Supplementary Material for: Misclassified exposure in epigenetic mediation analyses. Does DNA methylation mediate effects of smoking on birthweight?

This document contains the following ten sections:

A1. Direct and Indirect causal effects definition and estimation using regression when the exposure is perfectly measured

We let Y_a and M_a denote respectively the values of the outcome and mediator that would have been observed had the exposure A been set to level a. We let Y_{am} denote the value of the outcome that would have been observed had the exposure, A, and mediator, M, been set to levels a and m, respectively.

The average natural direct effect is then defined by $NDE_{a,a^*}(a^*) = E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}]$. The average natural indirect effect can be defined as $NIE_{a,a^*}(a) = E[Y_{aM_a} - Y_{aM_{a^*}}]$, which compares the effect of the mediator at levels M_a and M_{a^*} on the outcome when exposure A is set to a. Natural direct and indirect effects within strata of $C = c$ are then defined by: $CDE_{a,a^*|c}(m) = E[Y_{am} - Y_{a^*m}|c]$, $NDE_{a,a^*|c}(a^*) = E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|c]$ and $NIE_{a,a^*|c}(a) = E[Y_{aM_a} - Y_{aM_{a^*}}|c]$ respectively.

As discussed in the paper, identification assumptions (i)-(iv) will suffice to identify these direct and indirect effects. If we let $X \perp Y|Z$ denote that X is independent of Y conditional on Z then these four identification assumptions can be expressed formally in terms of counterfactual independence as (i) $Y_{am} \perp A|C$, (ii) $Y_{am} \perp M|\{A, C\}$, (iii) $M_a \perp A|C$, and (iv) $Y_{am} \perp M_{a^*}|C$. Assumptions (i) and (ii) suffice to identify controlled direct effects; assumptions (i)-(iv) suffice to identify natural direct and indirect effects (Pearl, 2001; VanderWeele and Vansteelandt, 2009). The intuitive interpretation of these assuptions as described in the text follows from the theory of causal diagrams (Pearl, 2001). Alternative identification assumptions have also been proposed (Imai 2010a; Hafeman and VanderWeele, 2011). However, it has been shown that the intuitive graphical interpretation of these alternative assumptions are in fact equivalent (Shpitser and VanderWeele, 2011). Technical examples can be constructed where one set of identifiation assumptions holds and another does not, but on a causal diagram corresponding to a set of non-parametric structural equations, whenever one set of the assumptions among those in VanderWeele and Vansteelandt (2009), Imai (2010a), and Hafeman and VanderWeele (2011) holds, the others will also.

Suppose that both the mediator and the outcome are continuous and that the following models fit the observed data:

$$M_i = \beta_0 + \beta_1 A_i + \beta_2^{'} C_i + \epsilon_{2i} \tag{1}$$

$$Y_i = \theta_0 + \theta_1 A_i + \theta_2 M_i + \theta_3 A_i * M_i + \theta_4^{'} C_i + \epsilon_{1i} \tag{2}$$

$$Y_i = \theta_0^{\dagger} + \theta_1^{\dagger} A_i + \theta_4^{\dagger'} C_i + \epsilon_{1i} \tag{3}$$

If the covariates C satisfied the no-unmeasured confounding assumptions (i)-(iv) above, then the average controlled direct effect and the average natural direct and indirect effects were derived by VanderWeele and Vansteelandt (2009).

In particular, if the regression models (1) and (2) are correctly specified and assumptions of no unmeasured confounding of exposure-outcome relationship (i) and no unmeasured confounding of the mediator-outcome relationship (ii) hold, then we could compute the controlled direct effect as follows:

$CDE = E[Y_{am} - Y_{a^*m}|C = c]$
$= \theta_1(a - a^*) + \theta_3 m(a - a^*)$.

If the regression models (1) and (2) are correctly specified and assumptions (i) and (ii) together with two additional assumptions of (iii) no unmeasured confounding of the exposure-mediator relationship and (iv) that there is no mediator-outcome confounder that is affected by the exposure hold, then we could compute the natural direct effects by:

$NDE = E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|C = c]$
$= (\theta_1 + \theta_3\beta_0 + \theta_3\beta_1 a^* + \theta_3\beta_2' c)(a - a^*)$.

Moreover under the same assumptions we can compute the natural indirect effects by:

$NIE = E[Y_{aM_a} - Y_{a^*M_{a^*}}|C = c]$
$= (\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*)$.

Standard errors for these estimators can ben obtained either via bootstrap procedure or by the delta method (VanderWeele and Vansteelandt, 2009).

In the absence of exposure-mediator interaction ($\theta_3 = 0$) the causal effects estimators are given by

$NDE = \theta_1(a - a^*)$
$NIE = \theta_2\beta_1(a - a^*)$.

Finally, note that by the property of effect decomposition

$NIE = TE - NDE = (\theta_1^\dagger - \theta_1)(a - a^*)$.

A2. Description of measurement error mechanism

The misclassification mechanism can be expressed in additive form. Let A denote the true binary exposure indicating whether the mother is consistently smoking during pregnancy ($A = 1$) and let A^* denote the misclassified exposure, then we can write

$$A^* = A + U$$

when the latent variable is binary the measurement error, U, is not normally distributed and can take values $(-1, 0, 1)$ under certain probabilities and restrictions. Moreover, in the case of misclassification of a binary variable, $Cov(U, A) \neq 0$ and $Cov(U, A^*) \neq 0$, that is the error must be correlated with both the true and the observed level of the exposure (Carroll et al. 2006).
The moments of the error can be completely characterized by the knowledge of the prevalence of the true exposure, the sensitivity and specificity parameters.

Let $p^* = P(A^* = 1)$, $p = P(A = 1)$, $q^* = 1 - p^*$, $q = 1 - p$. Moreover define the reclassification probabilities $\eta = P(A = 1|A^* = 0)$ and $\nu = P(A = 0|A^* = 1)$. Then the moments of the misclassification

error are given by (Aigner, 1973)

$E(U) = \nu p^* - \eta q^*$, $Var(U) = \nu p^* + \eta q^* - (\nu p^* - \eta q^*)^2$, and $Cov(A^*, U) = (\nu + \eta)p^* q^*$.

Note that reclassification probabilities can be re-expressed in terms of misclassification probabilities. Define the misclassification probabilities as $\gamma_0 = P(A^* = 1|A = 0)$ and $\gamma_1 = P(A^* = 0|A = 1)$. Then,

$$\begin{aligned} \nu &= P(A = 0|A^* = 1) = \gamma_0 \frac{q}{p^*} \\ \eta &= P(A = 1|A^* = 0) = \gamma_1 \frac{p}{q^*} \end{aligned}$$

Note that misclassification probabilities can be expressed in terms of sensitivity ($SN = P(A^* = 1|A = 1)$) and specificity ($SP = P(A^* = 0|A = 0)$). In particular,

$$\begin{aligned} \gamma_0 &= 1 - SP \\ \gamma_1 &= 1 - SN \end{aligned}$$

Finally, note that the prevalence of the observed exposure can be expressed in terms of misclassification probabilities and true prevalence of the exposure.

$$p^* = (1 - \gamma_1)p + \gamma_0 q$$

These facts will be used throughout.
To study the impact of exposure misclassification in the assessment of mediation in the context of epigenetic studies we first assume the absence of unmeasured confounding. We further assume that the outcome Y and the mediator M as well as the additional covariates $\boldsymbol{C}$ are correctly measured. We assume that the error is non differential (i.e. $Cov(U, Y) = 0$ and $Cov(U, M) = Cov(U, \boldsymbol{C}) = 0$). Moreover, in the context of mediation analysis mediator M and $\boldsymbol{C}$, which can be either continuous or categorical variables, can be correlated with the misclassified exposure.
We additionally make the following relaxable assumptions.

- If mother declares to be a smoker during pregnancy, assume that the reporting is correct ($PPV = 1$, $\nu = 0$, $U = 0$).
- If mother declares to be a non-smoker during pregnancy, assume that the reporting might be incorrect ($U = (-1, 0, 1)$).
- If the mother is a non-smoker during pregnancy, assume she will declare that she is a non-smoker (SP=1).
- No smoking-methylation interaction.
- No measurement error on the confounders.

Given these assumptions the moments of the misclassification error are given by:

$E(U) = -\eta q^*$, $Var(U) = \eta q^* - (1 + \eta q^*)$, and $Cov(A^*, U) = \eta p^* q^*$.

A3. Description of Numerical Study

We conducted a simulation study to assess the finite-sample performance of the naïve mediation analyses and the SIMEX correction approach compared to the true model with correctly measured exposure. In particular, we investigated the relative bias, variance, and Type I error rates of the test for $H_0 : NIE = 0$ and the bias in the estimates of NDE, NIE and TE. We considered two scenarios (I) exposure potentially misclassified due to mis-reporting; (II) exposure potentially misclassified when a cotinine is used to classify exposure status. The scenarios mimic the recent study of the effect of maternal smoking exposure on birth weight, potentially mediated by methylation of CpG sites by Küpers et al. (2015).
Scenario I: A total of $n = 500$ observations are generated for each simulation. The continuous mediator (DNA-methylation) M_i for $i = 1, ..., n$ is generated from the true model given in equation (1). The continuous outcome (birth weight) Y_i for $i = 1, ..., n$ is generated from the true model given in equation (2). We generate three confounders, namely maternal age at conception ($C_1 \sim N(30, 3)$, maternal weight at conception ($C_2 \sim N(24, 3.5)$ and a binary indicator for whether mother ended the studies before college ($C_3 \sim Be(0.7)$). The binary exposure A_i, smoking status during pregnancy follows a Bernoulli distribution conditionally on the confounders ($A\ Be(p_a)$, $p_a = exp(-2 - 0.3C_1 + 0.5C_2 + 0.45C_3)/(1 + exp(-2 - 0.3C_1 + 0.5C_2 + 0.45C_3)$. The misclassified exposure A^* is generated assuming $SP = 1$ (i.e. if the smoking exposure is absent the mother will correctly report the smoking status) and SN in the range $(0.6, 1)$. For the simulations of bias, we assume a normally distributed mediator and outcome and set the mediator regression parameters to $\beta_0 = 0.65$, $\beta_1 = -0.15$, $\beta_2^{'} = (-0.01, 0.01, 0.01)$,$\sigma_M^2 = 0.1$, and the outcome regression parameters to $\theta_0 = 3685$,$\theta_1 = -150$, $\theta_2 = 300$, $\theta_4^{'} = (-10, -200, 10)$, $\sigma_Y^2 = 300$ we run $5,000$ simulations for each setting. For the bias simulations we consider two settings under the null hypothesis of no indirect effect and under the alternative hypothesis. In both settings methylation is a strong biomarker for the exposure $\beta_1 \neq 0$). Therefore, under the null hypothesis of no indirect effect $H_0 : NIE = 0$ is equivalent to $\theta_2 = 0$. For simulations of Type I error rates, we consider two cases. In the first case the indirect effect is null because of no effect of methylation on birth weight ($\theta_2 = 0$) and no effect of smoking status on methylation ($\beta_1 = 0$). In the second case we allow for smoking to affect methylation ($\theta_2 = 0$, $\beta_1 \neq 0$) and therefore methylation is a biomarker for the exposure. We consider 10,000 simulations for each setting to obtain an accurate estimate of the rate of false positives.
Scenario II: A continuous variable, cotinine at gestational week 18, is generated following results from recent studies (Joubert et al., 2014; Kvalvik et al., 2012). From smoking status generated under scenario I, cotinine level for non-smokers follows a normal distribution conditionally on the confounders ($A_{nonsmoker} \sim N(\mu_{nonsmoker}, 2.5)$, $\mu_{nonsmoker} = 2.5 - 0.1C_1 + 0.5C_2 + 0.21C_3$). Cotinine distribution for smokers is generated from a negative skewed normal distribution ($A_{smoker} \sim N(\mu_{smoker}, 200)$, $\mu_{smoker} = (500 - 0.1C_1 + 0.5C_2 + 0.21C_3)$, $sk = 0.65$). The smoking exposure status A^* is obtained by dichotomizing the cotinine measure at the cut-off of 30 nmol/l and 60 nmol/l (Kvalvik et al., 2012). We generate a normally distributed mediator and outcome and set the mediator regression parameters as described in the previous scenario.

A4. Bias of the linear regressions for birthweight in the presence of misclassified exposure

Suppose that A is subject to non-differential misclassification and measured as A^* and that we fit the observed continuous outcome regression model where A is replaced by A^*. We study the asymptotic

limit of the naive estimators of the exposure, mediator and the exposure-mediator interaction coefficients and we denote them by θ_1^*, θ_2^* and θ_3^* respectively.
Let $(\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*)$ be the naive maximum likelihood estimators of the outcome regressors if A is replaced by A^*. Let $(\theta_1, \theta_2, \theta_3)$ be the true parameters of the regressors. Let U denote the misclassification error taking values $(-1, 0, 1)$ and $MU = M \times U$; let $X = (A, M, AM, \boldsymbol{C})$ and $X^* = (A^*, M, A^*M, \boldsymbol{C})$ denote the matrix of the true and observed centered covariates respectively. Let $\Delta^* = E[X^{*'}X^*]$ denote the variance-covariance matrix of the observed centered covariates and $\delta^y_{i,j}$ denote an element of the inverse of the variance-covariance matrix.
For a continuous outcome modeled using the linear regression the asymptotic limit of the outcome regression parameters in the presence of exposure-mediator interaction is given by,

$$\begin{aligned}
\theta_1^* &= \theta_1 - \theta_1(\delta^y_{A^*,A^*}Cov(A^*,U) + \delta^y_{A^*,A^*M}Cov(A^*M,U)) + \\
&- \theta_3(\delta^y_{A^*,A^*}E(A^*MU) + \delta^y_{A^*,M}E(M,MU) + \delta^y_{A^*,A^*M}E(A^*M,MU) + \boldsymbol{\delta^y_{A^*,C}}'E(\boldsymbol{C}MU)) \\
\theta_2^* &= \theta_2 - \theta_1(\delta^y_{A^*,M}Cov(A^*,U) + \delta^y_{M,A^*M}Cov(A^*M,U)) + \\
&- \theta_3(\delta^y_{M,A^*}E(A^*MU) + \delta^y_{M,M}E(M,MU) + \delta^y_{M,A^*M}E(A^*M,MU) + \boldsymbol{\delta^y_{M,C}}'E(\boldsymbol{C}MU)) \\
\theta_3^* &= \theta_3 - \theta_1(\delta^y_{A^*,A^*M}Cov(A^*,U) + \delta^y_{A^*M,A^*M}Cov(A^*M,U)) + \\
&- \theta_3(\delta^y_{A^*M,A^*}E(A^*MU) + \delta^y_{A^*M,M}E(M,MU) + \delta^y_{A^*M,A^*M}E(A^*M,MU) + \boldsymbol{\delta^y_{A^*M,C}}'E(\boldsymbol{C}MU))
\end{aligned}$$

Note that the asymptotic limits of the naive outcome regression parameters estimators are complex functions of the true outcome regression parameters, the covariance between the observed covariates and the misclassification error, and the correlation between the covariates. In the presence of exposure-mediator interaction, it is not clear the direction that the asymptotic bias of the naive outcome regression parameters could take.

In the absence of exposure-mediator interaction and for centered covariates the asymptotic limit of the regression parameters is given by

$$\begin{aligned}
plim\theta_1^* &= \theta_1 - \theta_1\delta^y_{A^*,A^*}Cov(A^*,U) = \theta_1 - \theta_1\delta^y_{A^*,A^*}(\nu+\eta)p^*q^* \\
plim\theta_2^* &= \theta_2 - \theta_1\delta^y_{A^*,M}Cov(A^*,U) = \theta_2 - \theta_1\delta^y_{A^*,M}(\nu+\eta)p^*q^*.
\end{aligned}$$

Proof
Let $\hat{\boldsymbol{\theta}}^*$ be the vector of MLE estimators of the parameters from the outcome regression. Rewrite the outcome regression (2.2) in terms of M^* exploiting the assumption of additive measurement error

$$\begin{aligned}
Y &= \theta_0 + \theta_1(a^* - u) + \theta_2 m + \theta_3(a^* - u)m + \theta_4'c + \epsilon \\
&= \theta_0 + \theta_1 a^* + \theta_2 m + \theta_3 a^* m + \theta_4'c + \epsilon - \theta_1 u - \theta_3\xi \\
&= \theta_0 + \theta_1 a^* + \theta_2 m + \theta_3 a^* m + \theta_4'c + \epsilon - \theta_1 u - \theta_3\xi \\
&= \theta_0^* + \theta_1^* a^* + \theta_2^* m + \theta_3^* a^* m + \theta_4^{*'}c + \epsilon^*
\end{aligned}$$

with $\xi = m \times u$.
Let $X^* = (A^*, M, A^*M, C)^T$. Assume covariates are centered. Then the vector of MLE estimators of the outcome linear regression parameters is given by,

$$\begin{aligned}\hat{\boldsymbol{\theta}}^* &= (X^{*'}X^*)^{-1}X'^*Y \\ &= (X^{*'}X^*)^{-1}X^{*'}(X^*\boldsymbol{\theta}^* + \epsilon^*) \\ &= (X^{*T}X^*)^{-1}X^{*T}(X^*\boldsymbol{\theta}^* + \epsilon - \theta_1 u - \theta_3\xi)\end{aligned}$$

By rearranging the equation and taking the limit we obtain obtain a formula for the asymptotic bias of the outcome regression parameters estimators when M is replaced by M^*

$$ABIAS(\hat{\boldsymbol{\theta}}^*) = -\Sigma^{-1}_{x^*x^*}\{\theta_1\Sigma_{x^*u} + \theta_3\Sigma_{x^*\xi}\}$$

where,

$$\begin{aligned}\Sigma^{-1}_{x^*x^*} &= plim(\frac{(X^{*T}X^*)^{-1}}{n^{-1}}) \\ \Sigma_{x^*u} &= plim(\frac{X^{*T}u}{n}) = (E(A^*u), 0, E(A^*Mu), 0, ..., 0)^T \\ \Sigma_{x^*\xi} &= plim(\frac{X^{*T}\xi}{n}) = (E(A^*\xi), E(M\xi), E(A^*M\xi), E(C_1\xi), ..., E(C_K\xi))^T\end{aligned}$$

The elements of the vectors Σ_{x^*u} and $\Sigma_{x^*\xi}$ depend upon the specification of sensitivity and specificity parameters, the marginal probability of the latent mediator and the joint probability of the mediator and the exposure.
We can rewrite the asymptotic bias as

$$ABIAS(\hat{\boldsymbol{\theta}}^*) = -\theta_1 \times \begin{pmatrix} \delta^y_{A^*,A^*}E(A^*U) + \delta^y_{A^*,A^*M}E(A^*MU) \\ \delta^y_{A^*,M}E(A^*U) + \delta^y_{M,A^*M}E(A^*MU) \\ \delta^y_{A^*,A^*M}E(A^*U) + \delta^y_{A^*M,A^*M}E(A^*MU) \\ \boldsymbol{\delta^y_{A^*,C}}E(A^*U) + \boldsymbol{\delta^y_{C,A^*M}}E(A^*MU) \end{pmatrix} +$$

$$-\theta_3 \times \begin{pmatrix} \delta^y_{A^*,A^*}E(A^*MU) + \delta^y_{A^*,M}E(M,MU) + \delta^y_{A^*,A^*M}E(A^*M,MU) + \boldsymbol{\delta^y_{A^*,C}}'E(\boldsymbol{C}MU) \\ \delta^y_{M,A^*}E(A^*MU) + \delta^y_{M,M}E(M,MU) + \delta^y_{M,A^*M}E(A^*M,MU) + \boldsymbol{\delta^y_{M,C}}'E(\boldsymbol{C}MU) \\ \delta^y_{A^*M,A^*}E(A^*MU) + \delta^y_{A^*M,M}E(M,MU) + \delta^y_{A^*M,A^*M}E(A^*M,MU) + \boldsymbol{\delta^y_{A^*M,C}}'E(\boldsymbol{C}MU) \\ \boldsymbol{\delta^y_{C,A^*}}E(A^*MU) + \boldsymbol{\delta^y_{C,M}}E(M,MU) + \boldsymbol{\delta^y_{C,A^*M}}E(A^*M,MU) + \boldsymbol{\delta^y_{C,C}}'E(\boldsymbol{C}MU) \end{pmatrix}$$

where $\delta^y_{\cdot A^*}$, $\delta^y_{\cdot M}$, $\delta^y_{\cdot A^*M}$ and $\boldsymbol{\delta^y_{\cdot C}}$ are columns of $\Sigma^{-1}_{x_1^*,x_1^*}$.

From the asymptotic bias formulae given above the probability limit of $\hat{\theta}_1^*$, $\hat{\theta}_2^*$, and $\hat{\theta}_3^*$ can be easily derived as

$$plim\theta_1^* = \theta_1 - \theta_1(\delta^y_{A^*,A^*}E(A^*U) + \delta^y_{A^*,A^*M}E(A^*MU)) +$$

$$
\begin{aligned}
& - \theta_3(\delta^y_{A^*,A^*}E(A^*MU) + \delta^y_{A^*,M}E(M,MU) + \delta^y_{A^*,A^*M}E(A^*M,MU) + \boldsymbol{\delta^y_{A^*,C}}'E(\boldsymbol{C}MU)) \\
plim\theta_2^* &= \theta_2 - \theta_1(\delta^y_{A^*,M}E(A^*U) + \delta^y_{M,A^*M}E(A^*MU)) + \\
& - \theta_3(\delta^y_{M,A^*}E(A^*MU) + \delta^y_{M,M}E(M,MU) + \delta^y_{M,A^*M}E(A^*M,MU) + \boldsymbol{\delta^y_{M,C}}'E(\boldsymbol{C}MU)) \\
plim\theta_3^* &= \theta_3 - \theta_1(\delta^y_{A^*,A^*M}E(A^*U) + \delta^y_{A^*M,A^*M}E(A^*MU)) + \\
& - \theta_3(\delta^y_{A^*M,A^*}E(A^*MU) + \delta^y_{A^*M,M}E(M,MU) + \delta^y_{A^*M,A^*M}E(A^*M,MU) + \boldsymbol{\delta^y_{A^*M,C}}'E(\boldsymbol{C}MU))
\end{aligned}
$$

The probability limit for the outcome regression coefficients in absence of exposure-mediator interaction is easily obtained setting $\theta_3 = 0$ and setting to zero all the covariance terms that involve the exposure-mediator interaction.

A5. Bias of the linear regressions for methylation in the presence of misclassified exposure

Suppose that A is subject to non-differential misclassification and measured as A^* and that we fit the observed continuous mediator regression model where A is replaced by A^*. We study the asymptotic limit of the naive estimators of the exposure, mediator and the exposure-mediator interaction coefficients and we denote them by β_0^*, β_1^* and β_2^* respectively.
Let $(\hat{\beta}_0^*,\ \hat{\beta}_1^*,\ \hat{\beta}_2^*)$ be the naive maximum likelihood estimators of the outcome regressors if A is replaced by A^*. Let $(\beta_0^*,\ \beta_1^*,\ \beta_2^*)$ be the true parameters of the regressors. Let U denote the misclassification error taking values $(-1, 0, 1)$; let $X = (1, A, \boldsymbol{C})$ and $X^* = (1, A^*, \boldsymbol{C})$ denote the matrix of the true and observed centered covariates respectively. Let $\Delta^* = E[X^{*'}X^*]$ denote the variance-covariance matrix of the observed centered covariates and $\delta^m_{i,j}$ denote an element of the inverse of the variance-covariance matrix.
For a continuous mediator modeled using the linear regression the asymptotic limit of the mediator regression parameters is given by,

$$
\begin{aligned}
plim\beta_0^* &= \beta_0 - \beta_1\delta^m_{1,A^*}Cov(A^*,U) \\
plim\beta_1^* &= \beta_1 - \beta_1\delta^m_{A^*,A^*}Cov(A^*,U) \\
plim\beta_2^* &= \beta_2 - \beta_1\delta^m_{\boldsymbol{C},A^*}Cov(A^*,U).
\end{aligned}
$$

Proof

The proof follows from the previous section.

A6. Asymptotic bias of direct and indirect effects naive estimators in the absence of exposure-mediator interaction

The asymptotic bias of naive direct and indirect effect estimators in the presence of misclassified exposure can be easily derived as:

$$
ABIAS(\widehat{NDE}^*) = \theta_1^* - \theta_1 = -\theta_1\delta^y_{A^*,A^*}Cov(A^*,U)
$$

Result 1: By the properties of variance-covariance matrices $\delta^y_{A^*,A^*} > 0$ and by the assumptions $Cov(A^*,U) = \eta p^* q^* > 0$. Therefore, the natural direct effect is under-estimated.

$$\begin{aligned}
ABIAS(\widehat{NIE}^*) &= \theta_2^*\beta_1^* - \theta_2\beta_1 = (\theta_2 - \theta_1\delta^y_{A^*,M}Cov(A^*,U))(\beta_1 - \beta_1\delta^m_{A^*,A^*}Cov(A^*,U)) - \theta_2\beta_1 \\
&= -\theta_2\beta_1\delta^m_{A^*,A^*}Cov(A^*,U) - \theta_1\beta_1\delta^y_{A^*,M}Cov(A^*,U) + \theta_1\beta_1\delta^y_{A^*,M}Cov(A^*,U)\delta^m_{A^*,A^*}Cov(A^*,U) \\
&= -\theta_2\beta_1\delta^m_{A^*,A^*}Cov(A^*,U) + \theta_1\beta_1\delta^y_{A^*,M}Cov(A^*,U)(\delta^m_{A^*,A^*}Cov(A^*,U) - 1) \\
&= Cov(A^*,U)\beta_1\{-\theta_2\delta^m_{A^*,A^*} + \theta_1\delta^y_{A^*,M}(\delta^m_{A^*,A^*}Cov(A^*,U) - 1)\}
\end{aligned}$$

The indirect effect could be biased in either directions.
The asymptotic bias could also be derived using the property of effect decomposition.

$$\begin{aligned}
ABIAS(\widehat{NIE}^*) &= (\theta_1^{\dagger *} - \theta_1^*) - (\theta_1^\dagger - \theta_1) = [(\theta_1^\dagger - \theta_1^\dagger\delta^{y\dagger}_{A^*,A}Cov(A^*,U)) - (\theta_1 - \theta_1\delta^y_{A^*,A}Cov(A^*,U))] - (\theta_1^\dagger - \theta_1) \\
&= -Cov(A^*,U)(TE \times \delta^{y\dagger}_{A^*,A^*} - NDE \times \delta^y_{A^*,A^*})
\end{aligned}$$

Result 2: if $\theta_2\delta^m_{A^*,A^*} < \theta_1\delta^y_{A^*,M}(\delta^m_{A^*,A^*}Cov(A^*,U) - 1)$ then the indirect effect is over-estimated.

Result 3: if $TE \times \delta^{y\dagger}_{A^*,A^*} < NDE \times \delta^y_{A^*,A^*}$ then the indirect effect is over-estimated.

Result 4: Under the null hypothesis of no direct effect in the presence of exposure misclassification the direct effect is unbiased.

Result 5: Under the null hypothesis of no effect of the exposure on the mediator in the presence of exposure misclassification the indirect effect is unbiased.

Result 6: Under the null hypothesis of no effect of the mediator on the outcome in the presence of exposure misclassification the indirect effect is biased.

A7. Type I error rate in the absence of exposure-mediator interaction

In the absence of exposure-mediator interaction, for the size of the naive test for indirect effect to be correct we required that

1. $\widehat{NIE}^* = 0$
2. the naive variance, $Var(\widehat{NIE}^*) = \hat{\beta}_1^2\hat{\sigma}^2_{\theta_2} + \hat{\theta}_2^2\hat{\sigma}^2_{\beta_1}$ is equal to the true variability of the naive MLE estimator given the data measured with error, $Var(\widehat{NIE}^*|X^*)$.

Based on our investigation of the bias in the previous section we know that condition (i) will be violated unless $\beta_1 = 0$. Moreover, under exposure misclassification the type I error of the test for θ_2 is not in general preserved.
Note, however that the of the test for β_1 and θ_1 is preserved, therefore a test of no direct effect is in general preserved in the absence of exposure-mediator interaction.

Result 7: In general in the absence of exposure-mediator interaction and in the presence of exposure misclassification, the naive variance estimator of the indirect effect will not respect the true variability of the naive MLE's. The Type I error of the naive test will be inflated, leading to an increased rate of spurious results.

Proof

Given $Var(\hat{\beta_1}^*) = \sigma^*_{\beta_1} X^{*T} X^*$ the type one error is preserved if

$$Var(\widehat{\beta_1}^* | X^*) = \sigma^*_{\beta_1} X^{*T} X^*$$

where $X^* = (A^*, C)$. By the property of iterated expectations:

$Var(\widehat{\beta}^* | X^*) = E[Var(M_i | X, X^*) | X^*] + Var[E(M_i | X, X^*) | X^*]$
$= E[\sigma | X^*] + Var[\beta X | X^*]$
$= \sigma + Var(\beta_0 + \beta_1 A + \beta_2 C | (A^*, C))$
$= \sigma + Var(\beta_1 A | (A^*, C)) = \sigma + \beta_1^2 Var(A | (A^*, C)) = \sigma^*_{\beta_1} X^{*T} X^*$

The type one error for β_1 is preserved under the null hypothesis of $\beta_1 = 0$.

Given $Var(\hat{\theta_2}^*) = \sigma^*_{\theta_2} X^{*T} X^*$ the type one error is preserved if

$$Var(\widehat{\theta_2}^* | X^*) = \sigma^*_{\theta_2} X^{*T} X^*$$

where $X^* = (A^*, M, C)$. By the property of iterated expectations:

$Var(\widehat{\theta}^* | X^*) = E[Var(Y_i | X, X^*) | X^*] + Var[E(Y_i | X, X^*) | X^*]$
$= E[\sigma | X^*] + Var[\theta X | X^*]$
$= \sigma + Var(\theta_0 + \theta_1 A + \theta_2 M + \theta_4 C | (A^*, M, C))$
$= \sigma + Var(\theta_1 A | (A^*, M, C)) = \sigma + \theta_1^2 Var(A | (A^*, M, C))$

In general, this variance will not be constant for all $i = 1, .., n$ because the conditional variance could depend on M_i and C_i. When the variance is not constant, this equation will not simplify and will not

be equal to $\sigma^*_{\theta_2} X^{*T} X^*$. Thus, in general the Type I error rates will not be preserved. Note, however, that the type one error for θ_2 is preserved under the null hypothesis of no direct effect (i.e. $H_0 : \theta_1 = 0$). Moreover, the size of a test for direct effect is preserved under the null of no direct effect.

A8. Commented Code of Mediation Analysis in MoBa Cohort

```
#####################################################################
### Mediation Analysis ##############################################
#####################################################################
#observed exposure:
Astar <- phenotypes$mosmkyn # self-reported smoking
# Astar <- phenotypes$smkcot # cotinine-based smoking
# Astar <- phenotypes$smk2cot # combined smoking

#mediator
tdat <- t(betas(mldat_gfi1)) # methylation beta values
#outcome
Y <- phenotypes$vekt # birth weight
#confounders
C1 <- phenotypes$kjonn # child's sex
C2 <- phenotypes$MORS_ALDER # maternal age
C3 <- phenotypes$meducation # maternal education
C4 <- phenotypes$gestAge # gestational age
C5 <- phenotypes$parity # parity
C6 <- phenotypes$case # selection
C7 <- phenotypes$bmi_pre # pre-pregnancy bmi
```

Naïve Code

```
f.mediation.naive.parallel <-
    function(methcol,M,Astar,Y,C1,C2,C3,C4,C5,C6,C7,nboot=500){

        Mi <- M[,methcol]

        data <- as.data.frame(cbind(Astar,Mi,Y,
            C1,C2,C3,C4,C5,C6,C7))

        #objects to save
        nde_naive <- rep(NA,nboot)
        nie_naive_prod <- rep(NA,nboot)
        pm_naive_prod <- rep(NA,nboot)

        for (i in 1:nboot){

            if(is.wholenumber(i/(nboot/10))) { # print progress
                cat("nboot=",i," ; ",sep="")
                }
            set.seed(1234+i)

            databoot <- data[sample(nrow(data),replace=TRUE),]
            databoot$Astar <- as.factor(databoot$Astar)

            #Linear outcome regression
            regym <- glm(Y~Astar+Mi+C1+C2+C3+C4+C5+C6+C7,
```

```
                data=databoot,x=TRUE,y=TRUE)

            regy <- glm(Y~Astar+C1+C2+C3+C4+C5+C6+C7,
                data=databoot,x=TRUE,y=TRUE)

            #linear mediator regression
            regm <- glm(Mi~Astar+C1+C2+C3+C4+C5+C6+C7,
                data=databoot,x=TRUE,y=TRUE)

            #naive effects
            nde_naive[i] <- regym$coeff[2]
            nie_naive_prod[i] <- regym$coeff[3]*regm$coeff[2]
            pm_naive_prod[i] <- nie_naive_prod[i]/
                (nie_naive_prod[i]+nde_naive[i])
        }
        return(list(Naive=data.frame(nde_naive=nde_naive,
            nie_naive_prod=nie_naive_prod,
            pm_naive_prod=pm_naive_prod)))
}

B=500 # number of bootstraps
detectCores()
system.time(
    results <- mclapply(setNames(seq_len(ncol(tdat)), dimnames(tdat)[[2
]]),
        f.mediation.naive.parallel,M=tdat,Astar,Y,C1,C2,C3,C4,C5,C6,C7,
        nboot=B,mc.cores=detectCores())
)

f.CI.quantile <- function(x,simex=FALSE...){
    # Naive
    ci_naive <- t(apply(x$Naive,2,quantile,probs=c(0.5,0.025,0.975),...
))

    # Simex
    if(simex=TRUE){
        ci_simex <- t(apply(x$Simex,2,quantile,probs=c(0.5,0.025,0.975)
,...))
        out <- list(Naive=ci_naive,Simex=ci_simex)
    } else {
        out <- list(Naive=ci_naive)
    }
    return(out)
}

conf_int <- lapply(results,f.CI.quantile,simex=FALSE,na.rm=T)
```

SIMEX Code

```
specificity <- 1
(sensitivity <- seq(0.6,0.9,0.1))
B=500 # number of bootstraps

f.mediation.simex.parallel <-
    function(methcol,M,Astar,Y,C1,C2,C3,C4,C5,C6,C7,
        sn=0.8,sp=1,nboot=500){

        Mi <- M[,methcol]

        data <- as.data.frame(cbind(Astar,Mi,Y,C1,C2,C3,C4,C5,C6,C7))

        #objects to save
        nde_naive <- rep(NA,nboot)
        nie_naive_prod <- rep(NA,nboot)
        pm_naive_prod <- rep(NA,nboot)

        nde_simex <- matrix(nrow=nboot,ncol=length(sn))
        nie_simex_prod <- matrix(nrow=nboot,ncol=length(sn))
        pm_simex_prod <- matrix(nrow=nboot,ncol=length(sn))

        for (i in 1:nboot){

            if(is.wholenumber(i/(nboot/10))) { # print progress
                cat("nboot=",i," ; ",sep="")
                }
            set.seed(1234+i)

            databoot <- data[sample(nrow(data),replace=TRUE),]
            databoot$Astar <- as.factor(databoot$Astar)

            #linear outcome regression
            regym <- glm(Y~Astar+Mi+C1+C2+C3+C4+C5+C6+C7,
                data=databoot,x=TRUE,y=TRUE)

            regy <- glm(Y~Astar+C1+C2+C3+C4+C5+C6+C7,
                data=databoot,x=TRUE,y=TRUE)

            #linear mediator regression
            regm <- glm(Mi~Astar+C1+C2+C3+C4+C5+C6+C7,
                data=databoot,x=TRUE,y=TRUE)

            #naive effects
            nde_naive[i] <- regym$coeff[2]
            nie_naive_prod[i] <- regym$coeff[3]*regm$coeff[2]
            pm_naive_prod[i] <- nie_naive_prod[i]/
                (nie_naive_prod[i]+nde_naive[i])
```

```
            for (k in 1:length(sn)){

                ###SIMEX PROCEDURE###
                Pi <- matrix(data = c(sp, 1-sp, 1-sn[k], sn[k]),
                    nrow = 2, byrow = FALSE)
                dimnames(Pi) <- list(levels(databoot$Astar),
                    levels(databoot$Astar))
                regy_simex <- mcsimex(regy,mc.matrix=Pi,
                    SIMEXvariable="Astar")
                regym_simex <- mcsimex(regym,mc.matrix=Pi,
                    SIMEXvariable="Astar")
                regm_simex <- mcsimex(regm,mc.matrix=Pi,
                    SIMEXvariable="Astar")

                #corrected effects
                nde_simex[i,k] <- regym_simex$coeff[2]
                nie_simex_prod[i,k] <-
                    regym_simex$coeff[3]*regm_simex$coeff[2]
                pm_simex_prod[i,k] <- nie_simex_prod[i,k]/
                    (nie_simex_prod[i,k]+nde_simex[i,k])
            }
        }
        colnames(nde_simex) <- paste("nde_simex_",sn,sep="")
        colnames(nie_simex_prod) <- paste("nie_simex_prod_",sn,sep="")
        colnames(pm_simex_prod) <- paste("pm_simex_prod_",sn,sep="")

        return(list(Naive=data.frame(nde_naive=nde_naive,
                nie_naive_prod=nie_naive_prod,pm_naive_prod=pm_naive_pr
od),
            Simex=data.frame(nde_simex,nie_simex_prod,pm_simex_prod)))
}

detectCores()
system.time(
    results <- mclapply(setNames(seq_len(ncol(tdat)), dimnames(tdat)[[2
]]),
        f.mediation.simex.parallel,M=tdat,Astar,Y,C1,C2,C3,C4,C5,C6,C7,
        sn=sensitivity,sp=specificity,nboot=B,mc.cores=detectCores())
)

f.descriptives<-function(x,...){
    c(mean=mean(x,...),quantile(x,...))
}

f.CI.quantile <- function(x,simex=FALSE,...){
    # Naive
    ci_naive <- t(apply(x$Naive,2,f.descriptives,...))
```

```
    # Simex
    if(simex==TRUE){
        ci_simex <- t(apply(x$Simex,2,f.descriptives,...))
        out <- list(Naive=ci_naive,Simex=ci_simex)
    } else {
        out <- list(Naive=ci_naive)
    }
    return(out)
}

conf_int <- lapply(results,f.CI.quantile,simex=TRUE,
    na.rm=T,probs=c(0.5,0.025,0.975))
```

A9. Supplementary Tables

Table S1: Bias and variance of SIMEX-corrected estimates of natural direct effect (NDE), natural indirect effect (NIE), total effect (TE), proportion mediated (PM). Misclassification setting assumes perfect specificity and $SN = (0.90, 0.925, 0.95, 0.975)$.

	$SN = 0.70$	$SN = 0.80$	$SN = 0.90$	$SN = 0.95$
H_1 SIMEX	Rel. Bias (var)	Rel. Bias (var)	Rel. Bias (var)	Rel. Bias (var)
TE	-0.20 (1746)	-0.12 (1594)	-0.03 (1420)	-0.00 (1268)
NDE	-0.30 (2164)	-0.15 (2168)	-0.05 (2076)	-0.00 (1847)
NIE	0.03 (285)	0.03 (366)	0.03 (462)	0.00 (473)
PM	0.45 (0.03)	0.28 (0.02)	0.10 (0.01)	0.03 (0.01)
H_0 SIMEX	Rel. Bias (var)	Rel. Bias (var)	Rel. Bias (var)	Rel. Bias (var)
TE	-0.20 (1690)	-0.12 (1569)	-0.03 (1399)	-0.00 (1252)
NDE	-0.30(2164)	-0.18 (2168)	-0.05 (2076)	-0.01 (1847)
NIE	-12/0 (271)	-8.8/0 (363)	-3.2/0 (455)	-1.2/0 (468)
PM	0.13/0 (0.09)	0.08/0 (0.04)	0.03/0 (0.02)	0.01 (0.02)

Table S2. Linear regression of cotinine-augmented self-reported sustained maternal smoking during pregnancy (i.e. potentially re-classified based on additional cotinine information) in relation to infant birth weight before and after adjustment for effects of maternal smoking on methylation at three CpGs in the *GFI1* gene *

	Coefficient[a]	SE	P-value
No mediator (CpG) adjustment	-116.04	40.79	0.0045
Adjusting for cg09935388	-90.04	44.37	0.0427
Adjusting for cg12876356	-102.14	42.78	0.0171
Adjusting for cg14179389	-108.87	43.17	0.0118

[a] regression coefficient interpretable as difference in birthweight, in grams between offspring of smoking mothers relative to nonsmokers.

* Linear regression model includes the following covariates: gestational age, child gender, maternal age, maternal education, parity, selection group, and maternal pre-pregnancy BMI. The beta is in units of grams of birth weight.

Table S3 Estimates of natural direct (NDE) and natural indirect effects (NIE) of sustained maternal smoking, assessed by self-report augmented by cotinine, on birthweight and proportion mediated (PM) by three methylation cites (CpGs) in *GFI1* in both naïve analyses and after SIMEX correction for measurement error. The SIMEX corrected values are presented for four different values for sensitivity (SN) of the self-reported maternal smoking exposure variable: 0.6, 0.70,0.80,0.90 where specificity=1. Median and 95% percentile confidence intervals for the bootstrap estimates are in units of grams of birth weight

CpG	SN	NDE (CI)	NIE (CI)	PM
cg09935388	Naïve	-89.4 (-179.5,2.6)	-27.8 (-59.7,5.1)	0.24
	0.6	-106.9 (-216.4,8.1)	-18.7 (-60.4,21.7)	0.15
	0.7	-102.1 (-205.9,16.2)	-21.5 (-60.8,17.6)	0.17
	0.8	-97.9 (-195.5,4.2)	-23.9 (-59.6,11.9)	0.20
	0.9	-92.2 (-184.0,-0.9)	-25.7 (-59.1,8.3)	0.22
cg12876356	Naïve	-101.3 (-185,-15.3)	-13.9 (-37.4,10.7)	0.12
	0.6	-118.5 (-209.3,-14.3)	-8.3 (-36.5,20.6)	0.06
	0.7	-111.9 (-211.1,-9.8)	-9.9 (-35.8,18.1)	0.08
	0.8	-111.4 (-201.1,-14.8)	-11.2 (-36.3,14.7)	0.09
	0.9	-104.7 (-192.0,-14.2)	-12.8 (-37.3,13.0)	0.11
cg14179389	Naïve	-108.3 (-191.7,-13.9)	-8.3 (-36.4,17.5)	0.07
	0.6	-125.8 (-223.6,-14.5)	-0.7 (-34.0,31.3)	0.00
	0.7	-119.8 (-213.7,-12.5)	-2.6 (-34.9,26.9)	0.02
	0.8	-115.0 (-212.1,-12.4)	-4.9 (-36.5,23.4)	0.04
	0.9	-112.2 (-200.6,-13.9)	-6.7 (-36.2,20.0)	0.06

Table S4 Linear regression of self-report of any maternal smoking during pregnancy in relation to infant birth weight before and after adjustment for effects of maternal smoking on methylation at three CpGs in the *GFI1* gene*

	Coefficient[a]	SE	P-value
No mediator (CpG) adjustment	-39.55	30.94	0.2014
Adjusting for cg09935388	-17.46	32.48	0.5910
Adjusting for cg12876356	-26.63	31.94	0.4047
Adjusting for cg14179389	-31.04	31.83	0.3298

[a] regression coefficient interpretable as difference in birthweight, in grams between offspring of smoking mothers relative to nonsmokers.

* Linear regression model includes the following covariates: gestational age, child gender, maternal age, maternal education, parity, selection group, and maternal pre-pregnancy BMI. The beta is in units of grams of birth weight.

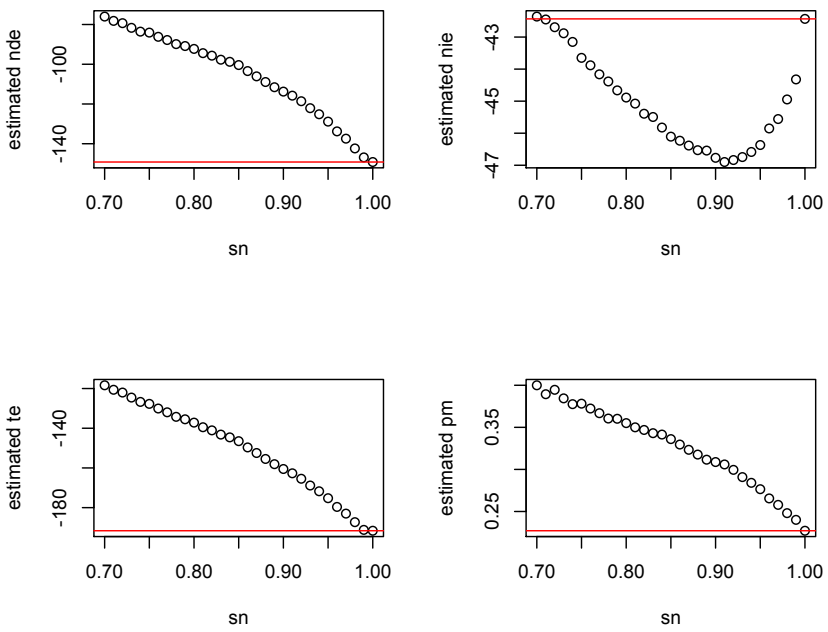
Table S5: Estimates of natural direct (NDE) and natural indirect effects (NIE) of any maternal smoking, assessed by self-report, on birthweight and proportion mediated (PM) by three methylation cites (CpGs) in *GFI1* in naïve analyses and after SIMEX correction for measurement error. The SIMEX corrected values are presented for four different values for sensitivity (SN) of the self-reported maternal smoking exposure variable: 0.6, 0.70,0.80,0.90 where specificity=1. Median and 95% percentile confidence intervals for the bootstrap estimates are in units of grams of birth weight.

CpG	SN	NDE (CI)	NIE (CI)	PM
cg09935388	Naïve	-16.7 (-84.6,52.9)	-23.3 (-41.4,-2.9)	0.58
	0.6	-21.1 (-112.1,64.6)	-26.8 (-51.6,-0.9)	0.55
	0.7	-20.7 (-103.8,65.4)	-25.8 (-48.0,-1.7)	0.55
	0.8	-18.7 (-96.0,63.5)	-25.0 (-46.1,-2.1)	0.57
	0.9	-18.3 (-90.0,54.3)	-24.1 (-43.4,-2.8)	0.57
cg12876356	Naïve	-25.9 (-91.3,42.8)	-13.1 (-27.0,2.1)	0.33
	0.6	-31.8 (-117.1,56.4)	-14.4 (-32.1,5.9)	0.31
	0.7	-31.6 (-113.5,56.6)	-13.9 (-30.6,4.8)	0.31
	0.8	-29.1 (-106.2,47.3)	-13.5 (-28.6,3.7)	0.32
	0.9	-27.5 (-99.9,47.4)	-13.3 (-28.2,3.0)	0.32
cg14179389	Naïve	-30.9 (-99.3,39.9)	-8.8 (-24.6,4.1)	0.22
	0.6	-38.8 (-124.1,57.4)	-9.0 (-29.8,8.5)	0.19
	0.7	-37.9 (-118.6,49.3)	-8.9 (-28.8,6.7)	0.19
	0.8	-34.2 (-110.2,46.8)	-9.0 (-27.2,5.7)	0.21
	0.9	-33.1 (-103.2,45.4)	-8.8 (-25.8,4.7)	0.21

A10. Supplementary Figures

Figure S1: Bias analysis for natural direct effect (nde), natural indirect effect (nie), total effect (te), proportion mediated (pm) under the alternative hypothesis (A: $\beta_1 \neq 0$, $\theta_2 \neq 0$) and under the null hypothesis (B: $\beta_1 \neq 0$, $\theta_2 = 0$)

(A)



(B)

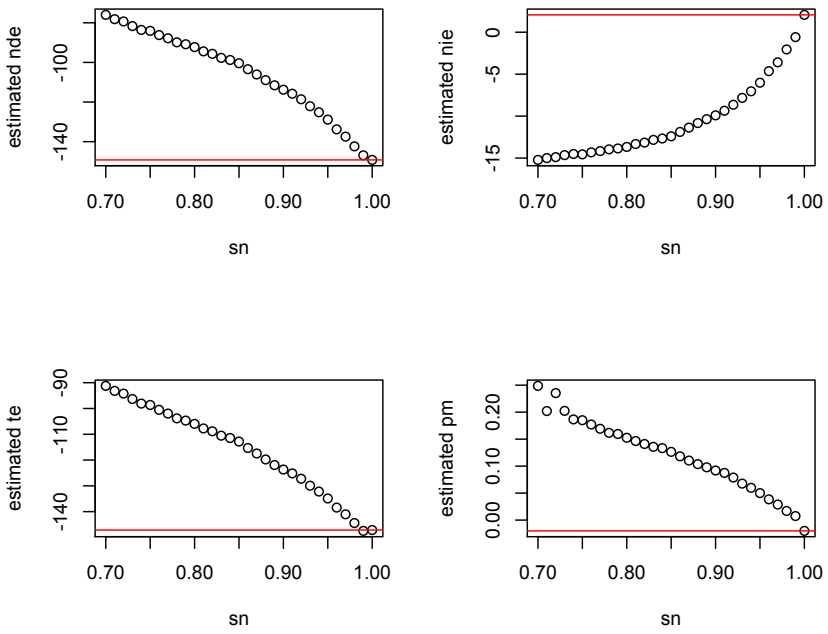
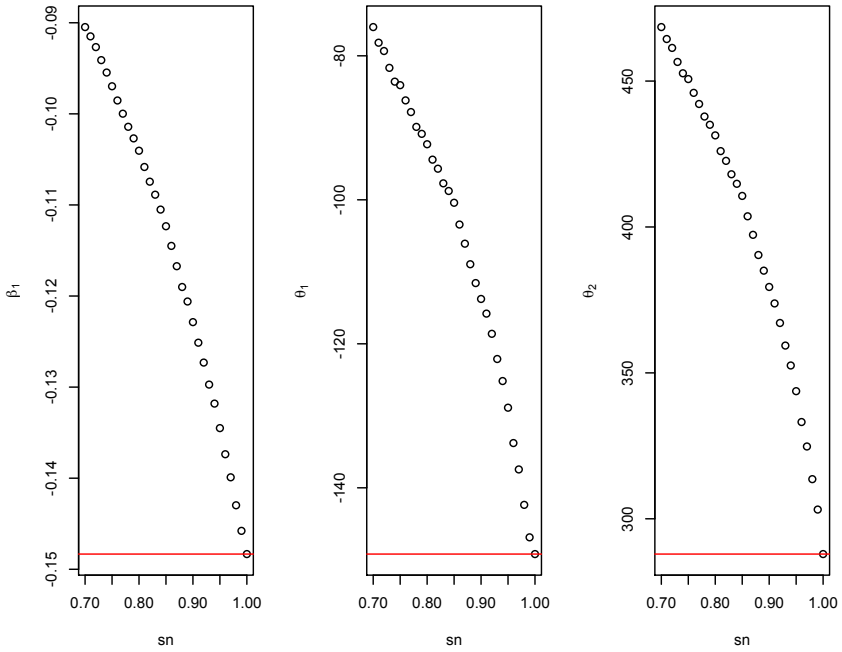


Figure S2: Bias analysis for β_1, θ_1, θ_2 under the alternative hypothesis (A: $\beta_1 \neq 0$, $\theta_2 \neq 0$) and under the null hypothesis (B: $\beta_1 \neq 0$, $\theta_2 = 0$)

(A)



(B)