Simpson's paradox

Kevin P. Murphy

Last updated November 24, 2006

1 Simpson's paradox

We will show a dramatic example of the dangers of not thinking causally. Suppose taking a drug (cause C) decreases recovery rate (effect E) in females (F) and males ($\neg F$)

$$\begin{aligned} P(E|C,F) &< P(E|\neg C,F) \\ P(E|C,\neg F) &< P(E|\neg C,\neg F) \end{aligned}$$

but in the combined population, the drug increases recovery rate

$$P(E|C) > P(E|\neg C)$$

By the rules of probability, this is perfectly possible, as the table of numbers below shows.

	Combined				Male				Female			
	E	$\neg E$	Total	Rate	E	$\neg E$	Total	Rate	E	$\neg E$	Total	Rate
C	20	20	40	50%	18	12	30	60%	2	8	10	20%
$\neg C$	16	24	40	40%	7	3	10	70%	9	21	30	30%
Total	36	44	80		25	15	40		11	29	40	

$$\begin{aligned} p(E|C) &= p(E,C)/p(c) = 20/40 = 0.5 && (1) \\ p(E|\neg C) &= 16/40 = 0.4 && (2) \\ p(E|C,F) &= 2/10 = 0.2 && (3) \\ p(E|\neg C,F) &= 9/30 = 0.3 && (4) \\ p(E|C,\neg F) &= 18/30 = 0.6 && (5) \\ p(E|\neg C,\neg F) &= 7/10 = 0.7 && (6) \end{aligned}$$

But the conclusion goes counter to intuition. Why? Put another way: given a new patient, do we use the drug or not? Novick wrote " The apparent answer is that when we know the gender of the patient, we do not use the drug, but if the gender is unknown, we should use the drug. Obviously that conclusion is ridiculous". (Quoted in [**?** , p175].)

We can resolve the paradox as follows. The statement that the drug C causes recovery E is

$$P(E|\text{do}(C)) > P(E|\text{do}(\neg C)) \quad (7)$$

whereas the data merely tell us

$$P(E|C) > P(E|\neg C) \quad (8)$$

This is not a contradiction. Observing C is positive evidence for E, since more males than females take the drug, and the male recovery rate is higher (regardless of the drug). Thus Equation 8 does not imply Equation 7.

If we assume that the drug C does not cause gender F, as in Figure 1(left), then we can prove that if taking the drug is harmful in each subpopulation (male and female), then it must be harmful overall. Specifically, if we assume

$$\begin{aligned} p(E|\text{do}(C),F) &< p(E|\text{do}(\neg C),F) && (9) \\ p(E|\text{do}(C),\neg F) &< p(E|\text{do}(\neg C),\neg F) && (10) \end{aligned}$$

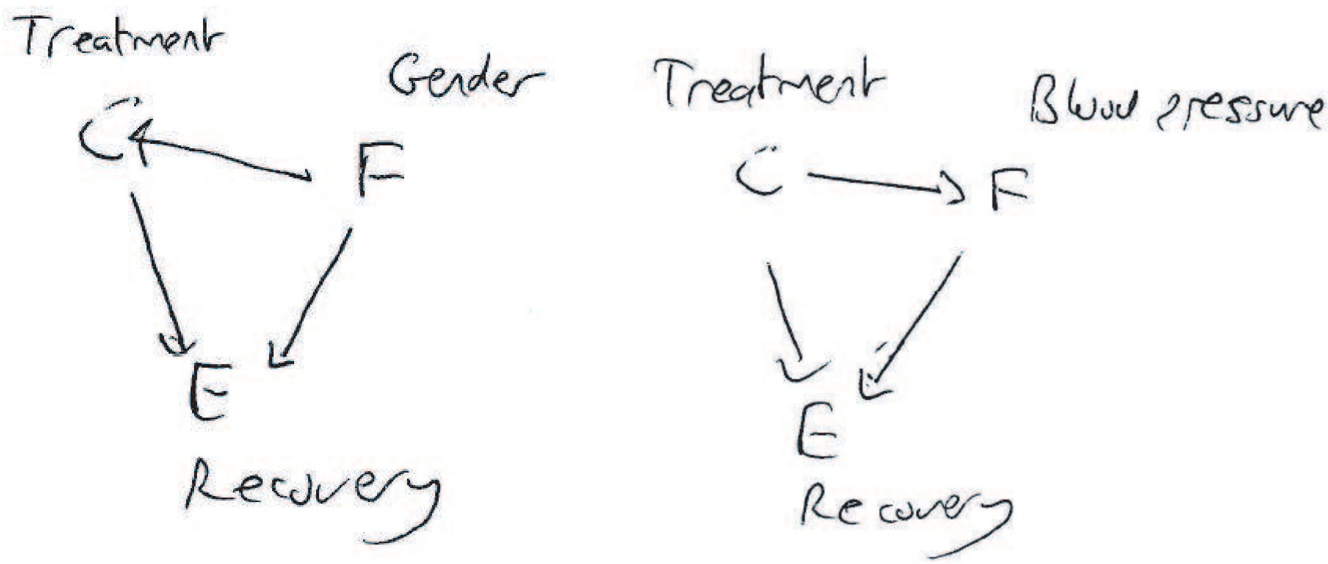

Figure 1: Two versios of the Simpson's paradox. Left: F is gender and causes C. Right: F is blood pressure and is caused by C.

then we can show

$$p(E|\text{do}(C)) < p(E|\text{do}(\neg C)) \tag{11}$$

The proof is as follows [**?** , p181]. First we assume that drugs have no effect on gender

$$p(F|\text{do}(C)) = p(F|\text{do}(\neg C)) = p(F) \tag{12}$$

Now using the law of total probability,

$$p(E|\text{do}(C)) = p(E|\text{do}(C), F)p(F|\text{do}(C)) + p(E|\text{do}(C), \neg F)p(\neg F|\text{do}(C)) \tag{13}$$

$$= p(E|\text{do}(C), F)p(F) + p(E|\text{do}(C), \neg F)p(\neg F) \tag{14}$$

Similarly,

$$p(E|\text{do}(\neg C)) = p(E|\text{do}(\neg C), F)p(F) + p(E|\text{do}(\neg C), \neg F)p(\neg F) \tag{15}$$

Since every term in Equation 14 is less than the corresponding term in Equation 15, we conclude that

$$p(E|\text{do}(C)) < p(E|\text{do}(\neg C)) \tag{16}$$

To assess the effect of C on E, we have to take into account that there is a **backdoor path** from E to C via F. Pearl [**?** , p79] proves that you have to adjust for (i.e., condition on) such backdoor variables. Intuitively, we need to be sure the effect of C on E is not due to their common cause, F. Thus we should check the $C \rightarrow E$ relationship for each value of F separately. In this example, the drug reduces E in both tables, so we should not take the drug regardless of gender.

Now consider a different cover story. Suppose we keep the data the same but interpret F as something that is affected by C, such as blood pressure. Thus F is now caused by C: see Figure 1(right). In this case, we can no longer assume

$$p(F|\text{do}(C)) = p(F|\text{do}(\neg C)) = p(F) \tag{17}$$

and the above proof breaks down. So $p(E|\text{do}(C)) - p(E|\text{do}(\neg C))$ may be positive or negaitve.

To assess the effect of C on E, we should look at the combined (C, E) table. We should not condition on F, since there is no backdoor path in this case. More intuitively, conditioning on F might block one of the causal pathways.

In other words, by comparing patients with the same post-treatment blood pressure, we may mask the effect of one of the two pathways by which the drug operates to bring about recover.

Thus we see that different causal assumptions lead to different actions. In this case, the models require distinguishing the direction of arcs into/ out of the latent variable F, so we need prior domain knowledge to choose the right one.