We give results for both conditional natural direct and indirect effects, given covariates $C$, and for marginal natural direct and indirect effects.

**Approach 1.** If (i) $Y_{alm} \perp A|C$, (ii) $Y_{alm} \perp (L, M)|\{A, C\}$, (iii) $(L_a, M_a) \perp A|C$, (iv) $Y_{alm} \perp (L_{a*}, M_{a*})|C$, then we have the following:

$$E[Y_{aL_{a*}M_{a*}}|c] = \sum_{l,m} E[Y_{alm}|c, L_{a*} = l, M_{a*} = m] P(L_{a*} = l, M_{a*} = m|c)$$

$$= \sum_{l,m} E[Y_{alm}|c] P(L_{a*} = l, M_{a*} = m|c) \text{ by (iv)}$$

$$= \sum_{l,m} E[Y_{alm}|c, a] P(L_{a*} = l, M_{a*} = m|c, a^*) \text{ by (i) and (iii)}$$

$$= \sum_{l,m} E[Y_{alm}|c, a, l, m] P(L_{a*} = l, M_{a*} = m|c, a^*) \text{ by (i)}$$

$$= \sum_{l,m} E[Y|c, a, l, m] P(l, m|c, a^*)$$

and thus

$$E[Y_{aL_{a*}M_{a*}}] = \sum_{c,l,m} E[Y|c, a, l, m] P(l, m|c, a^*) P(c).$$

Upon subtracting these expressions corresponding to specific choices of $a$ and $a^*$, one finds expressions for $E[Y_{aL_{a*}M_{a*}} - Y_{a^*L_{a*}M_{a*}}|c]$ and $E[Y_{aL_{a}M_{a}} - Y_{aL_{a*}M_{a*}}|c]$, and for $E[Y_{aL_{a*}M_{a*}} - Y_{a^*L_{a*}M_{a*}}]$ and $E[Y_{aL_{a}M_{a}} - Y_{aL_{a*}M_{a*}}]$.

A weighting-based estimator for the conditional natural direct and indirect effect can be obtained upon duplicating the dataset and adding an exposure variable $A^*$ which is 0 for the first replication and 1 for the second. For each individual, a weight is obtained by taking the product of the predicted probabilities (of the observed confounder and mediator values) from the two logistic regressions had the exposure been $A^*$, divided by the product of the corresponding predicted probabilities from the two logistic regressions had the exposure been as observed:

$$\frac{P(l|a^*, c) P(m|l, a^*, c)}{P(l|a, c) P(m|l, a, c)}.$$

If a model for the outcome is fitted conditional on the two exposures $A$ and $A^*$ and covariates
on the duplicated data set using weighted regression, the direct and indirect effects of interest are obtained as the coefficients of $A$ and $A^*$, respectively.

The validity of this proposed weighting estimator can be understood upon noting that it is obtained under a marginal structural model for the composite counterfactual $Y_{aL_a,M_a^*}$, e.g.:

$$E[Y_{aL_a,M_a^*}|c] = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 c,$$

where $\beta_1(a-a^*) = E[Y_{aL_a,M_a^*} - Y_{a^*L_a,M_a^*}|c]$ and $\beta_2(a-a^*) = E[Y_{aL_a,M_a} - Y_{aL_a^*M_a}|c]$. Upon letting $a^*$ take all possible values over the support of $A$ and noting that the expression for $E[Y_{aL_a,M_a^*}|c]$ can be equivalently rewritten as

$$\sum_{y,l,m} yP(y,l,m|c,a)P(l,m|c,a) = E\left(Y \frac{P(L,M|c,a^*)}{P(L,M|c,a)} | c, a\right),$$

the proposed weighting estimator is obtained. For marginal natural direct and indirect effects, we use that $E[Y_{aL_a,M_a^*}|c]$ can be equivalently rewritten as

$$\sum_{y,c,l,m} yP(y,l,m|c,a)P(c)P(l,m|c,a) = E\left(Y I(A = a) \frac{P(L,M|c,a^*)}{P(a|c)P(L,M|c,a)} \right),$$

following which the proposed estimators are obtained.

**Approach 2.** Let $M_{al}$ denote the value of $M$ that would be observed if $A$ were set to $a$ and $L$ to $l$. Formally, the conditional effects $E_{A\rightarrow Y}(c)$, $E_{A\rightarrow M\rightarrow Y}(c)$, $E_{A\rightarrow LY}(c)$ are defined as follows: $E_{A\rightarrow Y}(c) = E[Y_{aL_a,M_a^*} - Y_{a^*L_a,M_a^*}|c]$, $E_{A\rightarrow M\rightarrow Y}(c) = E[Y_{aL_a,M_{al}} - Y_{aL_a^*M_a}|c]$ and $E_{A\rightarrow LY}(c) = E[Y_{aL_a,M_a} - Y_{aL_a,M_{al}}|c]$. We then have the following effect decomposition:

$$Y_a - Y_{a^*} = Y_{aL_a,M_a} - Y_{a^*L_a,M_a},$$

$$= (Y_{aL_a,M_a} - Y_{aL_a,M_{al}}) + (Y_{aL_a,M_{al}} - Y_{aL_a^*M_a}) + (Y_{aL_a^*M_a} - Y_{a^*L_a^*M_a})$$

and thus $E[Y_a - Y_{a^*}|c] = E_{A\rightarrow LY}(c) + E_{A\rightarrow M\rightarrow Y}(c) + E_{A\rightarrow Y}(c)$.  

22
Suppose that the assumptions (i\textsuperscript{†})-(iv\textsuperscript{†}) in approach 1 hold that (i\textsuperscript{†}) $Y_{atm} \perp A|C$; (ii\textsuperscript{†}) $Y_{atm} \perp (L, M)|\{A, C\}$, (iii\textsuperscript{†}) $(L_a, M_a) \perp A|C$, (iv\textsuperscript{†}) $Y_{atm} \perp (L_{a*}, M_{a*})|C$, along with three further assumptions that (v\textsuperscript{†}) $Y_{atm} \perp (L_{a*}, M_{al})|C$ (vi\textsuperscript{†}) $M_{al} \perp L_{a*}|C$, (vii\textsuperscript{†}) $M_{al} \perp \perp (A, L)|C$. Note all seven of these assumptions would hold if Figure 2 is a causal diagram.\textsuperscript{3,6}

We have already shown that under (i\textsuperscript{†})-(iv\textsuperscript{†}) we have $E[Y_{aL_{a*}M_{a*}}|c] = \sum_{l,m} E[Y|c, a, l, m]P(l, m|c, a^*)$. Under (i\textsuperscript{†})-(vii\textsuperscript{†}), we have

$$E[Y_{aL_{a*}M_{a*}}|c] = \sum_{l,m} E[Y_{alM_{al}}|c, L_{a*} = l]P(L_{a*} = l|c)$$
$$= \sum_{l,m} E[Y_{al}|c, L_{a*} = l, M_{al} = m]P(M_{al} = m|c, L_{a*} = l)P(L_{a*} = l|c)$$
$$= \sum_{l,m} E[Y_{al}|c]P(M_{al} = m|c)P(L_{a*} = l|c) \text{ by (v\textsuperscript{†}) and (vi\textsuperscript{†})}$$
$$= \sum_{l,m} E[Y_{al}|c, a]P(M_{al} = m|c, a, l)P(L_{a*} = l|c, a^*) \text{ by (i\textsuperscript{†}) and (iii\textsuperscript{†}) and (vii\textsuperscript{†})}$$
$$= \sum_{l,m} E[Y_{al}|c, a, l, m]P(M_{al} = m|c, a, l)P(L_{a*} = l|c, a^*) \text{ by (ii\textsuperscript{†})}$$
$$= \sum_{l,m} E[Y|c, a, l, m]P(m|c, a, l)P(l|c, a^*).$$

Since $E[Y_{aL_{a}M_{a}}|c] = \sum_{l,m} E[Y|c, a, l, m]P(l, m|c, a)$ and $E[Y_{a*L_{a*}M_{a*}}|c] = \sum_{l,m} E[Y|c, a^*, l, m]P(l, m|c, a^*)$, we thus have that

$$E_{A \rightarrow Y}(c) = \sum_{l,m} \{E[Y|a, l, m, c] - E[Y|a^*, l, m, c]\}P(l, m|a^*, c)$$
$$E_{A \rightarrow M \rightarrow Y}(c) = \sum_{l,m} E[Y|c, a, l, m]\{P(m|c, a, l) - P(m|c, a^*, l)\}P(l|c, a^*)$$
$$E_{A \rightarrow L \rightarrow Y}(c) = \sum_{l,m} E[Y|c, a, l, m]\{P(l|c, a) - P(l|c, a^*)\}.$$

Marginal effects are similarly obtained, but require additional averaging over the distribution $P(c)$ of $C$.

Under approach 2, a weighting-based estimator for the conditional effects $E_{A \rightarrow L \rightarrow Y}(c)$, $E_{A \rightarrow M \rightarrow Y}(c)$ and $E_{A \rightarrow Y}(c)$ can be obtained upon merging three copies of the dataset and adding exposure variables $A^*$ and $A^{**}$, where $A^*$ equals the observed exposure for the first replication and $1 - A$ for the next two replications, and where $A^{**}$ equals the observed exposure for the first
two replications and 1 – A for the third replication. For each individual, a weight is now obtained by taking the product of the predicted probability (of the observed confounder value) from the first logistic regression had the exposure been A* and the predicted probability (of the observed mediator value) from the second logistic regression had the exposure been A**, divided by the product of the corresponding predicted probabilities from the two logistic regressions had the exposure been as observed:

\[ \frac{P(l|a^*, c)P(m|l, a^{**}, c)}{P(l|a, c)P(m|l, a, c)}. \]

If a model for the outcome is now fitted conditional on the three exposures A, A* and A** and covariates on the obtained data set using weighted regression, the effects E_{A \rightarrow Y}, E_{A \rightarrow LY} and E_{A \rightarrow M \rightarrow Y} of interest are obtained as the coefficients of A, A* and A**, respectively. The validity of this strategy can be understood by making reference to a marginal structural model for the composite counterfactual Y_{aL_0^*M_{a^{**}}L_{a^*}}, e.g.:

\[ E[Y_{aL_0^*M_{a^{**}}L_{a^*}}|c] = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 a^{**} + \beta_4 c, \]

from which it is easily verified that \( \beta_1(a - a^*) = E_{A \rightarrow Y}(c), \beta_2(a - a^*) = E_{A \rightarrow LY}(c) \) and \( \beta_3(a - a^*) = E_{A \rightarrow M \rightarrow Y}(c). \) Upon letting \( a^* \) and \( a^{**} \) take all possible values over the support of \( A \) and noting that by a similar reasoning as before:

\[ E[Y_{aL_0^*M_{a^{**}}L_{a^*}}|c] = \sum_{y,l,m} yP(y|c, a, l, m)P(m|c, a^{**}, l)P(l|c, a^*) \]

\[ = E \left( Y \frac{P(M|c, a^{**}, l)P(L|c, a^*)}{P(M|c, a, l)P(L|c, a)} | a, c \right), \]

the proposed weighting estimator is obtained. Marginal effects are similarly obtained, but require additional weighting by the reciprocal of \( P(a|c) \), as in Approach 1.

**Approach 3.** If (i) \( Y_{am} \perp A|C \), (ii*) \( Y_{am} \perp M|\{A, C, L\} \), and (iii) \( M_a \perp A|C \) then we have
the following:

\[ E(Y_{aG_{a^*}|c}) = \sum_m E[Y_{am}|c, G_{a^*|c} = m]P(G_{a^*|c} = m|c) \]

\[ = \sum_m E[Y_{am}|c]P(M_{a^*} = m|c) \]

\[ = \sum_m E[Y_{am}|a, c]P(M_{a^*} = m|a^*, c) \text{ by (i) and (iii)} \]

\[ = \sum_{l,m} E[Y_{am}|a, l, c]P(l|a, c)P(M_{a^*} = m|a^*, c) \text{ by (ii*)} \]

\[ = \sum_{l,m} E[Y|a, l, m, c]P(l|a, c)P(m|a^*, c). \]

Similarly, \( E(Y_{aG_{a|c}^*}|c) = \sum_{l,m} E[Y|a, l, m, c]P(l|a, c)P(m|a, c) \) and \( E(Y_{a^*G_{a^*|c}}|c) = \sum_{l,m} E[Y|a^*, l, m, c]P(l|a^*, c)P(m|a, c). \) Subtracting gives the expressions for \( E(Y_{aG_{a^*|c}}|c) - E(Y_{a^*G_{a^*|c}}|c) \) and \( E(Y_{aG_{a|c}}|c) - E(Y_{aG_{a|c}^*}|c) \). Marginal effects are similarly obtained, but require additional averaging over the distribution \( P(c) \) of \( C \). As further motivation for this third approach, note that Approaches 1 and 2 become essentially untenable when \( L \) is a high-dimensional confounder as they assume its association with the outcome to be unconfounded after adjustment for \( A \) and \( C \). This third approach, however, works even when the association between \( L \) and \( Y \) is confounded by unmeasured factors.

For conditional effects, Approach 3 works like Approach 1, but using the weights

\[ \sum_{l} \frac{P(m|l, a^*, c)P(l|a^*, c)}{P(m|l, a, c)} \]

instead. The resulting weighting estimator is obtained under a marginal structural model for the composite counterfactual \( Y_{aG_{a^*|c}} \), e.g.:

\[ E[Y_{aG_{a^*|c}}|c] = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 c, \]

from which it is easily verified that \( \beta_1(a - a^*) = E(Y_{aG_{a^*|c}}|c) - E(Y_{a^*G_{a^*|c}}|c) \) and \( \beta_2(a - a^*) = \)
$E(Y_{aG_{a^*c}}|c) - E(Y_{aG_{a^*c}}|c)$. Upon letting $a^*$ take all possible values over the support of $A$ and noting that the expression for $E(Y_{aG_{a^*c}}|c)$ can be equivalently rewritten as

$$\sum_{y,l,m} y P(y,l,m|c,a) \frac{P(l|a,c)P(m|a^*,c)}{P(l,m|c,a)} = E \left( Y \left. \frac{P(M|c,a^*)}{P(M|l,c,a)} \right| c, a \right),$$

the proposed weighting estimator is obtained. Marginal effects are similarly obtained, but require additional averaging over the distribution $P(c)$ of $C$. Note moreover that unlike approaches 1 and 2, the assumptions of approach 3 do not require that Figure 2 is a non-parametric structural equation model$^6$; the assumptions can hold under weaker interpretations of Figure 2 as a causal diagram$^8$.

**SAS implementation.** We describe how the proposed weighting approaches to marginal direct and indirect effects given above can be implemented in SAS statistical software (SAS Institute, Inc., Cary, North Carolina). Below we let $c, a, l, m$ and $y$ correspond to the observed confounders $C$, exposure $A$, exposure-induced confounder $L$, mediator $M$ and outcome $Y$. First, we consider approaches 1 and 3.

```sas
data mydata0;
set mydata;
a = 0; output;
run;
data mydata1;
set mydata;
a = 1; output;
run;
proc logistic data = mydata;
model a = c;
score data = mydata out = preda;
run;
```
data preda;
set preda;
pa1 = P_1;
run;

proc logistic data = mydata;
model l = a c;
score data = mydata1 out = predl1;
score data = mydata0 out = predl0;
run;

data predl1;
set predl1;
pl1 = P_1;
run;

data predl0;
set predl0;
pl10 = P_1;
run;

data mydata00;
set mydata;
a = 0; l = 0; output;
run;

data mydata10;
set mydata;
a = 1; l = 0; output;
run;

data mydata01;
set mydata;
a = 0; l = 1; output;
run;
data mydata11;
set mydata;
a = 1; l = 1; output;
run;
proc logistic data = mydata;
model m = a l c;
score data = mydata1 out = predm1;
score data = mydata0 out = predm0;
score data = mydata00 out = predm00;
score data = mydata01 out = predm01;
score data = mydata10 out = predm10;
score data = mydata11 out = predm11;
run;
data predm1;
set predm1;
pm1 = P_1;
run;
data predm0;
set predm0;
pm0 = P_1;
run;
data predm00;
set predm00;
pm00 = P_1;
run;
data predm10;
set predm10;
    pm110 = P_1;
run;
data predm01;
set predm01;
    pm101 = P_1;
run;
data predm11;
set predm11;
    pm111 = P_1;
run;
data mydataw;
merge preda predl1 predl0 predm1 predm0 predm00 predm10 predm11 mydata;
run;
data mydatanew;
set mydataw;
astar = a; w1 = a/pa1+(1-a)/(1-pa1); output;
astar = 1-a;
if a = 0 then w1 = (1/(1-pa1))*((1*pl1/pl10+(1-l)*(1-pl1)/(1-pl10))
    *((1-m)*(1-pm1)/(1-pm10) + m*pm1/pm10);
if a = 1 then w1 = (1/pa1)*((1*pl10/pl1+(1-l)*(1-pl10)/(1-pl1))
    *((1-m)*(1-pm10)/(1-pm1) + m*pm10/pm1)); output;
run;
data mydatanew;
set mydatanew;
if (a = 0) & (astar = 0) & (m = 1) then
\[ w_3 = \frac{1}{1-p_{a1}} \times \frac{(1-p_{m100})(1-pl_1)+p_{m101}pl_1}{pm_{10}}; \]

if (a = 0) & (astar = 0) & (m = 0) then
\[ w_3 = \frac{1}{1-p_{a1}} \times \frac{(1-p_{m100})(1-pl_1)+(1-p_{m101})pl_1}{(1-pm_{10})}; \]

if (a = 0) & (astar = 1) & (m = 1) then
\[ w_3 = \frac{1}{1-p_{a1}} \times \frac{p_{m110}(1-pl_1)+p_{m111}pl_1}{pm_{10}}; \]

if (a = 0) & (astar = 1) & (m = 0) then
\[ w_3 = \frac{1}{1-p_{a1}} \times \frac{(1-p_{m110})(1-pl_1)+(1-p_{m111})pl_1}{(1-pm_{10})}; \]

if (a = 1) & (astar = 0) & (m = 1) then
\[ w_3 = \frac{1}{pa_{1}} \times \frac{p_{m100}(1-pl_1)+p_{m101}pl_1}{pm_{1}}; \]

if (a = 1) & (astar = 0) & (m = 0) then
\[ w_3 = \frac{1}{pa_{1}} \times \frac{(1-p_{m100})(1-pl_1)+(1-p_{m101})pl_1}{(1-pm_{1})}; \]

if (a = 1) & (astar = 1) & (m = 1) then
\[ w_3 = \frac{1}{pa_{1}} \times \frac{p_{m110}(1-pl_1)+p_{m111}pl_1}{pm_{1}}; \]

if (a = 1) & (astar = 1) & (m = 0) then
\[ w_3 = \frac{1}{pa_{1}} \times \frac{(1-p_{m110})(1-pl_1)+(1-p_{m111})pl_1}{(1-pm_{1})}; \]

run;

The results from approach 1 (except for standard errors) can now be obtained from:

```plaintext
proc logistic data = mydatanew;
where astar = 0;
model y = a;
weight w1;
run;
```

for the natural direct effect, and

```plaintext
proc logistic data = mydatanew;
where a = 1;
model y = astar;
```
weight w1;
run;

for the natural indirect effect, where \(a_{star}\) corresponds to \(A^*\). The results from approach 3 (except for standard errors) can be obtained using the same commands, upon substituting \(w1\) by \(w3\). Finally, the results from approach 2 (except for standard errors) can be obtained from:

data mydatanew2;
set mydataw;
astar = a; astarstar = a; w2 = a/pa1+(1-a)/(1-pa1); output;
astar = 1-a; astarstar = a;
w2 = (a/pa1)*(l*pl10/pl1+(1-l)*(1-pl10)/(1-pl1)) +
((1-a)/(1-pa1))*(l*pl1/pl10+(1-l)*(1-pl1)/(1-pl10)); output;
astar = 1-a; astarstar = 1-a;
w2 = (a/pa1)*(l*(pl10/pl1)*(m*(pm101/pm111)+(1-m)*(1-pm101)/(1-pm111))
+(1-l)*((1-pl10)/(1-pl1))*(m*(pm100/pm110)+(1-m)*(1-pm100)/(1-pm110)))
+((1-a)/(1-pa1))*(l*pl1/pl10)*m*(pm111/pm101)+(1-m)*(1-pm111)/(1-pm101))
+(1-l)*((1-pl1)/(1-pl10))*m*(pm110/pm100)+(1-m)*(1-pm110)/(1-pm100)); output;
run;
proc logistic data = mydatanew;
where (astar = 0) & (astarstar = 0);
model y = a;
weight w2;
run;

for \(E_{A \rightarrow Y}\),

proc logistic data = mydatanew;
where (a = 1) & (astarstar = 0);
model y = astar;
weight w2;
run;

for $E_{A\rightarrow M \rightarrow Y}$ and

proc logistic data = mydatanew;
where (a = 1) & (astar = 1);
model y = astarstar;
weight w2;
run;

for $E_{A \rightarrow L \rightarrow Y}$.

When the mediator is not binary, then the above weight calculations change. For instance, when the mediator is continuous, then one may calculate the term $P(m|l, a, c)$, which appears in the inverse probability weights, via:

proc genmod data = mydata;
model m = l a c / error = n;
output out = mydatam p = predm;
run;
data mydatam;
set mydatam;
pm = pdf('normal', m, predm, sigma);
run;

where sigma must be replaced by the estimated residual standard deviation, as obtained from the SAS output. After running this, the data set mydatam will contain a column pm, which contains the estimates of $P(m|l, a, c)$.

Additional References