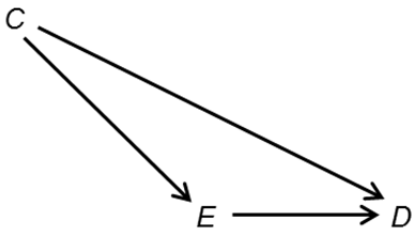
Insights on confounders under (time point) cause → (time point) effect

Confounding bias arises when the cause (*E*) and effect (*D*) of interest share a cause (*C*), as shown in Figure 1. It is therefore intuitive to call the culprit (*C*) a confounder.

Figure 1.



Causal reality, however, is more complicated than Figure 1 claims because a variable is a natural property of an object *at* a *time point* (t_i). A natural variable has no meaning – it takes no concrete value – without reference to time. At this moment for instance (say, t_0), as I am typing these words, I have a value of a property called blood pressure. And by the time you read this sentence (say, t_1), I almost certainly have a different value of blood pressure. Furthermore, the effect of my blood pressure at t_0 on some outcome at some future time, t_2, may be different from that of my blood pressure at t_1.

If you are entertaining the idea of combining time point causes into "a combined cause" – for example "the effect of treatment regime"[1] – please read a critique of that idea elsewhere.[2] You will be reading about the difference between the solving of a math problem and the studying of causal reality.

Classic confounding

How might Figure 1 be redrawn to allow for time point causes and time point effects?

One possible revision is founded on a proposed axiomatization of indeterministic causation.[2] Here are four relevant axioms:

1. All causation operates between time point variables: a variable at one time (e.g., E_0) affects a variable at a later time (e.g., D_1).

2. If $E \rightarrow D$, then $E_i \rightarrow D_j$ for any *i* and *j* where $j>i$.

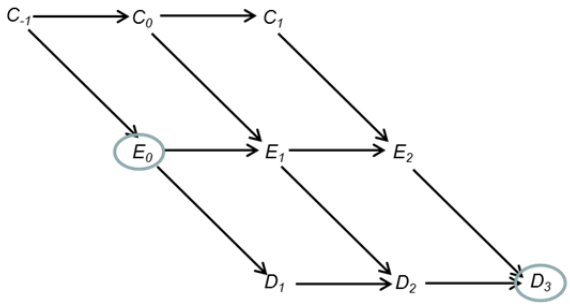
3. The causal parameter behind an arrow is constant over time. If $E_0 \rightarrow D_k$ with a parameter λ_k, then $E_t \rightarrow D_{k+t}$ with parameter λ_k for any *t*.

4. A variable at one time affects that variable at any future time (e.g., $E_i \rightarrow E_j$, where $j>i$).

Following these axioms the diagram in Figure 1 should be replaced by a diagram with time-indexed variables. For example, each variable may be replaced by three variables at evenly spaced time points (Figure 2).

The new diagram depicts three effects of *E* on *D* which may be estimated: E_0 on D_1; E_0 on D_2; and E_0 on D_3. Notice that all other possible $E_i \rightarrow D_j$ pairs are captured by one of the former. For example, the effect of E_1 on D_3 is identical to the effect of E_0 on D_2 (third axiom).

Figure 2.



Let's focus on the effect of E_0 on D_3 (circled variables) and assume it is not null. Figure 2 shows that this effect is composed of three causal paths:

$$E_0 \rightarrow E_1 \rightarrow E_2 \rightarrow D_3$$
$$E_0 \rightarrow D_1 \rightarrow D_2 \rightarrow D_3$$
$$E_0 \rightarrow E_1 \rightarrow D_2 \rightarrow D_3$$

Tellingly, the so-called direct effect of E_0 on D_3 – in fact, of *any* E_i on *any* D_j ($j>i$) – is partly mediated through interim *E* variables and through interim *D* variables. Of course, it may also be mediated through other intermediary variables between *E* and *D* (not shown).

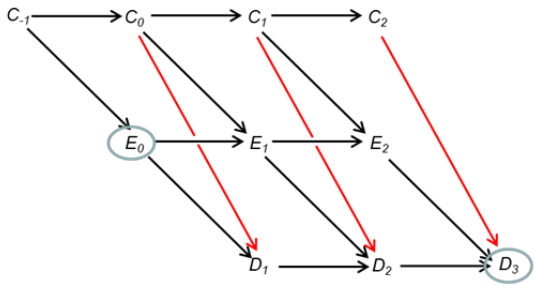
Figure 2 also provides deeper insight into the overly simplified confounding path $E \leftarrow C \rightarrow D$. Although the segment $C \rightarrow E$ (Fig. 1) was simply replaced by $C_{-1} \rightarrow E_0$

(Figure 2), the segment $C \rightarrow D$ is now composed of three causal paths:

$$C_{-1} \rightarrow C_0 \rightarrow E_1 \rightarrow D_2 \rightarrow D_3$$
$$C_{-1} \rightarrow C_0 \rightarrow E_1 \rightarrow E_2 \rightarrow D_3$$
$$C_{-1} \rightarrow C_0 \rightarrow C_1 \rightarrow E_2 \rightarrow D_3$$

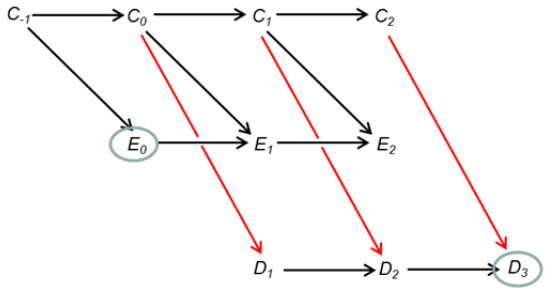
C may also affect *D* through paths that bypass *E* variables (Figure 3, red arrows), but a key point remains: At least some confounding paths travel through E_1 or E_2 – that is, through future time points of the time point exposure, E_0. Which reveals an overlooked point. Interim *E* variables between a time point cause and a time point effect play a dual role: they mediate the effect of the time point cause, and they also mediate some paths of confounding bias (Figure 3).[a]

Figure 3.



A notable exception is the null. If *E* has a null effect on *D*, interim *E* variables do not play any role in confounding either (Figure 4). In that case, confounding operates only through causal paths that bypass interim *E*, such as $E_0 \leftarrow C_{-1} \rightarrow C_0 \rightarrow C_1 \rightarrow D_2 \rightarrow D_3$.

Figure 4.

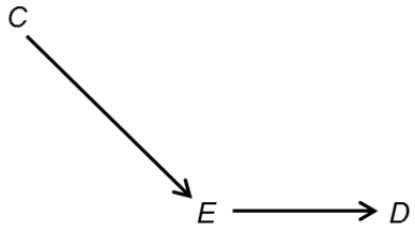


[a] To simplify, I omit confounding through prior exposure variables (e.g., E_{-1}) which operates through D_0. The topic is discussed elsewhere.[3]

A closer look at a cause of *E* "alone"

If a variable *C* affects a variable *D* only through *E* (Figure 5), then *C* is still a shared cause of *E* ($C \rightarrow E$) and *D* ($C \rightarrow E \rightarrow D$), but the structure does not appear to show confounding bias: *C* does not appear to contribute to the association between *E* and *D*. Accordingly, it is often taught that a variable that affects *D* only through *E* (Figure 5) is not a confounder.

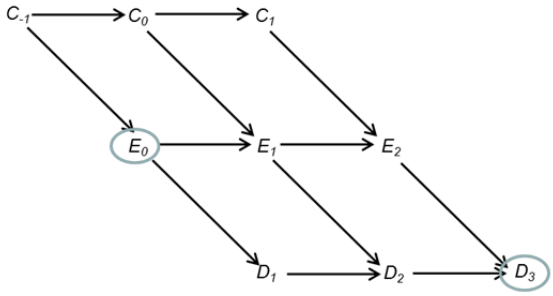
Figure 5.



Is it not a confounder, indeed?

Again, Figure 5 is lacking time indexing. Redrawing the diagram with time-indexed variables, as before, reveals an interesting causal structure (Figure 6).

Figure 6.

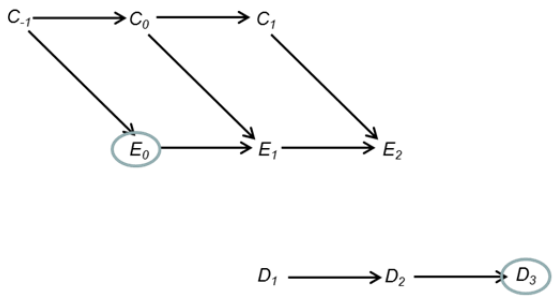


A cause of D_3 through E_0 "alone" is also a cause of D_3 through subsequent *E* variables (E_1, E_2). Moreover, we can now trace three confounding paths between E_0, the time point cause, and D_3, the time point effect:

$$E_0 \leftarrow C_{-1} \rightarrow C_0 \rightarrow E_1 \rightarrow D_2 \rightarrow D_3$$
$$E_0 \leftarrow C_{-1} \rightarrow C_0 \rightarrow E_1 \rightarrow E_2 \rightarrow D_3$$
$$E_0 \leftarrow C_{-1} \rightarrow C_0 \rightarrow C_1 \rightarrow E_2 \rightarrow D_3$$

Under "time point exposure→time point effect", *every cause of the exposure is a confounder*, except under the null. If *E* has a null effect on *D*, a cause of *E* alone is not a confounder (Figure 7).

Figure 7.



In summary:

- If *E* has a non-null effect on *D* (*E*→*D*), any variable, *C*, that has a non-null effect on *E* (*C*→*E*) is a confounder. Any cause of *E* is a source of confounding bias – even if it is a cause of *D* only through future *E* variables (Figure 6). Such a variable will be called "type 1 confounder".[b]

- If *E* has a null effect on *D*, only variables that have non-null effects on both *E* and *D* are confounders (Figure 4). Such variables will be called "type 2 confounders". The set of type 2 confounders is a subset of type 1 confounders.

Notice that in both cases we end up with the standard definition of a confounder: a shared cause of *E* and *D*. Only a cause of *E* that is also a cause of *D* will create an open path between the two variables and add confounding bias. As far as the definition is concerned (and as far as confounding bias is concerned), it makes no difference whether that cause of *E* affects *D* only through *E* variables or otherwise.

A conditioning conundrum: type 1 confounders or type 2 confounders?

Setting up a study on the effect of some E_0 on some D_1, you would not usually claim to have background knowledge on the effect of *E* on *D* (or you would not have set up a study…). In particular, you would not usually claim to know *whether E has a null effect on D*, so you do not know if your study requires conditioning on type 1 confounders or on type 2 confounders. What are you to do, then? Should you condition on variables that have a non-null effect on *E*, or only on variables that have a non-null effect on *both E* and *D*?

You may try to argue that all that matters is background knowledge on the causes of *E* and *D*. Conditioning on known, shared, causes of *E* and *D* will do no harm, whether the underlying causal structure corresponds to Figure 3, Figure 4, or Figure 6.

That may seem like a compelling argument, but it is problematic. If you act on the possibility that Figure 6 is true (*C* affects *D* only through *E* variables), you are actually declaring that *E*→*D* without any testing. You admit to have conditioned on *C because E has a non-null effect on D*… (That's not the case in Figures 3 and 4, where you condition on *C* whether *E*→*D* or not). And if you say that you don't condition on some cause of *E* because it is not a cause of *D*, then you are declaring – without any testing – that Figure 7 is true: *E* has a null effect on *D*. If so, there is no need to proceed with your study about the effect of *E* on *D*…

The choice between the two types of confounders is not inconsequential. In most cases conditioning on a variable increases the variance of the estimated association – the "cost" we must pay to remove confounding bias. On the other hand, *not* conditioning on a variable might leave confounding bias – the "cost" we must pay to constrain the variance. That's the classic tradeoff between bias and variance.

Choosing between type 1 confounders and type 2 confounders, we encounter alternative paradigms, alternative actions, and alternative consequences (Table 1):

Table 1.

Paradigm	Action	Null	Consequences	
			Bias	Variance
Null is true until "proven" otherwise	Condition on type 2 confounders	True	Removed	Necessarily increased
		False	Remained	Necessarily increased
Estimation	Condition on type 1 confounders	True	Removed	Unnecessarily increased
		False	Removed	Necessarily increased

One paradigm accepts the null as true (*E* has a null effect on *D*) until empirical evidence to the contrary, an axiom of causal inquiry aligned with criminal law (innocent until proven otherwise). A competing paradigm accepts the estimation of effects as the fundamental purpose of science and rejects any superior standing of the null (which is just one point on a continuum). The former paradigm prescribes conditioning only on type 2 confounders (causes of both *E* and *D*), whereas the latter prescribes conditioning on type 1 confounders (Table 1).

[b] *E* variables prior to E_0 (a time point cause) are also included in type 1 confounders. See elsewhere.[3]

As Table 1 shows, the choice between the two sets for conditioning amount to a choice between alternative penalties for a wrong decision: conditioning on type 2 confounders alone carries the risk of residual confounding, whereas conditioning on type 1 carries the risk of increasing the variance unnecessarily.

Stated differently:

- Choosing type 1 confounders (the large set) to condition on accepts the threat of larger-than-needed variance, but ensures that confounding bias (due to any known member of the set) will never remain.

- Choosing type 2 confounders (the subset) accepts the threat of remaining confounding bias (through interim *E* variables), but ensures that the variance will never be larger than it needs to be.

Once again we are witnessing the bias-variance tension, an unavoidable tradeoff in causal inquiry.

At first glance, it seems possible to reconcile testing with estimation by a two-step procedure: First, condition on type 2 confounders to test the null. If rejected, proceed to conditioning on type 1 confounders because confounding bias remains when the null is false (Table 1). But more may be said. It is not all-or-none bias versus larger-or-not-larger than needed variance. The dichotomized penalty is an oversimplification of a continuum. To realize the point, consider two extreme situations where the null is false: 1) *E* has a very weak effect on *D*; 2) *E* has a very strong effect on *D*.

If the effect of *E* on *D* is very weak, nearly null, and we choose to condition on type 2 confounders, we will indeed leave confounding bias but its magnitude will be very small because unblocked confounding paths contain a very weak associational link ($E_i \rightarrow D_j$). Residual confounding will be minuscule and the penalty for our wrong choice (type 2 instead of type 1) will be likewise very small. Conversely, a very strong effect of *E* on *D* will carry a large penalty in the bias domain. In between the extremes, the penalty will be prorated.

A comparable argument may be developed for the variance. If the null is true, and we choose to condition on type 1 confounders, the variance will be unnecessarily larger than needed, but the increment might be very small because the "extra variance" depends on characteristics of the data. For instance, conditioning on a variable whose distribution is highly skewed (say, 95% of the observations take a single value) will have a negligible effect on the variance, so the penalty for our wrong choice (type 1 instead of type 2) will be minuscule. Conversely, conditioning on a variable whose values are evenly distributed will carry a large penalty in the variance domain. Again, in between, the penalty is prorated.

The examples above tell us that the practical consequences of choosing between the two paradigms are more complex than the simplified dichotomy in Table 1. It is not bias versus variance. It is the magnitude of leftover bias versus the magnitude of incremental variance. Unfortunately, no logic and no computation can tell us which conditioning will be better. We are left to decide on the preferred axiom of scientific methodology.

For many years now, I have sided with the camp of estimation and offered arguments against the merit of announcing that the null is false – that *E* is a cause of *D*. Make your own choice, but make it an informed one.[4,5]

References

1. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11:550–560 A significant dialogue (on my home page)
2. Shahar E, Shahar DJ. Marginal structural models: much ado about (almost) nothing. *Journal of Evaluation in Clinical Practice* 2013 Feb;19(1):214-22
3. Shahar E, Shahar DJ. On which variable(s) should we condition to remove confounding bias? *British Journal of Medicine and Medical Research* 2015;10(4):1-9
4. P-values and confidence intervals? Think again (a commentary here).
5. A significant dialogue (on my home page)