simulation summary binary

Joakim Wallmark

Simulation study

Goal

Parametric estimators for natural direct- and indirect effects are implemented in the sensmediation package. The goal is to test the methods implemented in this package. We will investigate and compare bias from model misspecification (interaction term missing in outcome model) with bias from mediator-outcome confounding through simulations.

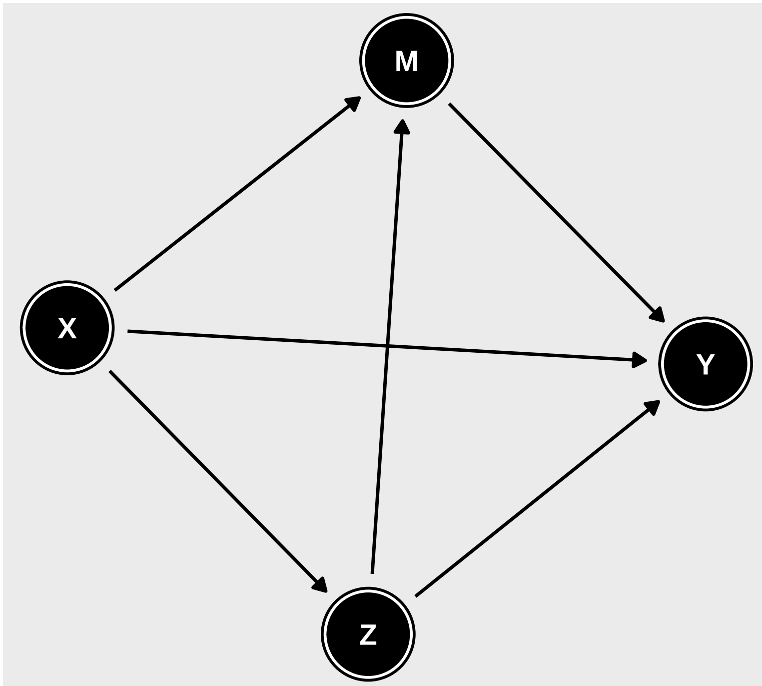
Simulation scenarios

Simulations of bias were done with model misspecification and confounding in separate comparable scenarios. 10000 iterations (1/10th of the linear case due to computation time) with a sample size of 2000 were used for each scenario.

Table 1: Variables used for simulations

variable	type	true model
X(additional covariate)	continuous	$X \sim gamma(8, 4.5)$
Z(exposure)	binary	$Z = I(Z* > 0)$ where $Z* \sim U_0 + U_1X + N(0, 1)$
M(mediator)	binary	$M = I(M* > 0)$ where
		$M* \sim \beta_0 + \beta_1 Z + \beta_2 X + N(0, 1)$
Y(outcome)	binary	$Y = I(Y* > 0)$ where
		$Y* \sim \theta_0 + \theta_1 Z + \theta_2 M + \theta_3 ZM + \theta_4 X + N(0, 1)$

DAG showing data generating process



Simulation scenarios (model misspecification)

To get realistic parameter values, models were estimated using data from riksstroke from n=60353 patients. Parameter values used for simulations:

- $U_0 = -3.416096$
- $U_1 = 0.036231$
- $B_0 = -1.6507546$
- $B_1 = 0.2683970$
- $B_2 = 0.0065543$
- $\theta_0 = -3.7220626$
- $\theta_1 = 0.2763912$
- $\theta_2 = 1.4729651$
- $\theta_3 =$ varied[-1, 1]
- $\theta_4 = 0.0283196$

Estimated mediator model was set to the correct one. Estimated outcome model was misspecified without ZM interaction: $Y* \sim Z + M + X$.

Simulation scenarios (mediator-outcome confounding)

Parameter values used for simulations:

$corr(\omega, \epsilon)$ =varied (-0.5,0.5)

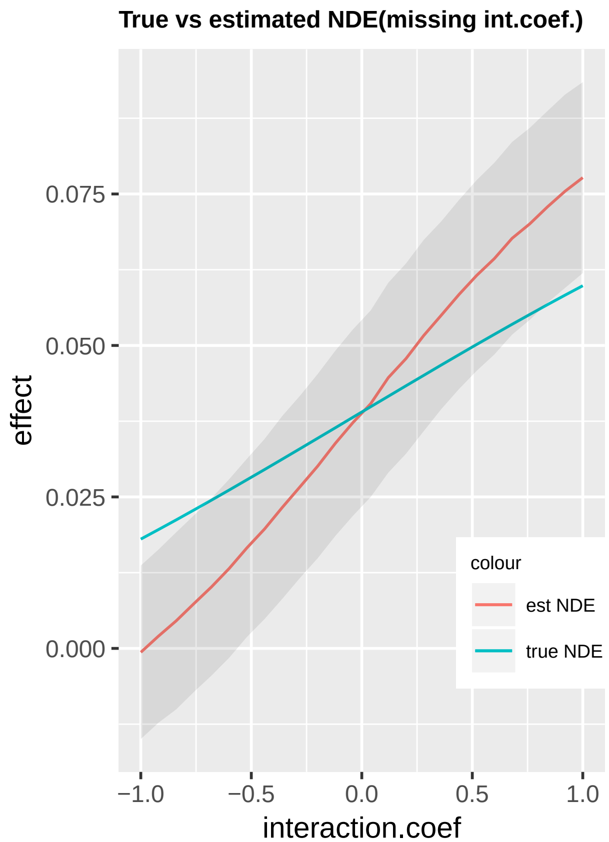
- $U_0 = -3.416096$
- $U_1 = 0.036231$

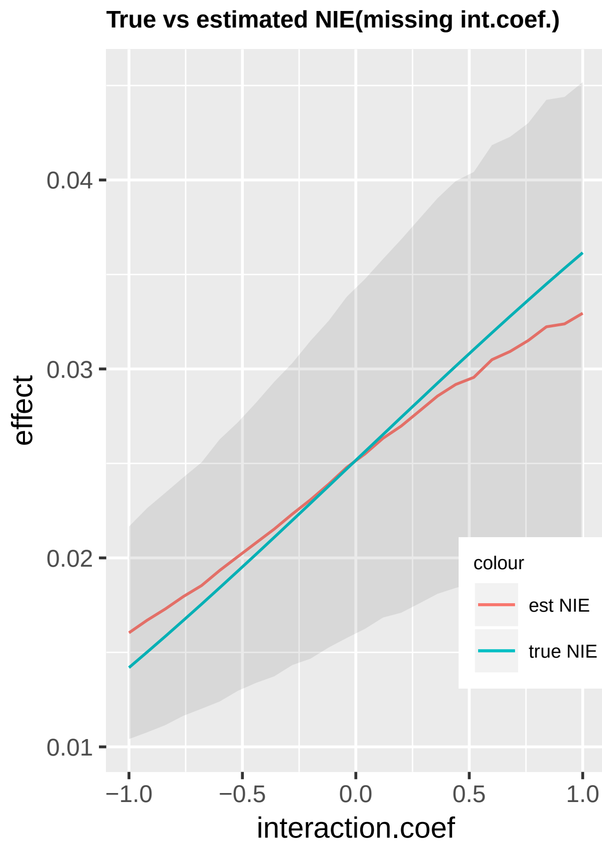
- $B_0 = -1.6507546$
- $B_1 = 0.2683970$
- $B_2 = 0.0065543$
- $\theta_0 = -3.7220626$
- $\theta_1 = 0.2763912$
- $\theta_2 = 1.4729651$
- $\theta_3 = -0.2583784$
- $\theta_4 = 0.0283196$

Estimated models were set to the correct ones. Bias was induced by changing error term correlation (between ω and ϵ).

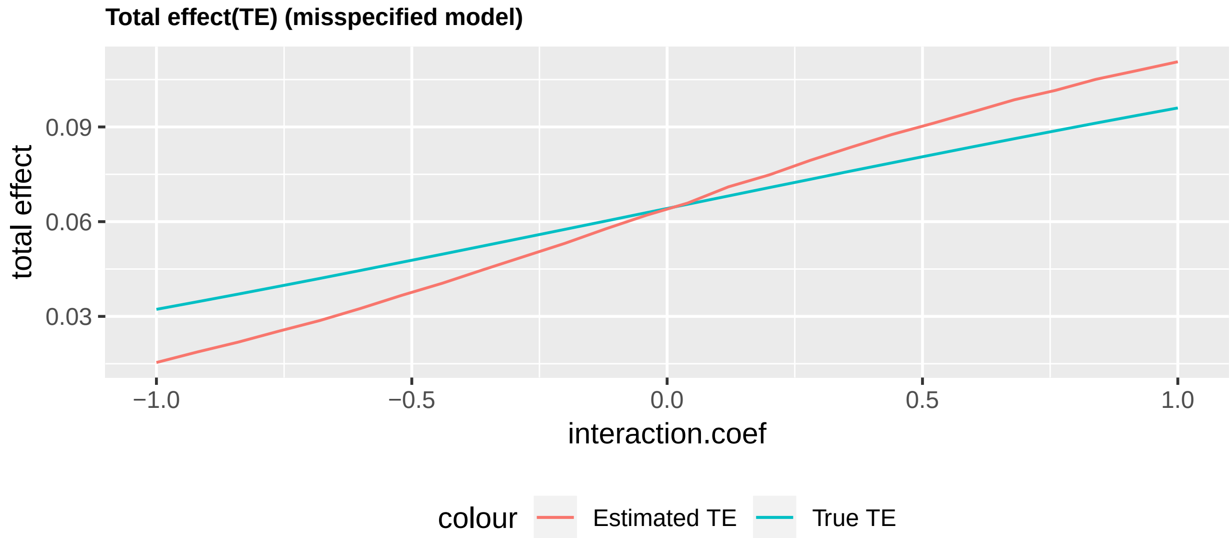
Results

Now NDE has more bias... true NDE is even outside of 1SE from estimate. NIE is only slightly biased:

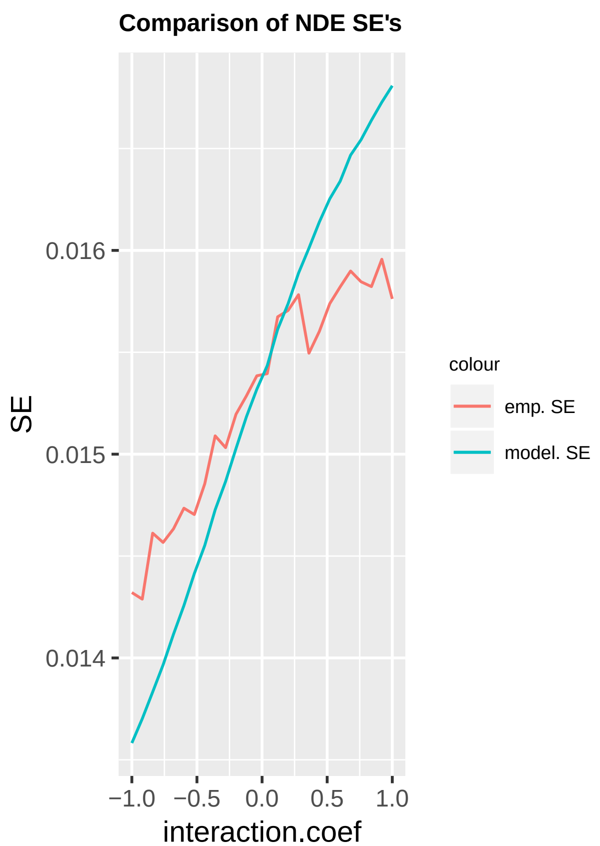


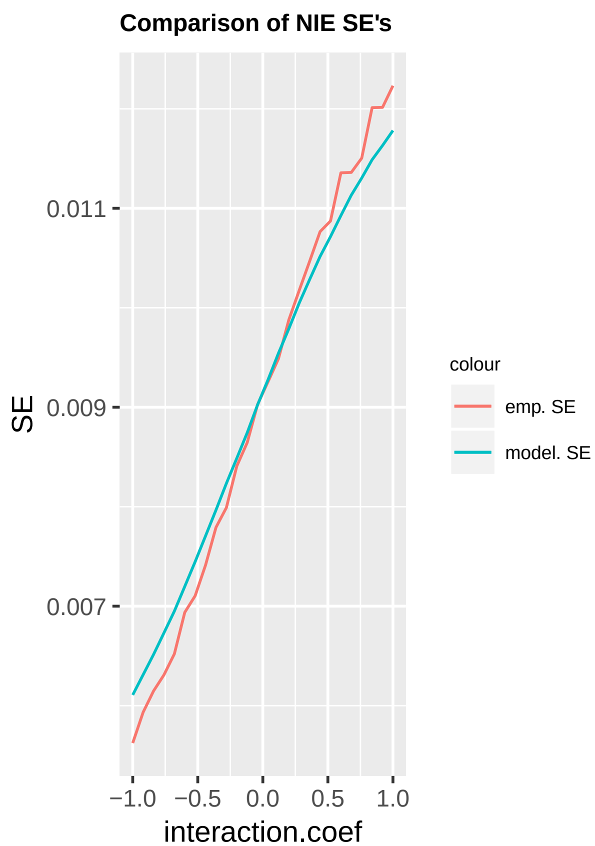


Total effect is biased. This differs from the situation where we used linear models:



When missing the interaction term in the outcome model, the empirical data shows that we slightly underestimate the SE for the NIE. The opposite is true for NDE. Empirical SE also varies a lot even though I use a lot of iterations. Might just be because y-scale is really small. All standard errors (both empirical and model) are pretty close in size for all values of the interaction:





Overall pretty good coverage here compared to the linear models since we have so much bigger SE's in relation to the effects and bias. For negative interaction the coverage is bigger than 95%. Not sure why. We overestimate the NIE and our model SE's are smaller as we increase the negative interaction strength. Smaller SE's make smaller interval which should give even worse coverage. Maybe I did something wrong here? (wording)If I make the interaction more negative, am I increasing or decreasing the interaction?

For the NDE case the bias is so small so the coverage gets better mostly because the SE gets bigger: