

Selection Bias with Outcome-dependent Sampling

 Arvid Sjölander

Abstract: In a seminal paper, Hernán et al. 2004 provided a systematic classification of selection biases, for scenarios where the selection is a collider between the exposure and the outcome. Hernán 2017 discussed another scenario, where the selection is statistically independent of the exposure, but associated with the outcome through common causes. In this note, we extend the discussion to scenarios where the selection is directly influenced by the outcome, but not by the exposure. We discuss whether these types of outcome-dependent selections preserve the sharp causal null hypothesis, and whether or not they allow for estimation of causal effects in the selected sample and/or in the source population.

Keywords: Causal diagrams; Causality; Counterfactual graphs; Outcome-dependent sampling; Selection bias

(*Epidemiology* 2023;34: 186–191)

Beside confounding, nonrandom selection is probably the most important source of bias in epidemiologic studies. Hernán et al.¹ provided a systematic classification of selection biases, based on causal diagrams. In their classification, the selection into the study is a “collider” on a noncausal path between the exposure and the outcome. One such scenario is depicted in Figure 1, where the exposure, outcome and selection indicator are denoted A , Y , and S , respectively. The conditioning on being selected into the study is illustrated with the square box around $S = 1$. This conditioning induces a noncausal association between the exposure and the outcome, even in the absence of a causal exposure effect, thus invalidating both hypothesis testing and effect estimation.^{2,3}

Hernán⁴ discussed another type of selection bias, where the selection is statistically independent of the exposure but

associated with the outcome through common causes, as in Figure 2⁴. He showed, using standard graphical rules, that this selection is more benign than those considered by Hernán et al.¹ in that it preserves the sharp causal null, that is, it does not induce an association between the exposure and the outcome unless the exposure has a causal effect on the outcome for at least some individuals. Thus, a hypothesis test of the sharp causal null remains valid. However, he showed with numerical examples that this selection may give bias away from the sharp causal null in that the exposure–outcome association in the selected population is not generally equal to the exposure effect in the source population, that is, the population from which the data were taken.

There are several issues with the discussion by Hernán.⁴ Although many readers may be content with the numerical examples that he gave, some may find the lack of formal proofs unsatisfactory. Furthermore, whereas the causal diagram in Figure 2 represents one possible mechanism of outcome-dependent selection, there are other such mechanisms. In particular, the causal diagram in Figure 3 represents the important scenario where the selection is directly influenced by the outcome, but not by the exposure. An obvious example is the unmatched case–control study, in which the selection is by definition influenced by the outcome status alone. To distinguish between the scenarios in Figures 2 and 3, we refer to them as “outcome-associated selection” and “outcome-influenced selection,” respectively. Finally, Hernán⁴ focused

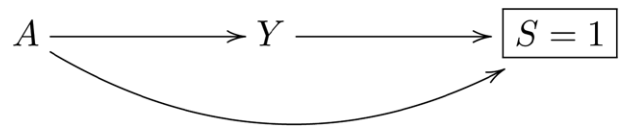


FIGURE 1. A causal diagram illustrating selection that is directly influenced by both the exposure and the outcome.

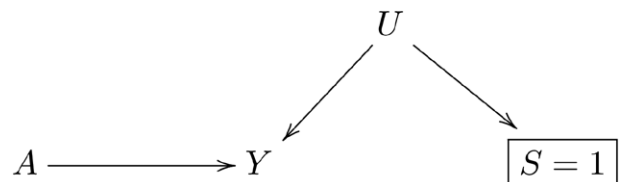


FIGURE 2. A causal diagram illustrating outcome-associated selection, where the selection is statistically independent of the exposure, but associated with the outcome through common causes.

Submitted July 10, 2022; accepted November 20, 2022

From the Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden.

Financially supported by The Swedish Research council, grant number 2020-01188.

The authors report no conflicts of interest.

This article does not include any real data analysis.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Arvid Sjölander, Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Nobels väg 12A, 171 77 Stockholm, Sweden. E-mail: arvid.sjolandar@ki.se.

Copyright © 2022 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 1044-3983/23/342-186-191

DOI: 10.1097/EDE.0000000000001567

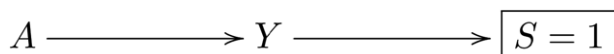


FIGURE 3. A causal diagram illustrating outcome-influenced selection, where the selection is directly influenced by the outcome, but not by the exposure.

on causal effects in the source population, that is, the population from which the observed data were selectively drawn. However, one may also wonder to what extent the observed data can help to estimate causal effects in the selected population, that is, the population defined by those selected into the study. To the best of our knowledge, this question has not been addressed in other literature either.

In this article, we address these issues. We consider both outcome-associated selection and outcome-influenced selection, and we use formal yet intuitive methods based on counterfactual diagrams⁵ to show whether causal effects, both in the source population and in the selected population, are estimable under these selection mechanisms. We will assume that the reader is familiar with causal diagrams, and in particular with the rules of d-separation.^{2,3} We will focus on selection bias and ignore other possible biases, confounding in particular. Real observational studies typically control for measured confounders to reduce the degree of confounding bias; all our arguments and results hold within levels of (conditional on) such measured confounders, provided that these are sufficient for confounding control. We restrict attention to binary exposures and outcomes, but some of our arguments and results carry over to other types of variables; we indicate this as we proceed. Throughout, we ignore uncertainty due to sampling variability.

POTENTIAL OUTCOMES AND CAUSAL TARGET PARAMETERS

We use standard potential outcome notation^{2,3} to define causal effects. Let Y_a denote the potential outcome of Y for a given subject if, possibly contrary to fact, exposed to level $A = a$. We say that the exposure has a causal effect for a given individual if $Y_0 \neq Y_1$ for that individual. Let $p(Y_a = 1)$ be the counterfactual probability of the outcome had everyone in the source population (both those selected into the study and those not selected) been exposed to level $A = a$. We say that the exposure has a causal effect in the source population if $p(Y_0 = 1) \neq p(Y_1 = 1)$. To measure the causal effect in the source population we may, for instance, use the causal risk difference $p(Y_1 = 1) - p(Y_0 = 1)$ or the causal risk ratio $p(Y_1 = 1)/p(Y_0 = 1)$. Finally, let $p(Y_a = 1|S = 1)$ be the counterfactual probability of the outcome had everyone in the selected population been exposed to level $A = a$. We say that the exposure has a causal effect in the selected population if $p(Y_0 = 1|S = 1) \neq p(Y_1 = 1|S = 1)$. To measure the causal effect in the selected population we may, for instance, use the causal risk difference $p(Y_1 = 1|S = 1) - p(Y_0 = 1|S = 1)$ or the causal risk ratio $p(Y_1 = 1|S = 1)/p(Y_0 = 1|S = 1)$.

We emphasize that, since we are not concerned with sampling variability, we use the term “selected population” in an asymptotic sense. That is, we do not use the term to refer to the limited sample of selected individuals in the particular study, but rather to an infinite “super-population” of individuals, generated under the same selection mechanism as the factual sample. Probabilities conditional on $S = 1$, such as $p(Y_0 = 1|S = 1)$ and $p(Y_1 = 1|S = 1)$, can be interpreted as proportions in this super-population.

CONSISTENCY AND EXCHANGEABILITY

We make the standard consistency assumption^{6–8} that the potential outcome Y_a is equal to the factual (observed) outcome Y for subjects who are factually exposed to level $A = a$: $A = a \Rightarrow Y = Y_a$ for all a .

(1)

From consistency (1) it follows that $p(Y = 1|A = a, S = 1) = p(Y_a = 1|A = a, S = 1)$ for all a ; that is, the probability of the outcome $Y = 1$ among those exposed to level $A = a$ is equal to the probability of the potential outcome $Y_a = 1$ among those exposed to level $A = a$, in the selected population. If we would further have that

$$p(Y_a = 1|A = a, S = 1) = p(Y_a = 1|S = 1) \text{ for all } a, \quad (2)$$

then we could interpret the exposure–outcome association in the selected population as the corresponding causal effect, for example, we could interpret the risk difference $p(Y = 1|A = 1, S = 1) - p(Y = 1|A = 0, S = 1)$ in the selected population as the causal risk difference $p(Y_1 = 1|S = 1) - p(Y_0 = 1|S = 1)$ in the selected population. Finally, if we would also have that

$$p(Y_a = 1|S = 1) = p(Y_a = 1) \text{ for all } a, \quad (3)$$

then we could interpret the causal effect in the selected population as the corresponding causal effect in the source population, for example, we could interpret the causal risk difference $p(Y_1 = 1|S = 1) - p(Y_0 = 1|S = 1)$ in the selected population as the causal risk difference $p(Y_1 = 1) - p(Y_0 = 1)$ in the source population.

The relation in (2) states that the potential outcome Y_a has the same distribution among those factually exposed to $A = a$ as among everyone in the selected population, or—equivalently—that Y_a is conditionally independent of A , given $S = 1$:

$$Y_a \perp A | S = 1 \text{ for all } a; \quad (4)$$

this is often referred to as “conditional exchangeability.”^{2,3} Similarly, the relation in (3) states that the potential outcome Y_a has the same distribution in the selected population as in the source population, or—equivalently—that Y_a is conditionally independent of S :

$$Y_a \perp S \text{ for all } a. \quad (5)$$

The concepts and definitions above are related to a recent study by Lu et al.⁹ These authors assumed that the causal effect

in the source population is the target parameter, and showed that the total bias of the exposure–outcome association in the selected population can be decomposed into two parts. They used the terms “type 1 selection bias” and “type 2 selection bias” for the bias components due to violations of (4) and (5), respectively. This decomposition is also related to the modern literature on transportability of causal effects, where we say that the causal effect in the selected population is “transportable” to the source population if it is equal to the causal effect in that population, that is, if there is no type 2 selection bias; see Barenboim and Pearl¹⁰ and the references therein.

ESTIMATION OF CAUSAL EFFECTS IN THE SELECTED POPULATION

Outcome-associated Selection

It is difficult to judge whether counterfactual independencies like (4) and (5) hold in a causal diagram using intuitive reasoning alone, because standard causal diagrams do not include potential outcomes like Y_a . Fortunately, there exists a simple method based on counterfactual diagrams.⁵ In this method, the causal diagram illustrating the factual world is augmented with a parallel diagram, illustrating the counterfactual world where the exposure is set to a certain level for everyone. The factual and counterfactual worlds are joined by exogenous error terms, corresponding to all (measured or unmeasured) factors that influence the variables under consideration, apart from those explicitly depicted on the original causal diagram. For instance, in the causal diagram of Figure 3 there is only one variable, A , that influences Y . However, there are of course always other factors (genetics, lifestyle etc) that influence Y as well, which are not explicitly depicted on the diagram; heuristically, the error term for Y is the whole set of all these implicit factors. Pearl² provides a formal connection between causal diagrams and potential outcomes through nonparametric structural equations. Once the counterfactual diagram has been constructed, counterfactual independencies like (4) and (5) can easily be evaluated using standard rules of d-separation.

The counterfactual diagram corresponding to the causal diagram in Figure 2 for outcome-associated selection is shown in Figure 4. The left part of the diagram represents the factual world where the exposure A varies randomly, and the right part represents the counterfactual world where the exposure is set to level a for everyone. The subindex a on Y_a reflects that

this is a potential outcome under the hypothetical intervention setting A to a . The variable ε_Y is the error term for Y . Because U , S , and ε_Y are not descendants of A , these are unaffected by interventions on A , and are thus shared between (that is, have the same value in) both worlds.

In Figure 4, there are two paths between Y_a and A , $Y_a \leftarrow \varepsilon_Y \rightarrow Y \leftarrow A$ and $Y_a \leftarrow U \rightarrow Y \leftarrow A$; however, both pass through the collider Y . Because we have not conditioned on Y in (4), only on $S = 1$, all paths between Y_a and A are blocked. It follows that conditional exchangeability (4) holds under the causal diagram in Figure 2, so that the conditional association between A and Y , given $S = 1$, can be interpreted as the corresponding causal effect of A on Y in the selected population. In the jargon of Lu et al.,⁹ we say that there is no type 1 selection bias. Because the counterfactual diagram in Figure 4 makes no assumption about A and Y being binary, this result holds for nonbinary exposures and outcomes as well.

Outcome-influenced Selection

Lu et al.⁹ considered a variation of outcome-influenced selection shown in Figure 5, where there is a covariate L that affects Y , but not A . Although the causal diagram in Figure 3 does not explicitly include such a covariate, there will always be other predictors for the outcome than the exposure of interest; these are usually subsumed into the implicit error term ε_Y . Thus, the scenarios in Figures 3 and 5 are essentially equivalent from our perspective. Lu et al.⁹ stated that, for the causal diagram in Figure 5, the exposure–outcome association in the selected population “suffers from type 1 selection bias by restricting to one level of a descendant of the collider Y , leading to a biased effect estimate on both risk difference and risk ratio scales.” This argument is somewhat unsatisfactory. Although it is technically correct that Y is a collider between A and L in Figure 5, it is not clear from the causal diagram why conditioning on its descendant S would give bias, because the causal diagram does not include the potential outcome Y_a that we ultimately care about.

To provide a more rigorous argument, we again use counterfactual diagrams. The counterfactual diagram corresponding to the causal diagram in Figure 3 for outcome-influenced selection is shown in Figure 6. Because S is now a descendant of, and thus influenced by, A in the causal diagram, S and S_a are not generally equal, and must be distinguished in the counterfactual diagram. In Figure 6, there are two

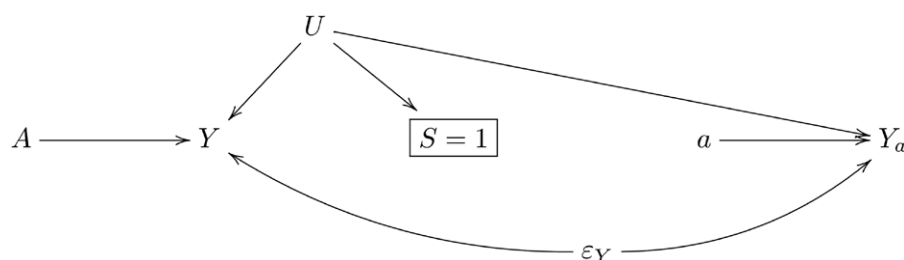
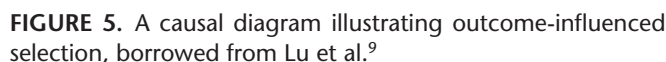


FIGURE 4. Counterfactual diagram corresponding to the causal diagram in Figure 2.



That causal effects in the selected population cannot be estimated under outcome-influenced selection does not mean that data are completely uninformative about such causal effects. A straightforward application of arguments by Robins¹¹ and Manski¹² leads to the conclusion that the counterfactual probability $p(Y_a = 1 | S = 1)$ is confined to the range

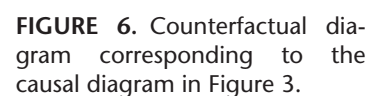
For completeness, we prove this relation in eAppendix 2, <http://links.lww.com/EDE/B995>. By maximizing $p(Y_0 = 1|S = 1)$ and minimizing $p(Y_1 = 1|S = 1)$ within the range in (6), we obtain lower bounds for the causal risk difference and risk ratio in the selected population.

$$\begin{aligned} & p(A = 1, Y = 1|S = 1)/\{p(A = 0, Y = 1|S = 1) + p(A = 1|S = 1)\} \\ & \leq p(Y_1 = 1|S = 1)/p(Y_0 = 1|S = 1) \leq \\ & \{p(A = 1, Y = 1|S = 1) + p(A = 0|S = 1)\}/p(A = 0, Y = 1|S = 1) \end{aligned} \quad (8)$$

Outcome-associated Selection

In fact, the situation is even worse. In eAppendix 3, <http://links.lww.com/EDE/B995>, we prove that the observed data have no information about causal effects in the source population under outcome-associated selection. Thus, whatever data we observe under outcome-associated selection, the causal risk difference in the source population can be anywhere between -1 and 1 , and the causal risk ratio in the source population can be anywhere between 0 and infinity. We only prove this result for binary variables, but we conjecture that a similar result holds for other types of variables as well.

This result may seem reasonable to some readers, because we have not assumed that the sampling fraction $p(S = 1)$ is known. Thus, it is possible that the selected population only constitutes a tiny (technically 0) proportion of the source population, in which case it makes intuitive sense that the observed data have no information about causal effects in the source population. However, this explanation cannot be trivially taken for granted, because it is not valid for all sampling schemes. Specifically, we show in the next section that the observed data do have some information about causal



effects in the source population under outcome-influenced selection, irrespectively of the sampling fraction.

To other readers, the fact that the observed data have no information about causal effects in the source population under outcome-associated selection may seem to contradict the result by Hernán,⁴ that we can test whether the sharp causal null holds under outcome-associated selection. However, we remind the reader that the sharp causal null means that the exposure has no effect for any single individual. A violation of the sharp causal null does not imply that the exposure has an effect on the population level, because the exposure may have positive effects for some individuals and negative effects for other individuals, which may cancel out in the population.

Outcome-influenced Selection

From the counterfactual diagram in Figure 6, we observe that Y_a and S are associated via the open path $Y_a \leftarrow \varepsilon_Y \rightarrow Y \rightarrow S$. Thus, the independence relation in (5) does not hold under outcome-influenced selection, which implies that the causal effect in the selected population is not equal to the causal effect in the source population; in the jargon of Lu et al.,⁹ we say that there is type 2 selection bias, in addition to the type 1 selection bias noted above. This bias occurs for nonbinary variables as well.

However, in contrast to outcome-associated selection, the data under outcome-influenced selection do have some information about causal effects in the source population. Under outcome-influenced selection, it can be shown that the causal odds ratio in the source population is equal to the odds ratio in the selected population,

$$\frac{p(Y_1 = 1)/p(Y_1 = 0)}{p(Y_0 = 1)/p(Y_0 = 0)} = \frac{p(Y = 1|A = 1, S = 1)/p(Y = 0|A = 1, S = 1)}{p(Y = 1|A = 0, S = 1)/p(Y = 0|A = 0, S = 1)} = \text{OR},$$

and can therefore be estimated; this is true irrespectively of the sampling fraction $p(S = 1)$. This has been shown by various authors^{13,14}; for completeness, we give a detailed proof in eAppendix 4, <http://links.lww.com/EDE/B995>. By using the causal odds ratio in the source population, it is possible to provide bounds on the causal risk difference and risk ratio in the source population. Specifically, it follows from results in King and Zeng¹⁵ that

$$\min \left(0, \frac{\sqrt{\text{OR}} - 1}{\sqrt{\text{OR}} + 1} \right) \leq p(Y_1 = 1) - p(Y_0 = 1) \leq \max \left(0, \frac{\sqrt{\text{OR}} - 1}{\sqrt{\text{OR}} + 1} \right) \quad (9)$$

and

$$\min(1, \text{OR}) \leq p(Y_1 = 1)/p(Y_0 = 1) \leq \max(1, \text{OR}). \quad (10)$$

The bounds in (7) for the causal risk difference in the selected population are qualitatively different from those in

(9) for the causal risk difference in the source population, in that the former include both positive and negative values, whereas the latter include either positive or negative values, but not both. Similarly, the bounds in (8) for the causal risk ratio in the selected population include both values above and below 1, whereas the bounds in (10) for the causal risk ratio in the source population include either values above or below 1, but not both. Thus, using these bounds we are able to tell the direction of the exposure effect in the source population, but not in the selected population.

CONCLUSIONS

In this note, we have contrasted outcome-associated selection and outcome-influenced selection. We have shown that causal effects in the selected population are estimable under outcome-associated selection but not under outcome-influenced selection. We have shown that data have no information about causal effects in the source population under outcome-associated selection, but that the causal odds ratio in the source population can be estimated, and the causal risk ratio and risk difference can be bounded, both in the selected population and in the source population, under outcome-influenced selection. For some of these results, we have used counterfactual diagrams, but we note that it may also be possible to prove these results with Single World Intervention Graphs (SWIGs).¹⁶

We have presented bounds for the causal risk difference and risk ratio in the source population and in the selected population under outcome-influenced sampling. Other authors have presented related bounds, but under somewhat different conditions. Kuroki et al.¹⁷ derived bounds for the causal risk difference and the causal risk ratio in the source population under case-control sampling, which is a special case of outcome-influenced sampling. Unlike us, though, these authors allowed for both confounding and biased selection, so we would expect their bounds to be less informative (i.e. wider) than our bounds in (9) and (10); we verify this in eAppendix 5, <http://links.lww.com/EDE/B995>. Gabriel et al.¹⁸ derived bounds for the causal risk difference in the source population in scenarios with missing data, which is analogous to selection. Unlike us though, these authors allowed simultaneous outcome-associated and outcome-influenced missingness or selection (their Figure 1A), and they assumed that the proportion of nonmissingness, corresponding to $p(S = 1)$ in our exposition, is known. Thus, their bounds are not directly comparable to our bounds. Neither of these authors considered causal effects in the selected population.

We have not taken a stance on which parameter is most relevant from a scientific perspective: a causal effect in the source population or in the selected population. We believe that most researchers would prefer to estimate causal effects in the source population, but given all sources of errors in real epidemiologic studies (e.g., selection bias, measurement bias, confounding bias), we conjecture that many researchers

would be content with an estimate of any causal effect that is at least approximately unbiased. Thus, the fact that outcome-associated selection admits estimation of causal effects in the selected population, whereas outcome-influenced selection does not, may be useful information to practitioners.

REFERENCES

- Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
- Pearl J. *Causality: Models, Reasoning and Inference*. 2nd ed. Cambridge University Press; 2009.
- Hernán MA, Robins JM. *Causal Inference: What If*. Chapman & Hall/CRC; 2020.
- Hernán MA. Invited commentary: selection bias without colliders. *Am J Epidemiol*. 2017;185:1048–1050.
- Shpitser I, Pearl J. What counterfactuals can be tested. Proceedings of the 23rd Annual Conference on Uncertainty in Artificial Intelligence. 2007:437–444.
- Cole S, Frangakis C. The consistency statement in causal inference: a definition or an assumption? *Epidemiology*. 2009;20:3–5.
- VanderWeele T. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20:880–883.
- Pearl J. On the consistency rule in causal inference: axiom, definition, assumption, or theorem? *Epidemiology*. 2010;21:872–875.
- Lu H, Cole SR, Howe CJ, Westreich D. Toward a clearer definition of selection bias when estimating causal effects. *Epidemiology*. 2022;33:699–706.
- Barenboim E, Pearl J. A general algorithm for deciding transportability of experimental results. *J Causal Inference*. 2013;1:107–134.
- Robins JM. The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A, eds. *Health service research methodology: a focus on AIDS*. US Public Health Service, National Center for Health Services Research; 1989:113–159.
- Manski CF. Nonparametric bounds on treatment effects. *Am Econ Rev*. 1990;8:319–323.
- Didelez V, Kreiner S, Keiding N. Graphical models for inference under outcome-dependent sampling. *Stat Sci*. 2010;25:368–387.
- Barenboim E, Pearl J. Controlling selection bias in causal inference. *Artif Intell Stat*. 2012;22:100–108.
- King G, Zeng L. Estimating risk and rate levels, ratios and differences in case-control studies. *Stat Med*. 2002;21:1409–1427.
- Richardson TS, Robins JM. 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*. 2013;128:1–148.
- Kuroki M, Cai Z, Geng Z. Sharp bounds on causal effects in case-control and cohort studies. *Biometrika*. 2010;97:123–132.
- Gabriel EE, Sjölander A, Sachs MC. Nonparametric bounds for causal effects in imperfect randomized experiments. *J Am Stat Assoc*. 2021. doi: 10.1080/01621459.2021.1950734.

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

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

and

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

Since $p(S = 1) = \sum_{a,y} p_{ay1}$ it follows that $\mathbf{p}_{\cdot 1} = \mathbf{p}_1/p(S = 1)$ and $\mathbf{q}_{\cdot 1} = \mathbf{q}_1/p(S = 1)$ 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 $\mathbf{p}_{\cdot 1}$, it is possible to find two valid distributions $p(A, Y_0, Y_1, S_0, S_1)$ which imply the same given $\mathbf{p}_{\cdot 1}$ but two different $\mathbf{q}_{\cdot 1}$.

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

$$v_{00} = (1 - v_0)(1 - v_1)$$

$$v_{01} = (1 - v_0)v_1$$

$$v_{10} = v_0(1 - v_1)$$

$$v_{11} = v_0v_1$$

so that (3) and (4) simplify to

$$p_{00.1} = (1 - p)(1 - v_0)r_0/r$$

$$p_{01.1} = (1 - p)v_0r_1/r$$

$$p_{10.1} = p(1 - v_1)r_0/r$$

$$p_{11.1} = pv_1r_1/r$$

(5)

and

$$q_{0.1} = [(1 - p)v_0r_1 + p\{v_0(1 - v_1)r_0 + v_0v_1r_1\}]/r$$

$$q_{1.1} = [pv_1r_1 + (1 - p)\{(1 - v_0)v_1r_0 + v_0v_1r_1\}]/r.$$

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 $\mathbf{p}_{\cdot 1}$ we may find two valid distributions $p(A, Y_0, Y_1, S_0, S_1)$ which imply the same given $\mathbf{p}_{\cdot 1}$ 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 is 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.

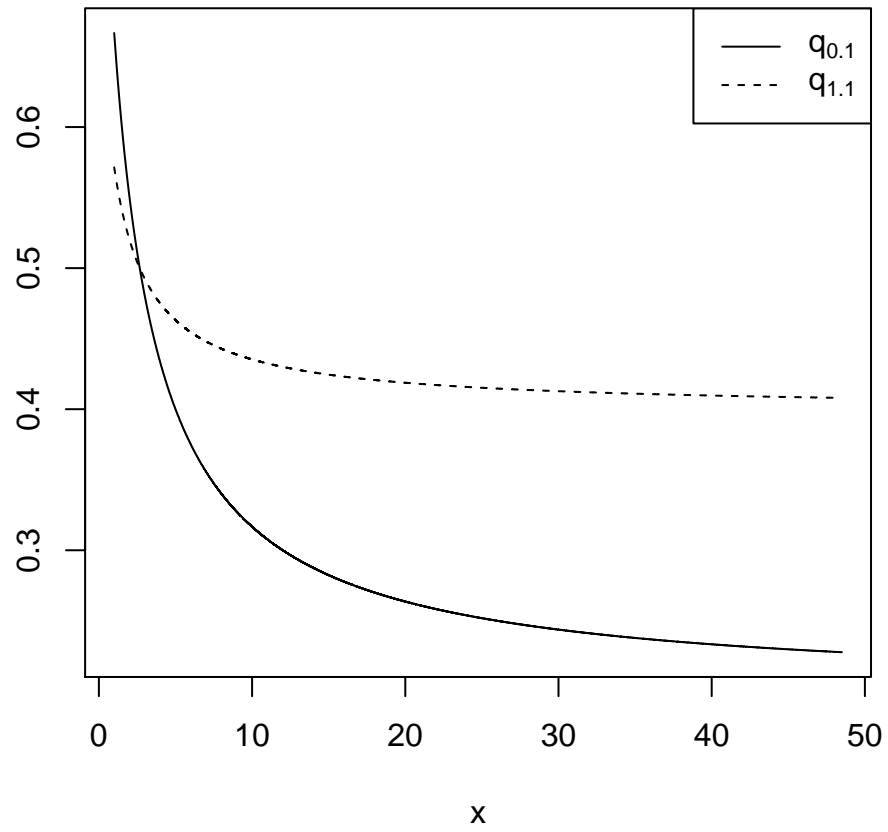


Figure 1: $q_{0.1}$ and $q_{1.1}$ as functions of x for $r = 0.05$, $p_{00.1} = 0.1$, $p_{01.1} = 0.2$, $p_{10.1} = 0.3$ and $p_{11.1} = 0.4$.

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

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

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

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

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:

$$\min(p_1 - 1, -p_0) \leq p(Y_1 = 1) - p(Y_0 = 1) \leq \max(p_1, 1 - p_0). \quad (8)$$

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

$$p_1 - 1 \leq -p_0, \quad (9)$$

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.