A Nearly Unavoidable Mechanism for Collider Bias with Index-Event Studies

W. Dana Flanders, a,b Ronald C. Eldridge, a and William McClellan a

Abstract: Factors suspected of causing certain chronic diseases and death are often associated with lower mortality among those with disease. For end-stage renal disease, examples include high cholesterol and homocysteine. Here, we consider obesity, thought to cause both end-stage renal disease and premature mortality, but which is associated with lower mortality among end-stage renal disease patients. Such seeming paradoxes could reflect collider (index event) bias due to selection of a diseased population for study. However, previous descriptions are incomplete, as they posit an uncontrolled factor causing both end-stage renal disease (the index event) and death. Here, we explicitly note that death can precede end-stage renal disease onset. The target population is obese persons with end-stage renal disease, effects of interest are seemingly controlled direct effects, the usual estimator is a conditional risk ratio, and remaining at risk until the onset of end-stage renal disease is a collider. Collider bias then expected if any mortality risk factor is uncontrolled, even if no factor also affects end-stage renal disease. The bias is similar to, but differs from, that associated with competing risks. Because control of every mortality risk factor is implausible, bias of the standard estimator is practically unavoidable. Better awareness of these issues by clinicians and researchers is needed if observational research is to usefully guide care of this vulnerable patient population.

(Epidemiology 2014;XX: 00–00)

Substantial evidence suggests that obesity causes both end-stage renal disease and premature mortality in unselected populations a–c but is typically associated with lower mortality among those with end-stage renal disease. d Such “reversed” associations have led some observers to suggest that the effect of obesity on mortality may differ between populations with and without end-stage renal disease. d–f

Similarly, reversed associations are commonly reported for other attributes and diseases. g–k Most studies in which such a reversal is found share certain basic characteristics. They usually involve selection of exposed and unexposed subjects with disease (the index event), l–m follow-up of those selected to estimate exposure-specific mortality risks for the post-index-event period, and comparison of these risks using a risk or rate ratio. For brevity, we refer to such studies as “index-event” studies.

Collider bias due to selection of a diseased population for study n–s can potentially explain the “reversed” associations often found in index-event studies. Collider bias posits that end-stage renal disease is caused by both obesity and an uncontrolled factor U, as in the causal graph t–v of Figure 1. End-stage renal disease is therefore a collider—a factor caused by 2 other factors. w–y As further described below, the observed risk ratio (RR) is then expected to be a biased effect estimator due to selecting on a collider.

However, previous considerations of bias in index-event studies are incomplete. There is also a second collider—survival until occurrence of end-stage renal disease. The importance of this collider is generally recognized t–x but overlooked in considerations of index-event studies. In these studies, bias can exist if any outcome risk factor is uncontrolled. This differs from previous conclusions concerning index-event studies wherein the uncontrolled factor(s) is posited to cause both the index event and death. Collider bias is well recognized t–x; that it is likely present in these studies even after controlling for every factor causing the index event directly is not. Bias is more likely than previously recognized.

FIGURE 1. Causal graph showing the usually recognized situation for collider bias. ESRD represents end-stage renal disease, the intermediate event; D represents death (presumably after disease onset [at age a]) and before age a); U* represents an uncontrolled risk factor with direct effects on both disease and death—the situation usually described for collider bias in this context.
Our goal is to highlight issues that affect obesity–end-stage renal disease–mortality and other index-event studies. Specifically, we define the causal effect of interest, note the relevance of considering time and temporal sequences, emphasize the importance of the additional collider, and describe biases of standard effect estimators. Despite challenges inherent in studying the effects of obesity on mortality, obesity merely illustrates the issues because the described biases are expected even if a well-defined point exposure had been randomized before occurrence of the index event.

BACKGROUND

In studies of obesity, end-stage renal disease, and mortality, the usual effect estimator is the observed hazard ratio or RR. It is calculated by selecting subjects with end-stage renal disease and comparing subsequent mortality among the obese with that among the nonobese.12 Most estimators are stratified by, or otherwise adjusted for, age at onset of end-stage renal disease to control potential confounding. Usually unspecified, the estimand is seemingly a controlled direct effect of obesity25 for the target population consisting of the obese with end-stage renal disease (eAppendix 1, http://links.lww.com/EDE/A800 provides additional details). Conceptually, this represents the effect of preexisting obesity on death after disease onset, controlling for age at disease onset. In a causal graph (Figure 1), the arrow from obesity (E) to death (D) represents a direct effect, and the arrows from obesity to ESRD (the “intermediate”) and from ESRD to death represent an indirect effect.26

PREVIOUSLY RECOGNIZED COLLIDER

Valid estimation of the controlled direct effect requires strong assumptions26—essentially no uncontrolled confounding of both the obesity–mortality and end-stage renal disease–mortality associations. Bias of controlled-direct-effect estimators like the conditional RR is expected if 2 conditions hold. First, obesity and another factor $U^*$ both affect end-stage renal disease, so that end-stage renal disease is a collider (Figure 1). Second, subjects are selected for study based on the collider. Such selection is normally built into index-event studies. Thus, if $U^*$ is uncontrolled and affects both end-stage renal disease and death, observed conditional RRs are expected to be biased (inconsistent) effect estimators.13,14,26

ANOTHER COLLIDER

Because death can precede the onset of end-stage renal disease, we now consider temporality and the possible sequences of events. Thus, we specifically consider preexisting obesity measured at age $a_0$ and the age stratum $a_1$ of subjects with disease onset at age $a_2$, where $a_1 > a_0$. Define $S = 1$ if selected for study, and 0 otherwise. We use subscripts to indicate age-specific death and reflect the sequence of events: $D_1$ and $D_2$ are dichotomous indicators of death before age $a_1$, and between $a_1$ and $a_2$, respectively. For age stratum $a_1$, selection is conditional on being at risk ($D_1 = 0$) and having end-stage renal disease occur at age $a_1$. We assume death is caused by obesity (the nonnull) and an additional factor $U$ (Figure 2).

Under these assumptions, $D_1$ is a collider and obesity tends to be associated with $U$ among those selected.20–22 Therefore, we expect standard effect estimators such as the conditional RR to be biased if $U$ is uncontrolled,14 reflecting well-known limitations of conditional risks.14,27 In contrast to previous descriptions, bias can exist even if end-stage renal disease is not a collider (eg, $U^*$ of Figure 1 is absent in Figure 2)—a deviation from the situation usually posited for collider bias in this context15 and in associated causal graphs.25

DISCUSSION

We emphasize here several conditions that have rarely been considered in discussions of studies of the relationships among obesity and end-stage renal disease, and mortality: a target consisting of exposed, at-risk, diseased subjects, the temporal sequence of events, particularly the possibility that the outcome can precede the index event; and the implications for bias. Careful consideration of temporal sequences in connection with causal relationships has been useful in other contexts, such as development of methods to address time-varying confounding28–30 and in the birth-outcomes literature.31,32, further evaluation might yield additional insights or solutions here.

Collider bias in index-event studies reflects general limitations of conditional risks.14,27 Although resembling potential biases associated with competing risks, this bias can exist for nonfatal outcomes and so can differ technically (“censoring” can occur through target selection alone, as discussed in eAppendix 1, http://links.lww.com/EDE/A800).

Although some clearly suspect controlled direct effect and related effect estimators will often be biased,33,34 recognition of this additional collider is important. It underscores the importance of considering temporal issues and implies that, to avoid bias in index-event studies, one must control or account for every mortality risk factor, not just the subset that

![FIGURE 2. Causal graph for age stratum $a_i$, reflecting temporality of events, and another collider ($S$). ESRD represents end-stage renal disease with onset at age $a_2$, the intermediate event; $D_1$ represents death before age $a_1$ and $D_2$ represents death between ages $a_1$ and $a_2$; $S$ represents selection into the study for age stratum $a_i$; $U$ represents an uncontrolled risk factor with direct effects only on $D_1$ and $D_2$. The arrow from $U$ to ESRD is omitted because $U$ has no direct effect on ESRD. The arrow from $D_1$ to ESRD is omitted because the bias described here can be present for nonfatal outcomes with no effect on the intermediate.](http://links.lww.com/EDE/A800)
causes both end-stage renal disease and death. This reinforces the suspicion that most studies of obesity–end-stage renal disease–mortality associations likely suffer from collider bias.

Preliminary work using accelerated life models with plausible parameters suggests that the bias is potentially large even if all factors directly affecting both end-stage renal disease and death are controlled. Furthermore, the bias can make a harmful effect appear beneficial, even if both the exposure and the index event were randomized (eAppendix 2, http://links.lww.com/EDE/A800). These biases could be avoided if exposure were randomized after end-stage renal disease onset or, in observational studies, if conditioning on survival could be avoided. Potential bias in these types of studies must be more widely recognized if results are to usefully guide care.

ACKNOWLEDGMENTS

We thank Miguel Hernán and Sander Greenland for their helpful comments on earlier versions of this manuscript.

REFERENCES

25. Banack HR, Kaufman JS. Perhaps the correct answer is: (D) all of the above. Epidemiology. 2014;25:7–9.