

**eAppendix. Empirical Formulae for Controlled Direct Effects and Natural Direct and Indirect Effects.**

If conditions (1) and (2) hold then, it follows from Robins' g-formula<sup>25,26</sup> that the average controlled direct effect is given by:

$$\begin{aligned}\mathbb{E}[Y_{az} - Y_{a^*z}] &= \sum_x \sum_w \mathbb{E}[Y|A = a, Z = z, X = x, W = w]P(W = w|A = a, X = x)P(X = x) \\ &\quad - \sum_x \sum_w \mathbb{E}[Y|A = a^*, Z = z, X = x, W = w]P(W = w|A = a^*, X = x)P(X = x).\end{aligned}$$

Note that if  $W = \emptyset$  so that the set of variables  $X$  that suffices to control for confounding for the treatment-outcome relationship also suffices to control for confounding for the mediator-outcome relationship then the above expression reduces to

$$\mathbb{E}[Y_{az} - Y_{a^*z}] = \sum_x \{\mathbb{E}[Y|X = x, A = a, Z = z] - \mathbb{E}[Y|X = x, A = a^*, Z = z]\}P(X = x).^{10}$$

For natural direct effects it follows from the work of Pearl<sup>8</sup> that if conditions (1)-(4) hold then  $\mathbb{E}[Y_{aZ_{a^*}} - Y_{a^*Z_{a^*}}]$

$$\begin{aligned}&= \sum_{x,w} \sum_z \{\mathbb{E}[Y|A = a, Z = z, X = x, W = w]P(W = w|A = a, X = x)P(Z = z|A = a^*, X = x)P(X = x) \\ &\quad - \sum_{x,w} \sum_z \mathbb{E}[Y|A = a^*, Z = z, X = x, W = w]P(W = w|A = a^*, X = x)\}P(Z = z|A = a^*, X = x)P(X = x).\end{aligned}$$

If  $W = \emptyset$  then the above expression reduces to

$$\mathbb{E}[Y_{aZ_{a^*}} - Y_{a^*Z_{a^*}}] = \sum_x \sum_z \{\mathbb{E}[Y|A = a, Z = z, X = x] - \mathbb{E}[Y|A = a^*, Z = z, X = x]\}P(Z = z|A = a^*, X = x)P(X = x).$$

If conditions (1)-(4) hold natural indirect effects can be computed by subtracting natural direct effect from total effects. Also if (1)-(4) hold then  $\mathbb{E}[Y_{aZ_a} - Y_{aZ_{a^*}}]$

$$\begin{aligned}&= \sum_{x,w} \sum_z \mathbb{E}[Y|A = a, X = x, Z = z, W = w]P(W = w|A = a, X = x)P(Z = z|A = a, X = x)P(X = x) \\ &\quad - \sum_{x,w} \sum_z \mathbb{E}[Y|A = a, X = x, Z = z, W = w]P(W = w|A = a, X = x)P(Z = z|A = a^*, X = x)P(X = x).\end{aligned}$$

If  $W = \emptyset$  then the above expression reduces to

$$\begin{aligned}\mathbb{E}[Y_{aZ_a} - Y_{aZ_{a^*}}] &= \sum_x \sum_z \mathbb{E}[Y|A = a, Z = z, X = x]P(Z = z|A = a, X = x)P(X = x) \\ &\quad - \sum_x \sum_z \mathbb{E}[Y|A = a, X = x, Z = z]P(Z = z|A = a^*, X = x)P(X = x).\end{aligned}$$