

Toward a Clearer Definition of Selection Bias When Estimating Causal Effects

 Haidong Lu,^a Stephen R. Cole,^b Chanelle J. Howe,^c and Daniel Westreich^b

Abstract: Selection bias remains a subject of controversy. Existing definitions of selection bias are ambiguous. To improve communication and the conduct of epidemiologic research focused on estimating causal effects, we propose to unify the various existing definitions of selection bias in the literature by considering any bias away from the true causal effect in the referent population (the population before the selection process), due to selecting the sample from the referent population, as selection bias. Given this unified definition, selection bias can be further categorized into two broad types: type 1 selection bias owing to restricting to one or more level(s) of a collider (or a descendant of a collider) and type 2 selection bias owing to restricting to one or more level(s) of an effect measure modifier. To aid in explaining these two types—which can co-occur—we start by reviewing the concepts of the target population, the study sample, and the analytic sample. Then, we illustrate both types of selection bias using causal diagrams. In addition, we explore the differences between these two types of selection bias, and describe methods to minimize selection bias. Finally, we use an example of “M-bias” to demonstrate the advantage of classifying selection bias into these two types.

Keywords: Selection bias; Collider bias; Effect measure modification; Effect heterogeneity; Causal diagram; Internal validity; External validity; Epidemiologic research

(*Epidemiology* 2022;33: 699–706)

When estimating causal effects, selection bias remains a subject of controversy in epidemiology.¹ The definition of selection bias is not as clear as that of confounding or

information bias. This controversy and ambiguity may stem from the fact that in the literature selection bias has sometimes been considered a threat to internal validity, while at other times it has been considered a threat to external validity.^{2–4} To improve communication and the conduct of epidemiologic research focused on estimating causal effects, we propose a definition of selection bias with two types.

This article is organized as follows. First, we review the concepts of the target population, the study sample and the analytic sample and provide a refined definition of selection bias. Next, we describe two types of selection bias: type 1 selection bias owing to restricting to one or more level(s) of a collider (or a descendant of a collider), and type 2 selection bias owing to restricting to one or more level(s) of an effect measure modifier. Then, we use an example to demonstrate the importance of classifying selection bias into these two types. Last, we describe the utility of defining selection bias as having two types and conclude with a brief discussion.

Before proceeding, it will be useful to state the following assumptions. First, we assume that causal consistency is satisfied.^{5,6} That is, here we will not consider interference⁷ or multiple treatment versions.⁶ For clarity and simplicity, hereafter we also assume no confounding of the exposure–outcome relationship, no measurement bias, and no random variability. For causal diagrams, we consider only four types of variables: binary exposure E, binary outcome D, selection S, and covariates L (e.g., L1, L2). Throughout, S = 1 means selection into the sample. It should also be noted that, in terms of treatment effect heterogeneity related to type 2 selection bias, two terms are involved—*effect measure modification* and *interaction*. Although there are differences between the meanings of these two terms⁸; hereafter, we use *effect measure modification* to describe the scenario of effect heterogeneity. Finally, in this article, we focus on the risk difference and risk ratio to avoid issues of noncollapsibility.^{9,10}

TARGET POPULATION, STUDY SAMPLE, AND ANALYTIC SAMPLE

Unlike confounding and information bias, selection bias results from a change in the sample under study. Thus, we need to define the populations that are involved in terms of selection bias: in our setting, the target population is the


Submitted March 19, 2020; accepted May 26, 2022

From the ^aPublic Health Modeling Unit and Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT;

^bDepartment of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC; and ^cDepartment of Epidemiology, School of Public Health, Brown University, RI.

This work was supported by grant(s) DP2HD084070 from National Institutes of Health.

The authors report no conflicts of interest.

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

Correspondence: Haidong Lu, Public Health Modeling Unit and Department of Epidemiology of Microbial Diseases, Yale School of Public Health, 350 George Street, Ste 3rd Floor, New Haven, CT 06511. E-mail: haidong.lu@yale.edu.

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

ISSN: 1044-3983/22/335-699-706

DOI: 10.1097/EDE.0000000000001516

population that inference is to be made about; the study sample (sometimes called study population) is the complete population that is included in the study, and it is used to make inference about the target population and may or may not be representative of the target population; and the analytic sample is the observed portion of the study sample that is used for analysis. The relationships between the target population, the study sample, and the analytic sample are visualized in Figure 1. It should be noted that in the absence of confounding and measurement bias, the observed association (i.e., associational risk difference or risk ratio) between the exposure and the outcome in the analytic sample is typically used as the effect estimate. Further, it is conceptually straightforward to imagine an epidemiologic study involving several steps: first, identify a target population, then select the study sample from that target population, and then select the analytic sample from the study sample. Following Westreich et al.,³ we define *internal validity* as the case when the effect estimated from the analytic sample is equal to the true causal effect in the study sample; and *external validity* as the case when the true causal effect in the study sample is equal to the true causal effect in the target population. Based on these definitions, external validity will be threatened by the degree to which the distribution of effect measure modifiers differs between the study sample and the target population.² Of note, here we focus on the target population from which the study sample was selected, and thus generalizability is of interest. However, broader definitions of the target population have been proposed.^{11,12} For example, sometimes researchers want to transport the study results to a target population that is partially or completely nonoverlapping with the study sample.¹¹ Transportability of study results is beyond the scope of this paper, we refer the reader to Westreich et al.¹¹ and Bareinboim and Pearl.¹³ It is also worth noting that here we focus on the scenarios where membership in the target population is not dependent on exposure or other variables. We avoid complexities that may arise when target population membership does have determinants, since addressing these complexities is beyond the scope of the present work.^{14–16}

To facilitate discussion, we further define the *referent population* as a population before the selection process, in contrast to the *selected sample*. Hence, when selecting the study sample from the target population, the referent population is the target population; when selecting the analytic sample (as a whole or to be within specific stratum) from the study sample, then the referent population is the study sample.

Broadly, we propose to unify the various existing definitions of selection bias in the literature by considering any bias away from the true causal effect in the referent population, due to selecting the sample from the referent population, as selection bias. That is, selection bias is defined as the difference between the true causal effect in the referent population and the effect estimate in the selected sample. For example, consider the risk

difference scale, and assume the true causal risk difference in the referent population is $P(D^{e=1} = 1) - P(D^{e=0} = 1)$ and the true causal risk difference in the selected sample is $P(D^{e=1} = 1 | S = 1) - P(D^{e=0} = 1 | S = 1)$. In the absence of confounding and measurement bias, the (observed) associational risk difference $P(D = 1 | E = 1, S = 1) - P(D = 1 | E = 0, S = 1)$ in the selected sample is typically used as the effect estimate. Then selection bias can be expressed in notation as follows (see eAppendix I; <http://links.lww.com/EDE/B942> for detail):

$$[P(D^{e=1} = 1) - P(D^{e=0} = 1)] - [P(D = 1 | E = 1, S = 1) - P(D = 1 | E = 0, S = 1)]$$

Given this unified definition, selection bias can be further categorized into two broad types: type 1 selection bias, which is due to restricting to one or more level(s) of a collider (or a descendant of a collider), and type 2 selection bias, which is due to restricting to one or more level(s) of an effect measure modifier. That is to say, if the selection (S in the causal diagrams) is a collider or an effect measure modifier, restricting to one or more level(s) of the selection S (e.g., $S = 1$) can lead to selection bias. Type 1 selection bias will result in a difference between the true causal effect in the selected sample and the effect estimate in the selected sample; type 2 selection bias will result in a difference between the true causal effect in the referent population and the true causal effect in the selected sample. Therefore, we can rewrite our above definition of selection bias in two parts:

$$\begin{aligned} & [P(D^{e=1} = 1) - P(D^{e=0} = 1)] - [P(D = 1 | E = 1, S = 1) - P(D = 1 | E = 0, S = 1)] \\ &= \frac{\{[P(D^{e=1} = 1) - P(D^{e=0} = 1)] - [P(D^{e=1} = 1 | S = 1) - P(D^{e=0} = 1 | S = 1)]\}}{\text{Type 2 selection bias}} \\ &+ \frac{\{[P(D^{e=1} = 1 | S = 1) - P(D^{e=0} = 1 | S = 1)] - [P(D = 1 | E = 1, S = 1) - P(D = 1 | E = 0, S = 1)]\}}{\text{Type 1 selection bias}} \end{aligned}$$

Type 1 Selection Bias

Type 1 selection bias is selection bias due to restricting to one or more level(s) of a collider (or a descendant of a collider). Hernán et al¹⁷ and Cole et al¹⁸ have explained type 1 selection bias in detail. Type 1 selection bias is sometimes called *collider stratification bias*,^{17,19} and sometimes, more specifically, called *collider restriction bias* when restricting to one level of a collider.^{20,21} Selection bias due to restricting to one level of a collider is a special case of collider *stratification bias*; here, we distinguish the two and chiefly address restriction. Briefly, type 1 selection bias arises if we restrict to one (or more) level(s) of a common effect of two causes, of which one is the exposure or a cause of the exposure, and the other is the outcome or a cause of the outcome.

The mechanism of type 1 selection bias is that restricting to one (or more) level(s) of a collider (or a descendant of a collider) opens a noncausal backdoor path between the

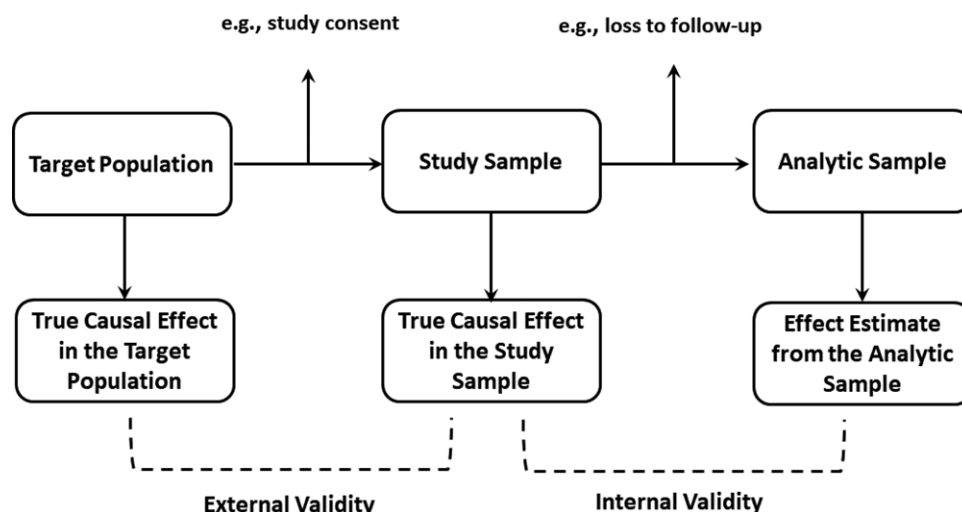


FIGURE 1. Relationships between target population, study sample, and analytic sample. Note that, in addition to selection bias, threats to internal validity also includes confounding and measurement bias.

exposure and the outcome. Here, we provide some common examples of type 1 selection bias using causal diagrams (Figure 2).²² For instance, Figure 2A is the typical example for Berkson's bias.²³ In this example, restricting to one level of selection S induces a noncausal association between exposure E and outcome D through two paths. First, it creates a noncausal association between exposure E and outcome D by opening the path $E \rightarrow [S=1] \leftarrow D$.¹⁷ Second, given that there may exist another cause of outcome D (i.e., $L1$) that is not shown in the causal diagram explicitly (because $L1$ is not a confounder for E - D relationship; we can consider it a hidden variable or error term not shown in the causal diagram²⁴), then outcome D becomes a collider for exposure E and covariate $L1$ as in Figure 2A. Restricting to one level of a descendant of the collider D , which is S , induces another noncausal association between E and D by opening the backdoor path E - $L1$ - D . Similarly, Figure 2B shows an example of a case-control study. It suffers from type 1 selection bias by restricting to one level of a descendant of the collider D , leading to a biased effect estimate on both risk difference and risk ratio scales.^{20,25} For the rules to identify sources of noncausal paths in the presence of hidden variables (i.e., error terms), we refer the reader to Daniel et al.²⁵ Figure 2C is a typical example of type 1 selection bias by restricting to one level of a descendant of the collider $L1$. Figure 2D is a common example of selection bias due to differential loss to follow-up.²⁶ It occurs when loss to follow-up is differential among exposure groups, and is also connected with the outcome by a common cause (i.e., $L1$). Figure 2E is the M-bias example where S is a collider in relation to the covariates $L1$ and $L2$ that affects the exposure and the outcome, respectively.

Type 1 selection bias can occur under the null hypothesis of no average causal effect or under the alternative (off-null) hypothesis, and even under the sharp null hypothesis of no causal effect for any individual. Importantly, restricting to one (or more) level(s) of a collider (or a descendant of a

collider) as in type 1 selection bias can produce an association measure that is biased not only for the referent population, but also for the selected sample. That is, type 1 selection bias will result in a difference between the true causal effect in the selected sample and the effect estimated from the selected sample. However, if only type 1 selection bias occurs (in the absence of other biases such as type 2 selection bias), the true causal effect in the selected sample will be equal to the true causal effect in the referent population. Therefore, when type 1 selection bias occurs during the selection of the study sample from the target population, the true causal effect in the study sample will be equal to the true causal effect in the target population, and thus external validity is not affected. However, the effect estimated from the study sample (which is also the analytic sample if no further selection processes occur) will not be equal to the true causal effect in the study sample, or in the target population. When type 1 selection bias occurs during the selection of the analytic sample from the study sample, the effect estimated from the analytic sample will not be equal to the true causal effect in the study sample. Therefore, we say type 1 selection bias only affects internal validity. Details are provided in eAppendix I; <http://links.lww.com/EDE/B942>.

Type 1 selection bias can be further classified into two subtypes based on whether the causal effect in the referent population is identifiable or not. Type 1A selection bias can be addressed, and the true causal effect recovered by measuring and adjusting for covariates that lie on the noncausal path that is opened by restricting to one (or more) level(s) of a collider via inverse probability weighting, g-computation, and sometimes stratification.²⁷ For example, one can adjust for $L1$ in Figure 2D and adjust for either $L1$ or $L2$ in Figure 2E to correct type 1 selection bias.

In contrast, type 1B selection bias occurs when there are no measured covariates that lie on the noncausal path opened by restricting to one (or more) level(s) of a collider or a descendant of a collider (e.g., shown in Figure 2A and

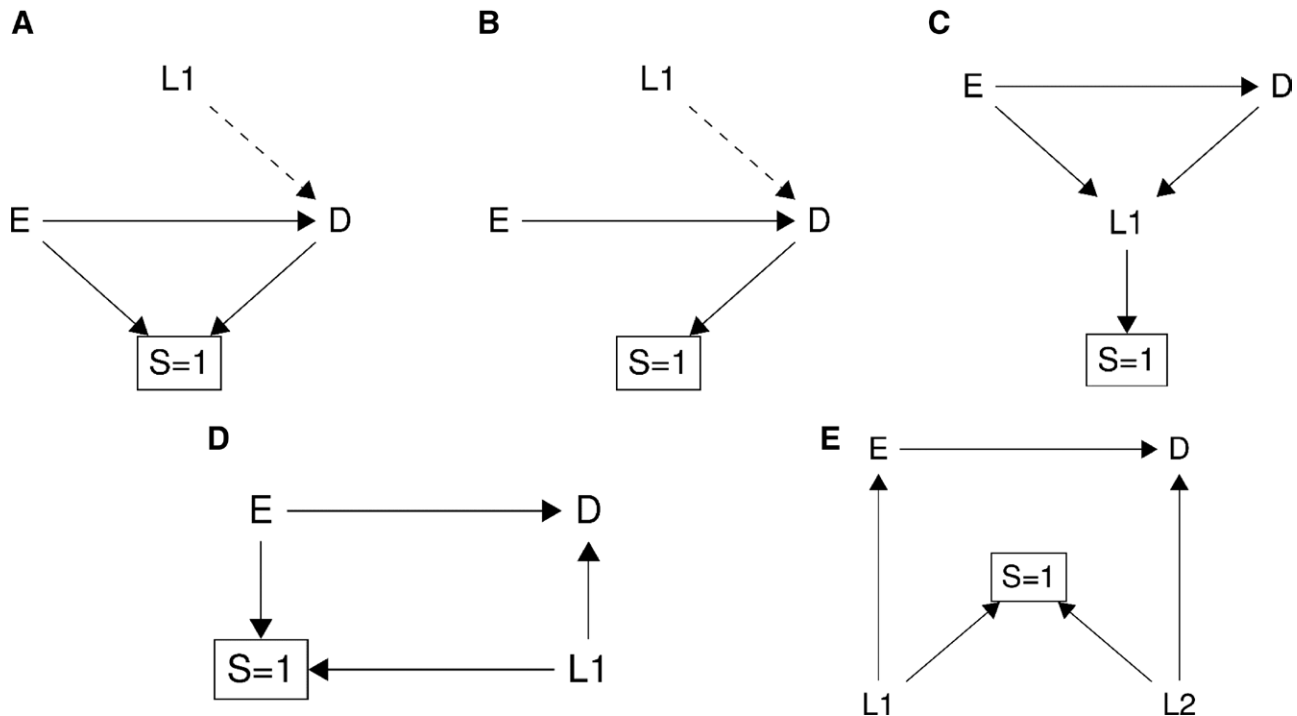


FIGURE 2. Examples of type 1 selection bias. E is exposure, D is outcome, S is selection, L (i.e., L1, L2) are covariates. The dashed line indicates the potential hidden cause.

C). Unlike type 1A selection bias, type 1B selection bias can generally not be addressed, and the causal effect in both the selected sample and the referent population is not identifiable unless the selection probability of each combination of exposure, covariates (if any) and outcome is known, which is typically unattainable in practice.²⁸ One important note is if the selection is outcome-dependent (that is, the selection is a descendant of the outcome, as for instance in case-control studies as shown in Figure 2B) and the selection probability is unknown, the causal effect in the referent population is not generally identifiable on risk difference or risk ratio scales, but is identifiable for the odds ratio measure once the covariates that affect the exposure E and the selection S, if any, are adjusted for.^{24,29,30} For those interested in the conditions and theorems to recover a causal effect under type 1 selection bias, we refer the reader to Bareinboim and colleagues.^{24,31}

Type 2 Selection Bias

In 1977, Greenland gave an example of selection bias that was distinct from type 1 selection bias, which recently received attention.³² First, we briefly review Greenland's example: in the context of no confounding, the relative risk of disease comparing the exposed with the unexposed among those who were uncensored and remained over follow-up period (i.e., the analytic sample in this example) was a biased estimate for the causal risk ratio in the entire population of interest including both the censored and the uncensored (i.e., the study sample). But interestingly, the exposed and unexposed

groups were equally likely to be lost to follow-up. How can this bias occur in the situation of the nondifferential loss to follow-up with regards to exposure status? Greenland explained that it is because the association between selection (i.e., censoring in this context) and the outcome varies across levels of the exposure.^{10,32} In another way, it means that the association between the exposure and the outcome varies across levels of the selection (i.e., the censored and the uncensored in this context). More recently, Hernán used causal diagrams, as shown in Figure 3A, to explain Greenland's example graphically.⁴ As in Figure 3A, even though there is no arrow from exposure E to selection S (in contrast to Figure 2D), which indicates the absence of restricting to one (or more) level(s) of a collider, selection bias can still arise. The presence of selection bias is due to effect measure modification of selection, S, for the relationship between exposure E and outcome D. By restricting to only one level of the effect measure modifier ($S = 1$) in this example, the observed effect of exposure E on outcome D in the analytic sample is not guaranteed to be equal to the causal effect in the study sample.

Greenland's and Hernán's explanations of selection bias without colliders^{4,32} indeed leads us to another type of selection bias: type 2 selection bias due to restricting to one or more level(s) of an effect measure modifier. That is, when the selection S is an effect measure modifier, restricting to one or more level(s) of the selection S (e.g., $S = 1$) can result in type 2 selection bias. To better understand type 2 selection bias, we need to review the structure of effect measure modification.

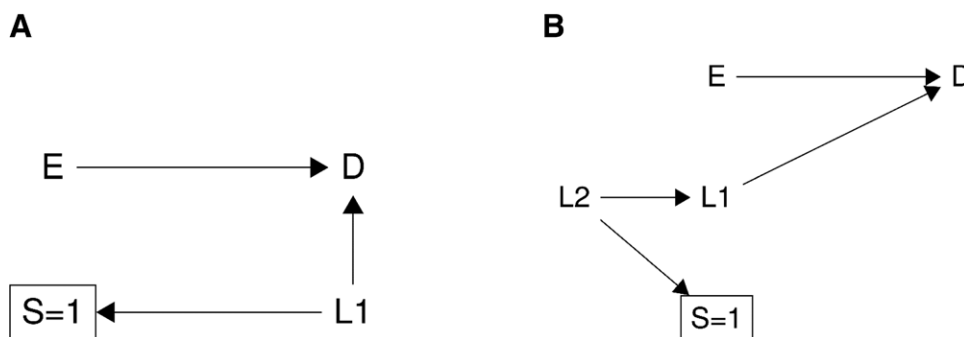


FIGURE 3. Two basic causal diagrams for type 2 selection bias. E is exposure, D is outcome, S is selection, L (i.e., L1, L2) is covariates.

VanderWeele and Robins³³ proposed four types of effect measure modification in causal diagrams (see eAppendix II; <http://links.lww.com/EDE/B942>; Figure 1): direct effect measure modification, indirect effect measure modification, effect measure modification by proxy, and effect measure modification by common cause. Briefly, a direct cause of the outcome is potentially a direct effect measure modifier (i.e., X in eAppendix II; <http://links.lww.com/EDE/B942>; Figure 1A); a variable that causes a direct effect measure modifier is an indirect effect measure modifier (i.e., C in eAppendix II; <http://links.lww.com/EDE/B942>; Figure 1B); a downstream surrogate of a direct effect measure modifier is an effect measure modifier by proxy (i.e., R in eAppendix II; <http://links.lww.com/EDE/B942>; Figure 1C); and a variable that is connected with the direct effect measure modifier by a common cause is an effect measure modifier by common cause (i.e., M in eAppendix II; <http://links.lww.com/EDE/B942>; Figure 1D). If we assume that the effect measure modifier is the selection S, and that selection S can only be affected by other variables (rather than affect other variables), we obtain two subtypes of type 2 selection bias, as shown in Figure 3A and B. In Figure 3A, the selection S is an effect measure modifier by proxy when L1 is a direct effect measure modifier, and type 2 selection bias can occur by restricting to one or more level(s) of the effect measure modifier by proxy; in Figure 3B, the selection S is an effect measure modifier by common cause when L1 is a direct effect measure modifier, and type 2 selection bias occurs by restricting to one or more level(s) of the effect measure modifier by common cause. In Greenland's example and elsewhere, the structural relationship between selection S and outcome D cannot usually be determined from the observed data, and thus either Figure 3A or B or even a more complex diagram can be the source of the type 2 selection bias (see eAppendix III; <http://links.lww.com/EDE/B942>; Figure 2A).

Type 2 selection bias has several properties. First, as Hernán explained,⁴ type 2 selection bias cannot occur under the sharp null hypothesis because no effect measure modification can exist if there is no causal effect of exposure E on outcome D for any individual. Second, type 2 selection bias is scale dependent (multiplicative vs. additive) since type 2 selection bias is dealing with restricting to one or more level(s) of an effect measure modifier.³⁴ Even if there is type 2 selection

bias for the causal risk ratio in the referent population, there is no guarantee that type 2 selection bias for the causal risk difference occurs. But it is worth noting the certainty that, when the sharp causal null does not hold, there always exists a type 2 selection bias for either the additive (e.g., risk difference) or multiplicative (e.g., risk ratio) scale if the selection is not conditionally independent of the outcome D within levels of the exposure E.^{3,35} Third, type 2 selection bias can produce an association measure that is biased only for the referent population, but not for the selected sample. That is, type 2 selection bias will result in a difference between the true causal effect in the referent population and the true causal effect in the selected sample, although the effect estimated from the selected sample will be equal to the true causal effect in the selected sample in the absence of other biases including type 1 selection bias. When selecting the study sample from the target population, type 2 selection bias is considered to affect external validity as the true causal effect in the study sample may not be equal to the true causal effect in the target population, and thus is often called *generalizability bias*.² When selecting the analytic sample from the study sample (e.g., loss to follow-up, withdrawal due to adverse effects, protocol deviation in per-protocol analyses,^{36,37} or other missing data scenarios), type 2 selection bias is considered to affect internal validity as the effect estimated from the analytic sample may not be equal to the true causal effect in the study sample. Details are provided in eAppendix I; <http://links.lww.com/EDE/B942>. Last, it should be noted that, if there is no type 2 selection bias that occurs during selection, the true causal effect in the selected sample should be equal to the true causal effect in the referent population, even in the presence of type 1 selection bias.

Type 2 selection bias can be minimized or even eliminated either during the design or analytic stage. Addressing type 2 selection bias during the analytic stage is possible if one can accurately measure and properly adjust for a sufficient set of covariates that affect selection S and outcome D to achieve d-separation between the selection and the outcome.³⁸ Such adjustment is analogous to the classic scenario of adjusting for a sufficient set of confounders to achieve conditional exchangeability and obtain an unbiased causal effect estimate of exposure on the outcome.² For example, one could adjust

for L1 in Figure 3A and adjust for either L1 or L2 in Figure 3B to address type 2 selection bias and recover the causal effect in the referent population. Technically, once the full distribution of covariates that affect the selection and the outcome are known, g-computation,³⁹ inverse probability (or odds) weighting (IPW),^{36,40} or augmented IPW^{41,42} can be employed to adjust for the covariates, and thus account for type 2 selection bias and recover the unconditional causal effect in the referent population under certain assumptions^{2,39} (i.e., causal consistency^{5,6} [no interference⁷ and treatment version irrelevance⁶], positivity,^{43,44} no measurement error, and correct model specification). For those interested in details of identification and estimation of causal effects in the referent population in the presence of type 2 selection bias, we refer the reader to Lesko et al² and Dahabreh et al.⁴²

AN ILLUSTRATIVE EXAMPLE: M-BIAS

Sometimes both type 1 selection bias and type 2 selection bias occur together in an epidemiologic study. That is, the selection node can act as both a collider and an effect measure modifier simultaneously. We use the “M-bias” diagram to further illustrate this point.¹⁹ M-bias is a bias caused by conditioning on a common effect of two causes, of which one is a cause of the exposure, and the other is a cause of the outcome, resulting in a so-called M-structure in its causal diagram. An example of M-bias is volunteer bias where the individuals’ characteristics affect their exposures and outcomes, and influence their decisions to participate in the study. As shown in Figure 4 (same as Figure 2E) as a standard M-bias diagram, the selection *S* is driven only by the covariates L1 and L2, which are causes of exposure *E* and outcome *D*, respectively, but neither of which is a confounder for the exposure-outcome relationship. Hence, the selection *S* is a collider, and restricting to one level of *S* (*S* = 1) induces a noncausal association

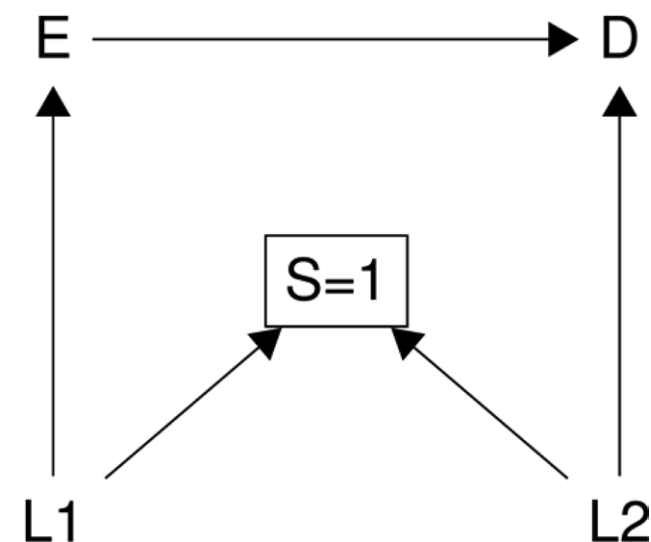


FIGURE 4. A typical example of “M-bias.” *E* is exposure, *D* is outcome, *S* is selection, *L* (i.e., L1, L2) is covariates.

between exposure *E* and outcome *D*. This implies that we must adjust for either L1 or L2 to explain away the noncausal association induced by type 1 selection bias and recover the causal effect in the referent population. One might think that adjusting for either one will work because either adjustment will block the noncausal path $E \leftarrow L1 \rightarrow [S=1] \leftarrow L2 \rightarrow D$ that is opened by restricting to one level of a collider due to selection. However, this perception is incorrect. In fact, it is critical to measure and adjust for covariate L2 rather than L1²⁵ because oftentimes selection *S* is not only a collider but also an effect measure modifier by proxy, as it is connected with a potential direct effect measure modifier L2. Measurement and adjustment of L1 instead of L2 may not remove type 2 selection bias. As previously mentioned, when the sharp null does not hold, if risk of the outcome changes across strata of L2, L2 must be an effect measure modifier of the relationship between *E* and *D* on either the risk difference or risk ratio scale, and therefore, there must exist type 2 selection bias for one scale (or both). Even though adjusting for either L1 or L2 enables us to address type 1 selection bias and recover the causal effect in the selected sample, we must accurately measure and properly adjust for L2 to account for type 2 selection bias to recover the causal effect in the referent population using one of the previously described approaches to addressing selection bias in the analytic stage. Specifically, the average potential outcome in the referent population can be identified from the observed data because

$$E(D^e) = \sum_{L2} E(D \mid E = e, L2 = l2, S = 1) P(L2 = l2)$$

A proof of this equivalence as well as code for a simulation of 1,000,000 individuals is given in eAppendix IV; <http://links.lww.com/EDE/B942>.

Similarly, in Figure 2D, the selection *S* acts as both a collider and an effect measure modifier when the covariate L1 is a direct effect measure modifier. In such cases, both type 1 and type 2 selection bias can arise. Fortunately, one can eliminate both biases if the distribution of L1 in the referent population is measured and accounted for.

DISCUSSION

The term “selection bias” is widely used in epidemiologic studies, but the distinction between different types of selection bias is usually not articulated. Varying use of the term by epidemiologists and others generates further confusion and impedes communication among medical researchers.⁴⁵ Here, using causal diagrams, we illustrate the two types of selection bias that can hinder accurate estimation of causal effects: type 1 selection bias owing to restricting to one or more level(s) of a collider (or a descendant of a collider), and type 2 selection bias owing to restricting to one or more level(s) of an effect measure modifier. That is, when the selection (*S* in the causal diagrams) is a collider or an effect measure modifier, restricting to one or more level(s) of the selection (e.g., *S* = 1) can

result in selection bias. This taxonomy may help improve the understanding, communication, and teaching of selection bias among epidemiologists and other health researchers.

A summary and comparisons of type 1 and type 2 selection bias are described in the Table. When estimating causal effects, type 1 selection bias, which is due to restricting to one or more level(s) of a collider (or a descendant of a collider), is the classic selection bias we often encounter in epidemiologic literature.^{17,18,27,45} It is often called *collider stratification bias* or *collider restriction bias*^{17,19,20} and can produce bias under the sharp null. Type 1 selection bias can be difficult to minimize analytically, especially when selection is the direct common effect of both the exposure and the outcome, or selection is dependent on the outcome and causal effects on risk difference or risk ratio scale are desired (i.e., type 1B selection bias).²⁴ But when selection is affected by a measured cause of the exposure or by a measured cause of an outcome and selection is not a descendant of the outcome (i.e., type 1A selection bias), analytically adjusting for type 1 selection bias is possible.

Here, we make a distinction between collider restriction bias and collider stratification bias. Although collider restriction bias is the bias introduced by restricting to one level of a collider, collider stratification bias is broadly defined as the bias introduced by conditioning on the collider.¹⁷ This includes not only bias due to restricting on a collider, but also bias introduced through (for example) the unnecessary inclusion of a collider in a regression model (analogous to restricting to more than one level of a collider¹⁷). Some consider collider stratification bias as a form of selection bias even in the absence of one-level restriction^{17,27,28}; others consider only collider restriction bias to be a selection bias and collider stratification bias by restricting to more levels of a collider as a form of overadjustment bias,²¹ in the sense that the inclusion of a collider in a regression does not change the overall sample under study or analysis despite the fact that including a collider in a regression does restrict estimation to be within strata of the collider.

Type 2 selection bias, which results from restricting to one or more level(s) of an effect measure modifier, is likely ubiquitous and underappreciated in epidemiologic studies.⁴⁶

More attention should be paid to type 2 selection bias and effect heterogeneity for several reasons. First, let us consider the scenario in which type 2 selection bias affects external validity when selecting the study sample from the target population. In both randomized trials and observational studies, it is rarely the case that the study sample is randomly selected from the target population, due in part to informed consent. Thus, we cannot assume that the effect in the study sample is the same as the effect in the target population in epidemiologic studies. Further, the conventional hierarchy in which internal validity, as previously defined, is given priority and external validity is considered secondary in clinical research is also problematic and misleading. Both internal and external validity are essential ingredients for achieving target validity (i.e., a joint measure of the validity of an effect estimate with respect to a specific-population of interest [target population]).³ A lack of either internal or external validity leads to bias with regards to the target population. Work by Breskin et al. demonstrates that the relative strength of the exchangeability assumptions for internal and external validity generally depend on the proportion of the target population that is selected into the study.⁴⁷ Establishing external validity does not necessarily require stronger assumptions than does internal validity. Last, as aforementioned, type 2 selection bias can affect internal validity as well, when selecting the analytical sample from the study sample (e.g., loss to follow-up or other missing data). Thus, one still needs to take type 2 selection bias into account when addressing the threats to internal validity. Fortunately, type 2 selection bias might be minimized if all effect measure modifiers that affect selection are measured.

Some caveats should be noted. First, the examples we illustrated are simple. Throughout the paper, we assume no confounding. However, the examples can be extended to include confounding bias as shown in eAppendix III; <http://links.lww.com/EDE/B942>; Figure 2B and C. For instance, eAppendix III; <http://links.lww.com/EDE/B942>; Figure 2B provides an example of a special type 2 selection bias through exposure when the confounder L also acts as an effect measure modifier.⁴⁸ Further, caution should be taken when confounding is present, since simultaneously adjusting for confounding and adjusting for type 1 and type 2 selection bias may induce new

TABLE. Summary of type 1 and type 2 selection bias

Selection bias	Type 1	Type 2
Definition	Restricting to one or more level(s) of a collider (or a descendant of a collider)	Restricting to one or more level(s) of an effect measure modifier
Other names	Collider stratification bias; collider restriction bias	generalizability bias
Can occur on sharp null?	Yes	No
Bias in the referent population?	Yes	Yes
Bias in the selected sample?	Yes	No
Can affect internal validity?	Yes	Yes
Can affect external validity?	No	Yes
Effect measure scale dependent?	No	Yes

collider bias. Second, here we only focus on causal effects measured on the risk difference and risk ratio scales. Some of our conclusions may not apply to other commonly used effect measures, for example, the odds ratio due to its unique invariance property.⁴⁹ Third, while we describe the two types of selection bias, we do not quantitatively compare (on different scales) the relative magnitude of type 1 and type 2 selection bias. Future work is needed to explore this issue in different scenarios. Last, here we only considered the case where membership in the target population is not dependent on the exposure or other variables. Sometimes when membership in the target population depends on the exposure or other variables, complexities arise, which are beyond the scope of the present work.^{14–16}

To conclude, in this work, we present a refined definition for selection bias with two types: type 1 selection bias is due to restricting to one or more level(s) of a collider (or a descendant of a collider), and type 2 selection bias is due to restricting to one or more level(s) of an effect measure modifier. This classification aims to facilitate understanding and communication, and thereby improve epidemiologic research.

REFERENCES

- Infante-Rivard C, Cusson A. Reflection on modern methods: selection bias—a review of recent developments. *Int J Epidemiol*. 2018;47:1714–1722.
- Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. *Epidemiology*. 2017;28:553–561.
- Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target validity and the hierarchy of study designs. *Am J Epidemiol*. 2019;188:438–443.
- Hernán MA. Invited commentary: selection bias without colliders. *Am J Epidemiol*. 2017;185:1048–1050.
- Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology*. 2009;20:3–5.
- Vander Weele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20:880–883.
- Hudgens MG, Halloran ME. Toward causal inference with interference. *J Am Stat Assoc*. 2008;103:832–842.
- VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology*. 2009;20:863–871.
- Greenland S. Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology*. 1996;7:498–501.
- Greenland S, Pearl J. Adjustments and their consequences—collapsibility analysis using graphical models. *Int Stat Rev*. 2011;79:401–426.
- Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. *Am J Epidemiol*. 2017;186:1010–1014.
- Dahabreh IJ, Robins JM, Haneuse S, JPA, Hernán MA. Generalizing causal inferences from randomized trials: counterfactual and graphical identification. 2019;34:719–722.
- Bareinboim E, Pearl J. A general algorithm for deciding transportability of experimental results. *J Causal Inference*. 2013;1:107–134.
- Howe CJ, Robinson WR. Survival-related selection bias in studies of racial health disparities: the importance of the target population and study design. *Epidemiology*. 2018;29:521–524.
- VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*. 2014;25:473–484.
- Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70–75.
- Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
- Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol*. 2010;39:417–420.
- Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–306.
- Westreich D. Berkson's bias, selection bias, and missing data. *Epidemiology*. 2012;23:159–164.
- Lu H, Cole SR, Platt RW, Schisterman EF. Revisiting overadjustment bias. *Epidemiology*. 2021;32:e22–e23.
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.
- Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Int J Epidemiol*. 2014;43:511–515.
- Bareinboim E, Tian J, Pearl J. Recovering from selection bias in causal and statistical inference. *Proc Twenty-Eighth Conf Artif Intell*. 2014;28:339–341.
- Daniel RM, Kenward MG, Cousens SN, De Stavola BL. Using causal diagrams to guide analysis in missing data problems. *Stat Methods Med Res*. 2012;21:243–256.
- Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ Jr. Selection bias due to loss to follow up in cohort studies. *Epidemiology*. 2016;27:91–97.
- Hernan M, Robins JM. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC; 2020.
- Smith LH. Selection mechanisms and their consequences: understanding and addressing selection bias. *Curr Epidemiol Reports*. 2020;7:179–189.
- Didelez V, Kreiner S, Keiding N. Graphical models for inference under outcome-dependent sampling. *Stat Sci*. 2010;25:368–387.
- Kenah E. A potential outcomes approach to selection bias. 2020;1–25. Available at: <http://arxiv.org/abs/2008.03786>
- Bareinboim E, Tian J. Recovering causal effects from selection bias. *Proc Natl Conf Artif Intell*. 2015;29: 3475–3481.
- Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol*. 1977;106:184–187. Available at: <http://aje.oxfordjournals.org/content/106/3/184.short>
- VanderWeele TJ, Robins JM. Four types of effect modification: a classification based on directed acyclic graphs. *Epidemiology*. 2007;18:561–568.
- VanderWeele TJ. Confounding and effect modification: distribution and measure. *Epidemiol Method*. 2012;1:55–82.
- Webster-Clark M, Breskin A. Directed acyclic graphs, effect measure modification, and generalizability. *Am J Epidemiol*. 2020;190:322–327.
- Lu H, Cole SR, Hall HI, et al. Generalizing the per-protocol treatment effect: the case of ACTG A5095. *Clin Trials*. 2019;16:52–62.
- Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med*. 2017;377:1391–1398.
- Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press; 2009.
- Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7:1393–1512.
- Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol*. 2010;172:107–115.
- Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J Am Stat Assoc*. 1995;90:106–121.
- Dahabreh IJ, Robertson SE, Hernán MA. On the relation between g-formula and inverse probability weighting estimators for generalizing trial results. *Epidemiology*. 2019;30:807–812.
- Westreich D, Cole SR. Invited commentary: positivity in practice. *Am J Epidemiol*. 2010;171:674–677; discussion 678.
- Howe CJ, Cole SR, Chmiel JS, Muñoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. *Am J Epidemiol*. 2011;173:569–577.
- Mansournia MA, Higgins JP, Sterne JA, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology*. 2017;28:54–59.
- Greenland S. The causal foundations of applied probability and statistics. In: Gefner H, Dechter R, Halpern J, eds. *Probabilistic and Causal Inference: The Works of Judea Pearl*. ACM Books; 2022:605–624.
- Breskin A, Westreich D, Cole SR, Edwards JK. Using bounds to compare the strength of exchangeability assumptions for internal and external validity. *Am J Epidemiol*. 2019;188:1355–1360.
- Ross RK, Breskin A, Westreich D. When is a complete-case approach to missing data valid? The importance of effect-measure modification. *Am J Epidemiol*. 189:1583–1589.
- Bareinboim E, Pearl J. Controlling selection bias in causal inference. *Proc Twenty-fifth AAAI Conf Artif Intell*. 2011;2:1754–1755.

ERRATUM

Erratum: Toward a Clearer Definition of Selection Bias When Estimating Causal Effects

Regarding the paper by Lu et al.¹ in the September 2022 issue of *Epidemiology*, the authors have identified two errors in the explanation and illustration of type 1 selection bias.

First, the sentence on page 700 “The mechanism of type 1 selection bias is that restricting to one (or more) level(s) of a collider (or a descendant of a collider) opens a noncausal backdoor path between the exposure and the outcome” is inaccurate. The corrected sentence simply deletes the word “backdoor.”²

Second, the sentence on page 701 “Unlike type 1A selection bias, type 1B selection bias [including the Figure 2C example of Lu et al.¹] can generally not be addressed, and the causal effect in both the selected sample and the referent population is not identifiable unless the selection probability of each combination of exposure, covariates (if any) and outcome is known, which is typically unattainable in practice” is ambiguous. In Figure 2C of Lu et al.,¹ type 1 selection bias can be addressed once the full distribution of the covariate *L1* in the referent population is measured and properly accounted for.

REFERENCES

1. Lu H, Cole SR, Howe CJ, Westreich D. Toward a clearer definition of selection bias when estimating causal effects. *Epidemiology*. 2022;33:699–706.
2. Mathur MB, Shpitser I. Simple graphical rules to assess selection bias in general-population and selected-sample treatment effects. *Am J Epidemiol*. 2023. In Press.

eAPPENDIX

eAPPENDIX I

Explanation on internal validity and external validity regarding type 1 selection bias

For simplicity and convenience, we here focus on the risk difference scale, but the work can also be generalized to the risk ratio scale. We define the causal risk difference of the exposure E on the outcome D in the referent population as $P(D^{e=1} = 1) - P(D^{e=0} = 1)$. We also define the causal risk difference in the selected sample as $P(D^{e=1} = 1|S = 1) - P(D^{e=0} = 1|S = 1)$, where $S=1$ means selection into the sample.

Definition of selection bias

We consider any bias away from the true causal effect in the referent population, due to selecting the sample from the referent population, as selection bias. Consider the risk difference scale. In the absence of confounding and measurement bias, the (observed) associational risk difference $P(D = 1|E = 1, S = 1) - P(D = 1|E = 0, S = 1)$ in the selected sample is typically used as the effect estimate. Then selection bias occurs when $P(D = 1|E = 1, S = 1) - P(D = 1|E = 0, S = 1) \neq P(D^{e=1} = 1) - P(D^{e=0} = 1)$. More specifically, selection bias can be defined as the difference between the true causal effect in the referent population and the effect estimate in the selected sample, which is expressed in notation as follows:

$$[P(D^{e=1} = 1) - P(D^{e=0} = 1)] - [P(D = 1|E = 1, S = 1) - P(D = 1|E = 0, S = 1)]$$

Given this unified definition, selection bias can be further categorized into two broad types: type 1 selection bias, which is due to restricting to one or more level(s) of a collider (or a descendant of a collider), and type 2 selection bias, which is due to restricting to one or more level(s) of an effect measure modifier. Type 1 selection bias will result in a difference between the true causal effect in the selected sample and the effect estimate in the selected sample; type 2 selection bias will result in a difference between the true causal effect in the referent population and the true causal effect in the selected sample. Therefore, we can rewrite our above definition of selection bias in two parts:

$$\begin{aligned} & [P(D^{e=1} = 1) - P(D^{e=0} = 1)] - [P(D = 1|E = 1, S = 1) - P(D = 1|E = 0, S = 1)] \\ &= \underbrace{\{[P(D^{e=1} = 1) - P(D^{e=0} = 1)] - [P(D^{e=1} = 1|S = 1) - P(D^{e=0} = 1|S = 1)]\}}_{\text{Type 2 selection bias}} \\ &+ \underbrace{\{[P(D^{e=1} = 1|S = 1) - P(D^{e=0} = 1|S = 1)] - [P(D = 1|E = 1, S = 1) - P(D = 1|E = 0, S = 1)]\}}_{\text{Type 1 selection bias}} \end{aligned}$$

Type 1 selection bias

Type 1 selection bias, due to restricting to one (or more) level(s) of a collider (or a descendant of a collider), produces an estimate that is biased not only for the referent population, but also for the selected sample. That is, the observed risk difference $P(D = 1|E = 1, S = 1) - P(D = 1|E = 0, S = 1)$ from the selected sample is not equal to the true causal risk difference $P(D^{e=1} = 1) - P(D^{e=0} = 1)$ in the referent population, and is also not equal to the true causal

risk difference $P(D^{e=1} = 1|S = 1) - P(D^{e=0} = 1|S = 1)$ in the selected sample. It is also worth noting that, if only type 1 selection bias occurs (in the absence of other biases such as type 2 selection bias), the true causal risk difference $P(D^{e=1} = 1|S = 1) - P(D^{e=0} = 1|S = 1)$ in the selected sample will be equal to the true causal risk difference $P(D^{e=1} = 1) - P(D^{e=0} = 1)$ in the referent population. In other words, type 1 selection bias will result in a difference between the true causal effect in the selected sample and the effect estimated from the selected sample. Mathematically, in the presence of only type 1 selection bias, $P(D = 1|E = 1, S = 1) - P(D = 1|E = 0, S = 1) \neq P(D^{e=1} = 1|S = 1) - P(D^{e=0} = 1|S = 1) = P(D^{e=1} = 1) - P(D^{e=0} = 1)$.

Therefore, when type 1 selection bias occurs during the selection of the study sample from the target population (here we denote $T \rightarrow O$ for this selection process), the referent population is the target population and the selected sample is the study sample, and the true causal risk difference $P(D^{e=1} = 1|S_{T \rightarrow O} = 1) - P(D^{e=0} = 1|S_{T \rightarrow O} = 1)$ in the study sample will be equal to the true causal risk difference $P(D^{e=1} = 1) - P(D^{e=0} = 1)$ in the target population. Thus external validity is not affected. However, the associational risk difference $P(D = 1|E = 1, S_{T \rightarrow O} = 1) - P(D = 1|E = 0, S_{T \rightarrow O} = 1)$ in the study sample (which is also the analytic sample if no further selection processes occur) is not equal to the true causal risk difference $P(D^{e=1} = 1|S_{T \rightarrow O} = 1) - P(D^{e=0} = 1|S_{T \rightarrow O} = 1)$ in the study sample. Therefore, internal validity is affected.

When type 1 selection bias occurs during the selection of the analytic sample from the study sample (here we denote $O \rightarrow A$ for this selection process), the referent population is the study sample and the selected sample is the analytic sample. For convenience, assume the true causal risk difference in the study sample is $P(D^{e=1} = 1) - P(D^{e=0} = 1)$, and then the true causal risk difference in the analytic sample is $P(D^{e=1} = 1|S_{O \rightarrow A} = 1) - P(D^{e=0} = 1|S_{O \rightarrow A} = 1)$. In the presence of type 1 selection bias, the associational risk difference $P(D = 1|E = 1, S_{O \rightarrow A} = 1) - P(D = 1|E = 0, S_{O \rightarrow A} = 1)$ in the analytic sample is not equal to the true causal risk difference $P(D^{e=1} = 1) - P(D^{e=0} = 1)$ in the study sample. Therefore, internal validity is affected.

Putting this all together, type 1 selection bias only affects internal validity.

Type 2 selection bias

Type 2 selection bias, due to restricting to one (or more) level(s) of an effect measure modifier, produces an estimate that is biased only for the referent population, but not for the selected sample. That is, the observed risk difference $P(D = 1|E = 1, S = 1) - P(D = 1|E = 0, S = 1)$ from the selected sample is not equal to the true causal risk difference $P(D^{e=1} = 1) - P(D^{e=0} = 1)$ in the referent population, but is equal to the true causal risk difference $P(D^{e=1} = 1|S = 1) - P(D^{e=0} = 1|S = 1)$ in the selected sample. In other words, type 2 selection bias will result in a difference between the true causal effect in the referent population and the true causal effect in the selected sample. Mathematically, in the presence of only type 2 selection bias, $P(D = 1|E = 1, S = 1) - P(D = 1|E = 0, S = 1) = P(D^{e=1} = 1|S = 1) - P(D^{e=0} = 1|S = 1) \neq P(D^{e=1} = 1) - P(D^{e=0} = 1)$.

Therefore, when type 2 selection bias occurs during the selection of the study sample from the target population (here we denote $T \rightarrow O$ for this selection process), the referent population is the target population and the selected sample is the study sample, and the true

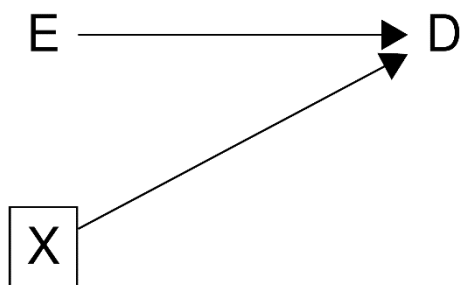
causal risk difference $P(D^{e=1} = 1|S_{T \rightarrow O} = 1) - P(D^{e=0} = 1|S_{T \rightarrow O} = 1)$ in the study sample may not be equal to the true causal risk difference $P(D^{e=1} = 1) - P(D^{e=0} = 1)$ in the target population. Thus external validity is affected. However, the associational risk difference $P(D = 1|E = 1, S_{T \rightarrow O} = 1) - P(D = 1|E = 0, S_{T \rightarrow O} = 1)$ in the study sample (which is also the analytic sample if no further selection processes occur) is equal to the true causal risk difference $P(D^{e=1} = 1|S_{T \rightarrow O} = 1) - P(D^{e=0} = 1|S_{T \rightarrow O} = 1)$ in the study sample. Therefore, internal validity is not affected.

When type 2 selection bias occurs during the selection of the analytic sample from the study sample (here we denote $O \rightarrow A$ for this selection process), the referent population is the study sample and the selected sample is the analytic sample. For convenience, assume the true causal risk difference in the study sample is $P(D^{e=1} = 1) - P(D^{e=0} = 1)$, and then the true causal risk difference in the analytic sample is $P(D^{e=1} = 1|S_{O \rightarrow A} = 1) - P(D^{e=0} = 1|S_{O \rightarrow A} = 1)$. In the presence of type 2 selection bias, the associational risk difference $P(D = 1|E = 1, S_{O \rightarrow A} = 1) - P(D = 1|E = 0, S_{O \rightarrow A} = 1)$ in the analytic sample is not equal to the true causal risk difference $P(D^{e=1} = 1) - P(D^{e=0} = 1)$ in the study sample. Therefore, internal validity is affected.

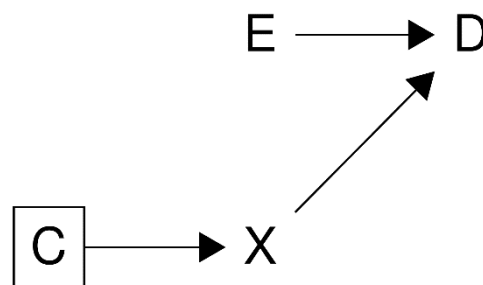
Therefore, we say type 2 selection bias affects external validity when selecting the study sample from the target population, and affects internal validity when selecting the analytic sample from the study sample.

eAPPENDIX II

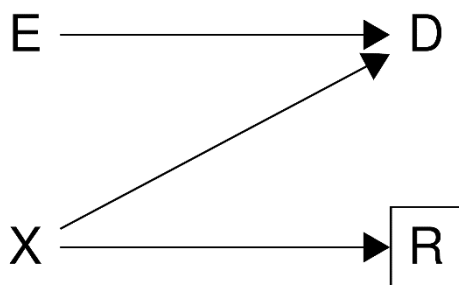
eAppendix Figure 1: four types of effect measure modification of the causal effect of E on D on at least one scale (RD or RR). E is exposure, D is outcome, and others (i.e., X, C, R, and M) are covariates. Note that, the structures of these causal diagrams do not imply that X is an effect measure modifier. To illustrate these four types of effect measure modification, it is assumed that X is an effect measure modifier.



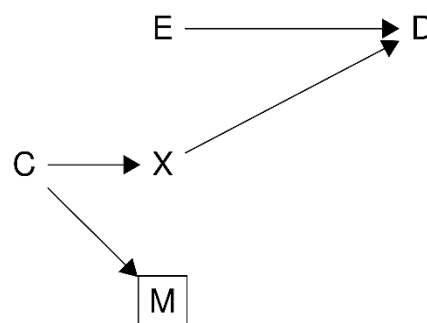
(a) X is a direct effect measure modifier



(b) C is an indirect effect measure modifier



(c) R is an effect measure modifier by proxy



(d) M is an effect measure modifier by common cause

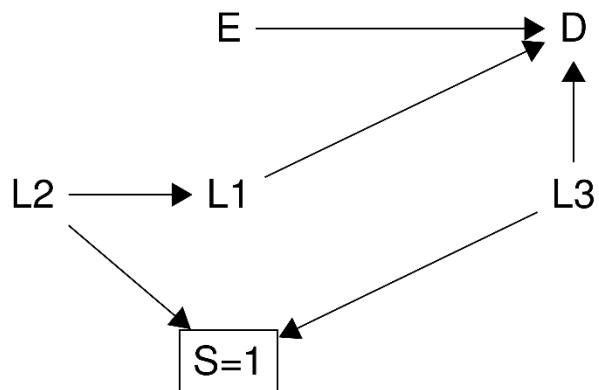
REFERENCE:

- 1) VanderWeele TJ, Robins JM. Four types of effect measure modification: a classification based on directed acyclic graphs. *Epidemiology*. 2007;18(5):561-568.

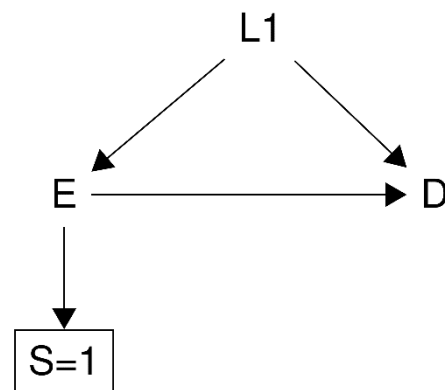
120 **eAPPENDIX III**

121 **eAppendix Figure 2:** Complex diagram for type 2 selection bias. E is exposure, D is outcome, S is selection, L (i.e., L1, L2, and L3) is
122 covariates

123 (a)



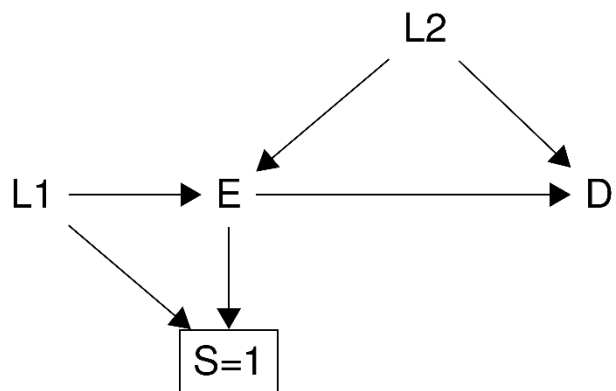
(b)



124

125

126 (c)



127

128

129

130 Note: eAppendix Figure 2 (a) suffers from type 2 selection bias by restricting to one level of the effect measure modifier by proxy as
131 well as restricting to one level of the effect measure modifier by common cause. (Coincidentally, it induces an open backdoor path

132 $L2 \rightarrow [S=1] \leftarrow L3$ by restricting to one level of the collider S . But this is not type 1 selection bias because the collider S is not the
133 common effect of the exposure (or a cause of the exposure), and the outcome (or a cause of the outcome.) eAppendix Figure 2 (b)
134 suffers from type 2 selection bias because there is an open path $L1 \rightarrow E \rightarrow [S=1]$. It means, the distribution of $L1$ in the selected
135 sample is different from that in the referent population. If $L1$ is an effect measure modifier on either scale, then type 2 selection bias
136 will arise. eAppendix Figure 2 (c) suffers from type 1 selection bias as well. The selection S is a collider itself, and also act as a
137 descendant of the collider E (which is the common effect of $L1$ and $L2$). By restricting to one level of the descendant of a collider, it
138 opens the path $[S=1] \leftarrow L1 \rightarrow E \leftarrow L2 \rightarrow D$. Similar to eAppendix Figure 2 (b), eAppendix Figure 2 (c) may also suffer from type 2 selection
139 bias if $L2$ is an effect measure modifier.
140

141 **eAPPENDIX IV**

142 **Proof for covariate adjustments in M-bias**

$$\begin{aligned}
 143 \quad E[D^e] &= \sum_{L2} E[D^e | L2 = l2] P(L2 = l2) && \text{(Law of total probability)} \\
 144 \quad &= \sum_{L2} E[D^e | E = e, L2 = l2] P(L2 = l2) && (E \perp\!\!\!\perp D^e | L2; \text{conditional exchangeability between exposed and unexposed}) \\
 145 \quad &= \sum_{L2} E[D | E = e, L2 = l2] P(L2 = l2) && \text{(causal consistency)} \\
 146 \quad &= \sum_{L2} E[D | E = e, L2 = l2, S = 1] P(L2 = l2) && (S \perp\!\!\!\perp D | L2, E)
 \end{aligned}$$

147
 148 Note: If we only have information on L1, we are not able to recover $E[D^e]$ using observed data since S is not conditionally
 149 independent of D given $L1$ and E , and thus the last step of derivation fails.

150

151

152

153 **SAS code for simulation study for the “M-bias” scenario**

```

154 /*Generate data*/
155 * M-bias;
156 DATA a;
157     CALL STREAMINIT(1234);
158     DO id=1 to 1000000;
159         PL1 = 0.5; L1 = rand("Bern", PL1); /* L1 */
160         PL2 = 0.6; L2 = rand("Bern", PL2); /* L2 */
161         PA=0.2 + 0.5*L1; A=rand("Bern", PA); /* Exposure */
162         PY=0.2 + 0.3*A - 0.1*L2 + 0.2*A*L2; Y=rand("Bern", PY); /* Outcome, with interaction between A and L2
163     */
164         PS = 0.3 + 0.2*L1 - 0.2*L2; S = rand("Bern", PS); /* Selection */
165 OUTPUT; END; RUN;
166
167 /*Truth: true causal effect in the referent population */
168 PROC GENMOD DATA=a DESC;
169     MODEL Y=A/ DIST=bin LINK=identity;
170     ODS SELECT modelinfo parameterestimates; TITLE "truth"; RUN;
171 /*Crude: effect estimate in the seleted sample */
172 PROC GENMOD DATA=a DESC;

```



```

173     MODEL Y=A / DIST=bin LINK=identity;
174     ODS SELECT modelinfo parameterestimates; TITLE "crude"; WHERE S=1; RUN;
175 /*Stratified on L1 */
176 PROC GENMOD DATA=a DESC;
177     MODEL Y=A L1/ DIST=bin LINK=identity;
178     ODS SELECT modelinfo parameterestimates; TITLE "stratified"; WHERE S=1; RUN;
179 /*Stratified on L2 */
180 PROC GENMOD DATA=a DESC;
181     MODEL Y=A L2/ DIST=bin LINK=identity;
182     ODS SELECT modelinfo parameterestimates; TITLE "stratified"; WHERE S=1; RUN;
183 /*Stratified on L1 and L2 */
184 PROC GENMOD DATA=a DESC;
185     MODEL Y=A L1 L2/ DIST=bin LINK=identity;
186     ODS SELECT modelinfo parameterestimates; TITLE "stratified"; WHERE S=1; RUN;
187 /* Conventional regression adjustment won't work */
188
189
190 /* Gformula */
191 /*Gformula using information on L2*/
192 TITLE "G-comp Models using information on L2";
193 ODS SELECT NONE;
194 PROC GENMOD DATA=a DESC;
195     WHERE S=1;
196     MODEL Y=A L2 A*L2/ DIST=bin LINK=identity;
197     ODS OUTPUT parameterestimates=yest(keep=parameter estimate stderr);
198 RUN;
199
200 PROC GENMOD DATA=a DESC;
201     MODEL L2= / DIST=bin LINK=identity;
202     ODS OUTPUT parameterestimates=west(keep=parameter estimate stderr);
203 RUN; ODS SELECT ALL;
204
205 DATA west;
206     SET west;
207     RETAIN z 1 intl2 intl2se;
208     IF parameter="Intercept" THEN DO; intl2=estimate; intl2se=stderr; OUTPUT; END;
209     KEEP z intl2 intl2se; RUN;
210
211 DATA yest;
212     SET yest;
213     RETAIN z 1 inty intyse yona yonase yonl2 yonl2se;

```

```

214     IF parameter="Intercept" THEN DO; inty=estimate; intyse=stderr; END;
215     IF parameter="A" THEN DO; yona=estimate; yonase=stderr; END;
216     IF parameter="L2" THEN DO; yonl2=estimate; yonl2se=stderr; END;
217     IF parameter="A*L2" THEN DO; yonint=estimate; yonintse=stderr; OUTPUT; END;
218     KEEP z inty intyse yona yonase yonl2 yonl2se yonint yonintse; RUN;
219
220 PROC PRINT DATA=west; TITLE "G-comp model P(L2)"; RUN;
221 PROC PRINT DATA=yest; TITLE "G-comp model P(Y|A,L2,S=1)"; RUN;
222 DATA b;
223     MERGE west yest;
224     BY z;
225     gcomp1=0; gcomp0=0;
226     DO A=0 TO 1;
227         DO L2=0 TO 1;
228             pl2 = intl2;
229             IF L2=0 THEN pl2=1-pl2;
230             IF A=0 THEN gcomp0=gcomp0 + pl2*(inty + yona*A + yonl2*L2 + yonint*A*L2);
231             ELSE gcomp1=gcomp1 + pl2*(inty + yona*A + yonl2*L2 + yonint*A*L2);
232         END;
233     END; gcomp=gcomp1-gcomp0;
234 DROP A L2 pl2; RUN;
235 PROC PRINT DATA = b; VAR gcomp; TITLE "gcomp using L2"; RUN;
236
237
238 /*Gformula using information on L1*/
239 TITLE "G-comp Models using information on L1";
240 ODS SELECT NONE;
241 PROC GENMOD DATA=a DESC;
242     WHERE S=1;
243     MODEL Y=A L1/ DIST=bin LINK=identity;
244     ODS OUTPUT parameterestimates=yest(keep=parameter estimate stderr);
245 RUN;
246
247 PROC GENMOD DATA=a DESC;
248     MODEL L1= / DIST=bin LINK=identity;
249     ODS OUTPUT parameterestimates=west(keep=parameter estimate stderr);
250 RUN; ODS SELECT ALL;
251
252 DATA west;
253     SET west;
254     RETAIN z 1 intl1 intl1se;

```

```

255     IF parameter="Intercept" THEN DO; intl1=estimate; intl1se=stderr; OUTPUT; END;
256     KEEP z intl1 intl1se; RUN;
257
258 DATA yest;
259     SET yest;
260     RETAIN z 1 inty intyse yona yonase yonl1 yonllse;
261     IF parameter="Intercept" THEN DO; inty=estimate; intyse=stderr; END;
262     IF parameter="A" THEN DO; yona=estimate; yonase=stderr; END;
263     IF parameter="L1" THEN DO; yonl1=estimate; yonllse=stderr; OUTPUT; END;
264     KEEP z inty intyse yona yonase yonl1 yonllse; RUN;
265
266 PROC PRINT DATA=west; TITLE "G-comp model P(L1)"; RUN;
267 PROC PRINT DATA=yest; TITLE "G-comp model P(Y|A,L1,S=1)"; RUN;
268 DATA b;
269     MERGE west yest;
270     BY z;
271     gcomp1=0; gcomp0=0;
272     DO A=0 TO 1;
273         DO L1=0 TO 1;
274             pl1 = intl1;
275             IF L1=0 THEN pl1=1-pl1;
276             IF A=0 THEN gcomp0=gcomp0 + pl1*(inty + yona*A + yonl1*L1);
277             ELSE gcomp1=gcomp1 + pl1*(inty + yona*A + yonl1*L1);
278         END;
279     END; gcomp=gcomp1-gcomp0;
280 DROP A L1 pl1; RUN;
281 PROC PRINT DATA = b; VAR gcomp; TITLE "gcomp using L1"; RUN;
282
283
284 /* Inverse probability weighting */
285 /*Inverse probability weighting using information on L2*/
286 /* Construct IP sampling weights */
287 PROC LOGISTIC DATA=a DESC NOPRINT;
288     MODEL S=;
289     OUTPUT OUT=num1 P=num1;
290 RUN;
291
292 PROC LOGISTIC DATA=a DESC NOPRINT;
293     CLASS A L2;
294     MODEL S=A L2;
295     OUTPUT OUT=den1 P=den1;

```

```

296 RUN;
297
298 DATA a31;
299     MERGE a num1 den1;
300     IF S=1 THEN SW = num1/den1;
301     ELSE DELETE;
302 RUN;
303 PROC GENMOD DATA=a31 DESC;
304     WEIGHT SW;
305     MODEL Y=A/ DIST=bin LINK=identity; run;
306     ODS SELECT modelinfo parameterestimates; TITLE "IPW on A and L2"; RUN;
307 /* Construct IP treatment weights for the selected sample, since in the selected sample, there is residual
308 bias due to open backdoor path A-L1-S-L2-Y */
309 PROC LOGISTIC DATA=a31 DESC NOPRINT;
310     MODEL A=;
311     OUTPUT OUT=num2 P=num2;
312 RUN;
313 PROC LOGISTIC DATA=a31 DESC NOPRINT;
314     CLASS L2;
315     MODEL A=L2;
316     OUTPUT OUT=den2 P=den2;
317 RUN;
318
319 DATA a32;
320     MERGE a31 num2 den2;
321     IF A=1 THEN TW = num2/den2;
322     ELSE TW=(1-num2)/(1-den2);
323     W = SW*TW;
324 RUN;
325
326 PROC MEANS DATA=a32; VAR W SW TW; run;
327 PROC GENMOD DATA=a32 DESC;
328     WEIGHT W;
329     MODEL Y=A/ DIST=bin LINK=identity; run;
330     ODS SELECT modelinfo parameterestimates; TITLE "IPW on A and L2"; RUN;
331
332
333 /* Inverse probability weighting using information on L1 */
334 PROC LOGISTIC DATA=a DESC NOPRINT;
335     MODEL S=;
336     OUTPUT OUT=num P=num;

```



```

337 RUN;
338
339 PROC LOGISTIC DATA=a DESC NOPRINT;
340     CLASS A L1;
341     MODEL S=A L1;
342     OUTPUT OUT=den P=den;
343 RUN;
344
345 DATA a21;
346     MERGE a num den;
347     IF S=1 THEN SW = num/den;
348     ELSE DELETE;
349 RUN;
350
351 PROC GENMOD DATA=a21 DESC;
352     WEIGHT SW;
353     MODEL Y=A/ DIST=bin LINK=identity;
354     ODS SELECT modelinfo parameterestimates; TITLE "IPW on A and L1"; RUN;
355 /* Construct IP treatment weights for the selected sample, since in the selected sample, there may be
356 residual bias due to the open backdoor path A-L1-S-L2-Y */
357 PROC LOGISTIC DATA=a21 DESC NOPRINT;
358     MODEL A=;
359     OUTPUT OUT=num2 P=num2;
360 RUN;
361 PROC LOGISTIC DATA=a21 DESC NOPRINT;
362     CLASS L1;
363     MODEL A=L1;
364     OUTPUT OUT=den2 P=den2;
365 RUN;
366
367 DATA a22;
368     MERGE a21 num2 den2;
369     IF A=1 THEN TW = num2/den2;
370     ELSE TW=(1-num2)/(1-den2);
371     W = SW*TW;
372 RUN;
373
374 PROC MEANS DATA=a22; VAR W SW TW; run;
375 PROC GENMOD DATA=a22 DESC;
376     WEIGHT W;
377     MODEL Y=A/ DIST=bin LINK=identity; run;

```

```
378 ODS SELECT modelinfo parameterestimates; TITLE "IPW on A and L1"; RUN;  
379  
380  
381  
382  
383
```