



Selection Bias in Health Research: Quantifying, Eliminating, or Exacerbating Health Disparities?

L. Paloma Rojas-Saunero¹ · M. Maria Glymour^{2,3} · Elizabeth Rose Mayeda¹

Accepted: 2 May 2023 / Published online: 30 August 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

Purpose of Review To summarize recent literature on selection bias in disparities research addressing either descriptive or causal questions, with examples from dementia research.

Recent Findings Defining a clear estimand, including the target population, is essential to assess whether generalizability bias or collider-stratification bias is a threat to inferences. Selection bias in disparities research can result from sampling strategies, differential inclusion pipelines, loss to follow-up, and competing events. If competing events occur, several potentially relevant estimands can be estimated under different assumptions, with different interpretations. The apparent magnitude of a disparity can differ substantially based on the chosen estimand. Both randomized and observational studies may misrepresent health disparities or heterogeneity in treatment effects if they are not based on a known sampling scheme.

Summary Researchers have recently made substantial progress in conceptualization and methods related to selection bias. This progress will improve the relevance of both descriptive and causal health disparities research.

Keywords Health disparities · Estimands · Selection bias · Generalizability · Collider-stratification bias · Competing events

Introduction

Drawing on recent scholarship [1•, 2•, 3–6], we discuss selection bias in health disparities research. We review selection bias in both descriptive and causal health disparities research, offer intuitive definitions and estimands, and explain how selection bias can result in misleading conclusions regarding magnitude and drivers of disparities. Throughout, we provide specific examples from research on disparities in Alzheimer’s disease and Alzheimer’s

disease-related dementias to highlight how selection bias can be introduced via study design, recruitment approaches, differential survival and retention, and analytic choices. We briefly discuss how these processes have emerged in dementia research and strategies to avoid or remediate selection bias in both primary data collection and secondary data analyses. We end with priorities for future research directions to minimize selection bias in disparities research.

Health Disparities Estimands

Health disparities refer to systematic and plausibly avoidable health differences across socially constructed groups that are related to social hierarchy, marginalization, and disadvantage [7, 8]. Disparities arise from inequitable and unjust historical and contemporary social, economic, and political structures that benefit a dominant group and disadvantage others [9, 10••, 11, 12••]. Several topics fall under the umbrella of health disparities research, including describing the health of marginalized populations, quantifying disparities in health, evaluating the extent to which health disparities vary by place or over time, understanding the drivers of disparities, identifying promising targets for interventions

✉ L. Paloma Rojas-Saunero
lp.rojassaunero@ucla.edu

M. Maria Glymour
maria.glymour@ucsf.edu

Elizabeth Rose Mayeda
ermayeda@ph.ucla.edu

¹ Department of Epidemiology, University of California, Los Angeles Fielding School of Public Health, Los Angeles, CA, USA

² Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA

³ Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

to promote health equity, and evaluating effects of interventions on health equity [9, 13, 14, 15••]. Health disparities research thus includes both descriptive studies and studies where causal inference is the goal.

Defining the target estimand—the parameter of interest in the target population—is necessary for both descriptive and causal research questions. Bias is defined with respect to a specific estimand, i.e., the expected divergence between the estimate derived in a study and the desired estimand. A clear *causal* estimand should specify the target population, the hypothetical interventions to be contrasted, the outcome, and the statistical summary used to compare the counterfactual outcome distributions (e.g., risk ratio and risk difference) [16, 17]. *Descriptive* disparities estimands must specify the target population and social groups to be contrasted, if that is the aim, even if that contrast is not to be interpreted causally. In both causal and descriptive research, the target population is the group of people for whom the research aims to support inferences, anchored in a place and time. In longitudinal studies, the estimand should reflect processes such as mortality or loss to follow-up that occur during the study [18••].

We acknowledge the debate about whether disparities research should be conceptualized as causal or descriptive [14, 15••, 19–22]. Many researchers object to invoking causal language in disparities research. This critique aims to emphasize the role of structural forces and upstream factors (such as racism or sexism) rather than individual identities (such as race or sex and gender) in creating disparities. These structural phenomena can be challenging to conceptualize within the limits of target trial emulation. We consider that social factors can be defined as exposures/interventions in causal research, but must be contextualized by understanding structural processes [23–25]. The apparent “effects” of individual identities, such as racialized groups, are contingent on the structural processes that create social hierarchies, such as racism. Structural racism does not affect everyone equally but has been constructed and sustained because it differentially harms some groups to the benefit of others [9, 10••, 26]. Conceptualizing inequities as emerging due to causal processes will help identify opportunities to dismantle such structures.

Recent scholarship offers a detente in this debate, first by rejecting the narrowest notion of counterfactual frameworks that would allow causal language only for treatments amenable to randomization within a conventional trial framework [27]. Although simple interventions offer convenient teaching examples, the framework of potential outcomes can be applied to complex structural phenomena, even when trials of such phenomena are not feasible. Another important advance is from researchers including Jackson [14, 15••] and VanderWeele and Robinson [19, 20] who have offered disparities estimands that link to rigorous causal methods. Even when the target estimand is a descriptive measure of

disparities, it is often valuable to adopt causal reasoning and causal inference tools, such as directed acyclic graphs (DAGs), to understand potential sources of bias [5, 12••, 19, 28]. Therefore, the discussion for selection bias below is relevant for evaluating disparities estimands within both causal and descriptive frameworks.

Processes That Lead to Selection Bias

We define selection bias as any deviation between the target estimand (i.e., the parameter of interest in the target population) and the expected value of the estimate in the sample, if that deviation arises due to the processes by which observations are included in the sample. Selection bias can affect internal and external validity. Modern frameworks for selection bias emphasize two phenomena, which we refer to as *collider-stratification bias* (*collider bias*) and *generalizability bias* [1•, 2•, 3]. Collider-stratification bias can lead to incorrect inferences about the people in the sample (internal validity). Generalizability bias can lead to inferences that may be correct for individuals in the sample but are not correct for the people about whom the researcher is attempting to draw inferences (external validity).

In Fig. 1 we present three DAGs to ground conceptual definitions around selection bias. In these DAGs, R represents the social groups to be contrasted or an exposure/intervention affecting a social mechanism that creates or targets health disparities. Y represents the outcome of interest. In graphs A and B, S represents selection (inclusion) into the analytic sample, either due to study recruitment or retention in longitudinal studies. In graph A, we do not include an arrow between R and S ; *the lack of an arrow* reflects the assumption that mechanisms for sample selection are the same across social groups [29]. Still, generalizability bias emerges when the distribution of effect measure modifiers for Y , represented as node LI , differs between the sample and target population (indicated by the arrow from LI to S) [1•, 3, 30, 31]. When causal inference is the goal, generalizability bias will not be present if the effect of R on Y (on the chosen scale) is identical for all individuals in the population, for example, if there is no effect of R on Y for any individual in the target population. Effect modification depends on the scale of the effect measure, e.g., whether the effect measure is additive versus multiplicative [32]. Thus, if R affects Y and the distribution of *any* other strong determinant of the outcome of interest differs between the sample and target population, the results may not generalize to the target population on at least one scale. The same concern degrades generalizability in descriptive research: if major determinants of the outcome differ in distribution between the sample and target population, results may not generalize to the target population [30].

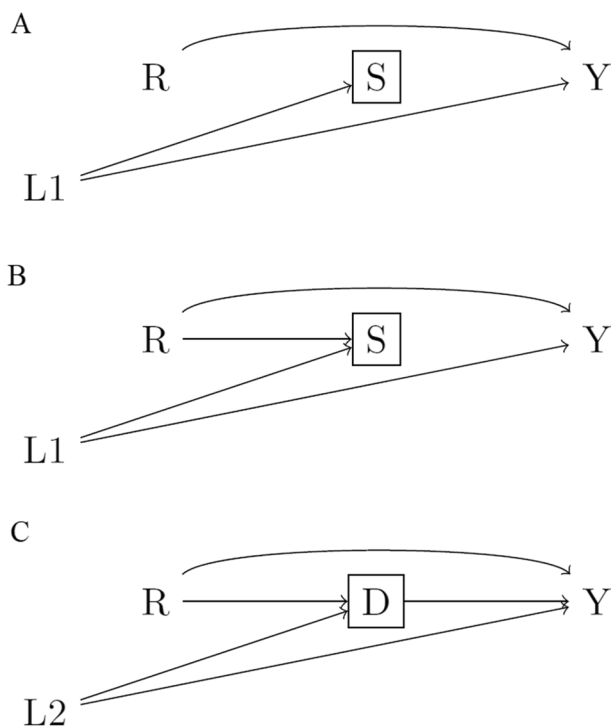


Fig. 1 Directed acyclic graphs representing selection mechanisms. *R* represents the social groups to be contrasted or an exposure/intervention around a social mechanism that creates or targets health disparities across social groups. *Y* represents the outcome of interest. In graphs **A** and **B**, *S* represents selection (inclusion) into the initial sample or into remaining in the sample in longitudinal studies. In graphs **A** and **B**, *L1* represents shared causes of *S* and *Y*. In graph **A**, this results in different distributions of *L1* in the sample and target population, which threatens generalizability. In graph **B**, where *R* influences selection into the sample, *S* becomes a collider and conditioning on $S=1$ creates an indirect pathway that induces a spurious association between *R* and *Y*. Graph **C** represents a scenario where *D* is death, a competing event that precludes the outcome of interest. The arrow between *D* and *Y* represents a special feature of competing events: when $D=1$, probability of *Y* is zero at future time points. In this scenario, *L2* represents shared causes of *D* and *Y*. Depending on the target estimand, the indirect pathway mediated by death either represents a mechanism included in the estimand (for a total effect) or a mechanism that induces selection bias (for a direct effect)

Collider-stratification bias emerges if selection into the sample is influenced by (1) the exposure or a determinant of the exposure *and* (2) the outcome or a determinant of the outcome [16, 33]. For example, if sample recruitment or retention is differential across social groups, we include an arrow between *R* and *S*, making *S* a collider, as observed in graph **B**. Conditioning on *S* creates an indirect pathway by which *R* is associated with *Y*. Unlike generalizability bias, collider-stratification bias can occur even if *R* does not have an effect on *Y* for anyone in the target population (i.e., the sharp null). Although conditioning can happen from sample restriction, stratification on a covariate, or adjustment for a covariate, in this work we refer to restricting to people

who are included in the initial sample or who are retained in longitudinal studies. In graph **C**, we make a particular distinction for competing events, which we discuss below.

In the following sections, we describe examples that connect to these graphical representations of selection bias under two broad processes: who is initially included in a study and who is retained in longitudinal studies.

Who Is Included in the Initial Sample?

Selective enrollment can result in generalizability bias and collider-stratification bias. Imagine a descriptive study aiming to estimate dementia incidence over a specified time period in the California population of older adults (age 65 years or older) across racial/ethnic groups and to estimate the incidence rate ratio between individuals who belong to a marginalized versus a privileged racialized or ethnic group, as a descriptive disparities estimand. Ideally, these estimates would be based on a sample drawn with known sampling probabilities from the target population. However, researchers must often rely on data sources in which the selection processes (sampling probabilities) are unknown. For example, Mayeda et al. used data from Kaiser Permanente Northern California (KPNC) health plan members to estimate dementia incidence across five racial/ethnic groups [34]. KPNC is advantageous for disparities research because the large and diverse membership enables research on multiple racial/ethnic groups living in northern California and health records provide high-quality clinical information. However, KPNC members may systematically differ with respect to major determinants of dementia, such as socioeconomic status (node *L1*), from the California population. Although selection into KPNC does not directly influence dementia incidence, socioeconomic status of participants influences both KPNC membership and dementia risk, creating an association between KPNC membership and dementia incidence as in graph **A**. This association can translate into generalizability bias for estimates of dementia incidence for the target population of older Californians. Note that immigration plays an important role in defining the target population for some racial/ethnic groups. For example, the target population of Asians in California represents differential immigration selection processes by Asian ethnicity [35].

If inclusion in KPNC differs by race/ethnicity (represented graphically by an arrow from *R* to *S*) and is influenced by other factors that affect dementia risk, this would create collider-stratification bias, as in graph **B** [33, 36, 37]. The bias will be larger if the strength of selection into KPNC by causes of dementia (*L1*) differs by race and ethnicity. This scenario is common in observational studies. Studies have shown that national dementia research registries have differential recruitment sources for Black and White participants. For example, White participants are recruited more

frequently from memory clinics, while Black participants are more often recruited through community-based strategies [38]. A recent systematic review showed that barriers to research participation in dementia research among marginalized social groups included fear of injury, mistrust of research or medical staff, insufficient information about the study, and geographic accessibility [39••]. A qualitative study described the awareness of the history of racism in health research in Black, Latinx, and Chinese populations in the US as a potential reason for not participating in dementia research studies and brain donation [40]. The intensive recruitment strategies fielded to address the lack of diversity in dementia research studies may lead to important biases when White participants and participants from minoritized groups are recruited with different outreach strategies.

Extensive literature has discussed implications of collider-stratification bias for causal effect estimation, but it has received less attention in the context of *describing* disparities. Defining a descriptive estimand and using causal diagrams to illustrate potential selection mechanisms clarify the relevance of collider-stratification bias for describing disparities. Selection processes may often be layered, so careful consideration of the target population for inferences is necessary. Studies of outcomes among people with a disease or condition—for example, racial disparities in post-stroke mortality or progression of pain in people with osteoarthritis—face a double-selection process. Selection into having the disease is intrinsic to the definition of the target population: people who have not had a stroke cannot be part of the target population for inferences on post-stroke mortality. In this case, the development of the disease or condition is a potential collider, and all analyses are conditioned on that collider. If racialized group and another factor *both* influence development of the disease, the disadvantaged racialized group may have *lower* prevalence of the other factor among people with the condition. A meaningful estimate of disparity should account for this bias, for example, by standardizing the distribution of the other variable so it is equivalent between racialized groups among the patient population. Jackson outlines considerations for deciding which factors should be included in such adjustments depending on how the disparity estimate is to be used [15••, 41].

Who Remains in the Study?

Loss to follow-up is a frequent challenge in longitudinal studies. Both descriptive and causal research typically aim to answer questions about the people who are initially included in the study. If loss to follow-up is influenced by determinants of the outcome, the follow-up sample is no longer representative of the initial sample. A simple version of this scenario is represented in graph A, where $S = 1$ represents remaining in the study and retention is unaffected by racialized group.

Many phenomena may cause loss-to-follow up, such as long or uncomfortable study visits, upfront costs of participation related to transportation or lost work (even if later reimbursed), alienating interactions with staff, or deterioration in the sense of value of the study [42, 43]. To generalize estimates (e.g., the cumulative incidence of dementia at 10 years of follow-up) to those in the initial sample, researchers have to rely on an exchangeability assumption for censoring. This means that the (unknown) dementia risk of participants lost to follow-up is similar to the observed dementia risk of individuals who remained in the study, conditional on measured covariate [18••, 44]. If loss to follow-up is also differential across social groups to be contrasted, as in graph B, this would result in collider-stratification bias. In this setting, the measure of disparity (e.g., the 10-year risk difference of dementia comparing two social groups) may be biased in either direction, depending on differences in participants who are lost to follow-up and those who remain.

Methods to satisfy the exchangeability assumption for censoring, such as inverse probability of censoring weights, have been developed for causal questions [45]. In this setting, the estimand reflects the *joint* effect of the exposure/intervention of interest *and* an intervention to retain everyone throughout follow-up. This is also relevant for descriptive estimands, so the estimand of interest may include a descriptive contrast with a causal component for preventing loss to follow-up. The causal component of the estimand relies on its own identifiability assumptions, including exchangeability, positivity, and consistency [18••, 44].

In contrast to loss to follow-up, competing events or truncating events, such as death prior to dementia diagnosis, usually cannot be prevented by study design. For time-to-event outcomes (e.g., time to dementia diagnosis), we refer to death as a competing event; for continuous outcomes (e.g., trajectories of cognitive change), we refer to death as a truncation event. When these events occur, the target estimand should account for them. In recent years, increasing attention has focused on what questions can be answered and the implications of different estimands for this setting, especially for causal questions [18••, 46–48]. Consider if graph C represents a research question about the effect of an intervention intended to reduce disparities in the cumulative incidence of dementia. Participants may die prior to dementia onset, so we represent death as node D . We additionally include shared causes for death and dementia as $L2$. This DAG includes an arrow from D to Y , which reflects the deterministic relationship between these variables: after death occurs, the probability of dementia at future time-points becomes zero [18••]. Assuming that the intervention also influences mortality, we include an arrow between R and D .

There is no single correct way to “account” for competing events. Choosing between estimands requires weighing interpretability and relevance for clinical and public health

decisions against the plausibility of identifying assumptions [18••, 48, 49]. One estimand of interest in this setting is a *controlled direct effect* of R on Y setting D to zero (treating death as a censoring event), i.e., the joint effect of the treatment/exposure and an intervention to prevent death during the study follow-up, just as in the setting of loss-to-follow up. Bias can emerge when we are interested in a direct effect of R on Y if we fail to satisfy the exchangeability assumption for censoring.

Alternatively, we may be interested in the *total effect* of R on Y ; in this case, the pathway mediated by death is not a source of bias but is conceptualized as an indirect effect of R on Y [18••]. If the exposure/treatment or causes of the exposure/treatment affect death, the total effect can reflect counterintuitive results (as discussed for the interpretation of subhazard ratios using Fine-Gray models [50]). For example, the total effect for smoking cessation may increase dementia risk only because it decreases mortality risk [49]. Recent scholarship has offered new causal estimands such as the *separable effects* [46]. Separable effects are relevant when it is possible to conceptualize the exposure as having distinct components; one of which influences the competing event, and the other influences the risk of dementia. We can then ask about effects of intervening on only one of these components [46]. These estimands have been extended to be implemented for continuous outcomes [47].

Another estimand is the *survivor average causal effects* (SACE), which considers a target population of individuals who would survive at least until the moment when the outcome is assessed regardless of treatment/exposure status or, in this case, regardless of social group membership [51]. The “always survivors” are conceptualized as a principal stratum of individuals to be distinguished from the “never survivors,” “survive only if treated,” or “survive only if not treated.” These labels are slight misnomers because the estimand does not require the existence of immortal individuals—it merely classifies individuals based on their potential survival up to the moment when the outcome of interest is to be evaluated. The SACE might be appealing in that the effect of a treatment on dementia risk among the “always survivors” seems more relevant than treatment effects on people in the other principal strata. A critique of the SACE is that it is impossible to know which people are “always survivors,” i.e., the group is unidentifiable [52]. Nevo et al. recently introduced new estimands based on principal stratification, offering alternative assumptions for partial identification bounds, using applied examples in the field of dementia [53].

For descriptive purposes, we recommend extending the notions of causal estimands with competing events. For example, if the aim is to describe risk of dementia at age 80 in a birth cohort of different social groups and participants die for other reasons without having developed dementia, one might prefer the cause-specific cumulative incidence

(crude risk) since this estimand reflects the fact that it is impossible to develop dementia after death [18••]. However, the cause-specific cumulative incidence may be lower in one group simply because the competing event is more common in that group. With this in mind, if the estimand is a descriptive disparity measure and we estimate the difference between the cause-specific risk of dementia for both social groups, we are allowing that the difference is partially explained by differential mortality distributions across social groups (as for a total effect estimate). Alternatively, if we contrast the marginal risk (net risk) of dementia and treat death as a censoring event, eliminating the path from D to Y would mean that we are conceptualizing a scenario where some intervention could have prevented death among all participants over follow-up (as for a direct effect estimate). This is similar to the idea adopted for loss to follow-up, but prevention of mortality throughout follow-up is rarely plausible. Nonetheless, such estimands are often more relevant for identifying causal mechanisms linking social inequalities to specific health events, mechanisms which may be obscured by evaluating total effects.

Selective Survival on Recruitment: When Two Biases Collide

There will be many cases where it is challenging to distinguish between generalizability bias and collider-stratification bias, especially in cross-sectional studies. However, it is not necessarily important to distinguish between these two forms of selection bias in a given study [54]. Specifying a clear question with a defined target population is the first step towards achieving target validity.

For example, in dementia research, cohort enrollment often conditions on survival to older ages. An example is the study of centenarians in the Netherlands, which identified Dutch citizens aged 100 years or older in 2017 as the target population [55]. If the aim is purely descriptive, for example, to estimate prevalence of dementia in the Dutch population aged 100 years or older, then there would be no source of selection bias. However, if the aim is to study the effect of an exposure or intervention in midlife on dementia risk at age 100, selecting on those who survived up to 100 years old may create collider-stratification bias. Mayeda et al. illustrated the potential magnitude of bias when estimating effects of education on later-life rate of cognitive decline under a range of assumptions about the determinants of survival to later life [37]. Comparing multiple causal scenarios, they demonstrated the potential for substantial bias if survival is influenced by both education and other determinants of cognitive decline and the influence of education and the other risk factor interact to influence mortality. For example, if higher education doubles odds of survival among people without the other risk factor but *more than doubles* odds of

survival among people with the other risk factor, education and the other risk factor will become increasingly strongly associated among survivors at older ages. If the other risk factor is not controlled for, this can be an important bias in analyses of the effects of education on cognitive aging among older adults. In settings where the causes have perfectly multiplicative effects on survival or fairly small effects on survival, the magnitude of collider-stratification bias is likely small [31, 33, 56–58].

Exacerbating Disparities due to Selective Sampling

The lack of diversity in study samples is a predominant concern about selection bias in dementia research and is relevant across health research domains. A landmark Lancet Commission review on dementia prevention illustrates the extremely limited racial/ethnic diversity in studies driving the current evidence base in dementia research [59]. This review identified several top-priority dementia risk factors, including alcohol use and hypertension control. Evidence on the effects of alcohol use was based on predominantly European studies with limited racial/ethnic diversity, such as from the Whitehall cohort of British civil servants, the French National Hospital Discharge database, and UK Biobank [59]. Evidence on hypertension control includes evidence from a prior meta-analysis using individual data from six cohort studies; in combination, this study included 18,967 non-Latinx White individuals, 31 Black individuals, and 2646 Japanese American men (from a single study) [60]. The Lancet Commission review contributed to the World Health Organization guidelines for dementia prevention. These results reflect how much of the scientific knowledge in dementia research used to develop worldwide guidelines is derived from non-Latinx White populations from the global north. Current evidence on more diverse samples is so scant; we simply do not know if results from these samples are generalizable to all older adults. Prioritizing risk factors evaluated in such a selected fraction of the human population could potentially exacerbate disparities in dementia. Any risk factors that are distinctively relevant for different ethnicities and social groups across the world would simply not be detectable in many major dementia studies [61–63].

The lack of inclusion in research extends to randomized trials, which often have minimal racial/ethnic diversity or representation of marginalized populations. For example, despite the national guidelines to include minoritized samples, the US Food and Drug Administration (FDA) approval for aducanumab highlights the systematic social exclusion of marginalized populations in research. The randomized trials informing the FDA approval only

included 0.6% Black participants—the data submitted to the FDA for drug approval included only 6 Black individuals who received the treatment ultimately approved by the FDA—providing no evidence of the safety or efficacy of this drug for this population (or any other marginalized racial/ethnic group) [64]. Randomization resolves potential collider-stratification bias at enrollment, but does not ameliorate concerns about effect heterogeneity and lack of generalizability. As a consequence of life-course exposure to racism, Black older adults have higher prevalence of many comorbidities, including diabetes and cardiovascular disease, than White older adults living in the US [65, 66]. These comorbidities could result in heterogeneity in benefits and harms of the medication; ignoring such heterogeneity may increase disparities in dementia. Phase III trial results for another medication approved by the FDA for people with early Alzheimer's disease (lecanemab) were published including 20 Black participants who received active treatment (of 859 total treated participants) [67]. Careful assessment of the processes driving lack of inclusion in the Anti-Amyloid in Asymptomatic Alzheimer's Disease (A4) trial implicated several common eligibility criteria and outreach approaches that were not essential, such as requiring a study partner [68, 69]. One troubling observation from this work is that the differences in recruitment strategies for Black compared to White participants imply that Black trial participants are likely to be less representative of the population of Black older adults than White trial participants are of White older adults.

Prior targeted recruitment efforts to compensate for lack of inclusion have fallen short, introducing new challenges. An early illustration of this occurred with the Atherosclerosis Risk in Communities (ARIC), which enrolled older adults in three predominantly White communities and included a fourth community (Jackson, Mississippi) in which only Black individuals were eligible [70]. This design—in which nearly all Black participants reside in a single recruitment city where no White participants live—makes it impossible to disentangle the effects of racialization and geographic context. Similar examples continue to emerge, especially as pressure for cohorts to diversify increases. For example, intensive outreach in predominantly Black neighborhoods or community settings may be used to enrich a cohort of mostly White individuals recruited from clinical settings. With the wisdom of 30 years of hindsight, it is clearly problematic to envision a cohort of nearly entirely White participants and then attempt to correct the cohort by recruiting only non-White individuals. We contrast this with research designs such as the Health and Retirement Study (HRS) [71], the Washington Heights Inwood Columbia Aging Project [72], and the Reasons for Geographic and Racial Disparities in Stroke (REGARDS)

cohort [73], in which the target populations from inception were both geographically and racially diverse.

To remediate egregious evidence gaps, studies that only include members of marginalized groups have emerged, such as the Black Women's Health Study [74] and the Hispanic Community Health Study/Study of Latinos–Investigation of Neurocognitive Aging (SOL-INCA) [75]. These studies ameliorate the urgent need for evidence on determinants of health among non-White groups. The long-history of essentially exclusively White research studies adds a sense that equity demands exclusively Black, Latinx, Indigenous, or Asian American studies. Focusing on a specific group may allow studies to improve recruitment and retention success. For example, the Black Women's Health Study launched by recruiting Black women readers of *Essence* magazine [74]. These studies are ideal for centering research on the priority group and often include many more individuals from the traditionally excluded group than multiethnic studies.

In addition to the value of studies that enroll only members of traditionally excluded groups, such studies also present a disadvantage because it is not possible to differentiate study design features from features of the traditionally excluded group. This makes it more difficult to evaluate whether prior research in predominantly white populations is relevant for the traditionally excluded group. Learning which past findings generalize to more diverse populations is a high priority in disparities research, and evaluating this in single-race studies requires strong assumptions about study design or enrollment strategies not contributing to apparent differences between study findings. Thus, researchers planning studies should weigh advantages and disadvantages of studies restricted to one racial or ethnic group with their specific research goals in mind.

Considerations to Prevent and Assess Selection Bias

As we have described, selection processes can introduce biases in measures of disparities or effects of interventions on disparities. Accurate information on the magnitude and drivers of disparities is essential to successfully pursue health equity, so selection bias must be addressed throughout the development of study design and statistical analysis. To prevent selection bias, it is essential that the target population is clearly specified. Once the target population is defined, primary data collection should be performed in a way that ensures accessibility for participants who are often marginalized. In most cases, a higher sampling fraction for members of marginalized groups will be needed to achieve adequate precision to evaluate drivers of health within the marginalized group and to test effect heterogeneity. Ideally, all social groups should be recruited from the same source, rather than creating a distinct

recruitment pipeline that draws from different populations to achieve diversity.

Studies that require longitudinal follow-up should design retention strategies similar to those for recruitment and define how potential competing events will be handled. Ongoing monitoring of attrition rates should allow researchers to implement retention measures to prevent differential loss to follow-up. With respect to potential competing events, researchers must choose between estimands, based on the question of interest. If the association between the intervention/exposure of interest and competing events is not strong, or the competing event is not common, different estimands may not diverge substantially. In some cases, it may be useful to estimate multiple estimands.

When either recruitment or retention processes lead to important differences between the sample and the target population, statistical tools for generalizing or transporting findings may still allow for inferences about the target population [76–79]. These methods typically borrow information from other data sources. Many methods have focused on transport from randomized trials to observational studies but they are equally applicable when transporting causal inferences from observational data to new samples. For example, a recent study on the prevalence of cognitive impairment in California residents used data from a cohort study (Kaiser Healthy Aging and Diverse Life Experiences Study) designed specifically to have approximately equal representation of Asian, Black, Latinx, and White older adults. Race/ethnicity-specific cognitive impairment prevalence estimates were weighted to ultimately extend findings to the California population of older adults represented by a population-representative sample (California Behavioral Risk Factor Surveillance System). The weighted prevalence for each racial/ethnic group varied in magnitude compared to unweighted results [80••]. Successful implementation of these methods requires deep understanding of potential effect measure modifiers and measures of the modifiers in both the analytic sample and the population-representative data sources.

Even when formal transportability tools cannot be adopted, combining information across data sets can be informative. To overcome the limitations on study samples that informed the Lancet Commission [59] review on modifiable risk factors for dementia, two recent articles use different US-nationally representative samples to obtain sex- and race/ethnicity-specific population attributable fraction estimates for the US population [81, 82].

Conclusions

Selection bias is pervasive and can lead to incorrect inferences about the extent to which disparities exist, the extent to which they vary across time or location, or about the

efficacy of alternative strategies to promote health equity. The consequences of selection bias have gained more attention in recent years, and strategies to avoid or rectify such biases are proliferating. Careful conceptualization of the target population, the target estimand, the recruitment and retention strategies focused on equitable inclusion, and analytic methods to correct selection processes are all important steps to avoiding selection bias in health disparities research and ultimately conducting meaningful research to promote health equity.

Funding This work was supported by the National Institutes of Aging (NIA) (grant numbers: R01AG074359; R01AG063969; R56AG069126; R01AG057869; R01AG052132).

Declarations

Conflict of Interest The authors declare no conflict of interest.

Human and Animal Rights and Informed Consent 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.● Lu H, Cole SR, Howe CJ, Westreich D. Toward a clearer definition of selection bias when estimating causal effects. *Epidemiology*. 2022;33:699–706. **Unifies and classifies the numerous existing definitions of selection bias.**
- 2.● Degtiar I, Rose S. A review of generalizability and transportability. *Annu Rev Stat Appl*. 2023;10(1):501–24. **Provides an extensive review for addressing external validity, discusses underlying assumptions for generalizability and transportability methods.**
3. Smith LH. Selection mechanisms and their consequences: understanding and addressing selection bias. *Curr Epidemiol Rep*. 2020;7:179–89.
4. Boone SC, le Cessie S, van Dijk KW, de Mutsert R, Mook-Kanamori DO. Avoiding selection bias in metabolomics studies: a tutorial. *Metabolomics*. 2019;15:7.
5. 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;29:521–4.
6. Infante-Rivard C, Cusson A. Reflection on modern methods: selection bias—a review of recent developments. *Int J Epidemiol*. 2018;47:1714–22.
7. Braveman P. Health disparities and health equity: concepts and measurement. *Annu Rev Public Health*. 2006;27:167–94.
8. Braveman PA, Kumanyika S, Fielding J, LaVeist T, Borrell LN, Manderscheid R, et al. Health disparities and health equity: the issue is justice. *Am J Public Health*. 2011;101:S149–55.
9. Bailey ZD, Krieger N, Agenor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. *Lancet*. 2017;389:1453–63.
- 10.●● Bailey ZD, Feldman JM, Bassett MT. How structural racism works — racist policies as a root cause of U.S. racial health inequities. Malina D, editor. *New England J Med*. 2021;384:768–73. **Discusses historical systemic processes that have contributed to structural racism and its consequences in people’s health. Highlights future directions to dismantle racism in health research.**
11. Martinez RAM, Andrabi N, Goodwin AN, Wilbur RE, Smith NR, Zivich PN. Conceptualization, operationalization, and utilization of race and ethnicity in major epidemiology journals 1995–2018: a systematic review. *Am J Epidemiol*. 2023;192(3):483–96.
- 12.●● Howe CJ, Bailey ZD, Raifman JR, Jackson JW. Recommendations for using causal diagrams to study racial health disparities. *Am J Epidemiol*. 2022;191(12):1981–9. **Provides several recommendations that translate into clearer research questions and study design decisions to prevent sources of bias in racial health disparities, using causal diagrams.**
13. Kaufman JS. Epidemiologic analysis of racial/ethnic disparities: some fundamental issues and a cautionary example. *Soc Sci Med*. 2008;66:1659–69.
14. Jackson JW, VanderWeele TJ. Decomposition analysis to identify intervention targets for reducing disparities. *Epidemiology*. 2018;29:825–35.
- 15.●● Jackson JW. Meaningful causal decompositions in health equity research: definition, identification, and estimation through a weighting framework. *Epidemiology*. 2021;32:282–90. **Provides causal estimands that address health disparities and conceptual ideas around the equity value judgments for choosing confounders (allowability).**
16. Hernán MA, Robins JM. *Causal inference: what if*. Boca Raton: Chapman & Hall/CRC; 2020.
17. ICH E9 (R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. EMA/CHMP/ICH/436221/2017.
- 18.●● 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;39:1199–236. **Formalizes the causal estimands for competing events that correspond to common statistical methods in survival analysis used in research. It outlines the underlying assumptions that allow identification of such, with directed acyclic graphs and single intervention world graphs.**
19. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*. 2014;25:473–84.
20. VanderWeele TJ, Robinson WR. Rejoinder: how to reduce racial disparities? *Epidemiology*. 2014;25:491–3.
21. Holland PW. *Statistics and causal inference*. Journal of the American Statistical Association. Taylor Francis. 1986;81:945–60.
22. Jackson JW, Arah OA. Making causal inference more social and (social) epidemiology more causal. *Am J Epidemiol*. 2020;189:179–82.
23. Breilh J. Latin American critical (“social”) epidemiology: new settings for an old dream. *Int J Epidemiol*. 2008;37:745–50.
24. Schwartz S, Prins SJ, Campbell UB, Gatto NM. Is the “well-defined intervention assumption” politically conservative? *Soc Sci Med*. 2016;166:254–7.
25. Krieger N, Davey Smith G. The tale wagged by the DAG: Broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol*. 45(6):1787–808.

26. Churchwell K, Elkind MSV, Benjamin RM, Carson AP, Chang EK, Lawrence W, et al. Call to action: structural racism as a fundamental driver of health disparities: a presidential advisory from the American Heart Association. *Circulation*. 2020;142:e454–68.
27. Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol*. 2016;26:674–80.
28. Robinson WR, Bailey ZD. Invited Commentary: What social epidemiology brings to the table—reconciling social epidemiology and causal inference. *Am J Epidemiol*. 2020;189:171–4.
29. Cinelli C, Pearl J. Generalizing experimental results by leveraging knowledge of mechanisms. *Eur J Epidemiol*. 2021;36:149–64.
30. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. *Epidemiology*. 2017;28:553–61.
31. Hernán MA. Invited Commentary: Selection bias without colliders. *Am J Epidemiol*. 2017;185:1048–50.
32. Huitfeldt A, Swanson SA, Stensrud MJ, Suzuki E. Effect heterogeneity and variable selection for standardizing causal effects to a target population. *Eur J Epidemiol*. 2019;34:1119–29.
33. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–25.
34. Mayeda ER, Glymour MM, Quesenberry CP, Whitmer RA. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. *Alzheimers Dement*. 2016;12:216–24.
35. Hayes-Larson E, Fong J, Mobley TM, Gilsanz P, Whitmer RA, Gee GC, et al. The role of nativity in heterogeneous dementia incidence in a large cohort of three Asian American groups and white older adults in California. *Alzheimer's & Dementia*. John Wiley & Sons, Ltd; 2022;18:1580–5.
36. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–6.
37. Mayeda ER, Filshtein TJ, Tripodis Y, Glymour MM, Gross AL. Does selective survival before study enrolment attenuate estimated effects of education on rate of cognitive decline in older adults? A simulation approach for quantifying survival bias in life course epidemiology. *Int J Epidemiol*. 2018;47:1507–17.
38. Gleason CE, Norton D, Zuelsdorff M, Benton SF, Wyman MF, Nystrom N, et al. Association between enrollment factors and incident cognitive impairment in Blacks and Whites: data from the Alzheimer's Disease Center. *Alzheimer's Dementia*. 2019;15:1533–45.
- 39.●● Gilmore-Bykovskiy AL, Jin Y, Gleason C, Flowers-Benton S, Block LM, Dilworth-Anderson P, et al. Recruitment and retention of underrepresented populations in Alzheimer's disease research: a systematic review. *Alzheimers Dement (N Y)*. United States; 2019;5:751–70. **Identifies common limitations in the current practice of recruitment/retention approaches in dementia research and provides recommendations for improvement recruitment science.**
40. Boise L, Hinton L, Rosen HJ, Ruhl M. Will my soul go to heaven if they take my brain? Beliefs and worries about brain donation among four ethnic groups. *Gerontologist*. 2017;57(4):719–34.
41. Jackson JW, Hsu Y-J, Greer RC, Boonyasai RT, Howe CJ. The observational target trial: a conceptual model for measuring disparity [Internet]. arXiv; 2022. Available from: <https://arxiv.org/abs/2207.00530>
42. Grill JD, Kwon J, Teylan MA, Pierce A, Vidoni ED, Burns JM, et al. Retention of Alzheimer disease research participants. *Alzheimer Dis Assoc Disord*. 2019;33:299–306.
43. Salazar CR, Ritchie M, Gillen DL, Grill JD. Strategies associated with retaining participants in the longitudinal National Alzheimer's Coordinating Center uniform data set study. *JAD*. 2022;87:1557–66.
44. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ. Selection bias due to loss to follow up in cohort studies. *Epidemiology*. 2016;27:91–7.
45. 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;56:779–88.
46. Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán MA. Separable effects for causal inference in the presence of competing events. *J Am Stat Assoc*. 2020;1–9.
47. Stensrud MJ, Hernán MA, TchetgenTchetgen EJ, Robins JM, Didelez V, Young JG. A generalized theory of separable effects in competing event settings. *Lifetime Data Anal*. 2021;27:588–631.
48. Young JG, Stensrud MJ. Identified versus interesting causal effects in fertility trials and other settings with competing or truncation events. *Epidemiology [Internet]*. 2021;32(4):569–72.
49. Rojas-Saunero LP, Young JG, Didelez V, Ikram MA, Swanson SA. Considering questions before methods in dementia research with competing events and causal goals. *Am J Epidemiol*. 2023;kwad090.
50. Austin PC, Fine JP. Practical recommendations for reporting fine-gray model analyses for competing risk data. *Stat Med*. 2017;36:4391–400.
51. Tchetgen Tchetgen EJ. Identification and estimation of survivor average causal effects. *Stat Med*. 2014;33(21):3601–28.
52. Dawid P, Didelez V. “Imagine a can opener”--the magic of principal stratum analysis. *Int J Biostat*. Germany; 2012;8:19.
53. Nevo D, Gorfine M. Causal inference for semi-competing risks data. *Biostatistics*. 2022;23:1115–32.
54. Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target validity and the hierarchy of study designs. *Am J Epidemiol*. 2019;188:438–43.
55. Holstege H, Beker N, Dijkstra T, Pieterse K, Wemmenhove E, Schouten K, et al. The 100-plus study of cognitively healthy centenarians: rationale, design and cohort description. *Eur J Epidemiol*. 2018;33:1229–49.
56. Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol United States*. 1977;106:184–7.
57. Glymour MM, Vittinghoff E. Commentary: selection bias as an explanation for the obesity paradox: just because it's possible doesn't mean it's plausible. *Epidemiol United States*. 2014;25:4–6.
58. Mayeda ER, Banack HR, Bibbins-Domingo K, Zeki Al Hazzouri A, Marden JR, Whitmer RA, et al. Can survival bias explain the age attenuation of racial inequalities in stroke incidence?: A simulation study. *Epidemiology*. United States; 2018;29:525–32.
59. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 2020;396:413–46.
60. Ding J, Davis-Plourde KL, Sedaghat S, Tully PJ, Wang W, Phillips C, et al. Antihypertensive medications and risk for incident dementia and Alzheimer's disease: a meta-analysis of individual participant data from prospective cohort studies. *Lancet Neurol*. 2020;19:61–70.
61. Maestre G, Carrillo M, Kalaria R, Acosta D, Adams L, Adoukonou T, et al. The Nairobi Declaration—reducing the burden of dementia in low- and middle-income countries (LMICs): declaration of the 2022 Symposium on Dementia and Brain Aging in LMICs. *Alzheimer's & Dementia*. John Wiley & Sons, Ltd; 2023;19:1105–8.
62. Parra MA, Garcia AM, Ibanez Sr. A, LAC-CD. Addressing dementia challenges through international networks: evidence

- from the Latin American and Caribbean Consortium on Dementia (LAC-CD). *Alzheimer's & Dementia*. John Wiley & Sons, Ltd; 2021;17:e055106.
63. Barnes LL. Alzheimer disease in African American individuals: increased incidence or not enough data? *Nat Rev Neurol*. 2022;18:56–62.
 64. Manly JJ, Glymour MM. What the aducanumab approval reveals about Alzheimer disease research. *JAMA Neurol*. 2021;78:1305.
 65. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and trends in diabetes among adults in the United States, 1988–2012. *JAMA*. 2015;314:1021.
 66. Carnethon MR, Pu J, Howard G, Albert MA, Anderson CAM, Bertoni AG, et al. Cardiovascular health in African Americans: a scientific statement from the American Heart Association. *Circulation*. 2017;136:e393–423.
 67. van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M, et al. Lecanemab in early Alzheimer's disease. *N Engl J Med*. 2023;388:9–21.
 68. Raman R, Quiroz YT, Langford O, Choi J, Ritchie M, Baumgartner M, et al. Disparities by race and ethnicity among adults recruited for a preclinical Alzheimer disease trial. *JAMA Netw Open*. 2021;4:e2114364.
 69. Manly JJ, Gilmore-Bykovskiy A, Deters KD. Inclusion of underrepresented groups in preclinical Alzheimer disease trials—opportunities abound. *JAMA Netw Open*. 2021;4:e2114606.
 70. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol United States*. 1989;129:687–702.
 71. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: the Health and Retirement Study (HRS). *Int J Epidemiol*. 2014;43:576–85.
 72. Tang MX, Cross P, Andrews H, Jacobs DM, Small S, Bell K, et al. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. *Neurology*. 2001;56:49–56.
 73. Moise N, Khodneva Y, Jannat-Khah DP, Richman J, Davidson KW, Kronish IM, et al. Observational study of the differential impact of time-varying depressive symptoms on all-cause and cause-specific mortality by health status in community-dwelling adults: the REGARDS study. *BMJ Open*. 2018;8:e017385.
 74. Palmer JR, Cozier YC, Rosenberg L. Research on health disparities: strategies and findings from the Black Women's Health Study. *Am J Epidemiol*. 2022;kwac022.
 75. González HM, Tarraf W, Fornage M, González KA, Chai A, Youngblood M, et al. A research framework for cognitive aging and Alzheimer's disease among diverse US Latinos: design and implementation of the Hispanic Community Health Study/Study of Latinos-Investigation of Neurocognitive Aging (SOL-INCA). *Alzheimers Dement*. 2019/11/20 ed. United States; 2019;15:1624–32.
 76. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol*. 2010;172:107–15.
 77. Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *Proc Natl Acad Sci USA*. 2016;113:7345–52.
 78. Dahabreh IJ, Hernán MA. Extending inferences from a randomized trial to a target population. *Eur J Epidemiol*. 2019;34:719–22.
 79. Dahabreh IJ, Haneuse SJ-PA, Robins JM, Robertson SE, Buchanan AL, Stuart EA, et al. Study designs for extending causal inferences from a randomized trial to a target population. *Am J Epidemiol*. 2021;190:1632–42.
 - 80.●● Hayes-Larson E, Mobley TM, Mungas D, Seamans MJ, Glymour MM, Gilsanz P, et al. Accounting for lack of representation in dementia research: generalizing KHANDLE study findings on the prevalence of cognitive impairment to the California older population. *Alzheimer's & Dementia*. 2022;18:2209–17. **Implements inverse odds of selection weights to generalize prevalence estimates from a cohort study to the target population, and illustrates the change in the prevalence differences between race/ethnicity groups using this method.**
 81. Nianogo RA, Rosenwohl-Mack A, Yaffe K, Carrasco A, Hoffmann CM, Barnes DE. Risk factors associated with Alzheimer disease and related dementias by sex and race and ethnicity in the US. *JAMA Neurol*. 2022;79:584–91.
 82. Lee M, Whitsel E, Avery C, Hughes TM, Griswold ME, Sedaghat S, et al. Variation in population attributable fraction of dementia associated with potentially modifiable risk factors by race and ethnicity in the US. *JAMA Netw Open*. 2022;5:e2219672.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.