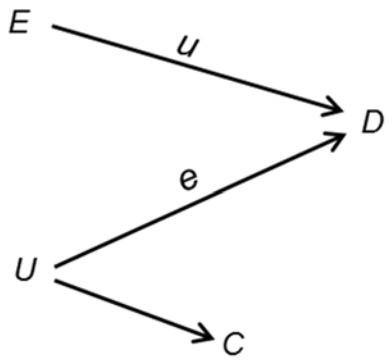
"Selection bias without colliders" was… effect-modification bias x 3

Effect modification

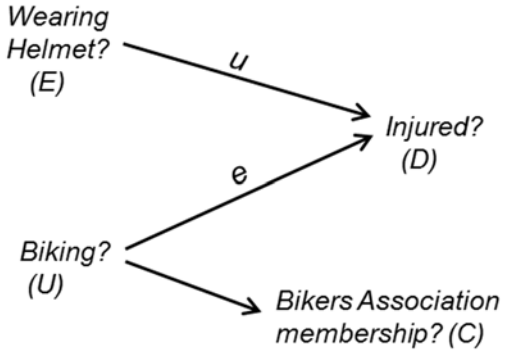
Figure 1 shows two causes of *D*, variables *E* and *U*, and notation for effect modification between them, say, on the probability ratio scale. The lower case "*u*" above *E*→*D* indicates dependence of the probability ratio on the value of *U*. Likewise, the lower case "*e*" above *U*→*D* indicates dependence on the value of *E*. (Effect modification is a symmetric property.[1,2]) Also shown is variable *C*, an effect of *U*. No confounders are in sight for *E*→*D* and *U*→*D*, but I hope it is okay to make up a simplistic story in the interest of teaching.

Figure 1.



To make the story lively, let's add explicit variables to the generic notation (Figure 2), all of which will be treated as binary (*no*, *yes*).

Figure 2.



Let me assume that you are interested in the modified effects. Or at least I am. For instance, I want to know the effect size of wearing a helmet ("no" or "yes") on injury when biking (presumably large), and the effect size of wearing a helmet when not biking (presumably small). That knowledge will help me decide whether to wear a helmet in each condition. I have no interest whatsoever in the association of wearing a helmet with injury, *ignoring biking status*. Nor have you, I am sure. I am also hoping that common sense tells you that the so-called "effect of wearing a helmet" – biking status omitted – is not only uninteresting, but *it is not an effect at all*. If Nature created a causal structure in which the effects of *E* on *D* depend on the value of *U* (Figure 1), it is self-deception to claim that we learn something about causal reality when *U* is ignored.

Effect-modification bias

There is more to the story, however, than self-deception. Under the causal structure in Figure 1, the marginal association between *E* and *D* is a biased estimator – twice over. It is a biased estimator of the effect *E*→*D* when *U*=*no*, and it a biased estimator of the effect *E*→*D* when *U*=*yes*. To claim that the marginal association unbiasedly estimates a causal parameter is to embrace *effect-modification bias*.[3] As a matter of gossip, you might be surprised to learn that some sharp minds downplay the causal idea of effect modification and ignore effect-modification bias. Wondering why? I have a few explanations to offer – mathematical, philosophical, and psychological – but that's for another day.[a]

Question: Suppose someone stubbornly computed the marginal association between *E* and *D*, obtained a probability ratio of 1.69, and called it "the effect". How is that number related to the *U*-modified effects of *E* on *D*?

Answer: It is some hybrid of *E*→*D* when *U*=*no*, and *E*→*D* when *U*=yes, a weighted average of two probability ratios. Keep this key point in mind; we'll return to it shortly.

[a] Related, in part, to foggy ideas such as "effect in a target population", "parameter of interest in population", "definition of causal parameters" – inevitable products of hidden determinism.

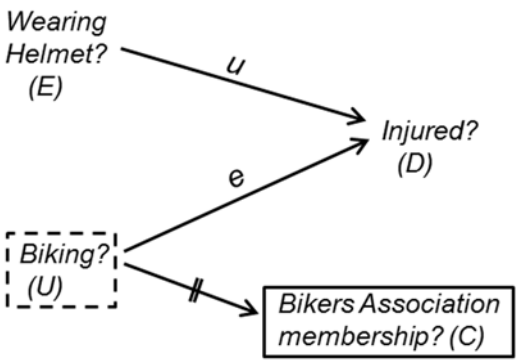
Conditioning and pseudo-conditioning

In statistical thought, classical conditioning on a variable means restricting a variable to one of its values (or stratifying on all values). Since conditioning manipulates the distribution of a variable, let me offer an extended idea, which is useful.

Any alteration of the natural distribution of a variable resembles conditioning. For instance, if the natural distribution of "Bikers Association membership?" (*C*) is 5% "yes" and 95% "no", then any change of that distribution may be called pseudo-conditioning on *C*. Again, the extreme case – classical conditioning – would be restricting to members of the bikers' association (100% "yes") or to non-members (100% "no"), but changing the sample to include 80% "yes" and 20% "no" may carry similar consequences. Classical conditioning is formally depicted by placing a solid frame around a variable (Figure 3). Why? Because such conditioning dissociates a variable (now a value) from incoming causes and outgoing effects. Pseudo-conditioning will be denoted by a dashed frame. It does not isolate a variable from the rest of the world.

The next crucial point is not widely appreciated. Conditioning on a variable, such as *C*, also alters the distribution of its causes, such as *U* (Figure 3). Though the new distribution of *U* is not known exactly, it is not necessarily restricted to one value of *U*. It is a case of pseudo-conditioning (Figure 3).

Figure 3.



To realize the point of the last paragraph, let's assume that the marginal distribution of "Biking?" is 5% "yes" and 95% "no". Suppose we condition on "Bikers Association membership?" by restriction to the value "yes". That is, we are observing only members. Do we still expect only 5% (of members!) to be bikers (Biking?=*yes*). Of course not. The percentage of Biking?=*yes* among members of the Bikers Association is surely much higher (but not necessarily 100%). Another example? smoking→lung cancer. We therefore expect a higher percentage of smokers among people with lung cancer, as compared with the percentage of smokers ignoring cancer status.

Not every averaging gives the same average

Back to the association between wearing a helmet and injury. Remember that someone computed a marginal association (probability ratio of 1.69) and called it "the effect" (of not wearing a helmet)? Remember that the marginal association was merely "some average" of two effects in the two strata of "Biking"? Well, this "some average" (a weighted average) depends on the distribution of "Biking". Most important, this average (a probability ratio of 1.69) will change when the distribution of "Biking" is altered – for example, by conditioning on "Bikers Association membership". It will no longer be 1.69.

If someone makes the mistake of calling 1.69 "the effect of not wearing a helmet", they will argue that conditioning on "Bikers Association membership" has produced a surprising bias – colliding ("selection") bias without conditioning on a collider (Figure 3). But you know better now. No new bias was created. It was effect-modification bias before conditioning, and it is effect-modification bias after conditioning. The magnitude of the bias has changed because the new estimate comes from averaging over a different distribution of "Biking", the modifier. Again, recall what has changed the distribution of this variable. It was pseudo-conditioning following classical conditioning on "Bikers Association membership" (Figure 3)… [b]

Stated differently, what looked like a change from "unbiased" to "a new bias" was simply a change from "some magnitude of old bias" to "some other magnitude of the same old bias".

Special cases

What happens if there is no effect modification between *E* and *U*?

The probability ratio will be identical in the marginal association, in the strata of *U* (and *C*), and in any distribution of *U* (and *C*). This phenomenon is known as the collapsibility property of the probability ratio.

[b] Perhaps "second-hand conditioning" may be another phrase to describe what has happened here.

What if *E* has a null effect on *D*?

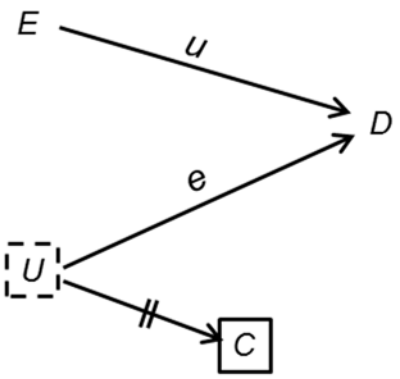
Well, in this case there is nothing to modify and the null effect will be evident in the marginal association, in the strata of *U* (and *C*), and in any other altered distribution of *U* (and *C*). Quite simple.

Invited commentary

Now go read an invited commentary on the pages of the *American Journal of Epidemiology*.[4] It is titled "Selection bias without colliders", which is also the principal claim of the article.

I have news to share with you: it was not an example of selection bias without colliders. It was an example of effect-modification bias before conditioning on *C* (Figure 1), and after conditioning on *C* (Figure 4). Neither the marginal probability ratio (1.69) nor one conditional probability ratio (2.05) nor another (0.70) was estimating any effect unbiasedly.

Figure 4.



The conditions for the bias were evident. The variable *U* (a cause of *C* and *D*) was a modifier of *E*→*D*, easily inferred from the first paragraph on page 1049, and even explicitly stated a few paragraphs later:

"*This heterogeneity of the C—D association across levels of E can be attributed to an interaction between E and U on the risk ratio scale*."[4] [Italics added].

And I will add the missing conclusion:

Therefore, any result from averaging over the values of U (1.69, 2.05, 0.70) did not estimate unbiasedly any of the U-modified effects of E on D.

To deny this conclusion is equivalent to claiming interest in the association of wearing a helmet with injury, averaging over biking?=*yes* and biking?=*no*…

Epilogue

It took a long time to replace the counterfactual/ exchangeability explanation of confounding with "an open, natural path between *E* (the exposure variable) and *D* (the outcome variable) due to a shared cause". It is time to abandon "selection bias" too, in the context of causal inquiry.[5] As you can tell for yourself, even sharp minds fall for that outdated, misleading term. And no, there is no colliding bias without conditioning on a collider. Never!

References

1. Shahar E, Shahar DJ. On the definition of effect modification. *Epidemiology* 2010;21:587

2. Shahar DJ, Shahar E. A theorem at the core of colliding bias. *Int J Biostat* 2017 Mar 31;13(1).

3. Shahar E, Shahar DJ. Causal diagrams and three pairs of biases. In: *Epidemiology – Current Perspectives on Research and Practice* (Lunet N, Editor)www.intechopen.com/books/epidemiology-current-perspectives-on-research-and-practice, 2012: pp. 31-62

4. Hernan MA. Invited commentary: Selection bias without colliders. *Am J Epidemiol* 2017 Jun 1;185(11):1048-1050

5. Colliding bias (part 1): misnomers and the missing dashed line. (a commentary on this website)