Current Epidemiology Reports (2020) 7:179–189
https://doi.org/10.1007/s40471-020-00241-6

EPIDEMIOLOGIC METHODS (P HOWARDS, SECTION EDITOR)

Selection Mechanisms and Their Consequences: Understanding and Addressing Selection Bias

Louisa H. Smith [1]

Published online: 9 August 2020

Abstract

Purpose of Review Epidemiologic research is rarely based on a random sample of a well-defined target population. We used causal directed acyclic graphs to demonstrate the types of bias that can result when selection into that sample is associated with the exposure or outcome of interest, or with both. These selection mechanisms can affect both the internal and external validity of a study. We reviewed approaches to selection mechanisms that affect valid causal inference.

Recent Findings We noted that selection bias can refer to a number of issues with different consequences. We identified strategies for addressing selection bias when designing studies, collecting data, conducting analyses, and assessing possible bias in those analyses.

Summary Understanding the way in which a study sample relates to the target population is critical for avoiding and addressing bias. Communication about selection bias is aided by the use of causal graphs.

Keywords Selection bias · Sampling · Collider bias · Causal inference · Target population · Directed acyclic graphs

Introduction

Selection bias is one of the primary biases in epidemiology, but it remains more of a mystery than confounding or measurement error. One reason may be the epidemiologic tradition of classifying selection biases by their origin stories—non-response bias, follow-up bias, prevalent-user bias, volunteer bias, self-selection bias, survivor bias, Berkson's bias, incidence-prevalence bias, control-selection bias, index-event bias—instead of as a single phenomenon. At the same time, the term "selection bias" itself appears overused. Different definitions have emerged out of different fields; what economists often call selection bias is what epidemiologists know as confounding. Other social scientists may primarily use the term to refer to the non-representativeness of a sample, and some epidemiologists strictly mean collider-stratification bias.

Nevertheless, progress has been made toward a unified theory of selection bias. While in the past epidemiologists used 2×2 tables to make their selection bias arguments [1–4], recent work has used causal directed acyclic graphs (DAGs). Notably, Hernán et al. used DAGs to demonstrate a structure shared by selection biases [5], Didelez et al. used graphs to derive conditions under which outcome-dependent sampling does not induce selection bias [6], and Bareinboim and colleagues have described graphical criteria under which causal effects can be recovered from sample selection [7–9]. Such work is critical as data sources and study designs evolve. An understanding of the structure of selection bias enables us to measure the appropriate covariates, avoid poor design and analysis choices, and conduct meaningful sensitivity analyses.

A recent review on selection bias described the conditions necessary for using data from a non-random sample of the target population to (a) estimate a valid causal effect and (b) generalize it to the target population [10•]. These two processes are often referred to as satisfying internal and external validity, respectively [11••]. This review will focus on providing intuition to differentiate the selection mechanisms that affect each type of validity, and on describing recent methodological developments to address internal validity.

This article is part of the Topical Collection on *Epidemiologic Methods*

✉ Louisa H. Smith
Louisa_h_smith@g.harvard.edu

[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave., Boston, MA 02115, USA

Preliminaries

Terminology and Notation

Suppose we are interested in describing the causal effect of exposure A on outcome Y in some target population. Because we cannot measure A, Y, and other random variables V on every person in the target population, we select a study sample from this population. If simple random sampling is used, we can say that the study sample is representative of the target population. Other systematic forms of sampling can also result in representativeness once the sampling design is accounted for, e.g., with inverse probability of sampling weights [12]. However, study samples are often not the result of explicit sampling of a target population, perhaps because participants were chosen out of convenience or some did not consent to participate. In other cases, it may be that the target population is defined by some cohort on which data has been collected, but loss to follow-up or other forms of missing data have resulted in an analytic sample that is no longer representative of that target population.

We can imagine that however the selected sample came to be, it is representative of some subset of the target population that has the same joint distribution of A, Y, and V, whether measured or unmeasured.[1] We will call this hypothetical population the selected population. The study sample may comprise the entire selected population, or we can think of it as a simple random sample from the selected population. For example, if younger people are less likely to participate in a survey, then the selected population is a subset of the target population with the same age structure as the study sample. Finally, we will assume that apart from possible selection bias, other assumptions for identifying causal effects have been met.

Directed Acyclic Graphs

We will use DAGs as a tool to understand and differentiate types of selection bias. A now classic paper by Hernán et al. described structural selection bias via DAGs [5]. DAGs continue to be an important tool for reasoning about selection bias [14]. In particular, they can be useful for conveying concepts such as the distinction between selection bias and confounding [15] and how terminology commonly used in the realm of trials relates to concepts in observational epidemiology [16••]. Complete graphical conditions for identifying causal effects from selection biased data and determining when and what external data is necessary to generalize those effects to a target population have been derived [7–9]. DAGs actively aid research into what estimands can be identified in the presence of selection into trials [17] and into surveys [18••].

Each random variable, or vector of random variables, is a node on the graph, and each directed edge represents a (possible) causal relationship between two nodes. Besides the variables of interest in the study, we can include selection nodes S on the DAGs, as in Figs. 1, 2, and 3. For a given observation, this may represent explicit selection into a study sample or simply whether or not a participant has complete data or is lost to follow-up. We will denote this with $S = 1$ for members of the selected population and $S = 0$ otherwise. In the figures, the selection nodes are boxed to represent the fact that the analysis is conditional on $S = 1$.

A Taxonomy of Selection Bias

An example of selection *without* bias can help distinguish the two overarching phenomena referred to as selection bias. Suppose that A has been randomly assigned in the target population, but that observations with some value of A are preferentially selected in the study sample (Fig. 1*a*). For example, consider a trial that recruits a large sample, then, due to resource concerns, limits follow-up ($S = 1$) to the entire treatment arm but a small number of participants in the placebo arm. Within this selected population, the exposure groups are still, on average, balanced with respect to measured and unmeasured baseline characteristics—there are just more of one group than the other. This grants us internal validity: it allows us to use the study sample to estimate the average causal effect of A on Y for the selected population. There is no selection bias because we can validly estimate a causal effect for the selected population.

Next, since selection is only affected by exposure—among the placebo arm, follow-up was random—it is random *within* exposure groups. Apart from exposure, measured and unmeasured characteristics are again balanced, now across the selected and non-selected groups. We therefore have external validity as well: we can use our estimate of the causal effect in the selected population to infer the presence and magnitude of a causal effect in the total population, so there is no selection bias.

Although the two levels of so-called selection bias are different, the same term may be used for both scenarios, although many epidemiologists prefer to refer to the second as a lack of generalizability. The latter may affect descriptive as well as causal measures. This review will focus on the type of selection bias affecting internal validity, but it is instructive to walk through a taxonomy of selection bias to gain intuition about when the methods discussed can and should be applied.

[1] The related problem of transportability refers to the situation in which the study sample is not a subset of the target population [13], but we will not consider this further.

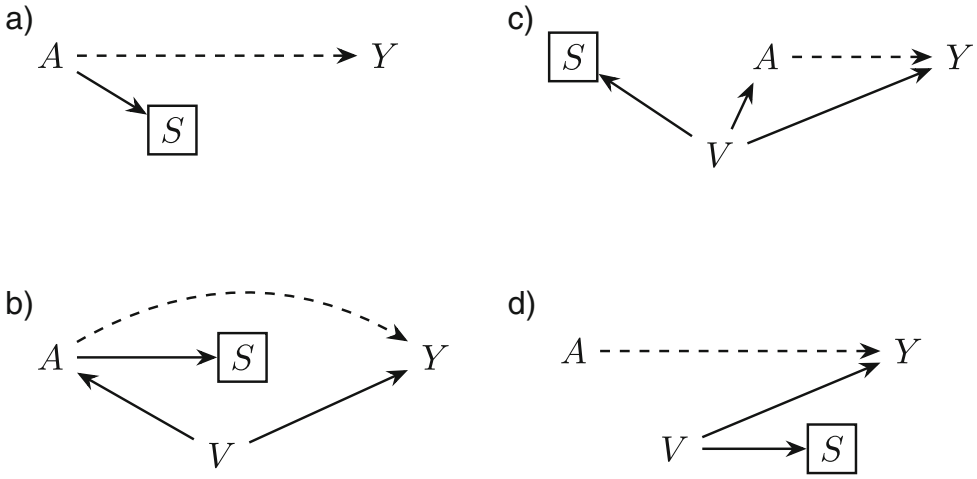

Fig. 1 Directed acyclic graphs (DAGs) depicting various selection mechanisms, where *A* is the exposure and *Y* the outcome of interest and *S* is an indicator of selection into the sample. *V* consists of other variables that may be measured or unmeasured. DAG *a* does not lead to biased inference in either the selected or the target population. Without proper adjustment for *V*, DAGs *b*–*d* may lead to biased inference in the target population but valid inference in the selected population

Failure to Generalize

In Fig. 1 *b* and *c*, the causal effect of exposure *A* on outcome *Y* is identifiable conditional on confounders *V* because *V* blocks all non-causal paths between *A* and *Y*. If there were no selection, we could, for example, use standardization to identify the average effect in the target population on the additive scale as $\sum_v \{E(Y|A=1,V=v) \quad -E(Y|A=0,V=v)\} \quad \Pr(V=v)$ (with the sum replaced by an integral for continuous *V*). That is, we average the difference in average outcomes between the exposure groups over the distribution of *V* in the population. In Fig. 1 *d*, which could represent a randomized trial, conditioning on *V* is not necessary but the same identifying expression can be used.

Because conditioning on *S* does not open any new paths between *A* and *Y* in any of the three DAGs, we can also identify a causal effect within the $S=1$ population, for example with $\sum_v \{E(Y|A=1,V=v,S=1) -E(Y|A=0,V=v, S=1)\} \Pr(V=v|S=1)$. However, even though the difference in average outcomes does not differ between the selected and the total population for any level of *V* (because $E(Y|\ A=a, V=v) = E(Y|\ A=a, V=v, S=1)$ for both exposure values), we would expect the distribution of *V* to differ in the selected population, so that $\Pr(V=v\ |\ S=1)$ does not equal $\Pr(V=v)$. This can occur because in each of graphs 1*b*–*d*, *V* is a cause of selection, directly or indirectly through *A*. As such, its distribution in the selected population will reflect this preferential selection of some value(s) of *V*. Unless the causal effect on the scale of interest is homogeneous across all values of *V*, the average effect in the total population will differ from that in the selected population.

External information on the distribution of *V* can be used to generalize the internally valid conditional causal effects estimated from the study sample. This requires every level of *V* in the target population to be represented in the selected population. It also requires, of course, the existence of such external data. Consider Fig. 1 *d*, where adjustment for *V* is not necessary to estimate valid causal effects for the selected population. In fact, *V* may include post-baseline characteristics (that aren't affected by exposure); as non-confounders, investigators may not have planned to measure them. Compared with the target population, the selected population will then have a different distribution of *V*, and if *V* is not known or measured, it may be impossible to characterize the selected population and generalize the causal effect to the target population. This situation has been referred to as selection bias "off the null" because unlike the situations that follow, bias will not occur if the exposure-outcome effect is null for every individual [19]. In other words, this structure allows for valid null hypothesis tests of the causal *A*–*Y* relationship, but not for valid estimation of the magnitude of effect in the total population.

Conditioning on a Collider

Conditioning on a collider generally biases inference even in the selected population. In the DAGs in Fig. 2, *S* is a collider, or a descendant of a collider, because there are two edges entering that node or one that causes it (DAG 2*b*). When *S* or any other collider is conditioned on, whether through restriction, stratification, or regression modeling, a non-causal association is usually (though not always [5, 20]) induced between variables causing it. The direction and magnitude of that association can be difficult to predict and has been a topic of great interest [21–23, 24•]. Even when this induced association is not between the exposure and outcome, it can create bias by opening up other non-causal pathways between the two variables, much as confounding does.

In Fig. 2 *a*, as in Fig. 1 *d*, selection is affected by a cause of the outcome *V*, but here it is also affected by the exposure. For example, consider a study in which education *A* reduces risk of early mortality (*S*), and only those alive in old age are eligible for a study on dementia (*Y*). However, those that survive differ with respect to risk factors for dementia (*V*).

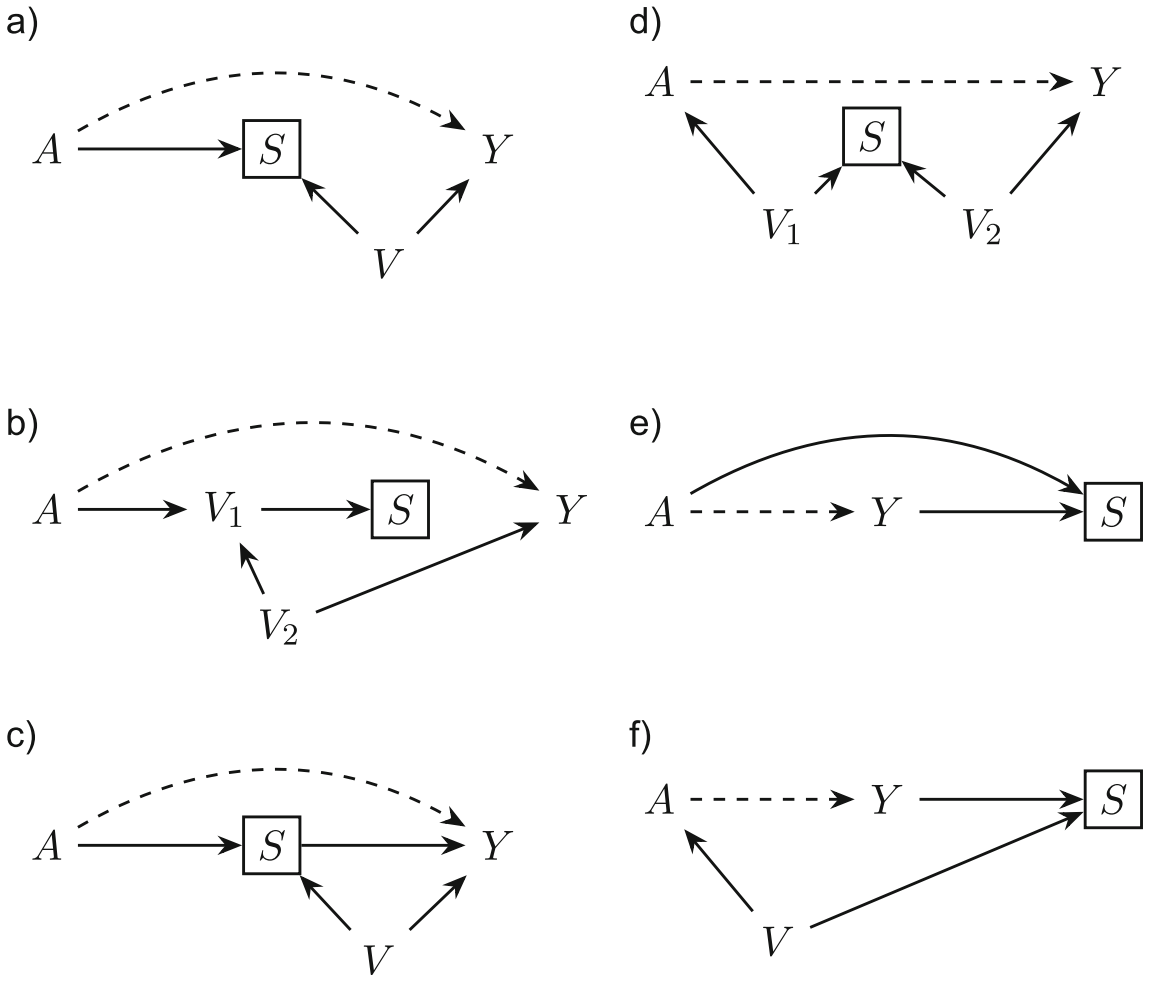

Fig. 2 Directed acyclic graphs (DAGs) depicting various selection mechanisms, where *A* is the exposure and *Y* the outcome of interest and *S* is an indicator of selection into the sample. *V* consists of other variables that may be measured or unmeasured. These DAGs depict situations that could lead to bias even in the selected population without proper adjustment. The main text describes several studies that correspond with these selection mechanisms (though additional nodes for confounders would likely exist on DAGs for these studies). DAG *a* could represent survival bias as described in the text, in which education *A* affects early mortality (*S*), and only those alive in old age are eligible for a study on dementia (*Y*). Common causes *V* of mortality and dementia can lead to bias, though the interpretation of an estimand that adjusts for that bias is not straightforward. DAG *b* could represent a prevalent-user design, in which *A* is initiation of a drug, V_1 is continued use, and *S* is selection into a study of prevalent users. Common causes V_2 of continued use and the outcome *Y*, such as underlying health conditions, could lead to selection bias if not taken into account. DAG *c* could represent an index-event study that gives rise to the phenomenon known as the obesity paradox. Here, *A* is obesity, *S* is the condition describing the population of interest (e.g., individuals with heart or renal disease), and *Y* is mortality. Failing to adjust for common causes of the index condition and mortality can lead to bias in the direct effect of obesity on mortality. DAG *d* could represent a study of the effect of antidepressants (*A*) on lung cancer (*Y*) conducted among people with coronary artery disease (*S*). If common causes of exposure and selection, such as depression (V_1), and common causes of selection and the outcome, such as smoking (V_2), are not adjusted for, M-bias can result. DAG *e* could represent a study of HIV treatment initiation timing on the risk of pre-term birth. *A* is treatment timing, which leads to selection *S* in that women who fail to initiate before or early in pregnancy are excluded, and *Y* is pre-term birth, which also leads to exclusion of women when birth occurs before treatment initiation. DAG *f* could represent a situation where a test-negative design could be useful. *A* could be influenza vaccination, *Y* influenza-like illness, *S* influenza testing, and *V*factors such as health-seeking behavior; restriction to patients with such behavior (i.e., conditioning on *V*) can avoid selection bias

Conditioning on selection—in this case, survival—induces an association between *V* and *A*, creating a non-causal path from *A* to *Y*. This means that education could appear to increase risk of dementia even if it had no causal effect for any person. If *V* is measured, proper adjustment can allow for valid inference in the selected population, and external data on the distribution of *V* in the target population can allow for generalization. However, when *S* represents survival, as in our example, this type of correction is equivalent to an intervention in which no one is subject to early mortality [25].

A similar structure is presented in Fig. 2*b*. In this case, although selection is a not collider, it is affected by one (V_1), which also induces bias. For example, a pharmacoepidemiologic study may compare current users of

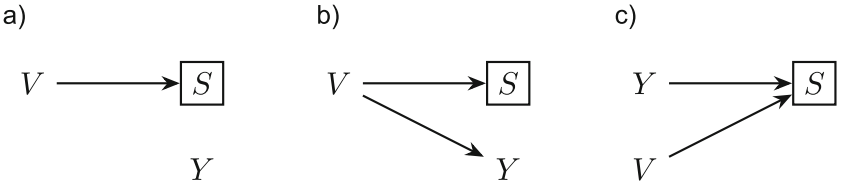

Fig. 3 Directed acyclic graphs (DAGs) depicting various selection mechanisms in studies of the distribution of *Y*. *S* is an indicator of selection into the sample. In DAG *a*, *Y* is independent of selection, so its distribution conditional on *S* is equal to its marginal distribution. In DAG *b*, the distribution of *Y*is expected to be different in the selected and non-selected populations, but external data on *V* can be used to standardize the distribution to the total population. In DAG *c*, the distribution of *Y* is expected to be different in the selected and non-selected populations, but cannot be corrected with simple standardization

a drug to those who have never used it, excluding those who no longer use it, perhaps due to side effects (V_1). If discontinuation of the drug was affected by causes of the outcome of interest, V_2, then selecting on a descendent of a collider by restricting to current or never users can open a non-causal path between exposure and outcome, and the drug could appear associated with the outcome even if its effect is null.

Selecting or conditioning on a variable on the causal path between A and Y, or on something caused by such a variable, can bias estimates of the *total* effect. However, Fig. 2 *c* is an example of a situation in which a *direct* effect may be the estimand of interest. That is, we are interested in the effect of *A* on *Y* while a fixing a mediating variable on the path between them to a specific value. Here, *S* represents the restriction of the study sample to participants with that value of the mediator. For example, there has been much epidemiologic interest in the effect of obesity (*A*) on mortality (*Y*), among people with heart disease (*S*) [25]. In some cases, a direct effect can be validly estimated by simply conditioning on the mediating variable, as would occur if the study sample was exclusively people with heart disease. However, in the presence of mediator-outcome confounders, as depicted in this causal diagram with the variable *V*, bias can result if they are not adjusted for.

M-bias is the name given to the structure in Fig. 2*d*. Researchers may restrict a study to participants with certain characteristics in an attempt to reduce confounding or may use a sample recruited for another purpose to investigate a secondary hypothesis. For example, consider a study of antidepressant use (*A*) on lung cancer (*Y*) among a sample of people with coronary artery disease (*S*) [26]. Restricting to this group may seem reasonable because *S* is associated with *A* and *Y* yet is not on the causal pathway. However, conditioning on *S* opens up a path between exposure and outcome through V_1 and V_2—for example, depression and smoking. Even if the study is not restricted to participants with $S = 1$, including a term for *S* in a model in an attempt to adjust for confounding can result in collider-stratification bias unless either of or both nodes of *V* are also conditioned on.

Finally, Fig. 2 *e* and *f* depict the structure of bias most likely to occur in case-control studies, in which selection is, by definition, based on outcome status. Figure 2 *e* is sometimes called Berkson's bias: selection is affected by both exposure and outcome. For example, in a study of HIV treatment initiation timing (*A*) on the risk of pre-term birth (*Y*), excluding women who fail to initiate treatment before birth also means excluding those with shorter pregnancies, as they are more likely to not have been treated [27]. Unlike the other forms of collider-stratification bias represented in previous graphs, the selection bias represented here cannot be corrected by properly adjusting for a covariate. Instead, the selection probabilities for every exposure-outcome combination must be known to validly estimate a causal effect. Figure 2 *f* occurs when a common cause of exposure and selection exists; measuring or restricting analysis to a certain value of *V* can avoid control-selection bias. A study of vaccine effectiveness may compare odds of vaccination (*A*) among those who test positive for the flu (*Y*) with healthy controls. If the controls are less likely both to get vaccinated and tested for the flu than the population from which the cases arose, selection bias can result due to the path opened through healthy behaviors *V*.

Descriptive Measures of Frequency

Equally important to epidemiology as inferring causation is describing a population. Descriptive measures may be measures of association like risk differences or odds ratios, but often we are interested in measures of frequency, such as prevalence or incidence. In the latter case, we are only concerned with the conditional distribution of *Y* given selection, and not the *A*–*Y* relationship. For example, we may be interested in using the study sample to estimate the prevalence of a disease in the target population. Unless we are trying to estimate this prevalence under a hypothetical intervention, we need only worry about generalizing descriptive inference, rather than causal inference, from the selected population to the target population.

The DAGs in Fig. 3 depict three simple situations in which estimating some aspects of the distribution of *Y* could be affected by selection. If *Y* is independent of selection (Fig. 3*a*), no correction is needed; if *Y*depends on selection via *V* (Fig. 3*b*), descriptive measures could be corrected using data on *V*, as with associational measures from the distributions depicted by the DAGs in Fig. 1*b*–*d*. If selection depends directly on *Y*

(Fig. 3*c*), even descriptive inference about the population requires additional assumptions.

Addressing Selection Bias

In this section, we will focus on the first stage of inference: from the study sample to a causal effect in the selected population. There also exists an extensive cross-disciplinary literature on generalizing both descriptive and causal measures from a sample to a target population that we will not describe. Several key articles in epidemiology include those by Cole and Stuart [28] and Lesko et al. [13], as well as others' contributions to assessment of generalizability [29], sensitivity analysis [30], and a tutorial on estimation [31••].

Design Considerations

Ideally, selection bias could simply be avoided in the design or data collection stage. Although some form of selection is almost inevitable when working with humans, changes to sampling methods and extra effort in tracking down participants or collecting data on suspected common causes of non-response and the outcome can help prevent bias. Additional variables to measure may also include negative control exposures and/or outcomes that aren't expected to induce the same selection bias as the study variables. Arnold and colleagues have described the characteristics of valid negative controls for selection bias using DAGs [32•].

As traditional approaches to recruiting participants and collecting information have declined, internet-based strategies have emerged. Such strategies provide benefits in speed, ease of data management, and reduced measurement error [33], although concerns about selection bias naturally follow. A number of longitudinal cohorts have been recruited via the Internet [34–36] and have reported reduced costs compared with offline recruitment [34] and limited selection bias [36].

Creative ways of sampling controls that represent exposure in the source population remain important in preventing selection bias in case-control studies. For example, there has been recent interest in test-negative designs [37] for vaccine (primarily influenza) effectiveness research. Such studies use patients presenting with influenza-like illness who test negative for influenza as controls and compare the odds of vaccination with that in those who test positive—in essence, controlling for factors that affect exposure and selection, as *V* does in Fig. 2*f*. Justification for the design has been given in the form of DAGs [38], and comparisons of multiple control groups have supported its use [39, 40], though concerns about bias and generalizability have been raised [41, 42].

A concern that particularly affects studies in pharmacoepidemiology is the inclusion of participants with prevalent vs. incident exposures, which can be addressed with a new user design [43]. Studies that include prevalent users of a drug of interest inevitably exclude former users who are no longer taking that drug due to side effects, the drug not working for them, or other reasons (see Fig. 2*b*). The contrast of only successful continued users of the drug to a comparison group can lead to bias, as those in the treatment group most at risk of poor outcomes may already have been selected out. Researchers can avoid this bias by including only new users of the treatment, which more closely replicates a trial where participants start treatment at baseline [44, 45].

Index-event studies are those in which subjects are recruited based on, or a sample is restricted to individuals with, an existing disease [46]. The "obesity paradox" is a well-known example: among people with conditions affected by obesity, such as renal or heart disease, obesity can appear to be protective against mortality [47]. Collider bias as depicted in Fig. 2*c* can result, as can an additional layer of selection due to survival until diagnosis of the index disease [46]. One solution is to measure all risk factors for mortality in such studies.

Survival bias is a specific type of selection bias that occurs when the selected population comprises individuals who have not yet died by the time of outcome ascertainment, or who have survived another competing event. For example, a study of education and dementia can only investigate dementia among people who have survived to old age, and early-life education may have affected that survival, as in Fig. 2*a*. While methods for general selection bias can be used (see next section), it is worth considering how the implied estimands can be interpreted [48, 49], as well as alternative explanations. For example, taking a composite outcome approach [50] to avoid survival bias might reframe the question to ask about the effect of education on death *or* a diagnosis of dementia. Survivor average causal effects, which describe effects among people who would hypothetically survive until outcome ascertainment whether exposed or unexposed, are another option for survival bias that some may find meaningful (see Tchetgen Tchetgen et al. [51] for straightforward explanation and estimation procedure, and Long et al. for an application [52•]). These are just two examples of possible approaches to survival bias from a broader literature; Young et al. [53••] provide a thorough explanation of causal effects in the presence of competing risks.

Finally, designing a study according to a target trial framework can help avoid selection and other biases [54]. For example, Stoner et al. describe an observational study design comparing risk of pre-term birth in pre-conception vs. post-conception antiretroviral therapy in HIV-infected pregnant women [27]. All women who give birth before initiating therapy are excluded. Because both exposure (timing of treatment initiation) and outcome (timing of birth) affect selection, as in Fig. 2*e*, bias can result. If we consider the equivalent trial, treatment would be assigned pre-conception and outcomes would be ascertained in each arm regardless of whether or

not women initiated therapy before giving birth. Several methods which use this target trial framework to address questions of treatment timing without inducing bias have been described [55, 56].

Analytic Strategies for Addressing Selection Bias

Options for approaching selection bias in the analysis depend on the data available. When the common causes of selection and the outcome have been measured, in some situations they can simply be included in the outcome regression model to estimate valid conditional causal effects [57]. In a missing data framework, these are the situations in which a "complete-case" analysis can be used without bias. Due to the nature of logistic regression models, this is more often the case when fitting those models [58]. However, since this technique relies on the same model to adjust for selection bias and confounding, model misspecification may be a concern. Also, time-varying exposures require more advanced methods.

Multiple imputation can also be used in missing data situations when a complete-case analysis would lead to selection bias. Moreno-Betancur et al. use DAGs to distinguish missing data situations in which various parameters of interest are recoverable with complete-case analysis and others which require multiple imputation or are not recoverable without bias [59]. Multiple imputation, which averages estimates over many imputed datasets, can be computationally intensive with large datasets, but is now widely available in statistical software packages. The reader is directed to other overviews of multiple imputation in epidemiology [60, 61]. In general, the missing data framework is a useful tool for understanding selection bias, and several authors have made explicit connections [62–64].

Inverse probability weighting methods [65] are a common, though possibly still underused [66], approach to correcting for selection bias when predictors of selection are measured in the non-selected as well as the selected group, as baseline characteristics are before loss to follow-up. As these methods have gained more popularity in epidemiology for confounding adjustment, particularly for time-varying confounding, they have proved to be useful in adjusting for selection bias as well. Weighting approaches (often referred to as inverse probability of censoring, attrition, or selection weighting) involve using the portion of the sample with known outcomes to stand in for those without data based on what was measured in both groups.

A number of examples of weighting for selection are available in the epidemiologic literature: for example, in a fertility study where some women stop treatment [67], to estimate racial disparities in a cohort with loss to follow-up [52•], and to correct for pre-hospital mortality in a study of mortality after acute cardiac events [68]. As with inverse probability weighting for confounding, the validity of the method depends on the extent to which the selection model succeeds at balancing measured covariates, and to which those covariates account for selection. It cannot, of course, balance unmeasured causes of censoring. Jackson [69•] provides a method for assessing measured covariate balance after applying weights.

The method can also incorporate multiple mechanisms for multiple stages of selection. For example, there may be multiple points at which participants are dropped or selected for the study [70]. Because predictors of selection at each stage might be different, they can be modeled separately. The estimated weights are then multiplied together such that the remaining observations are upweighted by the inverse of the *joint* probability that they made it through all rounds of selection. Other extensions include Sun and colleagues' approach for when the data are subject to non-monotone missingness, meaning that selection does not necessarily occur progressively [71]. Finally, doubly robust approaches use both weights and outcome regression to adjust for selection bias; validity requires that at least one of either the weight model or the outcome model is correctly specified [49].

Sensitivity Analysis

The previous section covered situations in which the predictors of missingness are known and measured, either in just the selected population or in the target population. Often, however, these variables may be unknown and unmeasured. Rather than qualitatively describing the bias that those variables can create, which is not always intuitive, it is useful to quantitatively assess the potential bias, both while planning the study and analyzing the data [72]. In fact, it may be useful to consider every analysis to be a sensitivity analysis, as the correction methods described previously rest on untestable assumptions [73].

A link from inverse probability weighting methods to sensitivity analysis was proposed by Thompson et al., who describe how to specify sensitivity parameters for the selection model that is used to estimate the probabilities in the weights [14, 74]. Other sensitivity analyses directly parameterize the selection probabilities [75, 76], a method used often in birth defects research, where studies are usually restricted to live births. For example, Patorno and colleagues provide a sensitivity analysis to explore the possible effect that differential termination of malformed fetuses would have on their study of lithium in pregnancy and cardiac malformations [77]. In that situation, there were a limited number of reasonable values for the selection probabilities. Other situations may be more complex, and by assigning probability distributions to various parameters describing selection, researchers can examine the estimate they would have observed under a range of plausible situations [78, 79]. For index-event studies, Stensrud et al. alternatively suggest fitting frailty models, which allow

exploration of the size of the bias due to selection based on the index condition [80]. These models posit that mortality risk is heterogeneous and that the source of that heterogeneity, "frailty," can be parameterized directly. In all of these analyses, employing various experts to describe the plausible values for the sensitivity parameters [81] may be useful to avoid a tendency to only present those that are favorable to the study hypothesis.

A number of different methods have emerged for bounding the size of selection. Huang and colleagues describe a bound for odds ratios estimated from case-control studies [82], while Smith and VanderWeele propose bounds for the bias for the causal effect in the total population [83], in the selected population only [83], and when the bias is due to improper selection of controls [84]. A related bound for M-bias can also be extended to other selection bias structures [85]. External information, such as that from a census, can also be used in a sensitivity analysis to estimate informative intervals [86]. Some of these bounds can be used to construct an *E* value–like measure [87] that summarizes the minimum bias needed to fully explain an estimate, if the true causal effect is null (or some other value) [83, 85].

Finally, another approach toward unmeasured predictors of selection was formulated by Heckman in 1979 [88], but the selection models he developed are not often used in epidemiology, possibly due to strong parametric assumptions and the need for a measured instrument-like variable that predicts missingness but not disease. An application to missing responses in HIV prevalence surveys is a notable exception [89], and the question has helped generate extensions to the method that relax some of the assumptions [90–92]. Additionally, West and McCabe compared several approaches for estimating prevalences of substance-use behaviors subject to non-response, including selection models, weighting, and imputation [93•]. They suggested researchers implement a series of these methods as a sensitivity analysis.

Conclusions

Selection bias can be seen either as a broad problem that occurs any time we have non-random sampling of a target population, or as a more narrow problem that does not allow for valid causal inference even under the null hypothesis. Whatever the lens used to view the phenomenon, it is worth carefully describing what is meant by "selection bias"; this may allow for more cross-disciplinary communication and sharing of strategies to avoid, correct for, and analyze selection bias of whatever form. As many other authors have noted, DAGs are a useful tool for doing so.

This review has taken a relatively narrow view of selection bias by primarily considering simple exposure-outcome relationships. Others have considered selection bias in instrumental variable studies [94, 95, 96•], including in the case of Mendelian randomization [97]. Other study designs and questions, such as self-matched designs [98], latent class analysis [99], and life expectancy estimation [100] have their own selection bias concerns. While randomized controlled trials can be affected by selection bias due to loss to follow-up, cluster-randomized designs [101] and multi-arm trials [102] can result in other types of selection bias if randomization occurs before participant selection or if treatment allocation can be predicted.

Because confounding is a more obvious culprit for non-causal associations and is often simpler to address, selection bias is often overlooked. While epidemiologists have long been trained on how non-response can affect studies, newcomers to data science in public health and elsewhere may be more likely to neglect the fact that only rarely is a study sample actually randomly sampled from a target population. As we become more creative with collecting data from a variety of sources and such data becomes more easily accessible, it will be increasingly important to understand the population on whom data is collected, and the population on which it is not.

Acknowledgements I am grateful to Joy Shi for the helpful comments.

Compliance with ethical standards

Conflict of Interest The author declares that there is no conflict of interest.

Human and Animal Rights This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Berkson J. Limitations of the application of fourfold table analysis to hospital data. Biom Bull. 1946;2(3):47–53.
2. Greenland S. Response and follow-up bias in cohort studies. Am J Epidemiol. 1977;106(3):184–7.
3. Criqui MH. Response bias and risk ratios in epidemiologic studies. Am J Epidemiol. 1979;109(4):394–9.
4. Austin MA, Criqui MH, Barrett-connor E, Holdbrook MJ. The effect of response bias on the odds ratio. Am J Epidemiol. 1981;114(1):137–43.
5. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615–25.
6. Didelez V, Kreiner S, Keiding N. Graphical models for inference under outcome-dependent sampling. Stat Sci. 2010;25(3):368–87.
7. Bareinboim E, Tian J, Pearl J. Recovering from selection bias in causal and statistical inference. In: Proceedings of the Twenty-

Eighth AAAI Conference on Artificial Intelligence. 2014. p. 2410–6.
8. Bareinboim E, Tian J. Recovering causal effects from selection Bias. In: Proceedings of the Twenty-Ninth Conference on Artificial Intelligence. 2015. p. 3475–81.
9. Correa JD, Tian J, Bareinboim E. Identification of causal effects in the presence of selection bias. Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence 2019;2744–51.
10.• Infante-Rivard C, Cusson A. Reflection on modern methods: selection bias—a review of recent developments. Int J Epidemiol. 2018;47(5):1714–22 **Translates recent work in algorithms for recoverability under selection for an epidemiology audience.**
11.•• Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target validity and the hierarchy of study designs. Am J Epidemiol. 2018;188(2):438–43 **Proposes the concept of target validity, which brings together internal and external validity.**
12. Kohler U, Kreuter F, Stuart EA. Nonprobability sampling and causal analysis. Ann Rev Stat Appl. 2019;6(1):149–72.
13. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results. Epidemiology. 2017;28(4):553–61.
14. Arah OA. Analyzing selection bias for credible causal inference: when in doubt, DAG it out. Epidemiology. 2019;30(4):517–20.
15. Haneuse S. Distinguishing selection bias and confounding bias in comparative effectiveness research. Med Care. 2016;54(4):e23–9.
16.•• Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. Epidemiology. 2017;28(1):54–9 **Translates bias terminology from and into the world of trials. Regarding selection bias, it is particularly helpful for explaining how bias can occur even under randomization.**
17. Dahabreh IJ, Robins JM, Haneuse SJ-PA, Hernán MA. Generalizing causal inferences from randomized trials: counterfactual and graphical identification [Internet]. 2019. Available from: http://arxiv.org/abs/1906.10792.
18.•• Schuessler J, Selb P. Graphical causal models for survey inference [Internet]. 2019. Available from: https://osf.io/preprints/socarxiv/hbg3m/. **A very clear introduction to directed acyclic graphs, translates concepts in causal inference to survey inference, with extensions including multiple selection nodes.**
19. Hernán MA. Invited commentary: selection bias without colliders. Am J Epidemiol. 2017;185(11):1048–50.
20. VanderWeele TJ, Robins JM. Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. Am J Epidemiol. 2007;166(9):1096–104.
21. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology. 2003;14(3):300–6.
22. Jiang Z, Ding P. The directions of selection bias. Stat Probabil Lett. 2017;125:104–9.
23. Nguyen TQ, Dafoe A, Ogburn EL. The magnitude and direction of collider bias for binary variables. Epidemiol Methods. 2019:1–29.
24.• Munafò MR, Tilling K, Taylor AE, Evans DM, Davey SG. Collider scope: when selection bias can substantially influence observed associations. Int J Epidemiol. 2018;47(1):226–35 **Provides intuitive explanation of collider bias as well as examples of selection bias in genetic studies.**
25. Chaix B, Evans D, Merlo J, Suzuki E. Commentary: weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. Epidemiology. 2012;23(1):129–31.
26. Wei L, Brookhart MA, Schneeweiss S, Mi X, Setoguchi S, Liu W, et al. Implications of M bias in epidemiologic studies: a simulation study. Am J Epidemiol. 2012;176(10):938–48.
27. Stoner MCD, Cole SR, Price J, Winston J, Stringer JSA. Timing of initiation of antiretroviral therapy and risk of preterm birth in studies of HIV-infected pregnant women. Epidemiology. 2018 Mar;29(2):224–9.
28. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. Am J Epidemiol. 2010;172(1):107–15.
29. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. J R Stat Soc Ser A (Stat Soc). 2011;174(2):369–86.
30. Nguyen TQ, Ebnesajjad C, Cole SR, Stuart EA. Sensitivity analysis for an unobserved moderator in RCT-to-target-population generalization of treatment effects. Ann Appl Stat. 2017;11(1):225–47.
31.•• Dahabreh IJ, Robertson SE, Steingrimsson JA, Stuart EA, Hernan MA. Extending inferences from a randomized trial to a new target population [Internet]. 2019. Available from: http://arxiv.org/abs/1805.00550. **Very clear tutorial, with code, for generalizing from a trial population using several methods.**
32.• Arnold BF, Ercumen A, Benjamin-Chung J, Colford JMJ. Brief report: Negative controls to detect selection bias and measurement bias in epidemiologic studies. Epidemiology. 2016;27(5):637–41 **Provides helpful guidance for using negative control exposures and outcomes, with DAGs and examples from the literature.**
33. Van Gelder MMHJ, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol. 2010;172(11):1292–8.
34. Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and efficiency of online and offline recruitment methods: a web-based cohort study. J Med Internet Res. 2017;19(3):e58.
35. Wise LA, Rothman KJ, Mikkelsen EM, Stanford JB, Wesselink AK, McKinnon C, et al. Design and conduct of an internet-based preconception cohort study in North America: pregnancy study online. Paediatr Perinat Epidemiol. 2015;29(4):360–71.
36. Hatch EE, Hahn KA, Wise LA, Mikkelsen EM, Kumar R, Fox MP, et al. Evaluation of selection bias in an internet-based study of pregnancy planners. Epidemiology. 2016;27(1):98–104.
37. Jackson ML, Nelson JC. The test-negative design for estimating influenza vaccine effectiveness. Vaccine. 2013;31(17):2165–8.
38. Sullivan SG, Tchetgen EJT, Cowling BJ. Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness. Am J Epidemiol. 2016;184(5):345–53.
39. Feng S, Cowling BJ, Kelly H, Sullivan SG. Estimating influenza vaccine effectiveness with the test-negative design using alternative control groups: a systematic review and meta-analysis. Am J Epidemiol. 2018;187(2):389–97.
40. Segaloff HE, Cheng B, Miller A V, Petrie JG, Malosh RE, Cheng C, et al. Influenza vaccine effectiveness in the inpatient setting; evaluation of potential bias in the test negative design by use of alternate control groups. Am J Epidemiol. 2019.
41. Westreich D, Hudgens MG. Invited commentary: beware the test-negative design. Am J Epidemiol. 2016;184(5):354–6.
42. Lewnard JA, Tedijanto C, Cowling BJ, Lipsitch M. Measurement of vaccine direct effects under the test-negative design. Am J Epidemiol. 2018;187(12):2686–97.
43. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. 2003;158(9):915–20.
44. Danaei G, Tavakkoli M, Hernán MA. Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins. Am J Epidemiol. 2012;175(4):250–62.
45. McCandless LC. Statin use and fracture risk: can we quantify the healthy-user effect? Epidemiology. 2013;24(5):743–52.

46. Flanders WD, Eldridge RC, McClellan W. A nearly unavoidable mechanism for collider bias with index-event studies. Epidemiology. 2014;25(5):762–4.
47. Banack HR, Kaufman JS. The "obesity paradox" explained. Epidemiology. 2013;24(3):461–2.
48. Tchetgen EJT, Glymour MM, Shpitser I, Weuve J. Rejoinder: to weight or not to weight?: on the relation between inverse-probability weighting and principal stratification for truncation by death. Epidemiology. 2012;23(1):132–7.
49. Shardell M, Hicks GE, Ferrucci L. Doubly robust estimation and causal inference in longitudinal studies with dropout and truncation by death. Biostatistics. 2015;16(1):155–68.
50. Hernán MA, Schisterman EF, Hernández-Díaz S. Invited commentary: composite outcomes as an attempt to escape from selection bias and related paradoxes. Am J Epidemiol. 2014;179(3):368–70.
51. Tchetgen Tchetgen EJ, Phiri K, Shapiro R. A simple regression-based approach to account for survival bias in birth outcomes research. Epidemiology. 2015;26(4):473–80.
52.• Long DL, Howard G, Long DM, Judd S, Manly JJ, McClure LA, et al. An investigation of selection bias in estimating racial disparity in stroke risk factors. Am J Epidemiol. 2019;188(3):587–97 **A clear, straightforward application of inverse probability weighting and a survivor average causal effect approach.**
53.•• Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. Stat Med. 2020;39(8):1199–236 **A thorough examination of what is implied by various estimands in the face of competing events.**
54. 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;79:70–5.
55. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. Int J Biostat. 2010;6(2):1–42.
56. Young JG, Cain LE, Robins JM, O'Reilly EJ, Hernán MA. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. Stat Biosci. 2011;3(1):119–43.
57. Little RJA. Regression with missing X's: a review. J Am Stat Assoc. 1992;87(420):1227–37.
58. Bartlett JW, Harel O, Carpenter JR. Asymptotically unbiased estimation of exposure odds ratios in complete records logistic regression. Am J Epidemiol. 2014;182(8):730–6.
59. Moreno-Betancur M, Lee KJ, Leacy FP, White IR, Simpson JA, Carlin JB. Canonical causal diagrams to guide the treatment of missing data in epidemiologic studies. Am J Epidemiol. 2018;187(12):2705–15.
60. Groenwold RHH, Donders ART, Roes KCB, Harrell FE, Moons KGM. Dealing with missing outcome data in randomized trials and observational studies. Am J Epidemiol. 2012;175(3):210–7.
61. Harel O, Mitchell EM, Perkins NJ, Cole SR, Tchetgen Tchetgen EJ, Sun B, et al. Multiple imputation for incomplete data in epidemiologic studies. Am J Epidemiol. 2018;187(3):576–84.
62. Haneuse S, Schildcrout J, Crane P, Sonnen J, Breitner J, Larson E. Adjustment for selection bias in observational studies with application to the analysis of autopsy data. Neuroepidemiology. 2009;32(3):229–39.
63. Westreich D. Berkson's bias, selection bias, and missing data. Epidemiology. 2012;23(1):159–64.
64. Howe CJ, Cain LE, Hogan JW. Are all biases missing data problems? Curr Epidemiol Rep. 2015;2(3):162–71.
65. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000;56(3):779–88.
66. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJJ. Selection bias due to loss to follow up in cohort studies. Epidemiology. 2016;27(1):91–7.
67. Modest AM, Wise LA, Fox MP, Weuve J, Penzias AS, Hacker MR. IVF success corrected for drop-out: use of inverse probability weighting. Hum Reprod. 2018;33(12):2295–301.
68. Banack HR, Harper S, Kaufman JS. Accounting for selection bias in studies of acute cardiac events. Can J Cardiol. 2018;34(6):709–16.
69.• Jackson JW. Diagnosing covariate balance across levels of right-censoring before and after application of inverse-probability-of-censoring weights. Am J Epidemiol. 2019;188(12):1–9 **Method for assessing whether models for selection succeed at balancing measured covariates when used to construct inverse probability weights, with example code.**
70. Haneuse S, Bogart A, Jazic I, Westbrook EO, Boudreau D, Theis MK, et al. Learning about missing data mechanisms in electronic health records-based research: a survey-based approach. Epidemiology. 2016;27(1):82–90.
71. Sun B, Perkins NJ, Cole SR, Harel O, Mitchell EM, Schisterman EF, et al. Inverse-probability-weighted estimation for monotone and nonmonotone missing data. Am J Epidemiol. 2018;187(3):585–91.
72. Fox MP, Lash TL. Quantitative bias analysis for study and grant planning. Ann Epidemiol. 2020 Mar;43:32–6.
73. Geneletti S, Mason A, Best N. Adjusting for selection effects in epidemiologic studies. Epidemiology. 2011;22(1):36–9.
74. Thompson CA, Arah OA. Selection bias modeling using observed data augmented with imputed record-level probabilities. Ann Epidemiol. 2014;24(10):747–53.
75. Greenland S. Basic methods for sensitivity analysis of biases. Int J Epidemiol. 1996;25(6):1107–16.
76. Khoury MJ, Flanders WD, James LM, Erickson JD. Human teratogens, prenatal mortality, and selection bias. Am J Epidemiol. 1989;130(2):361–70.
77. Patorno E, Huybrechts KF, Bateman BT, Cohen JM, Desai RJ, Mogun H, et al. Lithium use in pregnancy and the risk of cardiac malformations. N Engl J Med. 2017;376(23):2245–54.
78. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. Int J Epidemiol. 2014;43(6):1969–85.
79. Howards PP, Johnson CY, Honein MA, Flanders WD. Adjusting for bias due to incomplete case ascertainment in case-control studies of birth defects. Am J Epidemiol. 2015;181(8):595–607.
80. Stensrud MJ, Valberg M, Røysland K, Aalen OO. Exploring selection bias by causal frailty models: the magnitude matters. Epidemiology. 2017;28(3):379–86.
81. Shardell M, Scharfstein DO, Vlahov D, Galai N. Sensitivity analysis using elicited expert information for inference with coarsened data: illustration of censored discrete event times in the AIDS Link to Intravenous Experience (ALIVE) study. Am J Epidemiol. 2008;168(12):1460–9.
82. Huang T-HH, Lee W-CC. Bounding formulas for selection bias. Am J Epidemiol. 2015;182(10):868–72.
83. Smith LH, VanderWeele TJ. Bounding bias due to selection. Epidemiology. 2019;30(4):509–16.
84. Smith LH, VanderWeele TJ. Simple sensitivity analysis for control selection bias. Epidemiology. 2020;31(5).
85. Flanders WD, Ye D. Limits for the magnitude of M-bias and certain other types of structural selection bias. Epidemiology. 2019;30(4):501–8.
86. Tudball M, Zhao Q, Hughes R, Tilling K, Bowden J. An interval estimation approach to sample selection bias [Internet]. 2019. Available from: http://arxiv.org/abs/1906.10159.

87. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 2017;167(4):268–75.
88. Heckman JJ. Sample selection bias as a specification error. Econometrica. 1979;47(1):153–61.
89. Bärnighausen T, Bor J, Wandira-Kazibwe S, Canning D. Correcting HIV prevalence estimates for survey nonparticipation using Heckman-type selection models. Epidemiology. 2011;22(1):27–35.
90. McGovern ME, Bärnighausen T, Marra G, Radice R. On the assumption of bivariate normality in selection models. Epidemiology. 2015;26(2):229–37.
91. Marra G, Radice R, Bärnighausen T, Wood SN, McGovern ME. A simultaneous equation approach to estimating HIV prevalence with nonignorable missing responses. J Am Stat Assoc. 2017;112(518):484–96.
92. Marden JR, Wang L, Tchetgen Tchetgen EJ, Walter S, Glymour MM, Wirth KE, et al. Implementation of instrumental variable bounds for data missing not at random. Epidemiology. 2018;29(3):1.
93.• West BT, McCabe SE. Alternative approaches to assessing nonresponse bias in longitudinal survey estimates: an application to substance-use outcomes among young adults in the United States. Am J Epidemiol. 2017;185(7):591–600 **Application and comparison of multiple methods for estimating prevalence under selection, with a recommendation for a procedure for sensitivity analysis.**
94. Vansteelandt S, Walter S, Tchetgen ET. Eliminating survivor bias in two-stage instrumental variable estimators. Epidemiology. 2018;29(4):536–41.
95. Swanson SA. A practical guide to selection bias in instrumental variable analyses. Epidemiology. 2019;30(3):345–9.
96.• Hughes RA, Davies NM, Davey Smith G, Tilling K. Selection bias when estimating average treatment effects using one-sample instrumental variable analysis. Epidemiology. 2019;30(3):350–7 **Describes using DAGs selection mechanisms under which two-stage least squares is biased.**
97. Smit RAJJ, Trompet S, Dekkers OM, Jukema JW, le Cessie S. Survival bias in Mendelian randomization studies: a threat to causal inference. Epidemiology. 2019;30(6):813–6.
98. Mostofsky E, Coull BA, Mittleman MA. Analysis of observational self-matched data to examine acute triggers of outcome events with abrupt onset. Epidemiology. 2018;29(6):804–16.
99. Mooney SJ. Invited commentary: the tao of clinical cohort analysis—when the transitions that can be spoken of are not the true transitions. Am J Epidemiol. 2017;185(8):636–8.
100. Dowd JB, Hamoudi A. Is life expectancy really falling for groups of low socio-economic status? Lagged selection bias and artefactual trends in mortality. Int J Epidemiol. 2014;43(4):983–8.
101. Bolzern J, Mnyama N, Bosanquet K, Torgerson DJ. A review of cluster randomized trials found statistical evidence of selection bias. J Clin Epidemiol. 2018;99:106–12.
102. Uschner D, Hilgers RD, Heussen N. The impact of selection bias in randomized multi-arm parallel group clinical trials. PLoS One. 2018;13(1):1–18.