific statistical model in the particular case of a single mediator. The theoretical results in these articles are based on a strong set of assumptions known as *Sequential Ignorability*. These conditions are interpreted as the requirement that there must be no confounding of the T–Y, T–M and M–Y relationships after adjustment for the measured pretreatment covariates (i.e., confounders that are not affected by T) and T, and moreover that there must not be posttreatment confounding (i.e., confounders that are affected by T) between M and Y whatsoever, measured or unmeasured. In particular, [11, 12] proved that under Sequential Ignorability, the average indirect effect is nonparametrically identified, see Theorem 2.1 in the next section, and proposed a sensitivity analysis to assess the robustness of estimates to violations of Sequential Ignorability. Moreover they introduced estimation algorithms for the effects of interest that are implemented in the widely used mediation R package [13].

When multiple mediators are involved in the mediation model, three cases may arise, as shown in Figure 2: in Figure 2(a) mediators are conditionally independent given the treatment and measured covariates (not depicted here), in Figure 2(b) mediators are causally ordered, that is one affects the other; in Figure 2(c) mediators are conditionally dependent given the treatment and measured covariates without being causally ordered. In the latter situation, we will talk about *uncausally correlated* mediators as opposed to the situation of Figure 2(b) where mediators are causally correlated. We will also refer to the cases depicted in Figure 2(a) and (c) as mediation with multiple *causally unrelated* mediators.

Models in Figure 2(a) and (b) have been treated in the last few years [14–16] and will be commented further in the discussion section.

Figure 2(c) corresponds to an Acyclic Directed Mixed Graph (ADMG) as introduced by [17] and [18]. Bidirected dotted edges indicate a non-causal correlation, due for instance to a latent common cause, as in Figure 3. Shpitser and coauthors define districts as the connected components of the graph restricted to the bidirected edges and describe a necessary and sufficient condition for the effects to be identified, that is expressed in terms of observational data. In the case of multiple mediation, this condition says that the effect mediated by a set $\mathscr{S}$ of mediators can be written as a function of the observations if and only if $\mathscr{S}$ is the union of some districts. In the case of Figure 2(c), this means that the direct effect (mediated by neither M nor W) and the joint effect (mediated by both M and W) can be written in terms of observations, but that the effect mediated only by M cannot.

The estimation of such individual indirect effects, each specific to a given mediator, is however of practical importance. To do so, [19] extend their above mentioned approach to multiple mediators. When mediators are



(a) Independent (b) Causally correlated (c) Uncausally correlated

Figure 2: Three situations with multiple mediators M and W.

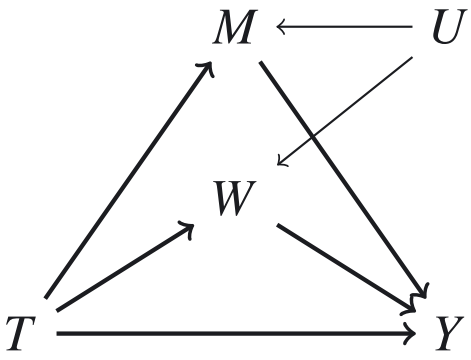


Figure 3: Correlation between mediators due to U.

causally unrelated, and Sequential Ignorability holds, they suggested to process several single mediator analyses in parallel, one mediator at the time. Obviously, this approach leads to a biased estimate of the direct effect, because it forces the indirect effects via all other mediators to contribute to the direct effect. More subtly, this approach is not appropriate when mediators are uncausally correlated due to an unmeasured covariate U causally affecting both mediators M and W as in Figure 3. As a matter of fact, in this situation, U is an unobserved confounder of the relationship between M and Y and Sequential Ignorability does not hold. This key fact was remarked by [19] and [14], but no explicit solution to the problem was proposed other than conducting the above mentioned sensitivity analysis. In this article, we suggest that a possible solution to this problem goes through the estimation of the multivariate law of the mediators conditionally on the treatment. This allows taking into account the spurious correlation among mediators induced by the unobserved variable U. A recent paper by Kim et al. [20] describes an alternative approach in which the dependence between mediators is characterised by a Gaussian copula together with marginal linear models; direct effect and indirect effects through each mediator are estimated imputing unobserved counterfactuals using a fully Bayesian approach. However, this approach has been specifically developed for continuous outcomes, while our method does not assume any particular form for the outcome as long as each marginal model is well specified.

In this article, we focus on the scenario of multiple causally unrelated mediators (i.e., either independent, Figure 2(a), or uncausally correlated, Figure 2(c), mediators). In Section 2, we start by reviewing definitions and results for simple mediation following [12]. Then, in Section 3, we extend these definitions and theoretical results to the scenario of multiple causally unrelated mediators. To do so, we introduce new identification hypotheses called SIMMA and compare them to Sequential Ignorability in the multiple cases as discussed by [19]. We show that under SIMMA the direct effect and the joint indirect effect through the vector of all mediators can be expressed by a formula involving observed variables only, while the indirect effect through each individual mediator is given by a formula involving both observed and counterfactual variables. The former formulae lead to an unbiased estimation of the direct and joint indirect effects, in compliance with [18]. Moreover, under an additional assumption, we propose a procedure based on the simulation of counterfactual distributions to estimate the indirect effects through individual mediators. In Section 4, we conduct an empirical study to show that the method results in unbiased estimates of the direct and indirect effects. The R implementation of our method is available on GitHub, https://github.com/AllanJe/multimediate. Finally, in Section 5, we apply our method to a real dataset from a large cohort to assess the effect of hormone replacement treatment on breast cancer risk through three uncausally correlated mediators, namely dense mammographic area, non-dense area and body mass index.

For the sake of clarity, we list here the notations used in this article:

- $T \in \{0, 1\}$: treatment
- $Z \in \mathbb{R}^K$: vector of all mediators
- $M^k \in \mathbb{R}$: kth mediator; when this is clear from the context, we will use the notation $M = M^k$

- $W^k \in \mathbb{R}^{K-1}$: complement of M^k in Z; when this is clear from the context, we will use the notation $W = W^k$
- $X \in \mathbb{R}^P$: vector of pretreatment confounders
- $Y \in \mathbb{R}$ or $\{0,1\}$: outcome
- $\delta^k(t)$: indirect effect of T mediated by M^k
- $\delta(t)$: indirect effect of T mediated by M
- $\zeta(t)$: direct effect of T
- δ,ζ: averages $(\delta(0)+\delta(1))/2$ and $(\zeta(0)+\zeta(1))/2$
- τ: total effect
- $PM^k(t) = \delta^k(t)/\tau$: proportion mediated by M^k
- Φ: the cumulative distribution function of the standard normal distribution $\mathcal{N}(0,1)$
- $A^\intercal$: the transpose of a matrix or vector A
- $A^{\intercal j}$: the transpose of the jth row of matrix A.

2 Brief review of simple mediation

We begin by recalling the main results by [12] in the case of a simple mediator and a binary treatment; we will adopt the same notations. Let Y be the variable denoting the observed outcome, T the treatment or exposure (coded as 1 for treated or exposed and 0 for non-treated or non-exposed) and M a single intermediate variable on the causal path from the T to Y. Finally let X represent a vector of pretreatment confounders. The causal diagram in Figure 4 depicts the causal relation between the four variables.

The causal approach to mediation analysis requires two types of counterfactual variables. On one hand, we consider the potential mediator when the treatment is set to t, denoted $M(t)$. On the other hand, we consider the potential outcome under the treatment status t and with the value of the mediator set to the potential value it would have under t', denoted $Y(t, M(t'))$. We recall the definition of counterfactuals in the supplementary materials.

The three quantities of interest in simple mediation analysis are the average causal indirect effect denoted $\delta(t)$, the average direct effect $\zeta(t)$, for $t \in \{0,1\}$, and the average total effect τ:

$$\delta(t) = \mathbb{E}[Y(t, M(1))] - \mathbb{E}[Y(t, M(0))]$$

$$\zeta(t) = \mathbb{E}[Y(1, M(t))] - \mathbb{E}[Y(0, M(t))]$$

$$\tau = \mathbb{E}[Y(1, M(1))] - \mathbb{E}[Y(0, M(0))].$$

Imai and collaborators showed that these effects can be identified regardless of a model assumption under two crucial hypotheses that go under the name of Sequential Ignorability Assumption (SIA):

$$\{Y(t', m), M(t)\} \perp\!\!\!\perp T | X = x \quad \forall\ t, t', m \tag{2.1}$$

$$Y(t', m) \perp\!\!\!\perp M(t) | T = t, X = x \quad \forall\ t, t', m. \tag{2.2}$$

Theorem 2.1. [12]. *Under SIA, the average indirect effect and the direct effect are identified non-parametrically and are given by*

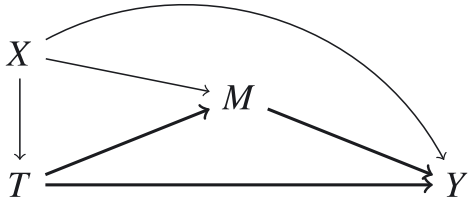


Figure 4: Simple mediation causal diagram.

$$\delta(t) = \int\int \mathbb{E}[Y|M=m, T=t, X=x]\mathrm{d}F_{M|T=1,X=x}(m) - \int \mathbb{E}[Y|M=m, T=t, X=x]\mathrm{d}F_{M|T=0,X=x}(m)\mathrm{d}F_X(x)$$
$$\zeta(t) = \int\int \mathbb{E}[Y|M=m, T=1, X=x]\mathrm{d}F_{M|T=t,X=x}(m) - \int \mathbb{E}[Y|M=m, T=0, X=x]\mathrm{d}F_{M|T=t,X=x}(m)\mathrm{d}F_X(x),$$

for $t \in \{0,1\}$.

In the setting of linear models, the two corollaries below follow, the first for a continuous outcome and the second for a binary outcome.

Corollary 2.2. [12]. *Under SIA and assuming the following linear structural equation model (LSEM)*

$$M = \alpha_2 + \beta_2 T + \xi_2^\Gamma X + \varepsilon_2$$
$$Y = \alpha_3 + \beta_3 T + \gamma M + \xi_3^\Gamma X + \varepsilon_3,$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ *for* $i \in \{2,3\}$, *the average indirect and direct effects are identified by* $\delta(0) = \delta(1) = \beta_2\gamma$ *and* $\zeta(0) = \zeta(1) = \beta_3$.

In the situation of a binary outcome, two main alternatives exist to model its conditional distribution. On the one hand, we can consider the probit regression

$$\mathbb{P}(Y=1|T,M,X) = \Phi_{\mathcal{N}(0,\sigma_3^2)}\left(\alpha_3 + \beta_3 T + \gamma M + \xi_3^\Gamma X\right),$$

where $\Phi_{\mathcal{N}(0,\sigma_3^2)}$ is the cumulative distribution function of the normal distribution $\mathcal{N}(0,\sigma_3^2)$.

On the other hand, we can assume the logistic regression

$$\text{logit}\left(\mathbb{P}(Y=1|T,M,X)\right) = \alpha_3 + \beta_3 T + \gamma M + \xi_3^\Gamma X.$$

Corollary 2.3. [12]. *Let Y be binary and assume the model*

$$M = \alpha_2 + \beta_2 T + \xi_2^\Gamma X + \varepsilon_2$$
$$Y = 1_{\{Y^*>0\}}, \text{ with } Y^* = \alpha_3 + \beta_3 T + \gamma M + \xi_3^\Gamma X + \varepsilon_3$$

where $\varepsilon_2 \sim \mathcal{N}(0,\sigma_2^2)$ and $\varepsilon_3 \sim \mathcal{N}(0,\sigma_3^2)$ (probit regression) or $\varepsilon_3 \sim \mathcal{L}(0,1)$ (logit regression), where $\mathcal{L}(0,1)$ denotes the standard logistic distribution.

Under SIA, the average indirect and direct effects are identified by

$$\delta(t) = \mathbb{E}\left[F_u(h_{t,1}) - F_u(h_{t,0})\right]$$
$$\zeta(t) = \mathbb{E}\left[F_u(h_{1,t}) - F_u(h_{0,t})\right]$$

where

$$h_{t,t'} = \alpha_3 + \beta_3 t + \gamma\left(\alpha_2 + \beta_2 \times t' + \xi_2^\Gamma X\right) + \xi_3^\Gamma X$$

and for a probit regression the function F_u is

$$F_u(z) = \Phi\left(\frac{z}{\sqrt{\gamma^2\sigma_2^2 + 1}}\right)$$

while for a logit regression we have

$$F_u(z) = \int_{-\infty}^{\infty} \Phi\left(\frac{z-y}{\gamma\sigma_2}\right) \frac{e^y}{(1+e^y)^2}\, \mathrm{d}y.$$

3 Extension to multiple causally unrelated mediators

In this subsection, we consider that K mediators intervene in the causal relationship between T and Y as in Figure 5. In particular, the following definitions and results apply when mediators are independent (Figure 2(a)) or uncausally correlated (Figure 2(c)).

3.1 Effect definitions

Let Z be the vector of all $K \geq 2$ mediators and M^k the mediator of interest. We denote by W^k the complement of M^k in Z, that is all mediators that are not of direct interest, and X the vector of pretreatment confounders.

The average indirect effect mediated by M^k was defined by [19] as

$$\delta^k(t) = \mathbb{E}\left[Y\left(t, M^k(1), W^k(t)\right)\right] - \mathbb{E}\left[Y\left(t, M^k(0), W^k(t)\right)\right].$$

As a measure of the average joint indirect effect, that is the indirect effect mediated by all the mediators, we take

$$\delta^Z(t) = \mathbb{E}[Y(t, Z(1))] - \mathbb{E}[Y(t, Z(0))].$$

Remark. *Note that the joint indirect effect can be decomposed as*

$$\delta^Z(t) = \frac{\sum_{k=1}^{K} \left(\delta^k(t) + \eta^k(t)\right)}{K}$$

where

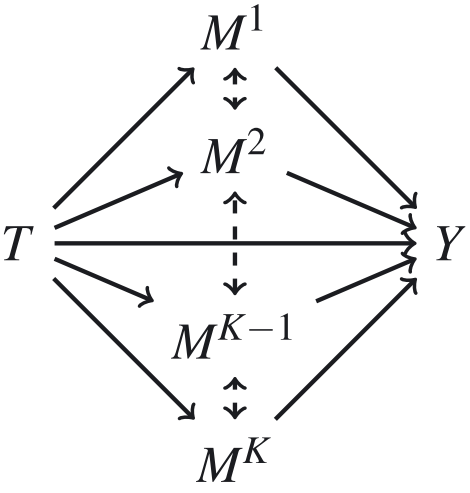


Figure 5: Multiple mediation causal diagram with possibly correlated mediators. The vector of pretreatment confounders X is not shown. Dashed lines represent possible non-causal correlations and solid lines causal relationships. Uncausal correlation is possible between each pair of mediators but this is not shown for improved readability of the figure.

$$\eta^k(t) = \mathbb{E}\left[Y\left(t, M^k(1-t), W^k(1)\right)\right] - \mathbb{E}\left[Y\left(t, M^k(1-t), W^k(0)\right)\right].$$

A proof of this result can be found in Appendix A.

Each of the 2^K direct effects is defined as

$$\begin{aligned}\zeta(t_1, \ldots, t_K) = \mathbb{E}\left[Y\left(1, M^1(t_1), \ldots, M^K(t_K)\right)\right] \\ -\mathbb{E}\left[Y\left(0, M^1(t_1), \ldots, M^K(t_K)\right)\right]\end{aligned}$$

where $t_k \in \{0,1\}$ for all $k \in \{1, \ldots, K\}$.

For the sake of simplicity, among all these direct effects, we will consider only $\zeta(0, \ldots, 0)$ and $\zeta(1, \ldots, 1)$, denoted $\zeta(t), t \in \{0,1\}$.

The total effect τ is

$$\tau = \mathbb{E}[Y(1, Z(1))] - \mathbb{E}[Y(0, Z(0))].$$

Note that τ is the sum of the joint indirect effect of treatment t and of the direct effect of treatment $1-t$:

$$\tau = \delta^Z(t) + \zeta(1-t).$$

3.2 Assumptions

Throughout the paper, we adopt the Stable Unit Treatment Value Assumption (SUTVA, [21] which implies that 1) there is no interference in the sense that potential mediator and outcome values of individual i do not depend on treatments of other individuals (i.e., $M_i^k(T) = M_i^k(T_i)$ and $Y_i(T, M^k, W^k) = Y_i(T_i, M_i^k, W_i^k)$ and 2) there are no multiple versions of treatments (i.e., $T_i = T_i'$ implies $M_i^k(T_i) = M_i^k(T_i')$ and $Y_i(T_i, M_i^k(T_i), W_i^k(T_i)) = Y_i(T_i', M_i^k(T_i'), W_i^k(T_i'))$). We augment the standard SUTVA to also assume that there are no multiple versions of mediators, that is if $M_i^k = M_i^{k'}$, then $Y_i(T_i, M_i^k, W_i^k) = Y_i(T_i, M_i^{k'}, W_i^k)$ [22].

Our results are based on the following hypotheses that we called Sequential Ignorability for Multiple Mediators Assumption (SIMMA):

$$\{Y(t,m,w), M(t'), W(t'')\} \perp\!\!\!\perp T | X = x, \tag{B.1}$$

$$Y(t',m,w) \perp\!\!\!\perp (M(t), W(t)) | T = t, X = x \tag{B.4}$$

$$Y(t,m,w) \perp\!\!\!\perp (M(t'), W(t)) | T = t, X = x \tag{B.5}$$

for all possible values of t, t', t'', m, w. A detailed explanation of SIMMA can be found in Appendix B.

Here, we recall that X is the vector of all the observed pretreatment covariates (by definition these variables are unaffected by the treatment). The first hypothesis implies that there must be no unobserved pretreatment confounders between the treatment and the outcome and between the treatment and the individual mediators after conditioning on all observed covariates. The second and third hypotheses exclude the existence of two distinct types of confounders between the mediators taken *jointly* and the outcome: the confounding by an unobserved pretreatment variable and the confounding by an observed or unobserved posttreatment variable.

Crucially, these hypotheses replace the second and third hypothesis that [19] make in the situation of multiple causally independent mediators, where a similar requirement applies to each counterfactual mediator separately and is interpreted as the randomisation of each mediator with respect to the outcome conditionally on the treatment arm (cf Appendix B). However, it is important to stress that assumption (B.4) is not more restrictive than Imai's hypotheses in the sense that it does not imply them, as we show in Appendix B. This hypothesis is the same as assumption 2) for multiple mediators in [14]. Our third assumption (B.5) is not included in [11] nor in [14] and is necessary to estimate the individual indirect effect of each mediator.

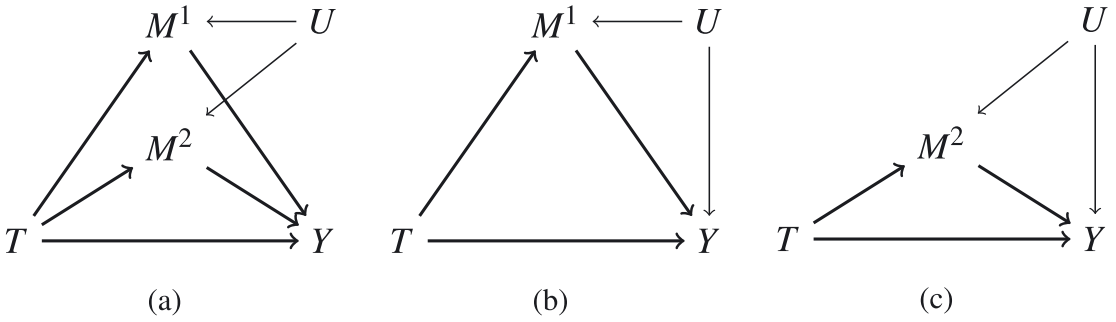


Figure 6: Multiple and simple mediation analyses, U observed. Data are simulated according to the model in (a).

The reason for replacing Imai's hypotheses with (B.4) and (B.5) is that we are interested in the situation where M and W are uncausally correlated, typically because of a pretreatment variable U affecting both as in Figure 6(a). Note that if U is unobserved (i.e., it is not part of the variables in X) conditions (B.4) and (B.5) are not violated because the joint distribution of the mediators incorporates the influence of U on the individual mediators. On the contrary, such a U would violate the corresponding hypothesis in [19] because it constitutes an unobserved confounder of the relations between W and Y and M and Y.

3.3 Identifiability

In the following, the mediator of interest M can be any of the K mediators, so that the results below can be applied to each mediator. In particular, this will allow to express the indirect effect mediated by each mediator taken individually.

Our first result extends Theorem 2.1 to multiple mediators, not only when mediators are causally independent as done by [19], but also when they are uncausally correlated.

Theorem 3.1. *Consider K mediators that can be either independent or uncausally correlated. Under SIMMA the following results hold.*

The average indirect effect of the mediator of interest is given by:

$$\delta(t) = \int\int_{\mathbb{R}^K} \mathbb{E}[Y|M=m, W=w, T=t, X=x] \left\{ \mathrm{d}F_{(M(1),W(t))|X=x}(m,w) - \mathrm{d}F_{(M(0),W(t))|X=x}(m,w) \right\} \mathrm{d}F_X(x). \tag{3.1}$$

Moreover the joint indirect effect, the direct effect and the total effect are identified non-parametrically respectively by:

$$\delta^Z(t) = \int\int_{\mathbb{R}^K} \mathbb{E}[Y|Z=z, T=t, X=x] \mathrm{d}F_{Z|T=1,X=x}(z) - \int_{\mathbb{R}^K} \mathbb{E}[Y|Z=z, T=t, X=x] \mathrm{d}F_{Z|T=0,X=x}(z) \mathrm{d}F_X(x),$$

$$\zeta(t) = \int\int_{\mathbb{R}^K} \mathbb{E}(Y|Z=z, T=1, X=x) - \mathbb{E}(Y|Z=z, T=0, X=x) \mathrm{d}F_{Z|T=t,X=x}(z) \mathrm{d}F_X(x),$$

$$\tau = \int \Big(\int_{\mathbb{R}^K} \mathbb{E}(Y|Z=z, T=1, X=x) \mathrm{d}F_{Z|T=1,X=x}(z) - \int_{\mathbb{R}^K} \mathbb{E}(Y|Z=z, T=0, X=x) \mathrm{d}F_{Z|T=0,X=x}(z) \Big).$$

In multiple mediation, Theorem 3.1 has the same role as Theorem 2.1 in simple mediation, because it shows that under proper assumptions, the (joint) indirect and direct effects are non-parametrically identified. In particular, from the equations above one can derive estimators for the joint indirect effect and for the direct effect, as already shown by [17]. In general, however, Eq. (3.1) does not allow to derive an estimator of the individual indirect effect of the mediator of interest, because the conditional distribution of $(M(t'), W(t))$ is not observable. Note that in the particular case where M is independent of W, Eq. (3.1) becomes

$$\delta(t) = \int\int \mathbb{E}[Y|M = m, T = t, X = x]\mathrm{d}F_{M|T=1,X=x}(m) - \int \mathbb{E}[Y|M = m, T = t, X = x]\mathrm{d}F_{M|T=0,X=x}(m)\mathrm{d}F_X(x),$$

which is the equation for $\delta(t)$ given by Theorem 2.1, thus allowing to identify the average indirect effect non-parametrically. This result was reported by [19]. A proof of Theorem 3.1 can be found in Appendix C.

The following two corollaries show identification formulae for the indirect and direct effects in the setting of the LSEM or when the mediating variables are Gaussian and Y is binary. Crucially, in the following corollaries, we assume that the correlations between the potential mediators are the same whatever the treatment governing the mediators:

$$cor\left(M^i(t), M^j(t')|T, X\right) = \rho_{ij}, \quad \forall\ t, t' \in \{0,1\}, \quad \forall\ i, j \in \{1, \ldots, K\}. \tag{3.2}$$

This hypothesis is indeed sufficient to identify the individual indirect effects through M from Eq. (3.1) in models where the joint distribution of the mediators is completely described by the expectation and covariance matrix, such as the multivariate Gaussian. In this particular situation, for all combinations of $t \neq t'$, the expectation of $(M(t), W(t'))|X = x$ is given by the vector $(E[M|T = t, X = x|], E(W|T = t', X = x|))$ and the covariance matrix is identified by the covariance matrix of $(M|T = t, X = x)$ and $(W|T = t', X = x)$ i.e., of $(M|T = t, X = x)$ and $(W|T = t, X = x)$.

3.4 Continuous outcome

Corollary 3.2. *With K mediators and P covariates, we assume the following linear model*

$$Z = \alpha_2 + \beta_2^\Gamma T + \xi_2^\Gamma X + \Upsilon_2 \tag{3.3}$$

$$Y = \alpha_3 + \beta_3 T + \gamma^\Gamma Z + \xi_3^\Gamma X + \varepsilon_3, \tag{3.4}$$

where $\alpha_2 = (\alpha_2^k)_{1\leq k\leq K}, \beta_2 = (\beta_2^k)_{1\leq k\leq K}, \gamma = (\gamma_k)_{1\leq k\leq K}$, $\xi_2 = (\xi_2^{kp})_{1\leq k\leq K, 1\leq p\leq P}$, $\xi_3 = (\xi_3^p)_{1\leq p\leq P}$, and $\Upsilon_2 = (\varepsilon_2^k)_{1\leq k\leq K} \sim \mathcal{N}(0, \Sigma_2)$ is the vector of residuals with covariance matrix $\sum_2 \in \mathbb{R}^K \times \mathbb{R}^K$ and $\varepsilon_3 \sim \mathcal{N}(0, \sigma_3^2)$, with $\sigma_3 \in \mathbb{R}$.

We assume that the K mediators are either independent or non-causally correlated. In the latter case, we assume that pairwise correlations between potential mediators do not depend on the treatments governing them, i.e., we assume condition (3.2). Under SIMMA, the indirect effect of the kth mediator is identified and given by:

$$\delta^k(0) = \delta^k(1) = \gamma_k \beta_2^k.$$

Moreover, the joint indirect effect is the sum of the average indirect effects by each mediator:

$$\delta^Z(t) = \sum_{k=1}^{K} \delta^k(t).$$

The direct effect is also identified and given by

$$\zeta(0) = \zeta(1) = \beta_3.$$

A proof of Corollary 3.2 can be found in the Supplementary material. Note that an equivalent result for the joint indirect effect is shown in [14]. Also note that the additivity of the individual indirect effects into the joint direct effect (i.e., $\delta^Z(t) = \sum_k \delta^k(t)$) holds only in the context of Corollary 3.2, otherwise it does not.

We have already observed that if the K mediators are independent, the equation for the marginal indirect effect given by Theorem 3.1 (multiple analysis) reduces to the equation given by Theorem 2.1 (simple analysis). In this situation, Corollary 3.2 implies that in the LSEM setting, the indirect effects given by simple analyses can be summed up to obtain the joint indirect effect. Obviously, simple analyses do not allow to assess a comprehensive direct effect, because depending on the mediator of interest, each simple analysis will lead to a different direct effect. All these aspects will be illustrated through simulations in Section 4.

3.5 Binary outcome

We now address the case of a binary outcome. As for simple mediation, we consider either the probit regression

$$\mathbb{P}(Y=1|T,Z,X)=\Phi_{\mathcal{N}\left(0,\sigma_3^2\right)}\left(\alpha_3+\beta_3 T+\gamma^{\Gamma}Z+\xi_3^{\Gamma}X\right),$$

or the logistic regression

$$\text{logit}\left(\mathbb{P}(Y=1|T,Z,X)\right)=\alpha_3+\beta_3 T+\gamma^{\Gamma}Z+\xi_3^{\Gamma}X.$$

Corollary 3.3. *Assume the following model with a binary outcome:*

$$Z=\alpha_2+\beta_2^{\Gamma}T+\xi_2^{\Gamma}X+\Upsilon_2, \tag{3.5}$$

$$Y^{\star}=\alpha_3+\beta_3 T+\gamma^{\Gamma}Z+\xi_3^{\Gamma}X+\varepsilon_3, \tag{3.6}$$

$$Y=1_{\{Y^{\star}>0\}} \tag{3.7}$$

where $\Upsilon_2\sim\mathcal{N}(0,\Sigma_2)$ and where $\varepsilon_3\sim\mathcal{N}(0,\sigma_3^2)$ or $\mathcal{L}(0,1)$. We assume that the K mediators are either independent or non-causally correlated. In the latter case, we assume that pairwise correlations between potential mediators do not depend on the treatments governing them as in condition (3.2). Under SIMMA, the effects of interest are given by:

$$\begin{aligned}
\delta^k(t)=\int & F_U\left(\left(\alpha_3+\sum_{j=1}^{K}\gamma_j\alpha_2^j\right)+\left(\beta_3+\sum_{j=1,j\neq k}^{K}\gamma_j\beta_2^j\right)t+\gamma_k\beta_2^k\times 1+\left(\xi_3+\sum_{j=1}^{K}\gamma_j\xi_2^{\Gamma j}\right)x\right)\\
&-F_U\left(\left(\alpha_3+\sum_{j=1}^{K}\gamma_j\alpha_2^j\right)+\left(\beta_3+\sum_{j=1,j\neq k}^{K}\gamma_j\beta_2^j\right)t+\gamma_k\beta_2^k\times 0+\left(\xi_3+\sum_{j=1}^{K}\gamma_j\xi_2^{\Gamma j}\right)x\right)\mathrm{d}F_X(x),\\
\delta^Z(t)=\int & F_U\left(\left(\alpha_3+\sum_{k=1}^{K}\gamma_k\alpha_2^k\right)+\beta_3\times t+\sum_{k=1}^{K}\gamma_k\beta_2^k\times 1+\left(\xi_3+\sum_{k=1}^{K}\gamma_k\xi_2^{\Gamma k}\right)x\right)\\
&-F_U\left(\left(\alpha_3+\sum_{k=1}^{K}\gamma_k\alpha_2^k\right)+\beta_3\times t+\sum_{k=1}^{K}\gamma_k\beta_2^k\times 0+\left(\xi_3+\sum_{k=1}^{K}\gamma_k\xi_2^{\Gamma k}\right)x\right)\mathrm{d}F_X(x),\\
\zeta(t)=\int & F_U\left(\left(\alpha_3+\sum_{k=1}^{K}\gamma_k\alpha_2^k\right)+\beta_3\times 1+\left(\sum_{k=1}^{K}\gamma_k\beta_2^k\right)\times t+\left(\xi_3+\sum_{k=1}^{K}\gamma_k\xi_2^{\Gamma k}\right)x\right)\\
&-F_U\left(\left(\alpha_3+\sum_{k=1}^{K}\gamma_k\alpha_2^k\right)+\beta_3\times 0+\left(\sum_{k=1}^{K}\gamma_k\beta_2^k\right)\times t+\left(\xi_3+\sum_{k=1}^{K}\gamma_k\xi_2^{\Gamma k}\right)x\right)\mathrm{d}F_X(x),
\end{aligned}$$

where for a probit regression we have

$$F_U(z)=\Phi\left(\frac{z}{\sqrt{\sigma_3^2+\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov\left(\varepsilon_2^k,\varepsilon_2^j\right)}}\right),$$

and for a logit regression we have

$$F_U(z) = \int_{\mathbb{R}} \Phi\left(\frac{z - e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K} \gamma_k \gamma_j cov\left(\varepsilon_2^k, \varepsilon_2^j\right)}} \right) \frac{e^{e_3}}{(1+e^{e_3})^2} \, \mathrm{d}e_3.$$

When the mediators are independent, we have for a probit regression

$$F_U(z) = \Phi\left(\frac{z}{\sqrt{\sigma_3^2 + \sum_{k=1}^{K} \gamma_k^2 \sigma_2^2}} \right),$$

and for a logistic regression

$$F_U(z) = \int_{\mathbb{R}} \Phi\left(\frac{z - e_3}{\sqrt{\sum_{k=1}^{K} \gamma_k^2 \sigma_2^2}} \right) \frac{e^{e_3}}{(1+e^{e_3})^2} \, \mathrm{d}e_3.$$

A proof of Corollary 3.3 can be found in the supplementary materials.

3.6 Estimation algorithm

The proof of Theorem 3.1 can be generalised to prove that, under SIMMA, the densities of the counterfactual outcomes can be expressed as follows:

$$f\left(Y(t, M(t'), W(t'))|X = x\right) = \int_{\mathbb{R}^K} f(Y|T = t, M = m, W = w, X = x) \mathrm{d}F_{(M,W)|T=t', X=x}(m, w) \tag{3.8}$$

$$f\left(Y(t, M(t'), W(t))|X = x\right) = \int_{\mathbb{R}^K} f(Y|T = t, M = m, W = w, X = x) \mathrm{d}F_{(M(t'),W(t))|X=x}(m, w). \tag{3.9}$$

Equation (3.8) justifies the Monte-Carlo estimation of the expectation $E[Y(t, Z(t'))] = E[Y(t, M(t'), W(t'))]$, and therefore of the direct, joint indirect and total effects. Moreover, under the additional condition (3.2) (and assuming that the joint distribution of mediators is completely determined by its expectation and covariance matrix), Eq. (3.9) makes it possible to sample $Y(t, M(t'), W(t))$ as well and therefore to estimate its expectation and the indirect effect through M. In particular, SIMMA and (3.2) allow to estimate the conditional covariance matrix of the counterfactual mediators for each possible combination of interventions as the covariance matrix of the mediators given the treatment and the pretreatment covariates.

Accordingly we adapt the quasi-Bayesian algorithm presented by [11], to the situation of multiple mediators uncausally related, i.e., for independent and uncausally correlated mediators.

Algorithm. *In order to estimate the effects of interest:*

(1) *Fit parametric models for the observed outcome (given all the mediators, treatment and covariates), and mediators (given all the treatment and covariates), denoted respectively as* $\widehat{\Theta}^Y$ *and* $\widehat{\Theta}^Z = (\widehat{\Theta}^1, \ldots, \widehat{\Theta}^K)$. *Obtain the estimate* $\widehat{\sum_2}$ *of the covariance matrix between mediators given the treatment and the covariates.*

(2) *For each model, sample J values for each of its parameters according to their multivariate sampling distribution, denoted as* $\widehat{\Theta}^Y_{(j)}, j = 1, \ldots, J$ *and* $\widehat{\Theta}^Z_{(j)} = (\widehat{\Theta}^1_{(j)}, \ldots, \widehat{\Theta}^K_{(j)})$. *As in [11], we use the approximation based on the multivariate normal distribution centered at the estimates of the parameters and with the estimated asymptotic covariance matrix between the estimators.*

(3) *For each* $j = 1, \ldots, J$, *repeat the followings steps:*

- *Simulate the potential values of each mediator. In particular, for each of the K mediators, each pair* $(t,t') \in \{0,1\}^2$*, and each individual* $i \in \{1,\dots,n\}$*, simulate R values of* $Z_{(ji)}^{(kr)}(t,t') = (M_{(ji)}^{(kr)}(t), W_{(ji)}^{(kr)}(t'))$*. When all mediators have the same treatment value, the vector of all mediators will be denoted as* $Z_{(ji)}^{(r)}(t) = Z_{(ji)}^{(kr)}(t,t)$*. Note that it is at this step that we take into account the correlation between mediators* $\widehat{\Sigma}_2$*.*
- *Simulate the potential outcomes given the simulated values of the potential mediators, denoted as* $Y_{(ji)}^{(r)}(t, Z_{(ji)}^{(kr)}(t',t''))$ *for each i, k and* $t,t',t'' \in \{0,1\}$*.*
- *Estimate the causal mediation effects*:

$$\widehat{\delta}_{(j)}^{k}(t) = \frac{1}{nR}\sum_{i=1}^{n}\sum_{r=1}^{R}\left\{Y_{(ji)}^{(r)}\left(t, Z_{(ji)}^{(kr)}(1,t)\right) - Y_{(ji)}^{(r)}\left(t, Z_{(ji)}^{(kr)}(0,t)\right)\right\}$$

$$\widehat{\delta}_{(j)}^{Z}(t) = \frac{1}{nR}\sum_{i=1}^{n}\sum_{r=1}^{R}\left\{Y_{(ji)}^{(r)}\left(t, Z_{(ji)}^{(r)}(1)\right) - Y_{(ji)}^{(r)}\left(t, Z_{(ji)}^{(r)}(0)\right)\right\}$$

$$\widehat{\zeta}_{(j)}(t) = \frac{1}{nR}\sum_{i=1}^{n}\sum_{r=1}^{R}\left\{Y_{(ji)}^{(r)}\left(1, Z_{(ji)}^{(r)}(t)\right) - Y_{(ji)}^{(r)}\left(0, Z_{(ji)}^{(r)}(t)\right)\right\}$$

$$\widehat{\tau}_{(j)}(t) = \frac{1}{nR}\sum_{i=1}^{n}\sum_{r=1}^{R}\left\{Y_{(ji)}^{(r)}\left(1, Z_{(ji)}^{(r)}(1)\right) - Y_{(ji)}^{(r)}\left(0, Z_{(ji)}^{(r)}(0)\right)\right\}.$$

(4) *From the empirical distribution of each effect above, obtain point estimates together with p-values and confidence intervals.*

Note that this algorithm does not implement the formulae given for the specific models of Corollaries 3.2 and 3.3.

We implemented this algorithm in the R package multimediate, currently available on GitHub. Our main function is based on the mediate() function of the package mediation [13] and makes it possible to work not only with continuous mediators but also binary and ordered categorical mediators using probit models.

4 Simulation studies

In this section, we validate our methodological results through empirical studies. In particular, we compare our estimates of the mediation causal effects to the true effects and to the estimates obtained by running simple mediation analyses, one for each mediator.

4.1 Data simulation method

Except for the LSEM framework, it is in general not straightforward to obtain the true mediation effect values from a causal generative model, that is from a set of causal structural equations. To overcome this difficulty, we start by simulating a large database of values for the treatment T and for all the counterfactual mediators $M^k(t)$, and outcomes $Y(t, M^1(t_1), \dots, M^K(t_K))$, see Table 1 for an example. Then we simply compute the indirect effects $\delta^k(t)$ and $\delta^Z(t)$ and the direct effect $\zeta(t)$ as means, according to the definitions given in Section 3.1. The large size of the dataset guarantees that these Monte-Carlo estimates can be taken as the true values. In this study we generate a dataset of 10^6 observations, so that the estimate error is as small as 0.2% of the standard deviation of the effect of interest.

In order to obtain a subset of observations to test the considered estimation methods, we sample N individuals (i.e., rows) $i = 1, \dots, N$ and for each of them we select only the values $Y(T_i, Z_i(T_i))$ and $Z_i(T_i)$ corresponding to the specific value of T_i. More precisely:

- if $T_i = 0$ we take $Z_i = (M_i^1, \dots, M_i^k) = (M_i^1(0), \dots, M_i^k(0)) = Z_i(0)$ and $Y_i = Y_i(0, Z_i(0))$,
- if $T_i = 1$ we take $Z_i = (M_i^1, \dots, M_i^k) = (M_i^1(1), \dots, M_i^k(1)) = Z_i(1)$ and $Y_i = Y_i(1, Z_i(1))$.

Table 1: Simulated counterfactuals with two independent mediators.

T	$M(0)$	$M(1)$	$W(0)$	$W(1)$	$Y(1, M(1), W(1))$	$Y(1, M(1), W(0))$	$Y(1, M(0), W(1))$
0	**0.28**	1.08	**0.53**	1.43	2.42	1.79	1.94
0	**0.42**	1.22	**−1.80**	−0.90	1.41	0.78	0.93
1	0.63	**1.43**	0.03	**0.93**	**1.87**	1.24	1.39
1	0.75	**1.55**	2.24	**3.14**	**2.95**	2.32	2.47

$Y(0, M(1), W(1))$	$Y(1, M(0), W(0))$	$Y(0, M(1), W(0))$	$Y(0, M(0), W(1))$	$Y(0, M(0), W(0))$
2.02	1.31	1.39	1.54	**0.91**
1.01	0.30	0.38	0.53	**−0.09**
1.47	0.76	0.84	0.99	0.36
2.55	1.84	1.92	2.07	1.44

Table 2: Simulated observed data with two independent mediators. Observations were extracted from Table 1.

T	M	W	Y
0	0.28	0.53	0.91
0	0.42	−1.80	−0.09
1	1.43	0.93	1.87
1	1.55	3.14	2.95

Tables 1 and 2 illustrate the simulation procedure.

For several simulation models, we estimate the different effects of interest by means of the general algorithm for multiple mediators described above in Section 3.6. We compare our estimates with both the true values and the estimates of two simple analyses (one for each mediator) obtained with the mediation package. Because in general $\delta^k(1) \neq \delta^k(0)$ and $\zeta(1) \neq \zeta(0)$, for the sake of simplicity we focus on average effects such as $\delta = (\delta(0) + \delta(1))/2$ and $\zeta = (\zeta(1) + \zeta(0))/2$. Note that for continuous outcome and in absence of interaction between treatment and mediators, Corollaries 2.2 and 3.2 imply that $\delta^k(1) = \delta^k(0)$ and $\zeta(1) = \zeta(0)$. For each mediator, we also show the proportion mediated $PM^k = \delta^k/\tau$.

For comparative purposes, we analyse the simulations with our multiple mediation method, and also with the approach consisting in running simple analyses in parallel [19], and with the method described by [14] which we refer to as V&V in the figures. In the latter case, we not only report the estimates of the joint indirect and direct effects, but also the estimates of the mediator-specific indirect effects, even though the authors clearly explain that correlation between mediators would lead to bias.

4.2 Limitations of repeated simple analyses when the common cause of mediators is not measured

In this section, data are generated under the model described in Figure 6(a), where the dependence between the two mediators is induced by the pretreatment variable U. More specifically, variables are simulated according to the following distributions ($N = 1000$):

- T follows a Bernoulli distribution $\mathscr{B}(0.3)$
- U follows a normal distribution $\mathscr{N}(0,1)$
- the conditional distribution of the counterfactual mediators

$$M^1(t,u) \sim \mathcal{N}(1 + 4t + 2u, 1)$$

$$M^2(t,u) \sim \mathcal{N}(2 + 6t + 3u, 1)$$

– the counterfactual outcomes follow the normal distributions

$$Y\left(t, M^1(t'), M^2(t'')\right) \sim \mathcal{N}(1 + 10t + 5M^1(t') + 4M^2(t''), 1).$$

Note that the correlation between the two mediators conditionally on the treatment (and not on U, Figure 2(c)), is equal to 0.7.

When we have two causally independent mediators and U is observed, the approach by [19] is to perform two simple analyses as in Figure 6(b) and (c). However, when U is unobserved, the situation is like in Figure 2(c) with mediators showing residual correlation. In this case, conducting separate simple analyses is not appropriate because Sequential Ignorability assumptions (B.2) and (B.3) are violated [19].

Here, we illustrate this problem through simulations. For comparison purposes, we also show the results obtained with our method for multiple analysis and with the method by [14].

As expected, Tables 3 and 4 show that simple analyses adjusted for U give precise and accurate estimates of the indirect effects (but obviously not of the direct effect), while they give biased estimates when U is not taken into account. On the contrary, our method gives precise and accurate estimates of all effects with or without taking into account U, showing that it is still possible to conduct a mediation analysis to estimate all effects even when U is unobserved.

In the following subsection, we suppose that U is unobserved, as it is often the case in practical situations.

4.3 Empirical study of the properties of the proposed estimators

The previous section illustrated our method on a single simulation run. In this section, we perform a simulation-based study to assess the properties of the proposed estimation procedure. More specifically, we

Table 3: Adjusting for U when all variables in Figure 6(a) are observed.

Effects	Value	Simple analysis M^1	Simple analysis M^2	V&V	Multiple analysis
δ^Z	44	NA	NA	45.20 [40.30; 50.10]	44.40 [43.30; 45.50]
PM^Z	0.81	NA	NA	0.82	0.81 [0.81; 0.82]
δ^1	20	19.40 [17.70; 21.10]	NA	20.40 [15.90; 24.90]	20.50 [19.60; 21.40]
PM^1	0.37	0.36 [0.33; 0.39]	NA	0.37	0.38 [0.36; 0.40]
δ^2	24	NA	21.60 [18.80; 24.60]	25.30 [18.90; 31.60]	23.90 [23.10; 24.66]
PM^2	0.44	NA	0.40 [0.34; 0.45]	0.45	0.44 [0.42; 0.45]
ζ	10	35 [33.40; 36.60]	32.80 [29.70; 35.65	9.90 [9.20 ;10.60]	9.90 [9.70; 10.20]
τ	54	54.40 [53.40; 55.40]	54.40 [53.30; 55.50]	55.10 [50.20; 59.90]	54.43 [53.20; 55.50]

Table 4: Not adjusting for U: data are generated as in Figure 6(a) but analysed as if U was unobserved.

Effects	Value	Simple analysis M^1	Simple analysis M^2	V&V	Multiple analysis
δ^Z	44	NA	NA	46.80 [40.30;53.40]	43.20 [40.20; 46.20]
PM^Z	0.81	NA	NA	0.80	0.81 [0.80; 0.82]
δ^1	20	38.40 [34.50; 42.30]	NA	41.60 [32.40;50.70]	20 [18.00; 22.20]
PM^1	0.37	0.72 [0.68; 0.75]	NA	0.73	0.37 [0.33; 0.42]
δ^2	24	NA	40.86 [37; 44.90]	44.70 [36.60;52.80]	23.20 [20.90; 25.50]
PM^2	0.44	NA	0.76 [0.73; 0.79]	0.78	0.43 [0.38; 0.48]
ζ	10	14.70 [13.20; 16.40]	12.46 [10.90; 13.90]	10.10 [9.90; 10.40]	9.90 [9.70; 10.20]
τ	54	53.20 [49.20; 57]	53.30 [49.50; 57.20]	57 [50.40; 63.60]	53.20 [50.20; 56.20]

compute bias, confidence interval coverage probability, mean square error (MSE) and variance of our estimators as means over 200 simulation runs for each considered parameter setting. We compare the estimates of several simple analyses, one for each mediator, to the estimates obtained with our multiple mediation analysis for different correlation levels. We consider two causal simulation models accounting for two types of outcome (continuous and logit binary), and two settings with two continuous causally unrelated mediators. Uncausally correlated mediators, Figure 2(c), are simulated from a multivariate normal distribution with fixed covariance matrix. The details of the simulation models can be found in Appendix D.

Simulations according to model 1 (continuous outcome) are run for different values of correlation between the mediators and increasing sample size (N = 50, 200, 500, 1000). Results for bias and coverage probability can be seen in Figures 7–10. These figures clearly show that our approach allows an unbiased estimation, contrary to the simple analyses, for both the direct and indirect effects.

The empirical 95% confidence interval given by our method contains the real value in approximatively 95% of the runs, for both the indirect and direct effects and whatever the correlation between the mediators. On the contrary, simple analysis obtains fair coverages only when the correlation is almost null. As expected, the estimators of the individual indirect effects obtained with the method of [14] have the same behaviour as simple analysis estimates. Moreover, the estimators of the joint indirect and direct effects by the method of [14] behave similarly as ours, except that the coverage probability is constant for their method. Our estimators have low variance and low MSE for sample sizes larger than 200.

Simulations were also run for model 2 (binary outcome) for different values of correlation between the mediators with 1000 observational data. As illustrated by Figure 13 in the Appendix, the results for bias, coverage probability, variance and MSE confirm that our estimators are unbiased and have low variance and the expected coverage probability, thus outperforming simple analysis. It is worth noting that for positive correlations, the coverage probability of the confidence intervals of the individual indirect effects is unsatisfactory. This is likely due to the very low variance of the estimators.

5 Application

We applied our method to estimate the amount of causal effect of hormone replacement therapy (HRT) on breast cancer (BC) risk that is mediated by mammographic density (MD) – specifically dense area (DA) and non-dense area (NDA) – and body mass index (BMI) in postmenopausal women. The data come from the E3N French cohort study [23]. Based on more than 5000 cases diagnosed between baseline and 2008 [24], a nested case–control study was designed using incidence density sampling. For 640 invasive breast cancer cases with known laterality and at least one mammogram taken between baseline and age at diagnosis, one control was randomly selected from women who had not been diagnosed with breast cancer at the age when the matched case was diagnosed (reference age). After excluding women with missing value, 489 cases and 489 controls were available for the analysis. HRT, prescribed to relief menopausal symptoms, consists in providing women with hormones whose production naturally decreases with menopause [25]. One of the consequences of taking HRT is that women do not experience the decrease of DA, the increase of NDA and the increase of BMI typically occurring at menopause [26]. HRT use has been since long recognised to be a risk factor for BC [27]. Independent BC risk factors are also high postmenopausal BMI and high per age and per BMI MD [28,29]. In order to better understand the mutual relationship between HRT, MD and BMI in BC carcinogenesis, it is important to determine whether and eventually to which extent the effect of HRT on BC risk is due to its action on MD and BMI (mediated effect) and to which extent it is independent of MD and BMI (direct effect).

Based on evidence from association studies on breast cancer risk [30,31], we can reasonably assume that BMI and mammography density are uncausally correlated, being their correlation likely due to common genetic traits, as suggested by twin studies [30,32] and Mendelian randomization analysis [33]. We make the implicit assumption that HRT precedes the mediators and that these precede BC; Figure 11 depicts the causal assumptions made for the following mediation analysis.

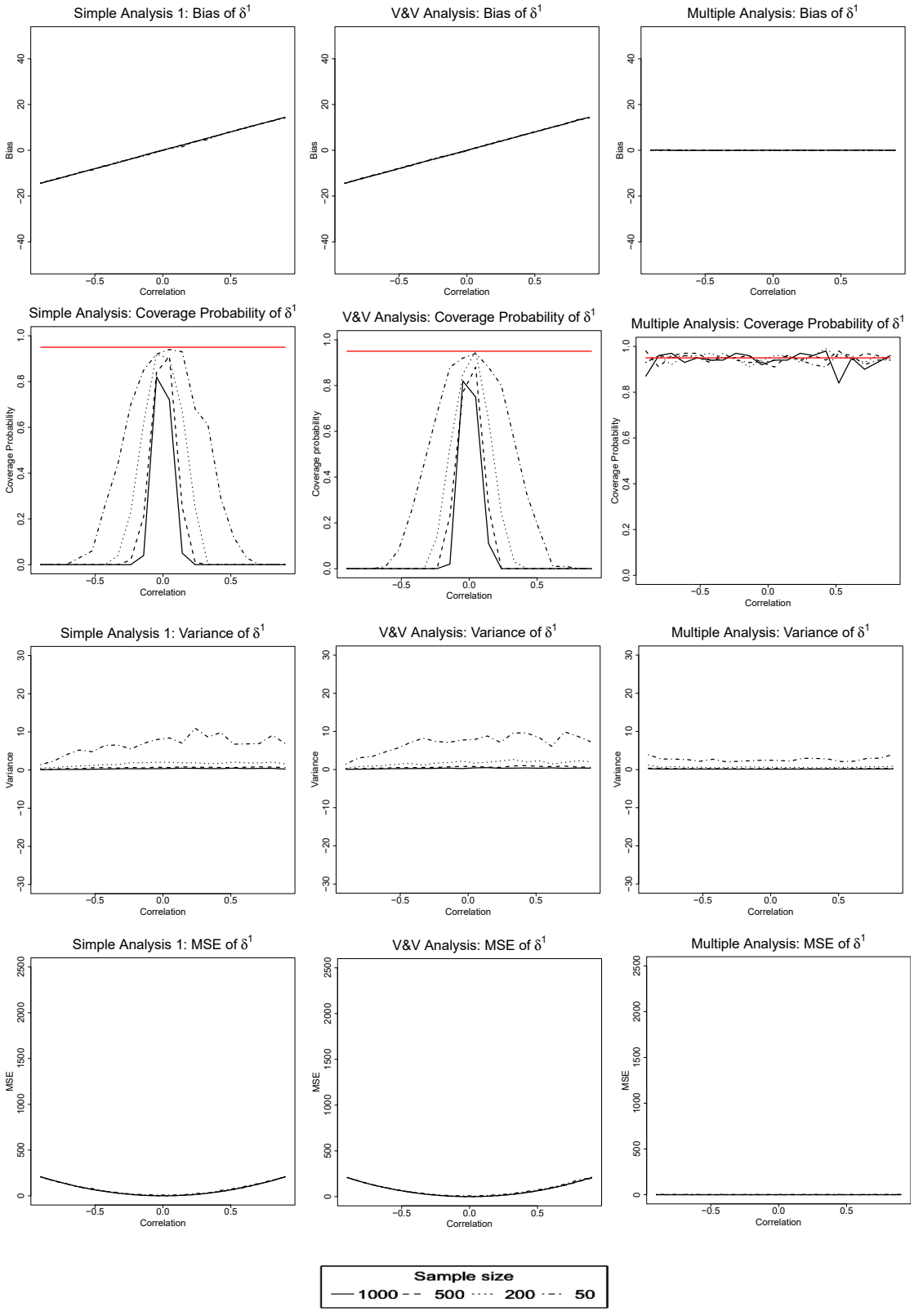


Figure 7: Model 1 (continuous outcome): bias, confidence interval coverage probability, mean square error (MSE), and variance of the indirect effect estimators $\widehat{\delta}^1$ calculated as means over 200 simulations when the correlation between mediators varies. The bias formula used here is $Bias = \Theta - \mathbb{E}[\widehat{\Theta}]$.

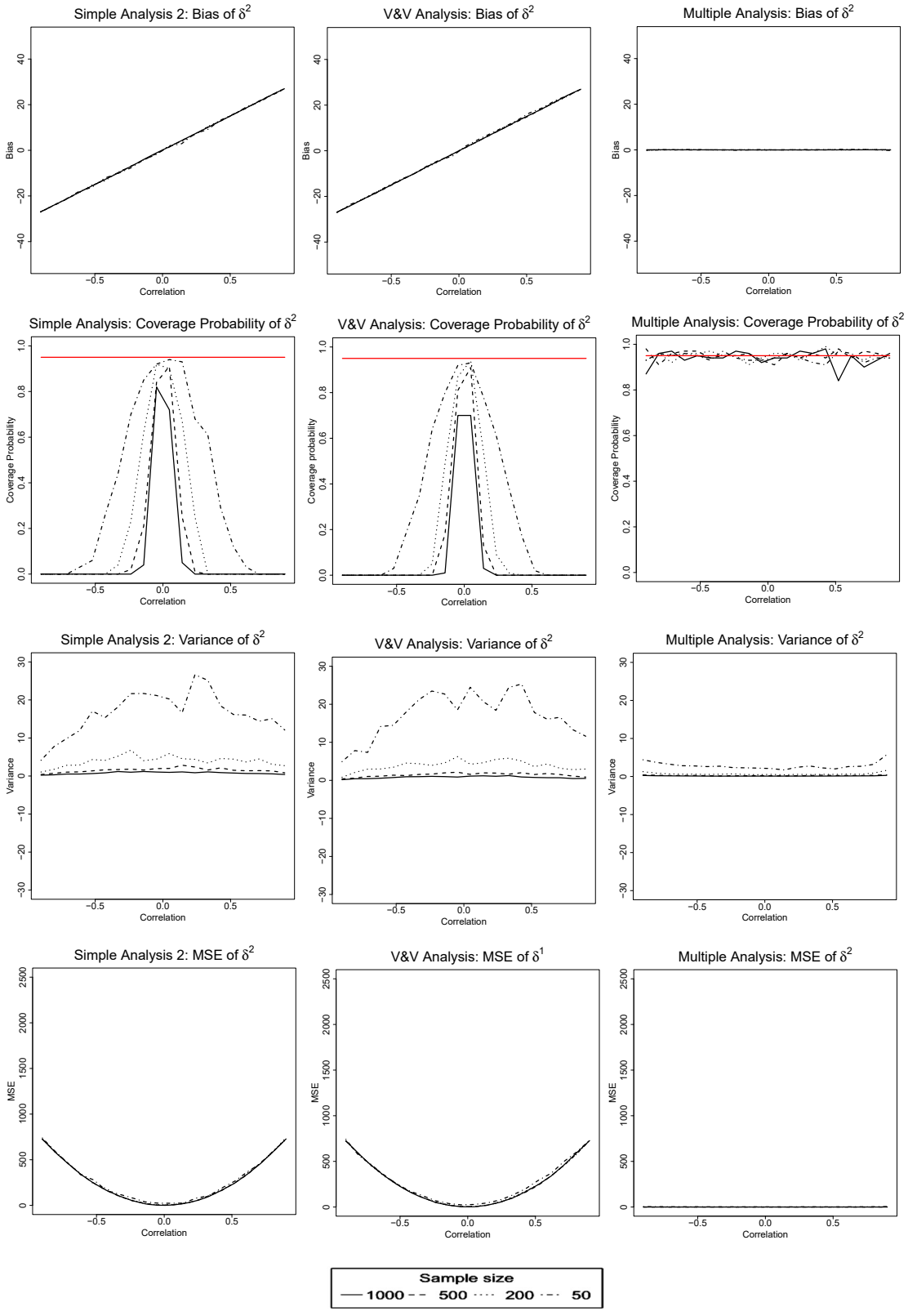


Figure 8: Model 1 (continuous outcome): bias, confidence interval coverage probability, mean square error (MSE), and variance of the indirect effect estimators $\widehat{\delta}^2$ calculated as means over 200 simulations when the correlation between mediators varies. The bias formula used here is $Bias = \Theta - \mathbb{E}[\hat{\Theta}]$.

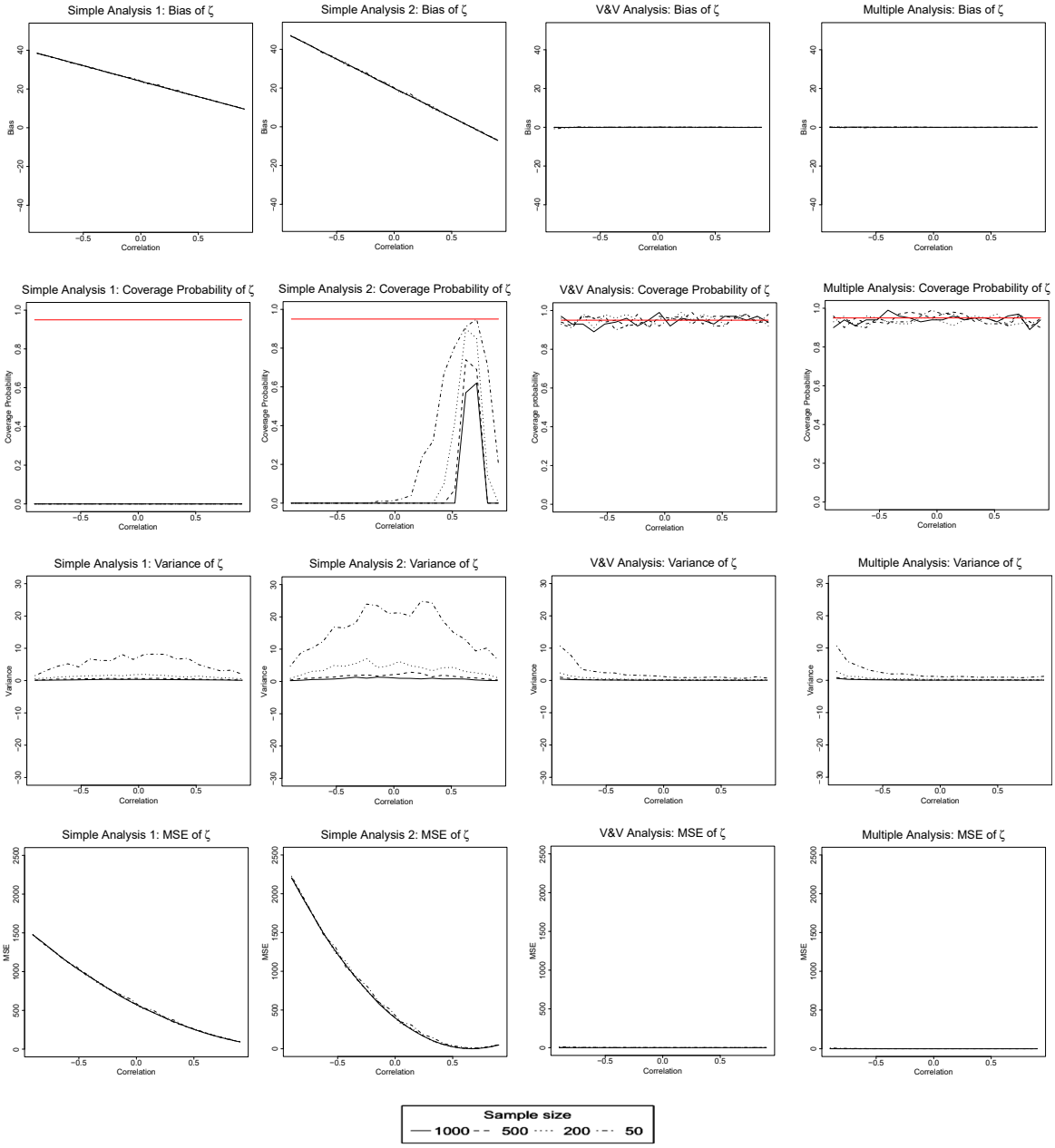


Figure 9: Model 1 (continuous outcome): bias, confidence interval coverage probability, mean square error (MSE), and variance of the direct effect estimator $\tilde{\zeta}$ calculated as means over 200 simulation runs when the correlation between mediators varies. The bias formula used here is $Bias = \Theta - \mathbb{E}[\hat{\Theta}]$.

5.1 Regression models

The continuous variables were normalised using the Box-Cox likelihood-like approach [34], $t(M) = \frac{M^{\lambda}-1}{\lambda}$, with λ equal to 0.38, 0.34 and –1.19 for DA, NDA and BMI respectively, as we can see in Figure 14. HRT was treated as a dichotomous variable whose levels were never versus ever users (past or current).

In preparation to our mediation analysis, we regressed each mediator on HRT and AGE (Table 5, models 1, 2 and 3 respectively) and BC on HRT and AGE with or without conditioning on the three mediators (respectively models 4a, 4b). As expected, HRT ever users had significantly higher values of DA and significantly lower of

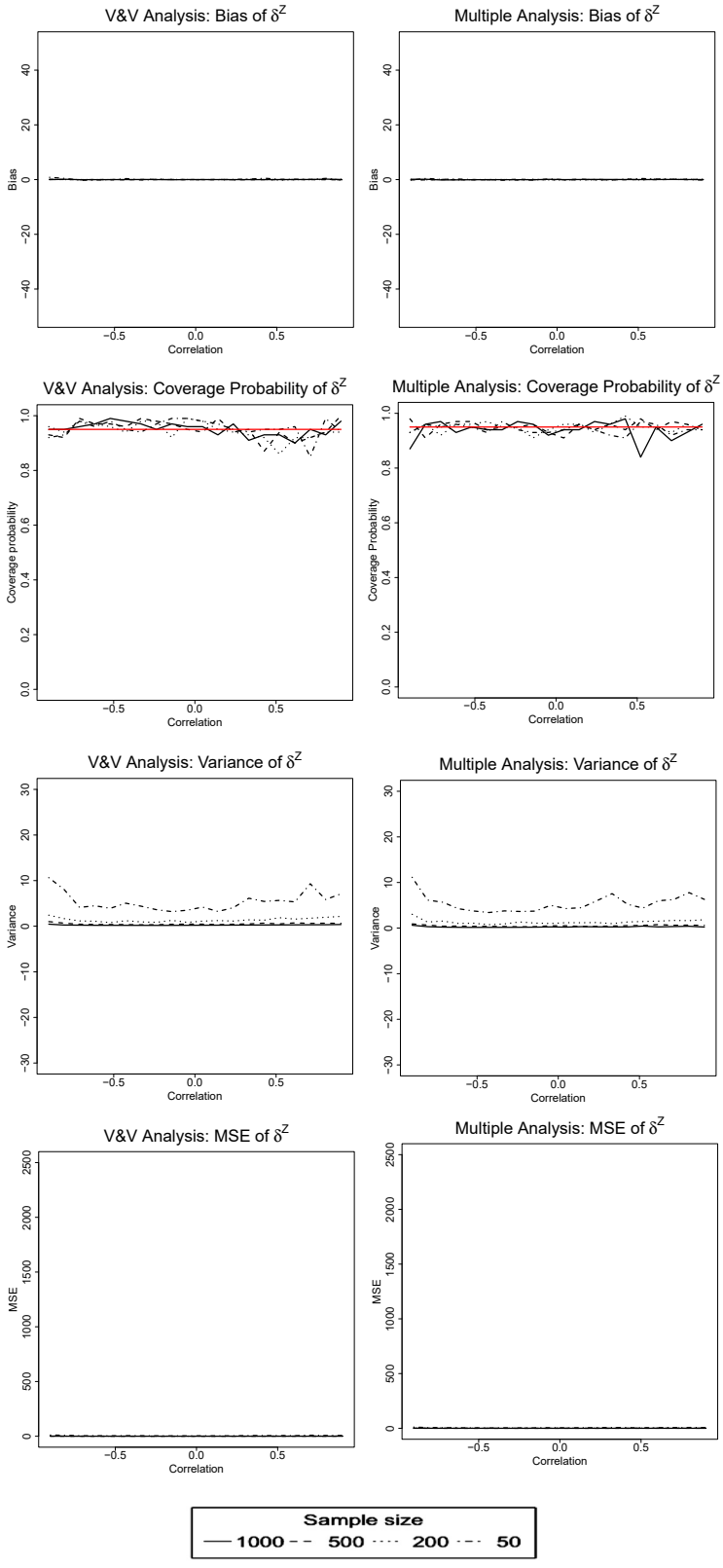


Figure 10: Model 1 (continuous outcome): bias, confidence interval coverage probability, mean square error (MSE), and variance of the joint direct effect estimator $\widehat{\delta}^Z$ calculated as means over 200 simulation runs when the correlation between mediators varies. The bias formula used here is $Bias = \Theta - \mathbb{E}[\hat{\Theta}]$.

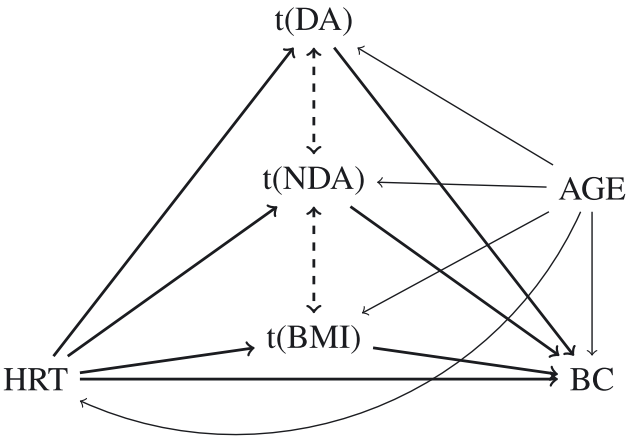


Figure 11: Causal diagram for the application.

NDA and BMI (Table 5); DA and BMI were positively associated with BC risk, whereas NDA was negatively associated with risk (Table 5). HRT was positively associated with BC risk and the association decreased of the 20% in the log-OR scale when accounting for DA, NDA and BMI into the model (Table 5 models 4a and 4b). Note that after adjusting for HRT and Age the residuals correlation between DA and BMI, NDA and BMI and DA and NDA are −0.04, −0.22 and 0.60 respectively.

5.2 Multiple mediation analysis

We applied our method with models 1 2, 3 and 4.b from Table 5 to estimate the causal mediated effect due to all mediators and the causal mediated effect due to each of them when accounting for their mutual correlation. As shown in Table 6 the causal mediated effects due to DA and NDA were positive, whereas the causal mediated effect due to BMI was negative; this resulted in a proportion of the total mediated effect of 22% (95% CI: 1 to

Table 5: Estimation of the regression coefficients. For example, the model for the Box-Cox transformed DA is $t(DA) \sim 10.09HRT - 0.22AGE$. Note that we have a logistic regression for BC.

Model		HRT	AGE	t(DA)	t(NDA)	t(BMI)
1	t(DA)	10.09 (5.38e-08)	−0.22(0.175)	–	–	–
2	t(NDA)	−4.80(5.07e-4)	0.60(1.08e-06)	–	–	–
3	t(BMI)	−4.901e-04(3.19e-2)	1.948e-05(0.33)	–	–	–
4.a	BC	0.49(2.52e-3)	3.757e-3(0.78)	–	–	–
4.b	BC	0.39(2e-2)	0.01(0.44)	0.01(4.93e-05)	−0.01(0.02)	102.14(2.09e-4)

Table 6: Multiple mediation analysis for $T \in \{0,1\}$ (i.e., never versus former/current HRT users).

	Estimate	95%IC
$\widehat{\delta}^{DA}$	2.51e-2	[1.21e-2; 4.14e-2]
$\widehat{\delta}^{NDA}$	1.22e-2	[0.19e-2; 2.55e-2]
$\widehat{\delta}^{BMI}$	−1.49e-2	[−3.05e-2; −0.38e-2]
$\widehat{\delta}^{Z}$	2.24e-2	[0.14e-2; 4.39e-2]
$\widehat{PM}^{Z}$	21.54e-2	[1.19e-2; 63.02e-2]
$\widehat{\zeta}$	8.00e-2	[1.60e-2; 14.71e-2]
$\widehat{\tau}$	10.24e-2	[3.58e-2; 16.60e-2]

63%). Our finding that the effect of HRT is partially mediated by MD is consistent with previous reports in the literature [35, 36]. So does the negative sign of the mediated effect by BMI [37, 38]. MacKinnon et al. [39] described a situation with opposite signs mediated effects as inconsistent mediation models, as the effects may cancel out each other. In the present case, the negative mediated effect of BMI is not large enough to make the relation between HRT and BC non-significant.

6 Discussion

This article addresses the problem of estimating direct and indirect effects, including indirect effects through individual mediators, in the framework of multiple mediation with uncausally related mediators. Theoretical work of Shpitser and coauthors proved that in presence of latent variables not all mediation quantities are identified [17, 18]. In particular, in the presence of a latent common cause between the mediators, indirect effects trough individual mediators cannot be expressed as functions of the observable data only. On the other hand, a common practice in multiple mediation is to perform several simple mediation analyses, one for each mediator, despite the introduction of a bias.

Most of the approaches to mediation analysis are based on strong assumptions such as Sequential Ignorability [11,40], and several authors have tried to address the problem through different techniques. In the framework of multiple mediation with uncausally related mediators, we define a set of hypotheses, called SIMMA, under which we express the direct and the joint indirect effect as functions of observed variables and the indirect effect through individual mediators in terms of both observed and counterfactual variables. Coupled to a choice of model and the quasi-Bayesian algorithm developed by [11]; the latter formula gives an estimation method for the individual indirect effects. Note that we restricted ourselves to models with the additional hypothesis that the correlation between counterfactual mediators is the same whatever the treatment governing them. The development of methods for addressing the situation in which this additional hypothesis is violated is left to future work, together with the development of a sensitivity analysis for assessing the robustness to departures from SIMMA.

The method is implemented in R. Currently our program makes it possible to work with parametric models with continuous or ordered categorical mediators and continuous or binary outcomes. A package has been published on Github.

We applied our R program to validate the proposed method empirically. This simulation study shows that our method provides an unbiased estimate of the direct effect, while, as expected, estimates obtained by running simple mediation analyses one mediator at the time are biased, even in the case of independent mediators. Moreover, when mediators share an unobserved common cause, we show that our multiple analysis provide estimates of the direct effects through individual mediators that are less biased than the ones obtained from simple analyses one mediator at the time. The reason behind this improvement, is that our method, by considering the joint law of the mediators conditionally on the treatment and the law of the outcome conditionally on all the mediators, automatically takes into account the influence that the unobserved common cause U has on the mediators and the outcome. On the contrary, doing a simple analysis one mediator at the time is not appropriate in this setting because U confounds the relationship between each mediator and the outcome. Moreover, we show empirically that, contrary to repeated simple analyses, the proposed quasi-Bayesian algorithm provides confidence intervals with the expected coverage property.

Repeated individual mediator analyses are still a popular approach despite a growing literature warning about its limitations. Indeed, the presence of an unobserved common cause for the mediators is not the only situation in which such an approach is problematic. VanderWeele and Vansteelandt [14] observed that, even when mediators are uncausally related, it is not possible to decompose the joint indirect effect in the sum of individual indirect effects when their effect on the outcome is characterised by an interaction in the additive scale, a situation we excluded in our theoretical results. In this situation, [41] provided a three way decomposition of the joint indirect effect into

individual natural indirect effects and an interactive effect. Interestingly, the assumptions required to show the identifiability of all the terms in this decomposition are similar to ours, with the only important difference that potential mediators are assumed to be conditionally independent given all observed covariates. More recently, [42] provided a decomposition of the total effect in the more general situation with both mediator-mediator and mediators-outcome interactions.

Another important setting where repeating simple analyses is the wrong approach to multiple mediation is when mediators are causally ordered as in Figure 2(b). In this situation, considering the vector of intermediate variables as one mediator and conducting a simple analysis will correctly estimate the joint indirect effect and the direct effect. However the former joint indirect effect is not equal to the sum of the individual indirect effects, each estimated with a simple analysis, because some paths are counted twice and the effect mediated by W is biased by M which acts as a posttreatment confounder of the W–Y relationship. More generally, unless strong conditions hold, it is not possible to identify all specific paths [43]. VanderWeele and Vansteelandt [14] introduced a sequential approach to identify the joint indirect effect, the direct effect, the effect mediated by M and the effect mediated by W but not M. The different steps in this strategy can be implemented using medflex, a recently introduced R package based on the *natural effect model* and imputation or weighting methods [44]. An alternative approach based on linear structural equations with varying coefficients was discussed by [19] and implemented in the mediation package. Nguyen et al. [45] presented a method based on the Inverse Odds Ratio Weighting (IOWR) approach introduced by [46]. This method is very flexible as it accommodates generalized linear models, quantile regression and survival models for the outcome and multiple continuous or categorical mediators; however, it does not allow to estimate the indirect effect through individual mediators, but only the joint indirect effect.

We conclude this brief overview of the literature around multiple mediation by underlining that our framework deals with *natural* direct and indirect effects. Vansteelandt and Daniel [47] recently introduced so-called *interventional* direct and path specific indirect effects that do add up to the total effect and are identifiable even when the mediators share unmeasured common causes or the causal dependence between mediators is unknown.

As an illustration of our method, we conducted a multiple mediation analysis on a real dataset from a large cohort to assess the effect of hormone replacement treatment on breast cancer risk through three non-sequential mediators, namely dense mammographic area, non-dense area and body mass index. The causal effects that we have estimated and reported can be interpreted as risk differences, that is differences in percentage points. For a binary outcome, it is however often preferred to measure risk changes in terms of odds ratios (OR). In a parallel work in progress aimed at the epidemiological community, we expand on the application of Section 5 and work out a method to compute the causal effects of interest in the OR scale following the definition by [8].

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: Allan Jérolon was supported by Fondation Sciences Mathématiques de Paris (FSMP) and the DIM Math Innov Région île de France funding programme.
Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Appendices

A Link between δ^Z and $\sum_k \delta^k$

Even though intuitively it would sound reasonable to think that the indirect effect via the kth mediator δ^k is the difference between the joint effect δ^Z and the indirect effect by all other mediators η^k, we show that this is not true in general.

We want to express δ^Z according to $\sum_{k=1}^{K} \delta^k$. To do so, we start from δ^k:

$$
\begin{aligned}
\delta^k(t) &= E\left[Y\left(t, M^k(1), W^k(t)\right) - Y\left(t, M^k(0), W^k(t)\right)\right] \\
&= \begin{cases} E\left[Y(1, Z(1)) - Y\left(1, M^k(0), W^k(1)\right)\right] & \text{if} \quad t = 1 \\ E\left[Y\left(0, M^k(1), W^k(0)\right) - Y(0, Z(0))\right] & \text{if} \quad t = 0 \end{cases} \\
&= \begin{cases} E\left[\tau + Y(0, Z(0)) - Y\left(1, M^k(0), W^k(1)\right)\right] & \text{if} \quad t = 1 \\ E\left[Y\left(0, M^k(1), W^k(0)\right) + \tau - Y(1, Z(1))\right] & \text{if} \quad t = 0 \end{cases} \\
&= \begin{cases} E\left[\tau + Y(1, Z(0)) - \zeta(0) - Y\left(1, M^k(0), W^k(1)\right)\right] & \text{if} \quad t = 1 \\ E\left[Y\left(0, M^k(1), W^k(0)\right) + \tau - \zeta(1) - Y(0, Z(1))\right] & \text{if} \quad t = 0 \end{cases} \\
&= \begin{cases} E\left[\delta^Z(1) - Y\left(1, M^k(0), W^k(1)\right) + Y(1, Z(0))\right] & \text{if} \quad t = 1 \\ E\left[\delta^Z(0) - Y(0, Z(1)) + Y\left(0, M^k(1), W^k(0)\right)\right] & \text{if} \quad t = 0 \end{cases} \\
&= \delta^Z(t) - \eta^k(t).
\end{aligned} \tag{1}
$$

η^k may be interpreted as the indirect effect by all mediators except the kth, when the treatment is fixed at t and the kth mediator is set to the value it would have under treatment $1 - t$. Summing over the K mediators, we have:

$$
\begin{aligned}
\sum_{k=1}^{K} \delta^k(t) &= \sum_{k=1}^{K} \left(\delta^Z(t) - \eta^k(t)\right) \\
&= K\delta^Z(t) - \sum_{k=1}^{K} \eta^k(t)
\end{aligned}
$$

Thus the joint indirect effect can be rewritten as:

$$
\delta^Z(t) = \frac{\sum_{k=1}^{K} \left(\delta^k(t) + \eta^k(t)\right)}{K}.
$$

B Assumptions

According to [19], the Sequential Ignorability Assumption in the situation of multiple mediators that are causally unrelated is:

$$
\{Y(t, m, w), M(t'), W(t'')\} \perp\!\!\!\perp T \big| X = x, \tag{B.1}
$$

$$
Y(t', m, W(t')) \perp\!\!\!\perp M(t) \big| T = t, X = x, \tag{B.2}
$$

$$
Y(t', M(t'), w) \perp\!\!\!\perp W(t) \big| T = t, X = x, \tag{B.3}
$$

where $\mathbb{P}(T = t|X = x) > 0$ et $\mathbb{P}(M = m, W = w|T = t, X = x) > 0$ for all x, t, t', m, w.

We replace assumptions (B.2) and (B.3) with the hypotheses

$$
Y(t', m, w) \perp\!\!\!\perp (M(t), W(t)) | T = t, X = x \tag{B.4}
$$

$$
Y(t, m, w) \perp\!\!\!\perp (M(t'), W(t)) | T = t, X = x \tag{B.5}
$$

to obtain the *Sequential Ignorability for Multiple Mediators Assumption (SIMMA)*:

$$
(Y(t, m, w), M(t'), W(t'')) \perp\!\!\!\perp T | X = x, \tag{B.1}
$$

$$
Y(t', m, w) \perp\!\!\!\perp (M(t), W(t)) | T = t, X = x \tag{B.4}
$$

$$
Y(t, m, w) \perp\!\!\!\perp (M(t'), W(t)) | T = t, X = x \tag{B.5}
$$

1 In fact $\tau = \delta^Z(t) + \zeta(1 - t)$.

It is important to stress that (B.4) does not imply (B.2) and (B.3): a consequence of (B.4) is that $Y(t', m, w) \perp\!\!\!\perp M(t) | T = t, X = x$ for all values of t', m, w and x but this does not imply that $Y(t', m, W(t')) \perp\!\!\!\perp M(t) | T = t, X = x$ because $Y(t',m,w)$ depends only on the residual causes of Y after setting $T = t', M = m$ and $W = w$ while $Y(t', m, W(t'))$ depends also on the residual causes of W after setting T to t'.

C Proof of Theorem 3.1

C.1 Joint indirect effect, direct effect and total effect

In order to demonstrate Theorem 3.1 for the joint indirect effect δ^Z, the direct effect ζ and the total effect τ, we start by rewriting the definitions in terms of counterfactuals:

$$\begin{aligned}\delta^Z(t) &= \mathbb{E}[Y(t, Z(1))] - \mathbb{E}[Y(t, Z(0))] \quad = \int \mathbb{E}[Y(t, Z(1))|X = x] - \mathbb{E}[Y(t, Z(0))|X = x] \mathrm{d}F_X(x) \\ &= \int \mathbb{E}[Y(t, M(1), W(1))|X = x] - \mathbb{E}[Y(t, M(0), W(0))|X = x] \mathrm{d}F_X(x)\end{aligned}$$

$$\begin{aligned}\zeta(t, \ldots, t) &= \mathbb{E}[Y(1, Z(t))] - \mathbb{E}[Y(0, Z(t))] \quad = \int \mathbb{E}[Y(1, Z(t))|X = x] - \mathbb{E}[Y(0, Z(t))|X = x] \mathrm{d}F_X(x). \\ &= \int \mathbb{E}[Y(1, M(t), W(t))|X = x] - \mathbb{E}[Y(0, M(t), W(t))|X = x] \mathrm{d}F_X(x)\end{aligned}$$

$$\begin{aligned}\tau &= \mathbb{E}[Y(1, Z(1))] - \mathbb{E}[Y(0, Z(0))] \quad = \int \mathbb{E}[Y(1, Z(1))|X = x] - \mathbb{E}[Y(0, Z(0))|X = x] \mathrm{d}F_X(x) \\ &= \int \mathbb{E}[Y(1, M(1), W(1))|X = x] - \mathbb{E}[Y(0, M(0), W(0))|X = x] \mathrm{d}F_X(x)\end{aligned}$$

It is then sufficient to demonstrate that:

$$\mathbb{E}[Y(t, M(t'), W(t'))|X = x] = \int_{\mathbb{R}^K} \mathbb{E}[Y|T = t, M = m, W = w, X = x] \mathrm{d}F_{(M,W)|T=t', X=x}(m, w).$$

It will then follow that:

$$\mathbb{E}[Y(t, Z(t'))|X = x] = \int_{\mathbb{R}^K} \mathbb{E}[Y|T = t, Z = z, X = x] \mathrm{d}F_{Z|T=t', X=x}(z)$$

We have:

$$\begin{aligned}\mathbb{E}\left[Y(t, M(t'), W(t'))|X = x\right] &= \int_{\mathbb{R}^K} \mathbb{E}\left[Y(t, M(t'), W(t'))\middle|M(t') = m, W(t') = w, X = x\right] \mathrm{d}F_{(M(t'),W(t'))|X=x}(m, w) \\ &= \int_{\mathbb{R}^K} \mathbb{E}\left[Y(t, m, w)\middle|M(t') = m, W(t') = w, X = x\right] \mathrm{d}F_{(M(t'),W(t'))|X=x}(m, w) \\ &= \int_{\mathbb{R}^K} \mathbb{E}\left[Y(t, m, w)\middle|T = t', M(t') = m, W(t') = w, X = x\right] \mathrm{d}F_{(M(t'),W(t'))|X=x}(m, w) \quad (2) \\ &= \int_{\mathbb{R}^K} \mathbb{E}\left[Y(t, m, w)\middle|T = t', X = x\right] \mathrm{d}F_{(M(t'),W(t'))|X=x}(m, w) \quad (3) \\ &= \int_{\mathbb{R}^K} \mathbb{E}[Y(t, m, w)|T = t, X = x] \mathrm{d}F_{(M(t'),W(t'))|T=t', X=x}(m, w) \quad (4) \\ &= \int_{\mathbb{R}^K} \mathbb{E}\left[Y(t, m, w)\middle|T = t, M(t) = m, W(t) = w, X = x\right] \mathrm{d}F_{(M,W)|T=t', X=x}(m, w) \quad (5) \\ &= \int_{\mathbb{R}^K} \mathbb{E}[Y|T = t, M = m, W = w, X = x] \mathrm{d}F_{(M,W)|T=t', X=x}(m, w).\end{aligned}$$

2 By (B.1) and the weak union property: $Y(t, m, w) \perp\!\!\!\perp T | M(t'), W(t), X = x$

3 By (B.4).

Note that in this proof we have only used assumptions (B.1) and (B.4).

C.2 Indirect effect via the mediator of interest

It follows from the definition that:

$$\begin{aligned}\delta(t) &= \mathbb{E}[Y(t, M(1), W(t))] - \mathbb{E}[Y(t, M(0), W(t))] \\ &= \int \mathbb{E}[Y(t, M(1), W(t))|X = x] - \mathbb{E}[Y(t, M(0), W(t))|X = x] \mathrm{d}F_X(x).\end{aligned}$$

It is then sufficient to demonstrate that:

$$\mathbb{E}[Y(t, M(t'), W(t))|X = x] = \int_{\mathbb{R}^K} \mathbb{E}[Y|T = t, M = m, W = w, X = x] \mathrm{d}F_{(M(t'),W(t))|X=x}(m, w).$$

We have:

$$\begin{aligned}\mathbb{E}\left[Y(t, M(t'), W(t))|X = x\right] &= \int_{\mathbb{R}^K} \mathbb{E}\left[Y(t, M(t'), W(t))\middle|M(t') = m, W(t) = w, X = x\right] \mathrm{d}F_{(M(t'),W(t))|X=x}(m, w) \\ &= \int_{\mathbb{R}^K} \mathbb{E}\left[Y(t, m, w)\middle|T = t, M(t') = m, W(t) = w, X = x\right] \mathrm{d}F_{(M(t'),W(t))|X=x}(m, w) \quad (6) \\ &= \int_{\mathbb{R}^K} \mathbb{E}[Y(t, m, w)|T = t, X = x] \mathrm{d}F_{(M(t'),W(t))|,X=x}(m, w) \quad (7) \\ &= \int_{\mathbb{R}^K} \mathbb{E}[Y(t, m, w)|T = t, M(t) = m, W(t) = w, X = x] \mathrm{d}F_{(M(t'),W(t))|X=x}(m, w) \quad (8) \\ &= \int_{\mathbb{R}^K} \mathbb{E}[Y|T = t, M = m, W = w, X = x] \mathrm{d}F_{(M(t'),W(t))|X=x}(m, w).\end{aligned}$$

Note that in this proof we have used all SIMMA assumptions. In the case, where M and W are independent, we have:

$$\begin{aligned}\mathrm{d}F_{(M(t'),W(t))|X=x}(m, w) &= f_{(M(t'),W(t))|X=x}(m, w)\mathrm{d}m\mathrm{d}w = f_{M(t')|X=x}(m)\mathrm{d}m f_{W(t)|X=x}(w)\mathrm{d}w \\ &= f_{M|T=t',X=x}(m)\mathrm{d}m f_{W|T=t,X=x}(w)\mathrm{d}w\end{aligned}$$

and therefore:

$$\begin{aligned}\delta(t) &= \int\int_{\mathbb{R}^K} \mathbb{E}[Y|T = t, M = m, W = w, X = x] \\ &\quad \left\{f_{M|T=1,X=x}(m)\mathrm{d}m f_{W|T=t,X=x}(w)\mathrm{d}w - f_{M|T=0,X=x}(m)\mathrm{d}m f_{W|T=t,X=x}(w)\mathrm{d}w\right\}\mathrm{d}F_X(x) \\ &= \int\int\int_{\mathbb{R}^{K-1}} \mathbb{E}[Y|T = t, M = m, W = w, X = x] f_{W|T=t,X=x}(w)\mathrm{d}w\left\{f_{M|T=1,X=x}(m) - f_{M|T=0,X=x}(m)\right\}\mathrm{d}m\mathrm{d}F_X(x) \\ &= \int\int \mathbb{E}[Y|T = t, M = m, X = x]\left\{f_{M|T=1,X=x}(m) - f_{M|T=0,X=x}(m)\right\}\mathrm{d}m\mathrm{d}F_X(x) \\ &= \int\int \mathbb{E}[Y|T = t, M = m, X = x]\{\mathrm{d}F_{M|T=1,X=x}(m) - \mathrm{d}F_{M|T=0,X=x}(m)\}\mathrm{d}F_X(x).\end{aligned}$$

D Models

We give here the models used for the simulation study in Section 4.3.

Model 1: Continuous outcome and continuous mediators

4 By (B.1).
5 By (B.4) with $t' = t$.
6 By (B.1) and the weak union property.
7 By (B.5).
8 By (B.4) with $t = t'$.

– T follows a Bernoulli distribution $\mathscr{B}(0.3)$
– the joint distribution of the counterfactual mediators is

$$\begin{pmatrix} M^1(1) \\ M^1(0) \\ M^2(1) \\ M^2(0) \end{pmatrix} \sim \mathscr{N}\left(\mu = \begin{pmatrix} 1+4\times 1 \\ 1+4\times 0 \\ 2+6\times 1 \\ 2+6\times 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 1 & \rho & \rho \\ 1 & 1 & \rho & \rho \\ \rho & \rho & 1 & 1 \\ \rho & \rho & 1 & 1 \end{pmatrix}\right)$$

– the counterfactual outcomes follow the normal distributions:

$$Y\left(t, M^1(t'), M^2(t'')\right) \sim \mathscr{N}\left(1 + 10t + 5M^1(t') + 4M^2(t''), 1\right).$$

In Table 7, we show the real causal effect values entailed by model 1.

Table 7: Real values of the causal effects entailed by model 1.

δ^Z	δ^1	δ^2	ζ	τ
44	20	24	10	54

Model 2: Binary outcome (logit) with continuous mediators

– T follows a Bernoulli distribution $\mathscr{B}(0.3)$
– the joint distribution of the counterfactual mediators is:

$$\begin{pmatrix} M^1(1) \\ M^1(0) \\ M^2(1) \\ M^2(0) \end{pmatrix} \sim \mathscr{N}\left(\mu = \begin{pmatrix} 0.1+0.6\times 1 \\ 0.1+0.6\times 0 \\ 0.2+0.8\times 1 \\ 0.2+0.8\times 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 1 & \rho & \rho \\ 1 & 1 & \rho & \rho \\ \rho & \rho & 1 & 1 \\ \rho & \rho & 1 & 1 \end{pmatrix}\right)$$

– the counterfactual outcomes follow the logistic distributions:

$$Y\left(t, M^1(t'), M^2(t'')\right) \sim B\left(\frac{1}{1+\exp\left(-2+0.4t+0.6M^1(t')+0.8M^2(t'')\right)}\right).$$

With this choice of parameters, 30% of the sampled observations are cases. As we can see in Corollary 3.3, with binary outcome, causal effects are related to the covariance of mediators. Figure 12 shows how the true causal values change when correlation changes.

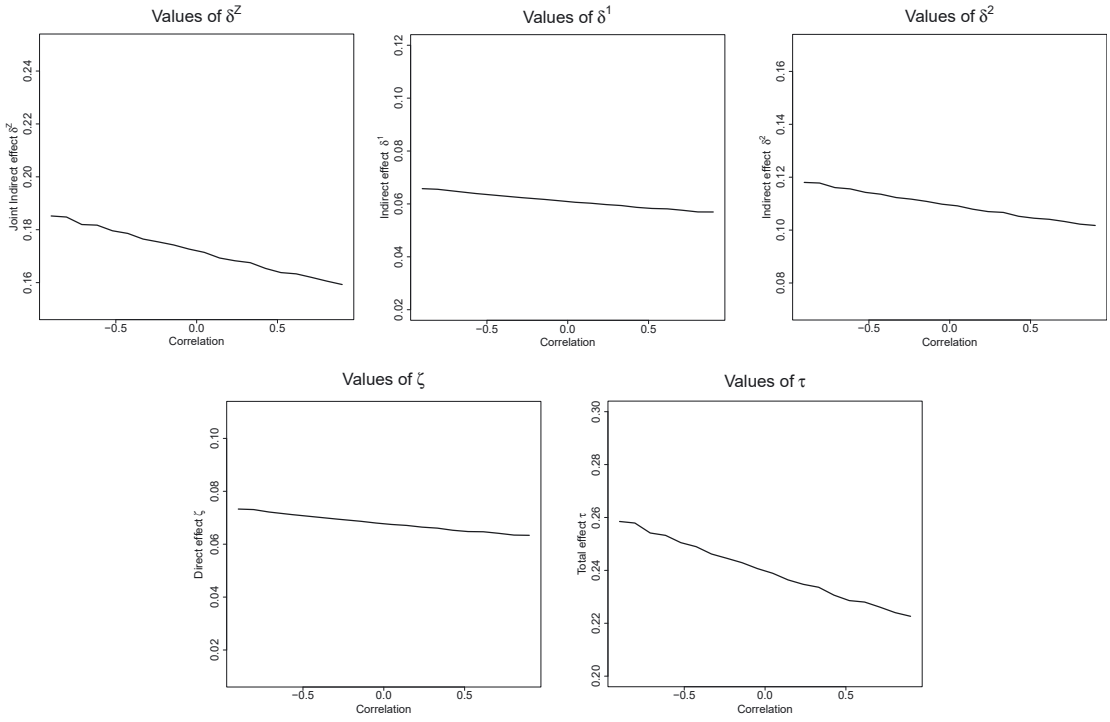


Figure 12: Binary outcome (model 2): variation in causal effects due to correlation.

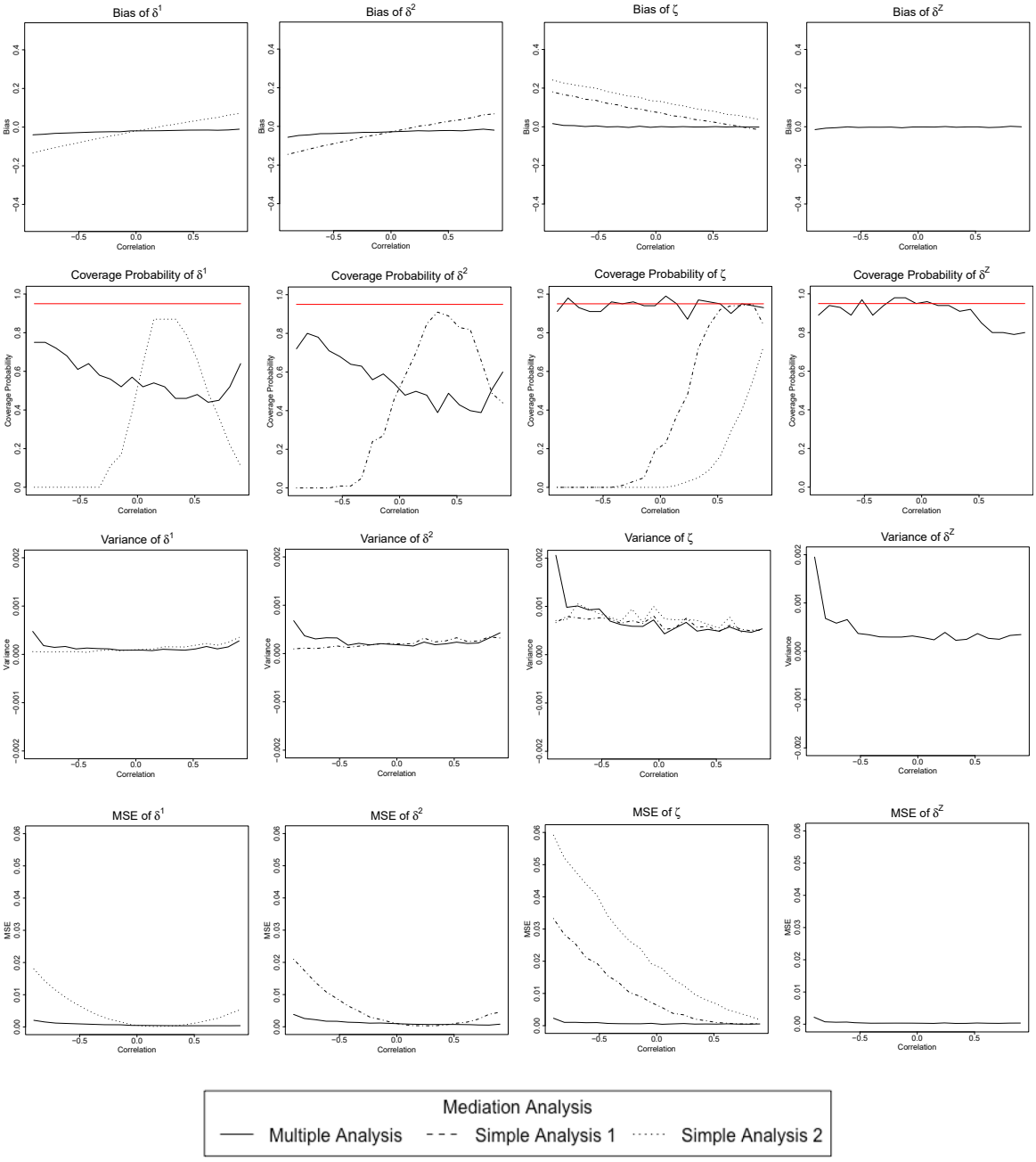


Figure 13: Model 2 (binary outcome): bias, coverage probability, variance, and MSE of mediation effect estimators when the correlation between mediators varies. These results have been obtained with 200 simulations. Each simulation consists in a dataset of size 1000.

E Normalization of mediators using Box-Cox likelihood-like approach

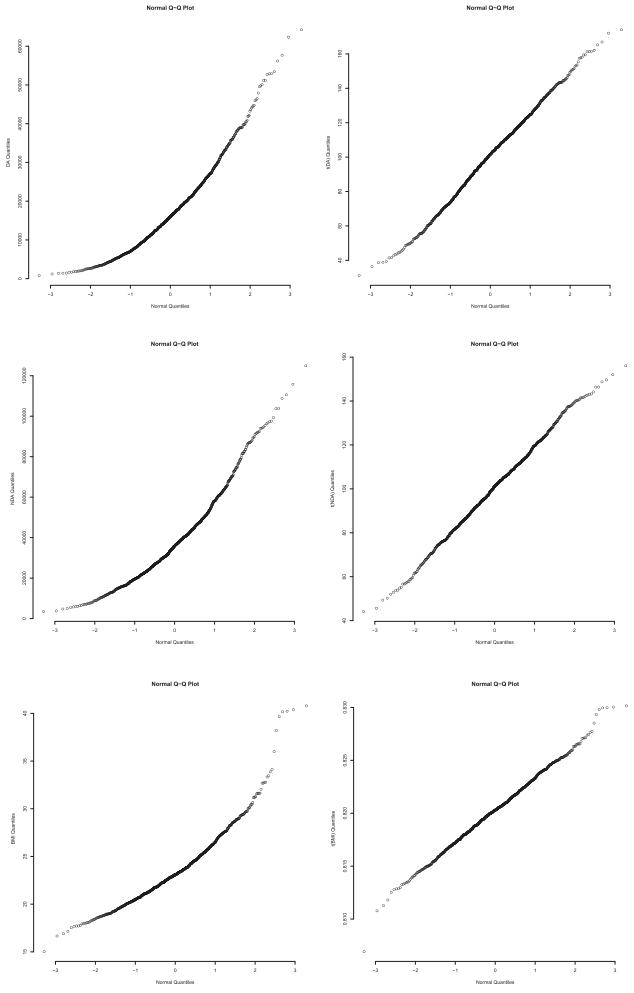


Figure 14: Normal q–q plots of mediators before and after transformation.

References

1. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986;51:1173–82.
2. James L, Mulaik S, Brett JM. Causal analysis: assumptions, models, and data. Acad Manag Rev 1982;9 https://doi.org/10.5465/amr.1984.4278125.
3. MacKinnon D. An introduction to statistical mediation analysis. New York: Lawrence Erlbaum Associates/Taylor & Francis Group; 2008:245 p.
4. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology 1992;3:143–55.
5. Pearl J. Direct and indirect effects. In: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, UAI'01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2001: 411–20 pp.
6. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. Epidemiology 2006;17:276–84.
7. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Stat Interface 2009;2:457–68.
8. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. Am J Epidemiol 2010;172:1339–48.

9. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. Am J Epidemiol 2012;176:190–5.
10. VanderWeele T. Explanation in causal inference: methods for mediation and interaction. Oxford: Oxford University Press; 2015.
11. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods 2010a;15:309–34.
12. Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. Stat Sci 2010b;25: 51–71.
13. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. J Stat Software 2014;59. https://doi.org/10.18637/jss.v059.i05.
14. VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. Epidemiol Methods 2014;2. https://doi.org/10.1515/em-2012-0010.
15. Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect effects through multiple pathways. Am J Epidemiol 2014;179:513–18.
16. Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. Biometrics 2015; 71:1–14.
17. Shpitser I. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. Cognit Sci 2013;37:1011–35.
18. Shpitser I, Sherman E. Identification of personalized effects associated with causal pathways. Uncertain Artif Intell 2018: 530–9.
19. Imai K, Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. Polit Anal 2013;21:141–71.
20. Kim C, Daniels MJ, Hogan JW, Choirat C, Zigler CM. Bayesian methods for multiple mediators: relating principal stratification and causal mediation in the analysis of power plant emission controls. Ann Appl Stat 2019;13:1927–56.
21. Rubin DB. Randomization analysis of experimental data: the Fisher randomization test comment. J Am Stat Assoc 1980;75:591.
22. Forastiere L, Mealli F, VanderWeele TJ. Identification and estimation of causal mechanisms in clustered encouragement designs: disentangling bed nets using Bayesian principal stratification. J Am Stat Assoc 2016;111:510–25.
23. Clavel-Chapelon F. Cohort profile: the French E3n cohort study. Int J Epidemiol 2015;44:801–9.
24. Binachon B, Dossus L, Danjou AMN, Clavel-Chapelon F, Fervers B. Life in urban areas and breast cancer risk in the French E3N cohort. Eur J Epidemiol 2014;29:743–51.
25. Miller VM, Harman SM. An update on hormone therapy in postmenopausal women: mini-review for the basic scientist. Am J Physiol Heart Circ Physiol 2017;313:H1013–21.
26. McTiernan A, Martin CF, Peck JD, et al. Estrogen-plus-progestin use and mammographic density in postmenopausal women: women's health initiative randomized trial. JNCI J Natl Cancer Inst 2005;97:1366–76.
27. Kim S, Ko Y, Lee HJ, Lim J-e. Menopausal hormone therapy and the risk of breast cancer by histological type and race: a meta-analysis of randomized controlled trials and cohort studies. Breast Canc Res Treat 2018;170:667–75.
28. Baglietto L, Krishnan K, Stone J, et al. Associations of mammographic dense and nondense areas and body mass index with risk of breast cancer. Am J Epidemiol 2014;179:475–83.
29. Maskarinec G, Dartois L, Delaloge S, Hopper J, Clavel-Chapelon F, Baglietto L. Tumor characteristics and family history in relation to mammographic density and breast cancer: the French E3n cohort. Cancer Epidemiol 2017;49:156–60.
30. Sung J, Song Y-M, Stone J, Lee K, Kim S-Y. Association of body size measurements and mammographic density in Korean women: the healthy twin study. Cancer Epidemiol Biomark Prev 2010;19:1523–31.
31. Boyd NF, Martin LJ, Sun L, et al. Body size, mammographic density, and breast cancer risk. Cancer Epidemiol Biomark Prev 2006;15:2086–92.
32. Wade TD, Zhu G, Martin NG. Body mass index and breast size in women: same or different genes?. Twin Res Hum Genet 2010; 13:450–4.
33. Ooi BNS, Loh H, Ho PJ, et al. The genetic interplay between body mass index, breast size and breast cancer risk: a Mendelian randomization analysis. Int J Epidemiol 2019;48:781–94.
34. Box GEP, Cox DR. An analysis of transformations. J Roy Stat Soc B 1964;26:211–52.
35. Rice MS, Tamimi RM, Bertrand KA, et al. Does mammographic density mediate risk factor associations with breast cancer? An analysis by tumor characteristics. Breast Canc Res Treat 2018;170:129–41.
36. Azam S, Lange T, Huynh S, et al. Hormone replacement therapy, mammographic density, and breast cancer risk: a cohort study. Cancer Causes & Control 2018;29:495–505.
37. Salpeter SR, Walsh JME, Ormiston TM, Greyber E, Buckley NS, Salpeter EE. Meta-analysis: effect of hormone-replacement therapy on components of the metabolic syndrome in postmenopausal women. Diabetes Obes Metabol 2006;8:538–54.
38. Cheraghi Z, Poorolajal J, Hashem T, Esmailnasab N, Doosti Irani A. Effect of body mass index on breast cancer during premenopausal and postmenopausal periods: a meta-analysis. PLoS One 2012;7:e51446.
39. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol 2007;58:593–614.
40. Hafeman DM, VanderWeele TJ. Alternative assumptions for the identification of direct and indirect effects. Epidemiology 2011; 22:753–64.

41. Taguri M, Featherstone J, Cheng J. Causal mediation analysis with multiple causally non-ordered mediators. Stat Methods Med Res 2015;27:3–19.
42. Bellavia A, Valeri L. Decomposition of the total effect in the presence of multiple mediators and interactions. Am J Epidemiol 2017;187:1311–18.
43. Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. In: IJCAI International Joint Conference on Artificial Intelligence; 2005: 357–63 pp.
44. Steen J, Loeys T, Moerkerke B, Vansteelandt S. Medflex: an r package for flexible mediation analysis using natural effect models. J Stat Software 2017;76. https://doi.org/10.18637/jss.v076.i11.
45. Nguyen QC, Osypuk TL, Schmidt NM, Glymour MM, Tchetgen Tchetgen EJ. Practical guidance for conducting mediation analysis with multiple mediators using inverse odds ratio weighting. Am J Epidemiol 2015;181:349–56.
46. Tchetgen Tchetgen EJ. Inverse odds ratio-weighted estimation for causal mediation analysis. Stat Med 2013;32:4567–80.
47. Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. Epidemiology 2017;28:258.

Supplementary material: The online version of this article offers supplementary material (https://doi.org/10.1515/ijb-2019-0088).