eAppendix

eAppendix to R.H.H. Groenwold, T.M. Palmer & K. Tilling: To adjust or not to adjust? When a "confounder" is only measured after exposure.

Identifiability of the causal effect

The directed acyclic graph presented in Figure 1 in the main text could also be expressed as a single world intervention graph (SWIG),[1] which provide a way to incorporate counterfactuals in causal diagrams. Figure A.1 shows the SWIG of an exposure (X), an outcome (Y), a mediator (M), and two unmeasured variables (U_1 and U_2). The node for exposure is split to distinguish between the random part (to the left of the $|$) and the fixed part (to its right).

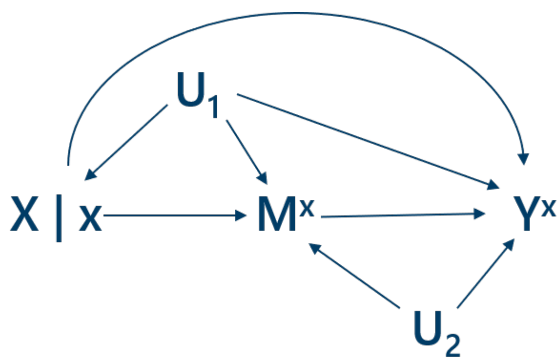

Figure A.1: Single world intervention graph (SWIG) of an exposure (X), an outcome (Y), a mediator (M), and two unmeasured variables (U_1 and U_2).

To identify the total effect of X on Y requires $Y^x \perp\!\!\!\perp X$. This requirement is not met, because X and Y are not d-separated, because of the path $X \leftarrow U_1 \rightarrow Y$. Therefore, the total causal effect of X on Y, e.g., $E[Y^{x=1}] - E[Y^{x=0}]$ cannot be identified from data on (X, Y, M).

To identify the direct effect of X on Y requires $Y^x \perp\!\!\!\perp X | M^x$. This requirement is not met, because the counterfactual value of M^x cannot be estimated from the data, one reason being that X and M are not d-separated (due to the path $X \leftarrow U_1 \rightarrow M^x$). Therefore, the (natural) direct causal effect of X on Y, e.g., $E[Y^{x=1,M^{x=1}}] - E[Y^{x=0,M^{x=1}}]$ cannot be identified from data on (X, Y, M).

Hence, in case of the causal structure depicted in Figure A.1, where U_1 and U_2 are considered unmeasured, neither the total or the direct effect of X on Y can be identified from data on (X, Y, M).

Derivation of bias expressions

We consider the causal diagram presented in Figure A.2. Details of the model are described in the section Notation and set-up of the main text.

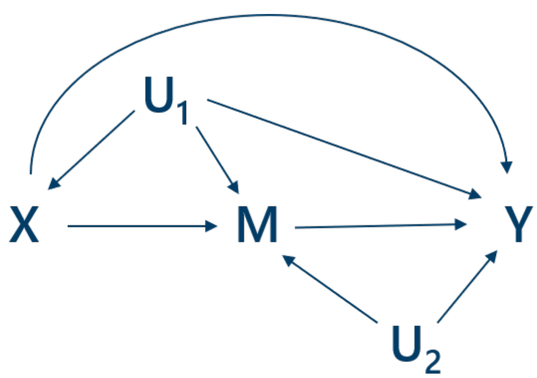

Figure A.2: Directed acyclic graph of an exposure (X), an outcome (Y), a mediator (M), and two unmeasured variables (U_1 and U_2).

Using the path tracing rules, the observed covariance between X and Y is, in expectation,

$$\mathrm{Cov}(X, Y) = \beta_{xy}\,\mathrm{Var}(X) + \beta_{xm}\beta_{my}\,\mathrm{Var}(X) + \beta_{u_1x}\,\mathrm{Var}(U_1)(\beta_{u_1y} + \beta_{u_1m}\beta_{my}).$$

It follows that

$$
\begin{aligned}
\widehat{\beta}_{xy} &= \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)} \\
&= \beta_{xy} + \beta_{xm}\beta_{my} + \frac{\beta_{u_1x}\operatorname{Var}(U_1)(\beta_{u_1y} + \beta_{u_1m}\beta_{my})}{\operatorname{Var}(X)} \\
&= \beta_{xy} + \beta_{xm}\beta_{my} + \frac{\beta_{u_1x}\sigma^2_{u_1}(\beta_{u_1y} + \beta_{u_1m}\beta_{my})}{\beta^2_{u_1x}\sigma^2_{u_1} + \sigma^2_x}.
\end{aligned}
$$

When conditioning on M the observed relation between X and Y is, in expectation,

$$
\widehat{\beta}_{xy|m} = \frac{\operatorname{Cov}(X,Y)\operatorname{Var}(M) - \operatorname{Cov}(Y,M)\operatorname{Cov}(X,M)}{\operatorname{Var}(X)\operatorname{Var}(M) - \operatorname{Cov}(X,M)^2}.
$$

Here,

$$
\begin{aligned}
\operatorname{Var}(X) &= \beta^2_{u_1x}\operatorname{Var}(U_1) + \sigma^2_x \\
\operatorname{Var}(M) &= \beta^2_{u_1m}\operatorname{Var}(U_1) + \beta^2_{u_2m}\operatorname{Var}(U_2) + \beta^2_{xm}\operatorname{Var}(X) + 2\beta_{u_1m}\beta_{xm}\operatorname{Cov}(X,U_1) + \sigma^2_m \\
\operatorname{Cov}(X,Y) &= \beta_{xy}\operatorname{Var}(X) + \beta_{xm}\beta_{my}\operatorname{Var}(X) + \beta_{u_1x}\operatorname{Var}(U_1)(\beta_{u_1y} + \beta_{u_1m}\beta_{my}) \\
\operatorname{Cov}(Y,M) &= \beta_{my}\operatorname{Var}(M) + \beta_{u_1m}\operatorname{Var}(U_1)\beta_{u_1y} + \beta_{u_2m}\operatorname{Var}(U_2)\beta_{u_2y} + \beta_{xm}\beta_{u_1x}\operatorname{Var}(U_1)\beta_{u_1y} + \\
&\quad \beta_{xm}\operatorname{Var}(X)\beta_{xy} + \beta_{u_1m}\operatorname{Var}(U_1)\beta_{u_1x}\beta_{xy} \\
\operatorname{Cov}(X,M) &= \beta_{xm}\operatorname{Var}(X) + \beta_{u_1x}\operatorname{Var}(U_1)\beta_{u_1m}.
\end{aligned}
$$

For completeness, we note that

$$
\begin{aligned}
\operatorname{Var}(U_1) &= \sigma^2_{u_1} \\
\operatorname{Var}(U_2) &= \sigma^2_{u_2} \quad \text{and} \\
\operatorname{Cov}(X,U_1) &= \beta_{u_1x}\operatorname{Var}(U_1).
\end{aligned}
$$

The bias expression above can be simplified when assuming all error terms are equal to 1, i.e., (σ_{u_1}, σ_{u_2}, etc. are 1). Then

$$\begin{aligned}
\mathrm{Var}(X) &= \beta_{u_1x}^2 + 1 \\
\mathrm{Var}(M) &= \beta_{u_1m}^2 + \beta_{u_2m}^2 + \beta_{xm}^2(\beta_{u_1x}^2 + 1) + 2\beta_{u_1m}\beta_{xm}\beta_{u_1x} + 1 \\
\mathrm{Cov}(X, U_1) &= \beta_{u_1x} \\
\mathrm{Cov}(X, Y) &= \beta_{xy}(\beta_{u_1x}^2 + 1) + \beta_{xm}\beta_{my}(\beta_{u_1x}^2 + 1) + \beta_{u_1x}(\beta_{u_1y} + \beta_{u_1m}\beta_{my}) \\
\mathrm{Cov}(Y, M) &= \beta_{my}\,\mathrm{Var}(M) + \beta_{u_1m}\beta_{u_1y} + \beta_{u_2m}\beta_{u_2y} + \beta_{xm}\beta_{u_1x}\beta_{u_1y} + \\
&\quad \beta_{xm}\beta_{xy}(\beta_{u_1x}^2 + 1) + \beta_{u_1m}\beta_{u_1x}\beta_{xy} \\
\mathrm{Cov}(X, M) &= \beta_{xm}(\beta_{u_1x}^2 + 1) + \beta_{u_1x}\beta_{u_1m},
\end{aligned}$$

in which case

$$\widehat{\beta}_{xy|m} = \frac{((\beta_{u_2m}^2 + 1)\beta_{xy} - \beta_{xm}\beta_{u_2m}\beta_{u_2y})(\beta_{u_1x}^2 + 1) + \beta_{u_1m}^2\beta_{xy} + (\beta_{u_2m}^2 + 1)\beta_{u_1x}\beta_{u_1y}}{(\beta_{u_1x}^2 + 1)(\beta_{u_2m}^2 + 1) + \beta_{u_1m}^2} - \frac{(\beta_{xm}\beta_{u_1y} + \beta_{u_1x}\beta_{u_2m}\beta_{u_2y})\beta_{u_1m}}{(\beta_{u_1x}^2 + 1)(\beta_{u_2m}^2 + 1) + \beta_{u_1m}^2}.$$

Verification of bias expressions

To check the expression for the bias with and without adjustment for M (i.e., expressions (1) and (2)), 10 000 scenarios were evaluated. These were obtained by random sampling of values of the parameters of the data generating model. Specifically, the values of each coefficient (β_{xy}, β_{xm}, etc.) and of each error term (σ_{u_1}, σ_{u_2}, etc.) were sampled from a uniform distribution ($\sim U(0, 2)$), except for σ_y, which was set to 1. For each scenario, data were then generated based on the data generating model. To reduce the role of sampling variability, the sample size was set to 200,000. For each scenario, the bias was calculated using expressions (1) and (2). These were compared to estimates obtained from OLS regression, with and without adjustment for M using the artificially generated data. As a reference, in each generated dataset also the total effect of X on Y was estimated with adjustment for U_1, which corresponds to the data generating model.

The median difference between the expression-based and estimation-based total effect of X on Y was 0.0000, with 95% of the differences in the interval $(-0.0461; 0.0441)$. The median difference between the expression-based and estimation-based crude effect (i.e., without adjustment for U_1 and M) was 0.0000, with 95% of differences lying in the interval $(-0.0173; 0.0185)$. For the adjusted effect (i.e., with adjustment for M, but no adjustment for U_1, nor for U_2), the median difference was 0.0000, with 95% of differences lying in the interval $(-0.0117; 0.0123)$. Based on these numbers we conclude that the bias-expressions appear to be correct.

References

1. Richardson, T. S. & Robins, J. M. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper* **128,** 2013 (2013).