1 Proof that causal effects in the selected population are not estimable under outcome-influenced selection

Define $p_{ay.1} = p(A = a, Y = y|S = 1)$, $q_{a.1} = p(Y_a = 1|S = 1)$, $\mathbf{p}_{.1} = \{p_{00.1}, p_{01.1}, p_{10.1}, p_{11.1}\}$ and $\mathbf{q}_{.1} = \{q_{0.1}, q_{1.1}\}$. Similarly, define $p_{ay1} = p(A = a, Y = y, S = 1)$, $q_{a1} = p(Y_a = 1, S = 1)$, $\mathbf{p}_1 = \{p_{001}, p_{011}, p_{101}, p_{111}\}$ and $\mathbf{q}_1 = \{q_{01}, q_{11}\}$. Causal effects in the selected population are contrasts between the elements of $\mathbf{q}_{.1}$. Let S_y denote the potential outcome of S for a given subject if, possibly contrary to fact, exposed to level $Y = y$. Note that $\mathbf{p}_1$ and $\mathbf{q}_1$ are obtained by marginalizing the distribution $p(A, Y_0, Y_1, S_0, S_1)$. Specifically, due to consistency (1) in the main text we have that

$$\begin{aligned} p_{ay1} &= p(A = a, Y_a = y, S_y = 1) \\ &= \sum_{y' \in \{0,1\}} p(A = a, Y_a = y, Y_{1-a} = y', S_y = 1) \end{aligned} \tag{1}$$

and

$$\begin{aligned} q_{a1} &= \sum_{a',y} p(A = a', Y_a = 1, Y_{1-a} = y, S = 1) \\ &= \sum_{y} p(A = a, Y_a = 1, Y_{1-a} = y, S_1 = 1) + p(A = 1 - a, Y_a = 1, Y_{1-a} = y, S_y = 1). \end{aligned} \tag{2}$$

Since $p(S=1)=\sum_{a,y} p_{ay1}$ it follows that $\mathbf{p}_{.1}=\mathbf{p}_1/p(S=1)$ and $\mathbf{q}_{.1}=\mathbf{q}_1/p(S=1)$ are obtained by first marginalizing then conditioning the distribution $p(A,Y_0,Y_1,S_0,S_1)$. To show that causal effects in the selected population are not estimable under the causal diagram in Figure 3 in the main text we thus show that, for any given $\mathbf{p}_{.1}$, it is possible to find two valid distributions $p(A,Y_0,Y_1,S_0,S_1)$ which imply the same given $\mathbf{p}_{.1}$ but two different $\mathbf{q}_{.1}$.

First, define $p=p(A=1)$, $v_{ij}=p(Y_0=i,Y_1=j)$, $r_y=p(S_y=1)$, $r=p(S=1)$ and $\theta=\{p,v_{00},v_{01},v_{10},v_{11},r_0,r_1\}$. Under the causal diagram in Figure in the main text, the distribution $p(A,Y_0,Y_1,S_0,S_1)$ factorizes into $p(A)p(Y_0,Y_1)p(S_0,S_1)$. From (1) and (2) we thus have that

$$\begin{aligned}
p_{00.1} &= (1-p)(v_{00}+v_{01})r_0/r \\
p_{01.1} &= (1-p)(v_{10}+v_{11})r_1/r \\
p_{10.1} &= p(v_{00}+v_{10})r_0/r \\
p_{11.1} &= p(v_{01}+v_{11})r_1/r
\end{aligned} \tag{3}$$

and

$$\begin{aligned}
q_{0.1} &= \{(1-p)(v_{10}+v_{11})r_1+p(v_{10}r_0+v_{11}r_1)\}/r \\
q_{1.1} &= \{p(v_{01}+v_{11})r_1+(1-p)(v_{01}r_0+v_{11}r_1)\}/r.
\end{aligned} \tag{4}$$

Define $v_a = p(Y_a = 1)$. If $Y_0 \perp Y_1$, then

$$\begin{aligned} v_{00} &= (1 - v_0)(1 - v_1) \\ v_{01} &= (1 - v_0)v_1 \\ v_{10} &= v_0(1 - v_1) \\ v_{11} &= v_0 v_1 \end{aligned}$$

so that (3) and (4) simplify to

$$\begin{aligned} p_{00.1} &= (1 - p)(1 - v_0)r_0/r \\ p_{01.1} &= (1 - p)v_0 r_1/r \\ p_{10.1} &= p(1 - v_1)r_0/r \\ p_{11.1} &= p v_1 r_1/r \end{aligned} \tag{5}$$

and

$$\begin{aligned} q_{0.1} &= [(1 - p)v_0 r_1 + p\{v_0(1 - v_1)r_0 + v_0 v_1 r_1\}]/r \\ q_{1.1} &= [p v_1 r_1 + (1 - p)\{(1 - v_0)v_1 r_0 + v_0 v_1 r_1\}]/r. \end{aligned}$$

Define $x = r_1/r_0$. Considering r, x and $\mathbf{p}_{.1}$ as fixed and solving (5) for

$\{p, v_0, v_1, r_0, r_1\}$ gives

$$
\begin{aligned}
p &= \frac{xp_{10.1} + p_{11.1}}{xp_{10.1} + p_{11.1} + xp_{00.1} + p_{01.1}} \\
v_0 &= \frac{p_{01.1}}{xp_{00.1} + p_{01.1}} \\
v_1 &= \frac{p_{11.1}}{xp_{10.1} + p_{11.1}} \\
r_0 &= \frac{(xp_{00.1} + p_{01.1} + xp_{10.1} + p_{11.1})r}{x} \\
r_1 &= (xp_{00.1} + p_{01.1} + xp_{10.1} + p_{11.1})r.
\end{aligned}
\tag{6}
$$

For this solution we have that $0 \leq \{p, v_0, v_1, r_0\} \leq 1$ if $x \geq 1$. We also have that $0 \leq r_1 \leq 1$ if

$$x \leq \frac{r^{-1} - p_{01.1} - p_{11.1}}{p_{00.1} + p_{10.1}} = \frac{r^{-1} - p_{01.1} - p_{11.1}}{1 - p_{01.1} - p_{11.1}}.$$

The range

$$1 \leq x \leq \frac{r^{-1} - p_{01.1} - p_{11.1}}{1 - p_{01.1} - p_{11.1}} \tag{7}$$

is non-empty, provided that $r < 1$. Thus, for any given $\mathbf{p}_{.1}$ we may find two valid distributions $p(A, Y_0, Y_1, S_0, S_1)$ which imply the same given $\mathbf{p}_{.1}$ by first setting r to an arbitrary number > 0 and < 1, then choosing two arbitrary values of x in the range (7) and solving (6) for $\{p, v_0, v_1, r_0, r_1\}$.

It remains to show that the obtained solutions $q_{0.1}$ and $q_{1.1}$ are non-constant functions of x. We show this by a numerical example. Figure 1 shows $q_{0.1}$ and $q_{1.1}$ as functions of x for $r = 0.05$, $p_{00.1} = 0.1$, $p_{01.1} = 0.2$, $p_{10.1} = 0.3$ and $p_{11.1} = 0.4$. Clearly, these are non-constant, which completes the proof.

2 Derivation of the Robins-Manski bounds for causal effects in the selected population under outcome-influenced selection

We have that

$$\begin{aligned} p(Y_a = 1|S = 1) &= p(Y_a = 1|A = a, S = 1)p(A = a|S = 1) \\ &+ p(Y_a = 1|A = 1 - a, S = 1)p(A = 1 - a|S = 1) \\ &= p(Y = 1|A = a, S = 1)p(A = a|S = 1) \\ &+ p(Y_a = 1|A = 1 - a, S = 1)p(A = 1 - a|S = 1), \end{aligned}$$

where the first equality follows from the law of total probability, and the second from consistency (1) in the main text. The right-hand side of this expression in minimized when $p(Y_a = 1|A = 1-a, S = 1) = 0$ and maximized when $p(Y_a = 1|A = 1 - a, S = 1) = 1$, which gives the bounds for $p(Y_a = 1|S = 1)$ in (6) in the main text.

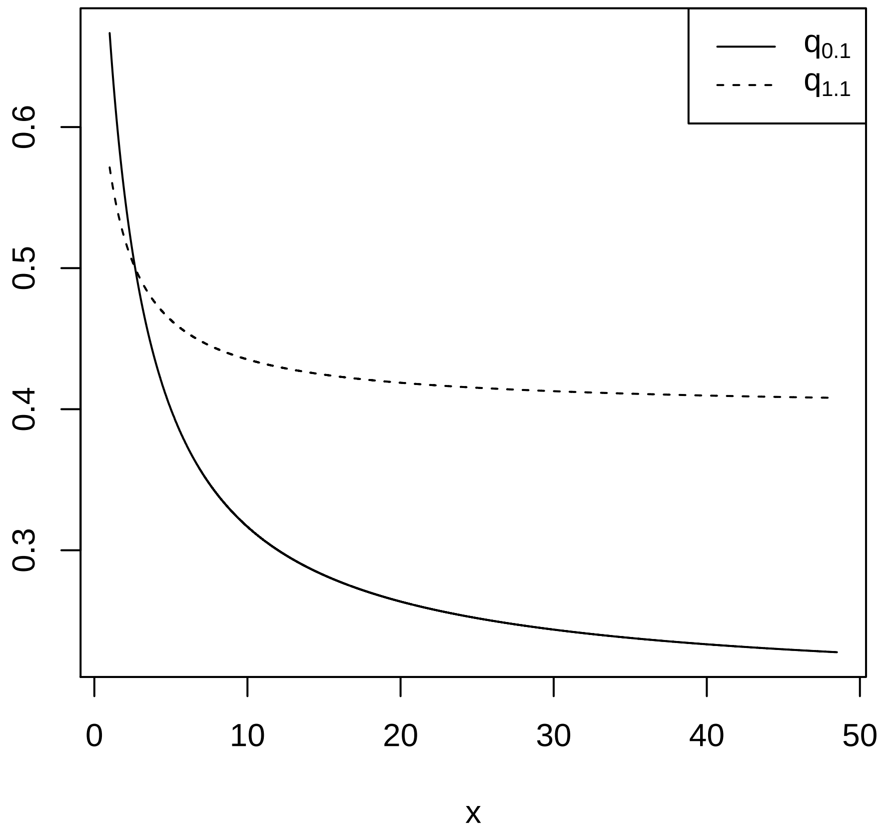

Figure 1: $q_{0.1}$ and $q_{1.1}$ as functions of x for $r = 0.05$, $p_{00.1} = 0.1$, $p_{01.1} = 0.2$, $p_{10.1} = 0.3$ and $p_{11.1} = 0.4$.

3 Proof that the observed data have no information about causal effects in the source population under outcome-associated selection

Consider an arbitrary distribution $p^*(A, Y|S = 1)$ and arbitrary counterfactual probabilities $p^*(Y_1 = 1)$ and $p^*(Y_0 = 1)$ under the causal diagram in Figure 2 in the main text. We first note that

$$\begin{aligned} p^*(Y_a = 1) &= p^*(Y_a = 1|A = a) \\ &= p^*(Y = 1|A = a), \end{aligned}$$

where the first equality follows from the fact that $Y_a \perp A$ under the counterfactual diagram in Figure 4 in the main text, and the second from consistency (1) in the main text. To prove the desired result we thus need to show that it is possible to construct a valid distribution $p(A, Y, U, S)$ that obeys the factorization $p(A, Y, U, S) = p(A)p(U)p(Y|A, U)p(S|U)$ implied by the causal diagram in Figure 2 in the main text, and marginalizes to $p^*(A, Y|S = 1)$, $p^*(Y_1 = 1) = p^*(Y = 1|A = 1)$ and $p^*(Y_0 = 1) = p^*(Y = 1|A = 0)$. We proceed through the following steps.

1. Set $p(A = 1) = p^*(A = 1|S = 1)$.

2. Let U be binary and set $U = S$.

3. If $p^*(Y=1|A=a) = p^*(Y=1|A=a, S=1)$ for $a \in \{0,1\}$, then set $p(U=1|A=a) = p(U=1) = 1/2$ for $a \in \{0,1\}$. Otherwise, set

$$
\begin{aligned}
& p(U=1|A=a) = p(U=1) = \\
& \quad = \min\left\{ \frac{p^*(Y=1|A=0)}{p^*(Y=1|A=0,S=1)}, \frac{p^*(Y=0|A=0)}{p^*(Y=0|A=0,S=1)}, \right. \\
& \quad \left. \frac{p^*(Y=1|A=1)}{p^*(Y=1|A=1,S=1)}, \frac{p^*(Y=0|A=1)}{p^*(Y=0|A=1,S=1)} \right\}
\end{aligned}
$$

   for $a \in \{0,1\}$.

4. Set $p(Y=1|A=a, U=1) = p^*(Y=1|A=a, S=1)$ for $a \in \{0,1\}$

5. Set

$$p(Y=1|A=a, U=0) = \frac{p^*(Y=1|A=a) - p(Y=1|A=a, U=1)p(U=1)}{p(U=0)}$$

   for $a \in \{0,1\}$, where we define 0/0=1.

The constructed distribution completely defines the joint distribution $p(A, Y, U, S)$. We have that

$$
\begin{aligned}
p(A, Y, U, S) &= p(A)p(U|A)p(Y|A, U)p(S|A, Y, U) \\
&= p(A)p(U)p(Y|A, U)p(S|U),
\end{aligned}
$$

where the second equality follows from steps 2 and 3. We further have that

$$\begin{aligned}
p(A = a, Y = y|S = 1) &= p(A = a|S = 1)p(Y = y|A = a, S = 1) \\
&= p(A = a|U = 1)p(Y = y|A = a, U = 1) \\
&= p(A = a)p(Y = y|A = a, U = 1) \\
&= p^*(A = a|S = 1)p^*(Y = y|A = a, S = 1) \\
&= p^*(A = a, Y = y|S = 1),
\end{aligned}$$

where the second equality follows from step 2, the third equality follows from step 3, and the fourth equality follows from steps 1 and 4. We further have that

$$\begin{aligned}
&p(Y = y|A = a) = p(Y = y|A = a, U = 0)p(U = 0|A = a) \\
&\quad + p(Y = y|A = a, U = 1)p(U = 1|A = a) \\
&\quad = p(Y = y|A = a, U = 0)p(U = 0) + p(Y = y|A = a, U = 1)p(U = 1) \\
&\quad = p^*(Y = 1|A = a),
\end{aligned}$$

where the second equality follows from step 3 and the third equality follows from step 5. It remains to show that the constructed distribution $p(A, Y, U, S)$ is valid. From step 1 we have that $0 \leq p(A = 1) \leq 1$; from step 3 we have that $0 \leq p(U = 1|A = a) \leq 1$ for $a \in \{0, 1\}$; from step 4 we have that $0 \leq p(Y = 1|A = a, U = 1) \leq 1$ for $a \in \{0, 1\}$; from step 2 we have that $p(S = 1|A, Y, U = 1) = 1$ and $p(S = 1|A, Y, U = 0) = 0$; from step 5 we have that $0 \leq p(Y=1|A=a,U=0) \leq 1$ if $p(U=1) \leq \min\{p^*(Y=1|A=a)/p(Y=1|A=a,U=1), p^*(Y=0|A=a)/p(Y=0|A=a,U=1)\}$, which is implied by steps 3 and 4.

4 Proof that the causal odds ratio in the source population is equal to the odds ratio in the selected population under outcome-associated selection

We have that

$$\begin{aligned}
\frac{p(Y_1=1)/p(Y_1=0)}{p(Y_0=1)/p(Y_0=0)} &= \frac{p(Y_1=1|A=1)/p(Y_1=0|A=1)}{p(Y_0=1|A=0)/p(Y_0=0|A=0)} \\
&= \frac{p(Y=1|A=1)/p(Y=0|A=1)}{p(Y=1|A=0)/p(Y=0|A=0)} \\
&= \frac{p(A=1|Y=1)/p(A=0|Y=1)}{p(A=1|Y=0)/p(A=0|Y=0)} \\
&= \frac{p(A=1|Y=1,S=1)/p(A=0|Y=1,S=1)}{p(A=1|Y=0,S=1)/p(A=0|Y=0,S=1)} \\
&= \frac{p(Y=1|A=1,S=1)/p(Y=0|A=1,S=1)}{p(Y=1|A=0,S=1)/p(Y=0|A=0,S=1)},
\end{aligned}$$

where the first equality follows from the fact that $Y_a \perp A$ under the counterfactual diagram in Figure 6 in the main text, the second from consistency (1) in the main text, the third from Bayes' theorem, the fourth from the fact that $A \perp S|Y$ under the causal diagram in Figure 3 in the main text, and the fifth from Bayes' theorem.

5 Comparison with the bounds by Kuroki et al. (2010)

Define $p_y = p(A = 1|Y = y, S = 1)$. Kuroki et al (2010) derived the following bounds for the causal risk difference under case-control sampling:

$$\min\left(p_1 - 1, -p_0\right) \leq p(Y_1 = 1) - p(Y_0 = 1) \leq \max\left(p_1, 1 - p_0\right). \tag{8}$$

Let l and u be the lower and upper bound in (8), and let $\tilde{l}$ and $\tilde{u}$ be our lower and upper bound in (9) in the main text. We show that $l \leq \tilde{l}$; that $u \geq \tilde{u}$ can be shown analogously.

By Bayes' theorem, the odds ratio OR in (9) in the main text can be expressed as

$$OR = \frac{p_1(1 - p_0)}{(1 - p_1)p_0}.$$

If $0 \leq \frac{\sqrt{OR}-1}{\sqrt{OR}+1}$, then $\tilde{l} = 0$. In this case, $l \leq \tilde{l}$, since $l \leq 0$. We thus proceed by considering the case when $0 > \frac{\sqrt{OR}-1}{\sqrt{OR}+1}$ so that $\tilde{l} = \frac{\sqrt{OR}-1}{\sqrt{OR}+1}$. Suppose first that

$$p_1 - 1 \leq -p_0, \tag{9}$$

so that $l = p_1 - 1$. We then have that $l \leq \tilde{l}$ if

$$p_1 - 1 \leq \frac{\sqrt{OR} - 1}{\sqrt{OR} + 1}.$$

After a bit of algebra, this relation can be simplified to

$$p_1(1 - p_1)p_0 \leq (1 - p_0)(2 - p_1)^2.$$

This relation holds, since (9) implies that $p_1 \leq 1 - p_0$, and $(1 - p_1)p_0 \leq 1$ whereas $(2 - p_1)^2 \geq 1$. Suppose next that

$$p_1 - 1 > -p_0, \tag{10}$$

so that $l = -p_0$. We then have that $l \leq \tilde{l}$ if

$$-p_0 \leq \frac{\sqrt{OR} - 1}{\sqrt{OR} + 1}.$$

After a bit of algebra, this relation can be simplified to

$$(1 - p_0)(1 - p_1)p_0 \leq p_1(1 + p_0)^2.$$

This relation holds, since (10) implies that $(1 - p_0) < p_1$, and $(1 - p_1)p_0 \leq 1$ whereas $(1 + p_0)^2 \geq 1$.

Finally, the bounds by Kuroki et al (2010) for the causal risk ratio are $[0, \infty)$, which are trivially wider than our bounds in (10) in the main text.