Supplement for *Simple sensitivity analysis for control selection bias*

In a case-control study we can estimate the odds ratio, conditional on covariates C,

$$\frac{\Pr(Y=1\mid A=1,S=1,c)}{\Pr(Y=0\mid A=1,S=1,c)}\bigg/\frac{\Pr(Y=1\mid A=0,S=1,c)}{\Pr(Y=0\mid A=0,S=1,c)}$$

where Y indicates case vs. control status, A is the binary exposure of interest, and S is an indicator of selection into the case-control study.

This quantity can be used to estimate the population odds ratio

$$\frac{\Pr(Y=1\mid A=1,c)}{\Pr(Y=0\mid A=1,c)}\bigg/\frac{\Pr(Y=1\mid A=0,c)}{\Pr(Y=0\mid A=0,c)}$$

without bias, as long as $\Pr(A=1\mid Y=0,S=1,c)=\Pr(A=1\mid Y=0,c)$ and $\Pr(A=1\mid Y=1,S=1,c)=\Pr(A=1\mid Y=1,c)$. In other words, selection of both cases and controls must be independent of exposure.

Although it may be straightforward to randomly sample the cases with respect to the distribution of their exposure, often because the cases can be fully enumerated, control selection is usually more difficult. When the sampled controls do not represent the exposure distribution in the source population, selection bias can result.

To quantify the possible size of this bias, consider the ratio of the observable odds ratio from case-control data to the odds ratio that would have been estimated had the entire cohort been sampled. Assume then that any bias from the case-control study is due to this selection.

We therefore have:

$$\begin{aligned}\text{bias} &= \left\{\frac{\Pr(Y=1\mid A=1,S=1,c)}{\Pr(Y=0\mid A=1,S=1,c)}\bigg/\frac{\Pr(Y=1\mid A=0,S=1,c)}{\Pr(Y=0\mid A=0,S=1,c)}\right\}\bigg/\\ &\qquad\qquad\left\{\frac{\Pr(Y=1\mid A=1,c)}{\Pr(Y=0\mid A=1,c)}\bigg/\frac{\Pr(Y=1\mid A=0,c)}{\Pr(Y=0\mid A=0,c)}\right\}.\end{aligned}$$

We can rewrite each odds ratio in terms of the probability of the exposure:

$$\begin{aligned}\text{bias} &= \left\{\frac{\Pr(A=1\mid Y=1,S=1,c)}{\Pr(A=0\mid Y=1,S=1,c)}\bigg/\frac{\Pr(A=1\mid Y=0,S=1,c)}{\Pr(A=0\mid Y=0,S=1,c)}\right\}\bigg/\\ &\qquad\qquad\left\{\frac{\Pr(A=1\mid Y=1,c)}{\Pr(A=0\mid Y=1,c)}\bigg/\frac{\Pr(A=1\mid Y=0,c)}{\Pr(A=0\mid Y=0,c)}\right\}.\end{aligned}$$

Now assume that the cases have been properly sampled independently of exposure status, such that $A \perp\!\!\!\perp S \mid Y = 1$, but that the independence does not hold for $Y = 0$:

$$\text{bias} \;= \frac{\Pr(A=1 \mid Y=0, c)}{\Pr(A=0 \mid Y=0, c)} \Bigg/ \frac{\Pr(A=1 \mid Y=0, S=1, c)}{\Pr(A=0 \mid Y=0, S=1, c)} \,.$$

Following the logic in Smith & Vanderweele 2019,[1] we see that

$$\begin{aligned}
\text{bias} \;&\leq \frac{\max_s \Pr(A=1 \mid Y=0, S=s, c)}{\min_s \Pr(A=0 \mid Y=0, S=s, c)} \Bigg/ \frac{\Pr(A=1 \mid Y=0, S=1, c)}{\Pr(A=0 \mid Y=0, S=1, c)} \\
&\leq \frac{\Pr(A=1 \mid Y=0, S=0, c)}{\Pr(A=0 \mid Y=0, S=0, c)} \Bigg/ \frac{\Pr(A=1 \mid Y=0, S=1, c)}{\Pr(A=0 \mid Y=0, S=1, c)} \\
&= \frac{\Pr(A=1 \mid Y=0, S=0, c)}{\Pr(A=1 \mid Y=0, S=1, c)} \Bigg/ \frac{\Pr(A=0 \mid Y=0, S=0, c)}{\Pr(A=0 \mid Y=0, S=1, c)} \,.
\end{aligned}$$

Suppose there exists some U such that $A \perp\!\!\!\perp S \mid Y = 0, C, U$. For notational simplicity we will assume discrete U. Then we can write, by Lemma A.3 in Ding & VanderWeele 2016:[2]

$$\begin{aligned}
\text{bias} \;&\leq \left\{ \frac{\sum_u \Pr(A=1 \mid Y=0, S=0, c, u) \Pr(U=u \mid Y=0, S=0, c)}{\sum_u \Pr(A=1 \mid Y=0, S=1, c, u) \Pr(U=u \mid Y=0, S=1, c)} \right\} \Bigg/ \\
&\qquad\qquad \left\{ \frac{\sum_u \Pr(A=0 \mid Y=0, S=0, c, u) \Pr(U=u \mid Y=0, S=0, c)}{\sum_u \Pr(A=0 \mid Y=0, S=1, c, u) \Pr(U=u \mid Y=0, S=1, c)} \right\} \\
&= \left\{ \frac{\sum_u \Pr(A=1 \mid Y=0, c, u) \Pr(U=u \mid Y=0, S=0, c)}{\sum_u \Pr(A=1 \mid Y=0, c, u) \Pr(U=u \mid Y=0, S=1, c)} \right\} \Bigg/ \\
&\qquad\qquad \left\{ \frac{\sum_u \Pr(A=0 \mid Y=0, c, u) \Pr(U=u \mid Y=0, S=0, c)}{\sum_u \Pr(A=0 \mid Y=0, c, u) \Pr(U=u \mid Y=0, S=1, c)} \right\} \\
&\leq \left\{ \frac{\text{RR}_{UA_1} \times \text{RR}_{S_0U}}{\text{RR}_{UA_1} + \text{RR}_{S_0U} - 1} \right\} \times \left\{ \frac{\text{RR}_{UA_0} \times \text{RR}_{S_1U}}{\text{RR}_{UA_0} + \text{RR}_{S_1U} - 1} \right\}
\end{aligned}$$

where

$$\begin{aligned}
\text{RR}_{UA_1} &= \frac{\max_u \Pr(A=1|Y=0, u, c)}{\min_u \Pr(A=1|Y=0, u, c)} \\
\text{RR}_{UA_0} &= \frac{\max_u \Pr(A=0|Y=0, u, c)}{\min_u \Pr(A=0|Y=0, u, c)} \\
\text{RR}_{S_1U} &= \max_u \frac{\Pr(U=u|Y=0, S=1, c)}{\Pr(U=u|Y=0, S=0, c)} \\
\text{RR}_{S_0U} &= \max_u \frac{\Pr(U=u|Y=0, S=0, c)}{\Pr(U=u|Y=0, S=1, c)} \,.
\end{aligned}$$

REFERENCES

1. Smith LH, VanderWeele TJ. Bounding bias due to selection. *Epidemiology.* 2019;30:509–516.

2. Ding P, VanderWeele TJ. Sharp sensitivity bounds for mediation under unmeasured mediator-outcome confounding. *Biometrika.* 2016;103:483–490.