Supplementary Materials

1 Definition of counterfactuals

For the sake of completeness, we briefly review basic facts about counterfactuals, using the same notations as in Wasserman (2010). Without loss of generality, we consider a binary treatment T as it is often the case in epidemiology. We introduce the potential outcomes $Y(1)$ et $Y(0)$. $Y(1)$ is the potential outcome status that would result under the treatment and $Y(0)$ the potential status under the absence of treatment. The relation between the outcome and the potential outcome is:

$$Y = \begin{cases} Y(0) & \text{if} \quad T = 0 \\ Y(1) & \text{if} \quad T = 1 \end{cases} = Y(T)$$

Note that, according to this *consistency* relationship, only one of the two potential outcomes is observable, the other is called counterfactual and corresponds to the random value that would have governed Y if T was fixed to the other value.
To define causal mediation effects we extend the previous notions. Consider the causal diagram in Figure 1.

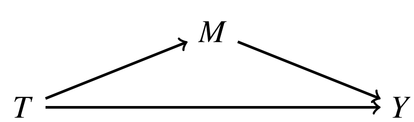

Figure 1: Simple mediation model with one mediator M and no confounding covariates.

As M is a post-treatment variable, we introduce two potential mediators, $M(0)$ and $M(1)$:

$$M = \begin{cases} M(0) & \text{if} \quad T = 0 \\ M(1) & \text{if} \quad T = 1 \end{cases} = M(T).$$

As Y is post-treatment and post-mediator, we introduce four counterfactuals. Two of them, namely $Y(0, M(0))$ and $Y(1, M(1))$ can be observed while two others, $Y(0, M(1))$ et $Y(1, M(0))$, cannot. The interpretation of the observed ones is trivial:

$$Y = \left\{ \begin{array}{lll} Y(0,M(0)) & \text{if} & T=0 \\ Y(1,M(1)) & \text{if} & T=1 \end{array} \right. = Y(T,M).$$

The interpretation of the two others is more complex: for $t \neq t'$, $Y(t,M(t'))$ is the potential value of the outcome had the treatment been set to t and had the mediator been fixed at the level it would have had under treatment t'.

2 Assumptions

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

$$\begin{aligned} Y(t,m,w), M(t'), W(t'') &\perp\!\!\!\perp T|X=x, && (2.1) \\ Y(t',m,w) &\perp\!\!\!\perp (M(t),W(t))\,|T=t,X=x && (2.2) \\ Y(t,m,w) &\perp\!\!\!\perp \big(M(t'),W(t)\big)\,|T=t,X=x && (2.3) \end{aligned}$$

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.

3 Identifiability

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

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

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

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

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

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

$$\begin{aligned}\tau &= \int \left(\int_{\mathbb{R}^K} \mathbb{E}(Y|Z=z,T=1,X=x) \mathrm{d}F_{Z|T=1,X=x}(z) \right. \\ &\quad \left. - \int_{\mathbb{R}^K} \mathbb{E}(Y|Z=z,T=0,X=x) \mathrm{d}F_{Z|T=0,X=x}(z) \right) \mathrm{d}F_X(x).\end{aligned}$$

4 Continuous Outcome

Corollary 4.1 (Corollary 3.2 in the main article)**.** *With K mediators and P covariables we assume the following linear model*

$$Z = \alpha_2 + \beta_2 T + \xi_2^\Gamma X + \Upsilon_2 \tag{4.1}$$

$$Y = \alpha_3 + \beta_3 T + \gamma^\Gamma Z + \xi_3^\Gamma X + \varepsilon_3, \tag{4.2}$$

where $\alpha_2, \beta_2, \gamma \in \mathbb{R}^K$, $\xi_2^\Gamma \in \mathbb{R}^K \times \mathbb{R}^P$, $\xi_3^\Gamma \in \mathbb{R}^P$, and $\Upsilon_2 \sim \mathcal{N}(0, \Sigma_2)$ is the vector of residuals with covariance matrix $\Sigma_2 \in \mathbb{R}^K \times \mathbb{R}^K$ 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 not causally correlated. In the latter case, we assume that pairwise correlations between potential mediators do not depend on the treatments governing them:

$$cor\left(M^i(t), M^j(t')|T,X\right) = \rho_{ij}, \forall t,t' \in \{0,1\}, \forall i,j \in [1,K]. \tag{4.3}$$

Under SIMMA the indirect effect of the k-th 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 of the k-th mediator is also identified and given by

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

The coefficients of the equations (4.1) and (4.2) are identified under the assumptions of Theorem 3.1. Indeed by rewriting equations (4.1) and (4.2) with counterfactuals we have :

$$\begin{aligned} Z(T) &= \alpha_2^\Gamma + \beta_2^\Gamma T + \xi_2^\Gamma X + \Upsilon_2(T) & (4.4) \\ Y(T, Z(T)) &= \alpha_3 + \beta_3 T + \gamma^\Gamma Z(T) + \xi_3^\Gamma X + \varepsilon_3(T, Z(T)) & (4.5) \end{aligned}$$

where $\mathbb{E}[\Upsilon_2(t)] = \mathbb{E}[\varepsilon_3(t,z)] = 0$.
Assumptions (2.1) and (2.2) implies the following independences:

$$\begin{aligned} \Upsilon_2(t) &\perp\!\!\!\perp T|X = x \\ \varepsilon_3(t,z) &\perp\!\!\!\perp T|X = x. \end{aligned}$$

Assumption (2.2) implies the following independence:

$$\varepsilon_3(t,z) \perp\!\!\!\perp Z(t)|T = t, X = x.$$

We have

$$\begin{aligned} \mathbb{E}[Z(T)|T = t, X = x] &= \alpha_2^\Gamma + \beta_2^\Gamma t + \xi_2^\Gamma x + \mathbb{E}[\Upsilon_2(T)|T = t, X = x] \\ &= \alpha_2^\Gamma + \beta_2^\Gamma t + \xi_2^\Gamma x + \mathbb{E}[\Upsilon_2(t)|T = t, X = x] \\ &= \alpha_2^\Gamma + \beta_2^\Gamma t + \xi_2^\Gamma x + \mathbb{E}[\Upsilon_2(t)] \\ &= \alpha_2^\Gamma + \beta_2^\Gamma t + \xi_2^\Gamma x. \end{aligned}$$

Then coefficients of (4.1) are identified.

Furthermore

$$\begin{aligned}
\mathbb{E}[Y(T,Z(T))|Z=z,T=t,X=x] &= \alpha_3+\beta_3 t+\gamma^{\Gamma}z+\xi_3^{\Gamma}x \\
&\quad +\mathbb{E}[\varepsilon_3(T,Z(T))|Z=z,T=t,X=x] \\
&= \alpha_3+\beta_3 t+\gamma^{\Gamma}z+\xi_3^{\Gamma}x \\
&\quad +\mathbb{E}[\varepsilon_3(t,z)|Z=z,T=t,X=x] \\
&= \alpha_3+\beta_3 t+\gamma^{\Gamma}z+\xi_3^{\Gamma}x \\
&\quad +\mathbb{E}[\varepsilon_3(t,z)] \\
&= \alpha_3+\beta_3 t+\gamma^{\Gamma}z+\xi_3^{\Gamma}x.
\end{aligned}$$

Then coefficients of (4.2) are identified.

According to Theorem 3.1 we have:

$$\begin{aligned}
\delta^1(t) &= \int\int_{\mathbb{R}^K}\mathbb{E}[Y|M=m,W=w,T=t,X=x] \\
&\quad \{\mathrm{d}F_{(M(1),W(t))|X=x}(m,w)-\mathrm{d}F_{(M(0),W(t))|X=x}(m,w)\}\mathrm{d}F_X(x) \\
&= \int\int_{\mathbb{R}^K}\left(\alpha_3+\beta_3 t+\sum_{j=1}^{K}\gamma_j m^j+\xi_3^{\Gamma}x\right) \\
&\quad \{\mathrm{d}F_{(M(1),W(t))|X=x}(m,w)-\mathrm{d}F_{(M(0),W(t))|X=x}(m,w)\}\mathrm{d}F_X(x) \\
&= \int\int_{\mathbb{R}^K}\left(\alpha_3+\beta_3 t+\xi_3^{\Gamma}x\right)\{\mathrm{d}F_{(M(1),W(t))|X=x}(m,w)-\mathrm{d}F_{(M(0),W(t))|X=x}(m,w)\} \\
&\quad +\int_{\mathbb{R}^K}\sum_{j=1}^{K}\gamma_j m^j\{\mathrm{d}F_{(M(1),W(t))|X=x}(m,w)-\mathrm{d}F_{(M(0),W(t))|X=x}(m,w)\}\mathrm{d}F_X(x).
\end{aligned}$$

$$\begin{aligned}
\delta^1(t) &= \int\int_{\mathbb{R}^K}\sum_{j=1}^{K}\gamma_j m^j\{\mathrm{d}F_{(M(1),W(t))|X=x}(m,w)-\mathrm{d}F_{(M(0),W(t))|X=x}(m,w)\}\mathrm{d}F_X(x) \\
&= \int\int_{\mathbb{R}^K}\sum_{j=1}^{K}\gamma_j m^j\mathrm{d}F_{(M(1),W(t))|X=x}(m,w) \\
&\quad -\int_{\mathbb{R}^K}\sum_{k=1}^{K}\gamma_j m^j\mathrm{d}F_{(M(0),W(t))|X=x}(m,w)\mathrm{d}F_X(x).
\end{aligned}$$

By the following substitutions $m^1=\alpha_2^1+\beta_2^1 t'+\xi_2^{\Gamma 1}x+e_2^1$ and $m^j=\alpha_2^j+\beta_2^j t+\xi_2^{\Gamma j}x+e_2^j, \forall j\in[2,K]$, we have :

$$((M(t'),W(t))|X=x) \sim \mathcal{N}(\mu_{(t',t)},\Sigma_2) \text{ with } \mu_{(t',t)} = \begin{pmatrix} \alpha_2^1+\beta_2^1 t'+\xi_2^{\Gamma 1}x \\ \alpha_2^2+\beta_2^2 t+\xi_2^{\Gamma 2}x \\ \alpha_2^3+\beta_2^3 t+\xi_2^{\Gamma 3}x \\ \vdots \\ \alpha_2^K+\beta_2^K t+\xi_2^{\Gamma K}x \end{pmatrix}$$

and where it follows from assumption (4.3) that Σ_2 is the covariance matrix of ε_2^j.
It follows that :

$$\begin{aligned} f_{(M(t'),W(t))|X=x}(m,w) &= \frac{1}{\sqrt{2\pi}|\Sigma_2|^{1/2}} \exp\left(\frac{1}{2}\left(\begin{pmatrix} m \\ w \end{pmatrix} - \mu_{(t',t)}\right)\Sigma_2^{-1}\left(\begin{pmatrix} m \\ w \end{pmatrix} - \mu_{(t',t)}\right)\right) \\ &= \frac{1}{\sqrt{2\pi}|\Sigma_2|^{1/2}} \exp\left(\frac{1}{2}\begin{pmatrix} e_2^1 \\ \vdots \\ e_2^K \end{pmatrix}\Sigma_2^{-1}\begin{pmatrix} e_2^1 \\ \vdots \\ e_2^K \end{pmatrix}\right) \\ &= f_\Upsilon\begin{pmatrix} e_2^1 \\ \vdots \\ e_2^K \end{pmatrix}. \end{aligned}$$

By plugging the joint density of the counterfactual mediators into the previous equation of the indirect effect $\delta^1(t)$ we have:

$$\begin{aligned} \delta^1(t) &= \int\int_{\mathbb{R}^K} \gamma_1\left(\alpha_2^1+\beta_2^1\times 1+\xi_2^{\Gamma 1}x+e_2^1\right)+\sum_{j=2}^K \gamma_j\left(\alpha_2^j+\beta_2^j t+\xi_2^{\Gamma j}x+e_2^j\right) \\ &\qquad \mathrm{d}F_\Upsilon\begin{pmatrix} e_2^1 \\ \vdots \\ e_2^K \end{pmatrix} \\ &\quad -\int_{\mathbb{R}^K} \gamma_1\left(\alpha_2^1+\beta_2^1\times 0+\xi_2^{\Gamma 1}x+e_2^1\right)+\sum_{j=2}^K \gamma_j\left(\alpha_2^j+\beta_2^j t+\xi_2^{\Gamma j}x+e_2^j\right) \\ &\qquad \mathrm{d}F_\Upsilon\begin{pmatrix} e_2^1 \\ \vdots \\ e_2^K \end{pmatrix}\mathrm{d}F_X(x) \\ &= \int\int_{\mathbb{R}^K} \gamma_1\beta_2^1 \mathrm{d}F_\Upsilon\begin{pmatrix} e_2^1 \\ \vdots \\ e_2^K \end{pmatrix}\mathrm{d}F_X(x) \\ &= \gamma_1\beta_2^1. \end{aligned}$$

We then conclude that in general we have for $k \in [1,K]$, $\delta^k(t) = \gamma_k \beta_2^k$.

$$\begin{aligned}
\zeta(t) &= \int\int \mathbb{E}(Y|Z=z,T=1,X=x) - \mathbb{E}(Y|Z=z,T=0,X=x) \\
&\quad \mathrm{d}F_{Z|T=t,X=x}(z)\mathrm{d}F_X(x) \\
&= \int\int \{\alpha_3 + \beta_3 + \gamma^\Gamma z + \xi_3^\Gamma x - \alpha_3 - \gamma^\Gamma z - \xi_3^\Gamma x\}\mathrm{d}F_{Z|T=t}(z)\,\mathrm{d}F_X(x) \\
&= \int\int \beta_3 \mathrm{d}F_{Z|T=t}(z)\,\mathrm{d}F_X(x) \\
&= \beta_3.
\end{aligned}$$

where $\gamma^\Gamma = (\gamma_1, \ldots, \gamma_K)^\Gamma$. Note that:

$$\begin{aligned}
\eta^1(t) &= \int\int_{\mathbb{R}^K} \mathbb{E}[Y|M=m,W=w,T=t,X=x] \\
&\quad \{\mathrm{d}F_{(M(1-t),W(1))|X=x}(m,w) - \mathrm{d}F_{(M(1-t),W(0))|X=x}(m,w)\}\mathrm{d}F_X(x) \\
&= \int\int_{\mathbb{R}^K} \left(\alpha_3 + \beta_3 t + \sum_{j=1}^K \gamma_j m^j + \xi_3^\Gamma x\right) \\
&\quad \{\mathrm{d}F_{(M(1-t),W(1))|X=x}(m,w) - \mathrm{d}F_{(M(1-t),W(0))|X=x}(m,w)\}\mathrm{d}F_X(x) \\
&= \int\int_{\mathbb{R}^K} \sum_{j=1}^K \gamma_j m^j \mathrm{d}F_{(M(1-t),W(1))|X=x}(m,w) \\
&\quad - \int_{\mathbb{R}^K} \sum_{j=1}^K \gamma_j m^j \mathrm{d}F_{(M(1-t),W(0))|X=x}(m,w)\mathrm{d}F_X(x)
\end{aligned}$$

By the following substitutions $m^1 = \alpha_2^1 + \beta_2^1(1-t) + \xi_2^{\Gamma 1} x + e_2^1$ and $m^j = \alpha_2^j + \beta_2^j t' + \xi_2^{\Gamma j} x + e_2^j, \forall j \in [2,K]$, we have:

$$\begin{aligned}
\eta^1(t) &= \int\int_{\mathbb{R}^K} \gamma_1 \left(\alpha_2^1 + \beta_2^1(1-t) + \xi_2^{\Gamma 1} x + e_2^1\right) \\
&\qquad + \sum_{j=2}^K \gamma_j \left(\alpha_2^j + \beta_2^j \times 1 + \xi_2^{\Gamma j} x + e_2^j\right) \mathrm{d}F_\Upsilon \begin{pmatrix} e_2^1 \\ \vdots \\ e_2^K \end{pmatrix} \\
&\quad - \int_{\mathbb{R}^K} \gamma_1 \left(\alpha_2^1 + \beta_2^1(1-t) + \xi_2^{\Gamma 1} x + e_2^1\right) \\
&\qquad + \sum_{j=2}^K \gamma_j \left(\alpha_2^j + \beta_2^j \times 0 + \xi_2^{\Gamma j} x + e_2^j\right) \mathrm{d}F_\Upsilon \begin{pmatrix} e_2^1 \\ \vdots \\ e_2^K \end{pmatrix} \mathrm{d}F_X(x),
\end{aligned}$$

and therefore

$$\begin{aligned}
\eta^1(t) &= \int\int_{\mathbb{R}^K} \sum_{j=2}^{K} \gamma_j \beta_2^j \mathrm{d}F_{\Upsilon} \begin{pmatrix} e_2^1 \\ \vdots \\ e_2^K \end{pmatrix} \mathrm{d}F_X(x) \\
&= \sum_{j=2}^{K} \gamma_j \beta_2^j \\
&= \sum_{j=2}^{K} \delta^j(t).
\end{aligned}$$

We conclude that in general for $k \in [1,K]$, $\eta^k(t) = \sum_{j=1, j\neq k}^{K} \delta^j(t)$. We have :

$$\begin{aligned}
\delta^Z(t) &= \frac{\sum_{k=1}^{K} \left(\delta^k(t) + \eta^k(t)\right)}{K} \\
&= \frac{\sum_{k=1}^{K} \left(\delta^k(t) + \sum_{j=1, j\neq k}^{K} \delta^j(t)\right)}{K} \\
&= \frac{K \sum_{k=1}^{K} \delta^k(t)}{K} \\
&= \sum_{k=1}^{K} \delta^k(t).
\end{aligned}$$

5 Binary outcome

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

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

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 5.1 (Corollary 3.3 in the main text). *Assume the following model with a binary outcome :*

$$
\begin{aligned}
Z &= \alpha_2 + \beta_2^\Gamma T + \xi_2^\Gamma X + \Upsilon_2, & (5.1)\\
Y^* &= \alpha_3 + \beta_3 T + \gamma^\Gamma Z + \xi_3^\Gamma X + \varepsilon_3, & (5.2)\\
Y &= \mathbb{1}_{\{Y^*>0\}} & (5.3)
\end{aligned}
$$

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 not 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 (4.3). Under SIMMA, the effects of interest are given by:

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

where for a probit regression we have

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

and for a logit regression we have

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

By injecting (5.1) in (5.2), we have:

$$Y^* = (\alpha_3 + \sum_{k=1}^{K}\gamma_k\alpha_2) + (\beta_3 + \sum_{k=1}^{K}\gamma_k\beta_2^k)T + (\varepsilon_3 + \sum_{k=1}^{K}\gamma_k\varepsilon_2^k).$$

We set $\varepsilon = \sum\limits_{k=1}^{K}\gamma_k\varepsilon_2^k$ and $U = \varepsilon_3 + \varepsilon$.

The errors of the mediators form a gaussian vector $\Upsilon = \begin{pmatrix}\varepsilon_2^1\\ \vdots \\ \varepsilon_2^K\end{pmatrix}$. This implies that $\varepsilon = \sum_{k=1}^{K}\gamma_k\varepsilon_2^k$ follows a Gaussian law.
First, let us determine the distribution function of U, denoted F_U for both a probit and logit model.

Probit modeling. In the case of a probit model, $\varepsilon_3 \sim \mathcal{N}(0, \sigma_3^2)$, and, because of equations (2.1) and (2.2), $\varepsilon_3 \perp\!\!\!\perp \Upsilon$. We therefore deduce that U follows a Gaussian law and the distribution functions of U are:

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

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

Note that $1 - F_U(-z) = F_U(z)$ by symmetry of the density function.

Logit modeling. In the case of a logit model, $\varepsilon_3 \sim \mathscr{L}(0,1)$ and $f_{\varepsilon_3}(z) = \frac{\exp(-z)}{(1+\exp(-z))^2}$. To determine the density function of U, consider a generic continuous function G with compact support. The density of U is then the function $f_U(u)$ such that $E[G(U)] = \int G(u) f_U(u) \mathrm{d}u$. According to (2.2), $\varepsilon_3 \perp\!\!\!\perp \Upsilon$ and then $\varepsilon_3 \perp\!\!\!\perp \varepsilon$.

$$\begin{aligned}
\mathbb{E}[G(U)] &= \mathbb{E}[G(\varepsilon + \varepsilon_3)] \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} G(e+e_3) f_\varepsilon(e) f_{\varepsilon_3}(e_3) \,\mathrm{d}e\mathrm{d}e_3.
\end{aligned}$$

We make the following substitutions:

$$\begin{aligned}
z &= e + e_3 \\
e &= z - e_3 \\
\mathrm{d}e &= \mathrm{d}z.
\end{aligned} \tag{5.4}$$

Then we have:

$$
\begin{aligned}
E[G(U)] &= \int_{\mathbb{R}}\int_{\mathbb{R}} G(z) f_{\varepsilon}(z-e_3) f_{\varepsilon_3}(e_3)\,\mathrm{d}z\mathrm{d}e_3 \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} G(z) \frac{\exp\left(\dfrac{(z-e_3)^2}{2(\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j))}\right)}{\sqrt{2\pi}\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}} f_{\varepsilon_3}(e_3)\,\mathrm{d}z\mathrm{d}e_3 \\
&= \int_{\mathbb{R}} G(z) \frac{1}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}} \int_{\mathbb{R}} f_{\mathcal{N}(0,1)}\left(\frac{z-e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}}\right) \\
&\quad f_{\varepsilon_3}(e_3)\,\mathrm{d}e_3\mathrm{d}z.
\end{aligned}
$$

We deduce the density function of U:

$$
f_U(u) = \frac{1}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}} \int_{\mathbb{R}} f_{\mathcal{N}(0,1)}\left(\frac{u-e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}}\right) f_{\varepsilon_3}(e_3)\,\mathrm{d}e_3.
$$

Then the distribution function of U is:

$$\begin{aligned}
F_U(z) &= \int_{-\infty}^{z} f_U(u)\,\mathrm{d}u \\
&= \int_{-\infty}^{z} \frac{1}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}} \int_{\mathbb{R}} f_{\mathcal{N}(0,1)}\left(\frac{u-e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_{i2}^j)}}\right) \\
&\quad f_{\varepsilon_3}(e_3)\,\mathrm{d}e_3\mathrm{d}u \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}} \int_{-\infty}^{z} f_{\mathcal{N}(0,1)}\left(\frac{u-e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}}\right) \\
&\quad \mathrm{d}u f_{\varepsilon_3}(e_3)\,\mathrm{d}e_3.
\end{aligned}$$

We make the following substitution:

$$\begin{aligned}
t &= \frac{u-e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}} \\
\mathrm{d}u &= \mathrm{d}t\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}
\end{aligned}$$

The bounds of the integral thus become:

u	-∞	z
t	-∞	$\frac{z-e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}}$

We set $b = \dfrac{z - e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k, \varepsilon_2^j)}}$, then we have:

$$
\begin{aligned}
F_U(z) &= \int_{\mathbb{R}} \frac{1}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k, \varepsilon_2^j)}} \int_{-\infty}^{b} f_{\mathcal{N}(0,1)}(t)\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k, \varepsilon_2^j)} \quad \mathrm{d}t\, f_{\varepsilon_3}(e_3)\,\mathrm{d}e_3 \\
&= \int_{\mathbb{R}} \int_{-\infty}^{b} f_{\mathcal{N}(0,1)}(t)\,\mathrm{d}t\, f_{\varepsilon_3}(e_3)\,\mathrm{d}e_3 \\
&= \int_{\mathbb{R}} \Phi(b)\, f_{\varepsilon_3}(e_3)\,\mathrm{d}e_3 \\
&= \int_{\mathbb{R}} \Phi\left(\frac{z - e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k, \varepsilon_2^j)}}\right) f_{\varepsilon_3}(e_3)\,\mathrm{d}e_3 \\
&= \int_{\mathbb{R}} \Phi\left(\frac{z - e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k, \varepsilon_2^j)}}\right) \frac{\exp(-e_3)}{(1+\exp(-e_3))^2}\,\mathrm{d}e_3.
\end{aligned}
$$

As we have

$$
\begin{aligned}
f_{\varepsilon_3}(e_3) &= \frac{\exp(-e_3)}{(1+\exp(-e_3))^2} \\
&= \frac{\exp(e_3)}{(1+\exp(e_3))^2},
\end{aligned}
$$

we conclude that

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

Moreover, even in this situation we can prove that $1 - F_U(-z) = F_U(z)$, as we observed for the probit modeling.

$$\begin{aligned}
1 - F_U(-z) &= 1 - \int_{\mathbb{R}} \Phi\left(\frac{-z-e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}}\right)\frac{\exp(e_3)}{(1+\exp(e_3))^2}\,\mathrm{d}e_3 \\
&= 1 - \int_{\mathbb{R}} \Phi\left(-\frac{z+e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}}\right)\frac{\exp(e_3)}{(1+\exp(e_3))^2}\,\mathrm{d}e_3 \\
&= 1 - \int_{\mathbb{R}} \left[1-\Phi\left(\frac{z+e_3}{\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}}\right)\right]\frac{\exp(e_3)}{(1+\exp(e_3))^2}\,\mathrm{d}e_3 \\
&= 1 - \int_{\mathbb{R}} \frac{\exp(e_3)}{(1+\exp(e_3))^2}\,\mathrm{d}e_3 \\
&\quad + \int_{\mathbb{R}} \Phi\left(\frac{z+e_3}{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}\right)\frac{\exp(e_3)}{(1+\exp(e_3))^2}\,\mathrm{d}e_3 \\
&= 1 - 1 + \int_{\mathbb{R}} \Phi\left(\frac{z+e_3}{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}\right)\frac{\exp(e_3)}{(1+\exp(e_3))^2}\,\mathrm{d}e_3
\end{aligned}$$

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

$$= \int_{\mathbb{R}} \Phi\left(\frac{z-e_3}{\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_k\gamma_j cov(\varepsilon_2^k,\varepsilon_2^j)}\right)\frac{\exp(e_3)}{(1+\exp(e_3))^2}\mathrm{d}e_3 \text{ [(a)]}$$

$$= F_U(z).$$

Now that we have determined the distribution function of U for both probit and logit outcomes, we can prove the equalities in the corollary. To facilitate the development of the proof we set:

$$\delta^Z(t) = \int A_{t1} - A_{t0}\mathrm{d}F_X(x)$$
$$\zeta(t) = \int B_{1t'} - B_{0t'}\mathrm{d}F_X(x),$$

where:

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

For the indirect effect $\delta(t)$ we have :

$$\begin{aligned} A_{tt'} &= \int_{\mathbb{R}^K} \mathbb{E}[Y|T=t, M=m, W=w, X=x]\,\mathrm{d}F_{(M(t'),W(t))|X=x}(m,w) \\ &= \int_{\mathbb{R}^K} \mathbb{E}[\mathbb{1}_{\{Y^*>0\}}|T=t, M=m, W=w, X=x]\,\mathrm{d}F_{(M(t'),W(t))|X=x}(m,w) \\ &= \int_{\mathbb{R}^K} \mathbb{P}[Y^*>0|T=t, M=m, W=w, X=x]\,\mathrm{d}F_{(M(t'),W(t))|X=x}(m,w). \end{aligned}$$

We replace Y^* by its expression and obtain:

[(a)] Here we substitute e_3 by $-e_3$.

$$A_{tt'} = \int_{\mathbb{R}^K} \mathbb{P}[(\alpha_3 + \beta_3 t + \sum_{k=2}^{K} \gamma_k m^k + \gamma m + \xi_3^{\Gamma} x + \varepsilon_3 > 0] \, \mathrm{d}F_{(M(t'),W(t))|X=x}(m,w).$$

By the substitutions $m^1 = \alpha_2^1 + \beta_2^1 t' + \xi_2^{\Gamma 1} x + e_2^1$ and $m^k = \alpha_2^k + \beta_2^k t + \xi_2^{\Gamma k} x + e_2^k \, \forall k \geq 2$, we have:

$$\begin{aligned}
A_{tt'} &= \int_{\mathbb{R}^K} \mathbb{P}\left[(\alpha_3 + \sum_{k=1}^{K} \gamma_k \alpha_2^k) + (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t + \gamma \beta_2^1 t' + (\xi_3^{\Gamma} + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x + (\varepsilon_3 + \sum_{k=1}^{K} \gamma_k e_2^j) > 0\right] \\
&\quad \mathrm{d}F_{\Upsilon}\begin{pmatrix} e_2^1 \\ \vdots \\ e_2^K \end{pmatrix} \\
&= \int_{\mathbb{R}} \mathbb{P}\left[(\alpha_3 + \sum_{k=1}^{K} \gamma_k \alpha_2^k) + (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t + \gamma \beta_2^1 t' + (\xi_3^{\Gamma} + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x + (\varepsilon_3 + e) > 0\right] \mathrm{d}F_{\varepsilon}(e) \\
&= \int_{\mathbb{R}} \mathbb{E}\left[\mathbb{1}_{\left\{(\alpha_3 + \sum_{k=1}^{K} \gamma_k \alpha_2^k) + (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t + \gamma \beta_2^1 t' + (\xi_3^{\Gamma} + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x + (\varepsilon_3 + e) > 0\right\}}\right] \mathrm{d}F_{\varepsilon}(e) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\left\{(\alpha_3 + \sum_{k=1}^{K} \gamma_k \alpha_2^k) + (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t + \gamma \beta_2^1 t' + (\xi_3^{\Gamma} + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x + (e_3 + e) > 0\right\}} \mathrm{d}F_{\varepsilon_3}(e_3) \mathrm{d}F_{\varepsilon}(e) \\
&= \int_{\mathbb{R}} \mathbb{1}_{\left\{(\alpha_3 + \sum_{k=1}^{K} \gamma_k \alpha_2^k) + (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t + \gamma \beta_2^1 t' + (\xi_3^{\Gamma} + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x + u > 0\right\}} \mathrm{d}F_U(u) \\
&= \mathbb{E}\left[\mathbb{1}_{\left\{(\alpha_3 + \sum_{k=1}^{K} \gamma_k \alpha_2^k) + (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t + \gamma \beta_2^1 t' + (\xi_3^{\Gamma} + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x + U > 0\right\}}\right] \\
&= \mathbb{P}\left((\alpha_3 + \sum_{k=1}^{K} \gamma_k \alpha_2^k) + (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t + \gamma \beta_2^1 t' + (\xi_3^{\Gamma} + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x + U > 0\right) \\
&= \mathbb{P}\left(U > -(\alpha_3 + \sum_{j=1}^{J} \gamma_k \alpha_2^k) - (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t - \gamma \beta_2^1 t' - (\xi_3 + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x\right) \\
&= 1 - \mathbb{P}\left(U \leq -(\alpha_3 + \sum_{k=1}^{K} \gamma_k \alpha_2^k) - (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t - \gamma \beta_2^1 t' - (\xi_3^{\Gamma} + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x\right) \\
&= 1 - F_U\left(-(\alpha_3 + \sum_{k=1}^{K} \gamma_k \alpha_2^k) - (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t - \gamma \beta_2^1 t' - (\xi_3^{\Gamma} + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x\right)
\end{aligned}$$

Because $1 - F_U(-z) = F_U(z)$, the last equation becomes:

$$A_{tt'} = F_U\left((\alpha_3 + \sum_{k=1}^{K} \gamma_k \alpha_2^k) + (\beta_3 + \sum_{k=2}^{K} \gamma_k \beta_2^k) t + \gamma \beta_2^1 t' + (\xi_3^{\Gamma} + \sum_{k=1}^{K} \gamma_k \xi_2^{\Gamma k}) x\right)$$

By replacing $A_{tt'}$ given by this last expression in $\delta(t) = \int A_{t1} - A_{t0} \mathrm{d}F_X(x)$, we obtain the result stated in the corollary. Analogous proofs hold for the other mediation effects in the corollary.

6 Complementary results

One of assumptions needed for our results is that correlations between the potential mediators are the same whatever the treatment governing the mediators: $cor(M^i(t), M^j(t')|T,X) = \rho_{ij}$ for all t, t'. In the two following data generating models for the mediators we do not consider this assumption. $M^1(t)$ and $M^2(t)$ follow a bivariate normal distribution with covariance Σ for each $t = 0,1$ separately. Then, clearly the simulated data has $cor(M^1(1), M^2(0)) = 0$. Analysing the data generated with models 1 and 2, we check the robustness of our method to a departure from our assumption. Note that we have choose the same value of parameter that the model used in subsection 4.3.

Results for bias and coverage probability can be seen in Figures 3 and 4. This figure clearly shows that our approach allows an unbiased estimation, contrary to the simple analyses, for both direct and indirect effects. The interpretations of these results are very close to those of the analysis with the model respecting our assumption in the article.

Model 1: Continuous outcome and continuous mediators

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

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

- The counterfactual outcome follows the normal distribution

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

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

δ^Z	δ^1	δ^2	ζ	τ
44	20	24	10	54

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

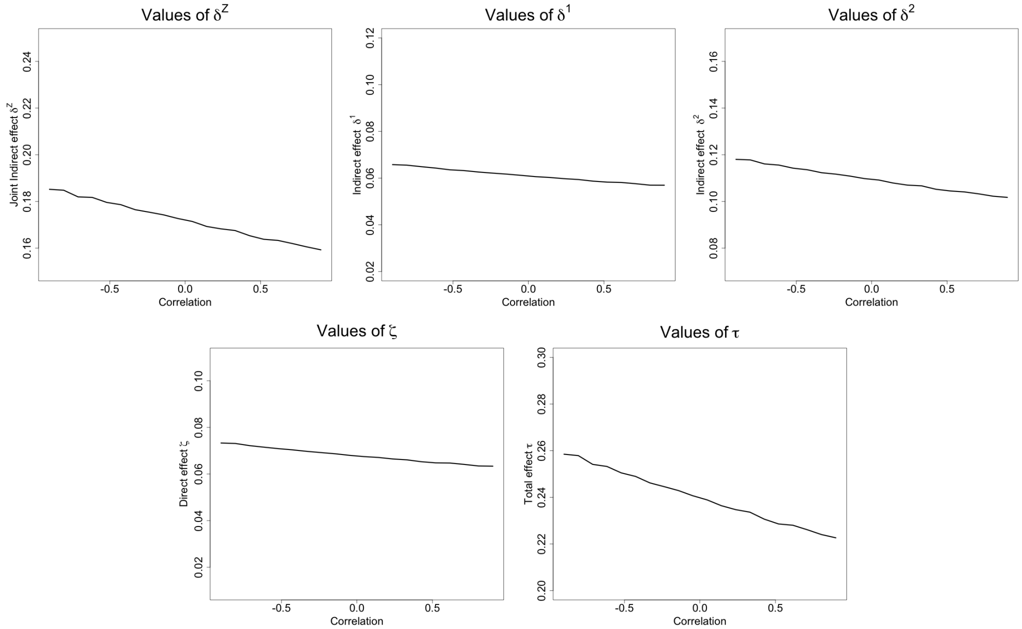

Figure 2: Binary outcome (model 2): Variation in causal effects due to correlation

Model 2: Binary outcome (logit) with continuous mediators

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

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

- The counterfactual outcome follows the logistic regression:

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

With this choice of parameters, 30% of the sampled observations are cases. As we can see in Corollary 5.1, with binary outcome, causal effects are related to the covariance of mediators. Figure 2 shows how the true causal values change when correlation changes.

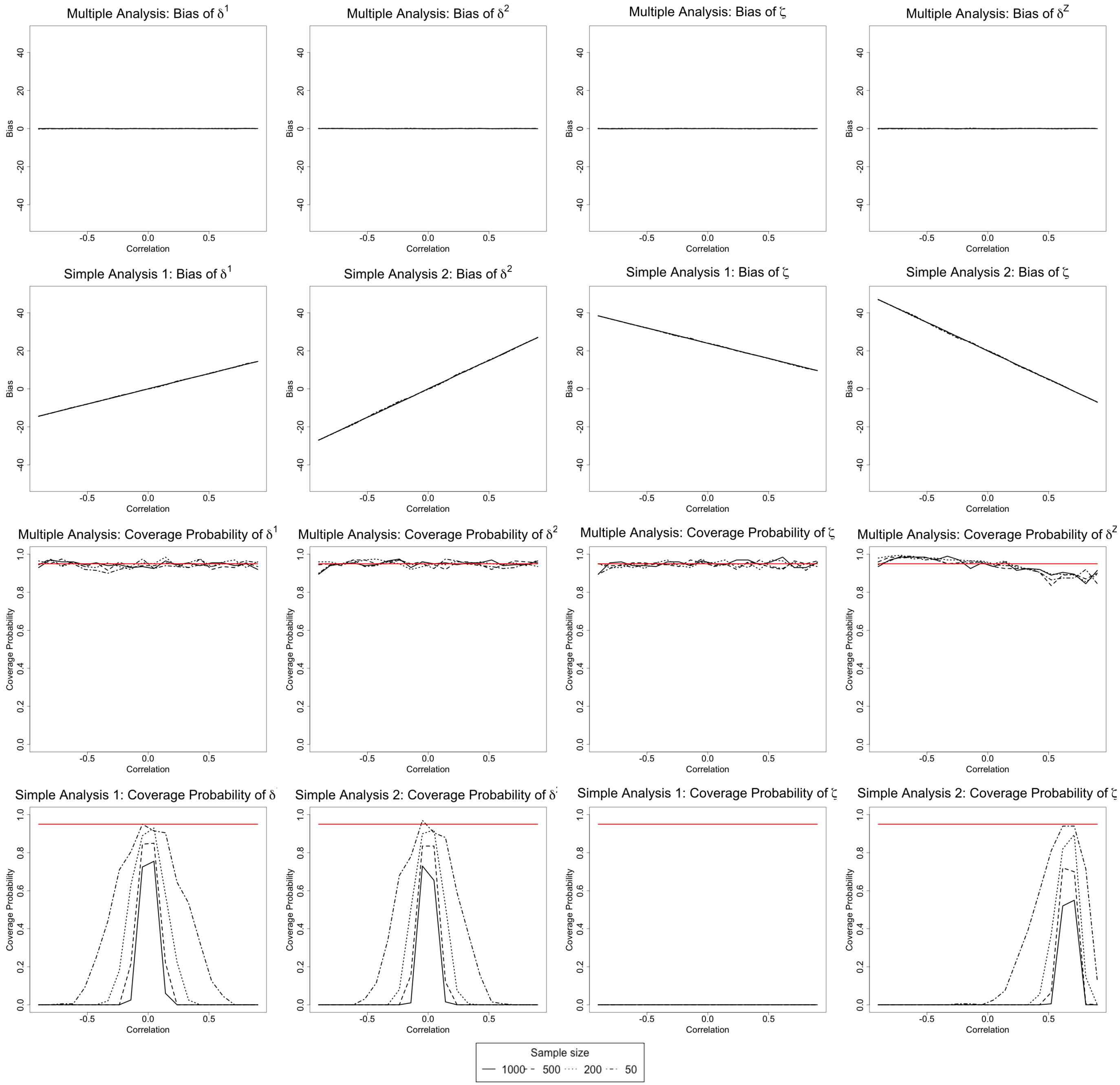

Figure 3: Model 1 (continuous outcome): bias, coverage probability, variance, and MSE of mediation effect estimators when the correlation between mediators varies. These results have benn obtained with 200 simulations. Each simulation consists in a dataset of size 1000.

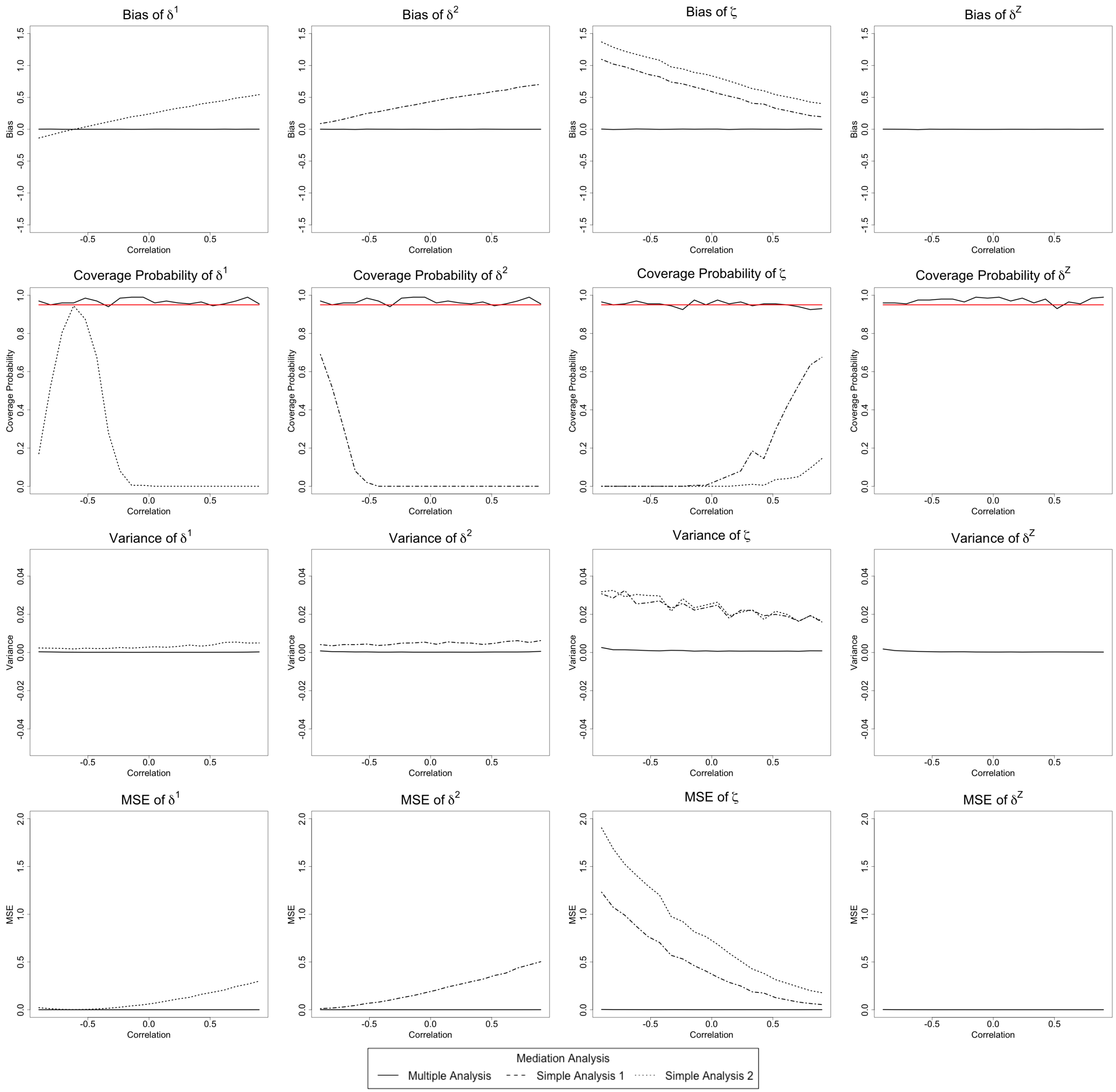

Figure 4: Model 2 (binary outcome): bias, coverage probability, variance, and MSE of mediation effect estimators when the correlation between mediators varies. These results have benn obtained with 200 simulations. Each simulation consists in a dataset of size 1000.

Model 3: Continuous outcome and continuous mediators

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

$$\begin{pmatrix} M^1(t) \\ M^2(t) \end{pmatrix} \sim \mathscr{N}\left(\mu = \begin{pmatrix} \frac{1}{2} + \frac{3}{2}t \\ 2 + 6t \end{pmatrix}, \Sigma\right).$$

- The counterfactual outcome follows the normal distribution

$$Y\left(t, M^1(t'), M^2(t'')\right) \sim \mathscr{N}(4 + 35t + 2M^1(t') + 3M^2(t''), 1).$$

Model 4: Binary outcome (probit) with continuous mediators

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

$$\begin{pmatrix} M^1(t) \\ M^2(t) \end{pmatrix} \sim \mathscr{N}\left(\mu = \begin{pmatrix} 0.2 + 0.7t \\ 0.4 + 0.7t \end{pmatrix}, \Sigma\right).$$

- The counterfactual outcome follows the probit distribution:

$$Y(t, M^1(t'), M^2(t'')) \sim \mathscr{B}\left(\Phi\left(-0.5 + 0.8t + 0.7M^1(t') + 0.7M^2(t'')\right)\right).$$

Model 5: Binary outcome (logit) with continuous mediators

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

$$\begin{pmatrix} M^1(t) \\ M^2(t) \end{pmatrix} \sim \mathscr{N}\left(\mu = \begin{pmatrix} 0.2 + 0.7t \\ 0.4 + 0.7t \end{pmatrix}, \Sigma\right).$$

- The counterfactual outcome follows the logistic distribution:

$$Y(t, M^1(t'), M^2(t'')) \sim B\left(\frac{1}{1 + \exp(0.5 - 0.8t - 0.7M^1(t') - 0.7M^2(t''))}\right).$$

Correlated and independent mediators

For all models, we consider two situations:

- model I when $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (independent mediators)
- model C when $\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$ (correlated mediators).

For instance Model 4 C refers to probit binary outcomes with correlated continuous mediators.

Table 2 summarizes the estimates provided by our multiple mediation analysis and by two simple mediation analyses, one for each mediator, for data simulated according to the six models. We can see that when mediators are independent, both the two simple analyses and our multiple analysis estimate correctly the individual indirect effects, the total effect and the proportion mediated. As expected the estimates of the direct effect provided by the two simple analyses is distant from the true value which does not even belong to the confidence intervals. On the contrary, our multiple analysis provide an accurate and precise estimate of the direct effect. When analyzing data with correlated mediators generated by Model 3 C, simple analysis provide wrong estimates of both the direct and indirect effects. Proportions mediated are largely overestimated. Unlike Model 3 I, where mediators are independent, the sum of the indirect effects estimated by the two simple analysis does not correspond to the joint indirect effect, thus showing one of the limits of applying multiple simple analyses in parallel. On the contrary, our multiple analysis provides accurate and precise estimates of the joint indirect effect when mediators are correlated. For data simulated from Model 4 C and Model 5 C, simple analyses provide wrong estimates of the indirect effects and accurate and precise estimate of the direct effect. Our method produces accurate and precise estimates of all effects.

	Effects	δ^Z	PM^Z	δ^1	PM^1	δ^2	PM^2	ζ	τ
Mod 3.I	Value	27	0.43	3	0.05	24	0.39	35	62
	Simple Analysis M^1			2.68 [1.98;3.52]	0.04 [0.03;0.06]			59.22 [58.05;60.34]	61.9 [60.82;63.00]
	Simple Analysis M^2					23.69 [21.81;25.52]	0.38 [0.36;0.41]	38.2 [36.64;39.84]	61.89 [60.86;62.98]
	Multiple Analysis	26.63 [25.35;27.85]	0.43 [0.41;0.44]	2.78 [2.26;3.27]	0.05 [0.04;0.05]	23.85 [22.7;24.97]	0.38 [0.37;0.40]	35.27 [34.53;36.02]	61.89 [60.71;62.95]
Mod 3.C	Value	27	0.43	3	0.05	24	0.39	35	62
	Simple Analysis M^1			8.30 [6.95;9.72]	0.13 [0.11;0.15]			53.6 [53.04;54.24]	61.9 [60.39;63.36]
	Simple Analysis M^2					34.83 [33.21;36.5]	0.56 [0.55;0.58]	27.06 [26.16;27.99]	61.9 [60.45;63.32]
	Multiple Analysis	27.07 [25.36;28.75]	0.44 [0.42;0.46]	2.94 [2.35;3.58]	0.05 [0.04;0.06]	24.13 [22.33;25.95]	0.39 [0.36;0.42	34.83 [33.61;36.2]	61.9 [60.75;63.07]
Mod 4.I	Value	0.23	0.55	0.11	0.27	0.11	0.27	0.18	0.41
	Simple Analysis M^1			0.09 [0.05;0.15]	0.23 [0.11;0.39]			0.3 [0.19;0.4]	0.4 [0.28;0.49]
	Simple Analysis M^2					0.13 [0.07;0.19]	0.31 [0.17;0.49]	0.28 [0.16;0.4]	0.41 [0.29;0.51]
	Multiple Analysis	0.23 [0.16;0.3]	0.59 [0.39;0.82]	0.09 [0.04;0.14]	0.23 [0.11;0.42]	0.13 [0.08;0.18]	0.33 [0.2;0.54]	0.16 [0.06;0.27]	0.39 [0.28;0.49]
Mod 4.C	Value	20	0.55	0.09	0.24	0.09	0.24	0.16	0.36
	Simple Analysis M^1			0.19 [0.11;0.27]	0.55 [0.35;0.79]			0.16 [0.06;0.26]	0.35 [0.23;0.46]
	Simple Analysis M^2					0.2 [0.12;0.27]	0.54 [0.33;0.75]	0.18 [0.07;0.28]	0.38 [0.25;0.48]
	Multiple Analysis	0.20 [0.15;0.27]	0.56 [0.4;0.8]	0.08 [0.02;0.16]	0.21 [0.04;0.46]	0.12 [0.05;0.2]	0.32 [0.12;0.65]	0.16 [0.06;0.27]	0.37 [0.26;0.47
Mod 5.I	Value	0.18	0.55	0.09	0.27	0.09	0.27	0.15	32
	Simple Analysis M^1			0.11 [0.05;0.17]	0.32 [0.16;0.56]			0.23 [0.11;0.35]	0.33 [0.21;0.46]
	Simple Analysis M^2					0.08 [0.03;0.13]	0.23 [0.1;0.44]	0.25 [0.11;0.37]	0.32 [0.19;0.44]
	Multiple Analysis	0.19 [0.12;0.26]	0.57 [0.35;0.93]	0.1 [0.05;0.17]	0.31 [0.15;0.61]	0.08 [0.03;0.13]	0.23 [0.09;0.48]	0.14 [0.02;0.26]	0.33 [0.2;0.44]
Mod 5.C	Value	0.16	0.55	0.07	0.25	0.07	0.25	0.13	0.30
	Simple Analysis M^1			0.15 [0.1;0.2]	0.51 [0.34;0.75]			0.14 [0.05;0.23]	0.29 [0.19;0.38]
	Simple Analysis M^2					0.14 [0.09;0.18]	0.48 [0.31;0.7]	0.15 [0.06;0.24]	0.29 [0.2;0.38]
	Multiple Analysis	0.15 [0.11;0.2]	0.53 [0.37;0.77]	0.08 [0.03;0.15]	0.28 [0.1;0.59]	0.07 [0.01;0.12]	0.23 [0.04;0.48]	0.14 [0.05;0.22]	0.29 [0.2;0.38]

Table 2: Comparison between simple mediation analyses one mediator at the time and our method for multiple mediation analysis; estimates and 95% confidence intervals of the mediation effects (in columns) for each simulation model (in blocks of four lines). The lines *Value* contain the true values, calculated from a large number of simulated counterfactuals. The lines *Simple Analysis* M^k and *Multiple Analysis* give respectively the estimates of the simple analysis done for the k-mediator and of our method accounting for all mediators.

References

Wasserman, L. (2010): *All of statistics: a concise course in statistical inference*, Springer texts in statistics, New York, NY: Springer, corrected second printing, 2005 edition, oCLC: 837651382.