Causation and counterfactuals

21 JANUARY 2020

COLIN FISCHBACHER:
COLIN.FISCHBACHER@NHS.NET

Outline

Causes

Counterfactuals

The notation for counterfactual models

Exchangeability

Randomisation

Conditional exchangeability

Standardisation

causes and counterfactuals

Defining causes

Fergus was driving too fast and had had too much to drink when his car hit the ice. He crashed into a tree and was killed.

Why do you think Fergus crashed?

Did the ice cause Fergus's crash?

What would it take to convince you that the ice caused Fergus's crash?

Defining causes

Suppose we knew that if the road had not been icy (all other things being equal, going just as fast, just as much to drink etc) Fergus would **not** have crashed.

Would this help you decide whether the ice caused Fergus's death?

Counterfactuals

The situation with no ice is a counterfactual.

It is an imaginary or potential situation, but it clarifies what we are trying to understand in relation to a cause

We often use counterfactuals in everyday speech (eg “if I hadn’t . . . “)

In making causal statements we define a counterfactual and compare the actual outcome with the outcome in the counterfactual

Defining causes

Graeme was driving too fast and had had too much to drink when his car hit the ice. He crashed into a tree and was killed.

Suppose we knew the outcome in the counterfactual situation: if the road had **not** been icy (all other things being equal, going just as fast, just as much to drink) Graeme would **still** have crashed and been killed.

Did the ice cause Graeme's death?

Causal notation: actual data

- In causal notation A represents the exposure or the treatment and Y represents the outcome (eg death)
- Y|A=1 means the actual outcome (Y) **given that** someone's exposure status was actually 1 (exposed)
- so we know that for Fergus Y|A=1 = 1 but we don't know Y|A=0
- what do we know for Graeme?

	A (the ice)	Y (died)
Fergus	1	1
Graeme	1	1

1 = yes - factor present, or outcome happened
0 = no – factor absent, or outcome did not happen

Causal notation: what we'd like to know

	A (the ice)	Y|A=0	Y|A=1
Fergus *(f)*	1		1
Graeme *(g)*	1		1

1 = yes - factor present, or outcome happened
0 = no – factor absent, or outcome did not happen

$Y_i|A=1$ means the outcome (Y) given that the exposure status for person i was actually 1 (exposed)

$Y_i^{a=1}$ means the outcome (Y) for person i when their exposure status (perhaps counter to the fact) is set to 1 (ie exposed)

Unobserved counterfactuals

	A (the ice)	$Y^{a=0}$	$Y^{a=1}$
Fergus *(f)*	1	**0**	1
Graeme *(g)*	0	**1**	1

1 = yes - factor present, or outcome happened
0 = no – factor absent, or outcome did not happen

The unobserved or potential outcomes are bold

If we knew them, we could say that:

the ice was causal for Fergus because $Y_f^{a=1} \neq Y_f^{a=0}$

the ice was not causal for Graeme because $Y_g^{a=1} = Y_g^{a=0}$

(All other things being equal)

Individuals

we don’t know the counterfactual (potential) outcomes for individuals because we don’t observe them (the data are missing)

so we can’t determine causal effects for individuals

Average causal effects in groups

However, we can estimate average causal effects in groups

Exposure A has an **average causal effect** when the risk of the outcome when everyone is exposed is different from risk of the outcome when no-one is exposed

or in maths notation:

- $Pr[Y^{a=1} = 1]$ (the probability of observing the outcome **had everyone been *exposed***) is **NOT** equal to
- $Pr[Y^{a=0} = 1]$ (the probability of observing the outcome **had everyone *not* been exposed**)

Y: the outcome
A: the exposure

Average causal effects

can be expressed as a risk difference, a risk ratio, an odds ratio, a hazard ratio etc

average causal effects can be determined accurately using perfect infinitely large randomised controlled trials

otherwise, average causal effects can be determined using observational studies with some strong assumptions

Association and causation

Association: you observe different average risks of outcomes depending on whether **actually** exposed or not

- eg people who live in care homes are more likely to have dementia than those who live in student residences – so do care homes cause dementia?

$$Pr[Y|A=0] \neq Pr[Y|A=1]$$

Association and causation

Causation: there's a difference between the average risk of outcome **if** everyone had been exposed versus **if** no-one had been exposed

- eg if everyone had lived in the student residences the risk of dementia would be just the same as if everyone had lived in the care home, so care homes do not cause dementia

$$Pr[Y^{a=0}] = Pr[Y^{a=1}]$$

Association and causation

Association risk ratio:

$$\Pr[Y=1|A=1] / \Pr[Y=1|A=0]$$

- Causal risk ratio:

$$\Pr[Y^{a=1} = 1] / \Pr[Y^{a=0} = 1]$$

- but we usually can't calculate this because we don't observe Y^a for everyone
- eg we didn't observe $Y^{a=0}$ (no ice) for Fergus or Graeme

Y: the outcome
A: the exposure

Causation from association

We usually know $Pr[Y=1|A=1]$ and $Pr[Y=1|A=0]$ because we can observe it in a study

But we can't observe $Pr[Y^{a=1} = 1]$ or $Pr[Y^{a=0} = 1]$ because no-one experiences both $^{a=1}$ **and** $^{a=0}$

So how can we estimate $Pr[Y^{a=1} = 1]$ and also $Pr[Y^{a=0} = 1]$?

We assume the groups are **exchangeable** – the exposure status is different, but everything else about them is the same

in traditional terms – there is no confounding

randomisation

Randomised trials

Randomly assign a very large number of study subjects (say with pneumonia) to two groups, group 1 and group 2

Give the treatment (say an antibiotic) to group 1 and no treatment (or a placebo treatment) to group 2

The exposed group is exposed to the antibiotic (A=1); the unexposed group is not exposed to the antibiotic (A=0)

The risk of having an outcome (Y=1, say death) is compared between the two groups

Randomisation

You decide to treat people in group 1 with an antibiotic (A=1), but not those in group 2 (A=0)

You estimate the probability of death (Pr[Y=1]) in the exposed group (A=1)

You find that $Pr[Y=1|A=1] = 0.32$

Suppose you had instead treated people in group 2 but not those in group 1 . .

What would you expect the risk of Y=1 to be in group 2?

Exchangeability

- When people are randomly assigned to groups, the probability of the outcome in the unexposed group if they had been exposed equals the probability of the outcome we observed in the group actually exposed
- The probability of death when treated with an antibiotic should be the same in both of the two groups we randomised, both those actually treated and those actually not treated
- Now we know the unobserved counterfactual,

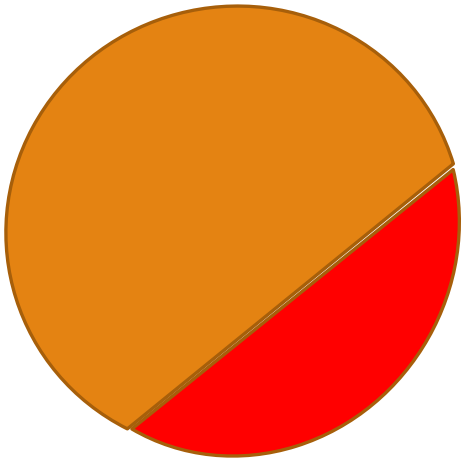
$$\Pr[Y^{a=1} = 1]$$

for the unexposed group . .

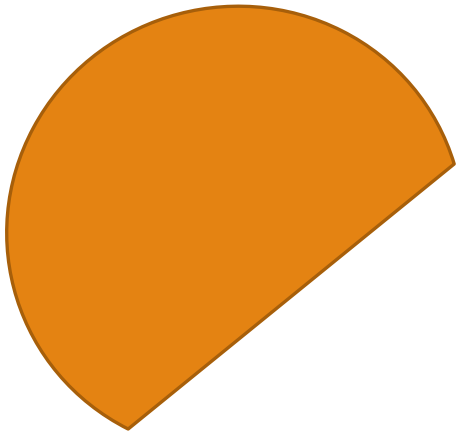
- ... it's just the risk we observed in the exposed group

$$\Pr[Y=1 \mid A=1]$$

Blue unexposed, red exposed to the risk factor



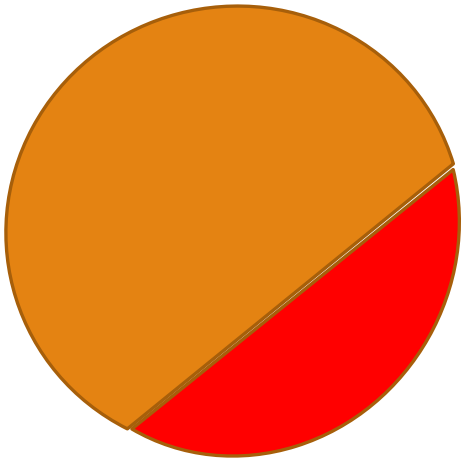
ASSOCIATION



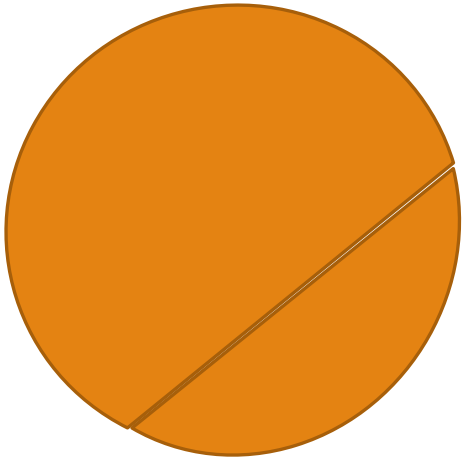
Pr[Y=1|A=0]

Pr[Y=1|A=1]

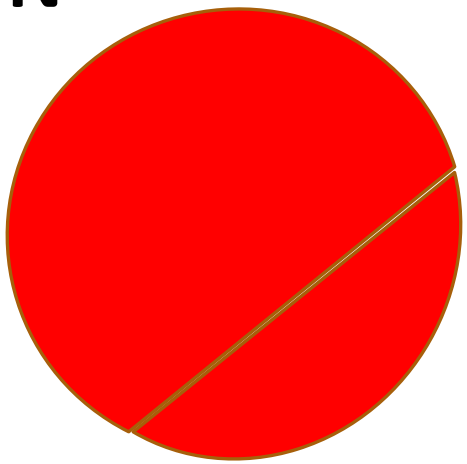
Blue unexposed, red exposed to the risk factor



CAUSATION



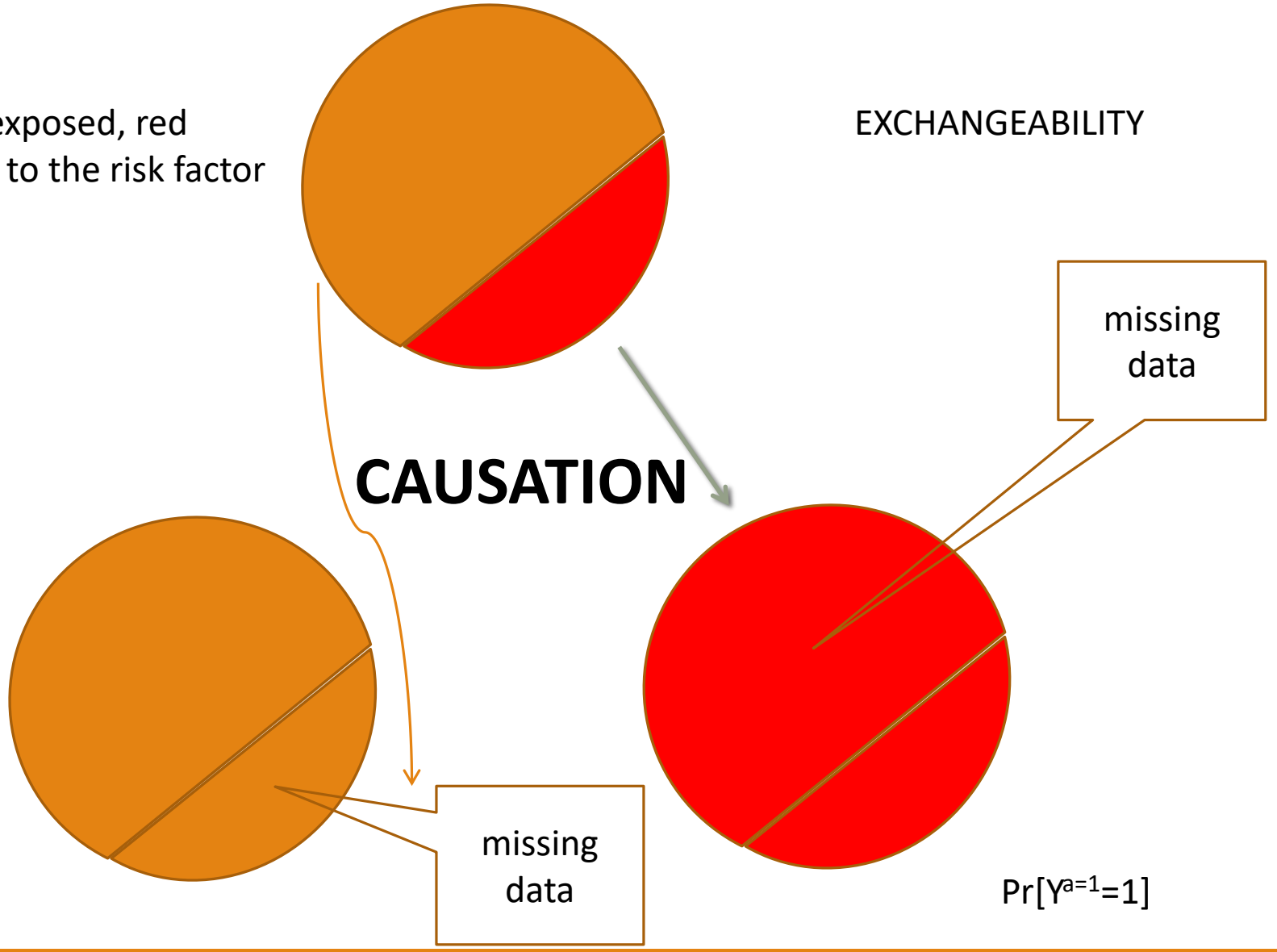
$Pr[Y^{a=0}=1]$



$Pr[Y^{a=1}=1]$

Blue unexposed, red exposed to the risk factor

EXCHANGEABILITY



exchangeability

Definition of exchangeability

The risk in the treated equals the counterfactual risk in the untreated

$Pr[Y^{a=1}] = Pr[Y = 1|A=1]$ and

$Pr[Y^{a=0}] = Pr[Y = 1|A=0]$

So, to summarise, Y^a is the same for both the exposed and the unexposed

Y^a is independent of which group you are in

Not exchangeable?

What do we do if the groups we are comparing are not exchangeable, perhaps because they weren't randomly allocated?

Maybe they are different in some other respect to the one we are interested in - for example people getting the antibiotic might be much sicker so at greater risk of death then those who didn't get the antibiotic

We can't assume that if the untreated had in fact been given the antibiotic (A=1) they would have had the same risk of death (Y=1) as those who were actually given it

In other words $Pr[Y^{a=0} = 1] \neq Pr[Y=1|A=0]$

Imaginary experiment

- 120 patients with pneumonia, 60 very sick and 60 not so sick
- Randomly select 50/60 very sick patients and 10/60 not so sick patients for treatment with a new antibiotic.

	Y=1 (died)	Y=0 (lived)	total
A=1 (antibiotic)	11	49	60
A=0 (no antibiotic)	14	46	60

Risk ratio (antibiotic compared with no antibiotic):

11/60 with antibiotic, 14/60 no antibiotic, RR = 0.79

The risk of death is 21% lower in the group that got the antibiotic

Imaginary experiment

- But the results are different if you look separately at those who were very sick and those who were not so sick.
- Note that mortality is higher overall in the very sick group, as you would expect (14/60 compared with 11/60)

	X=0 (not so sick)			X=1 (very sick)		
	Y=1 (died)	Y=0 (lived)	total	Y=1 (died)	Y=0 (lived)	total
A=1 (antibiotic)	1	9	10	10	40	50
A=0 (no antibiotic)	10	40	50	4	6	10
All	11	49	60	14	46	60

Risk ratio (antibiotic versus no antibiotic) among the not so sick (X=0) group: 1/10 (antibiotic), 10/50 (no antibiotic), RR = 0.5

Risk ratio (antibiotic versus no antibiotic) among the very sick (X=1) group: 10/50 (antibiotic), 4/10 (no antibiotic), RR = 0.5

Imaginary experiment

	X=0 (not so sick)			X=1 (very sick)		
	Y=1 (died)	Y=0 (lived)	total	Y=1 (died)	Y=0 (lived)	total
A=1 (medicine)	1	9	10	10	40	50
A=0 (no medicine)	10	40	50	4	6	10

medicine RR 0.5 in both groups

	Y=1 (died)	Y=0 (lived)	total
A=1 (medicine)	11	49	60
A=0 (no medicine)	14	46	60

medicine RR 11/60 div 14/60 = 0.79

- in the pink table the two groups are not comparable/ exchangeable
- if the medication group hadn't had medication they **wouldn't** have had the same outcome as the no medication group
- this is because 5/6 of the medication group are very sick, but only 1/6 of the non-medication group
- we can't estimate the counterfactual risk for the medication group using the non-medication group; this is because without treatment the medication group would have had a worse outcome than the non-medication group; this is because they include more sick people

Exchangeability and confounding

the traditional way of saying this is that the association between medication and death is **confounded** by severity of sickness

exchangeability can be thought of as the counterfactual equivalent of confounding

Conditional exchangeability

what can we do about non-exchangeability in this case?

we can say that **within** the sick and not so sick groups the antibiotic and non-antibiotic groups are exchangeable (we know this because they were randomly allocated to antibiotic / no antibiotic)

$$\Pr[Y^a = 1 | A=1, X=x] = \Pr[Y^a = 1 | A=0, X=x]$$

Within each group there is exchangeability

Standardisation

How do we achieve exchangeability in our combined group?

We weight the risk in the groups we are comparing – the **treated** and the **untreated** - by the distribution of illness in a standard population (or we could use the overall distribution in our whole study population)

We then sum those risks to get a standardised risk in the treated and a standardised risk in the untreated

Summing up

Counterfactuals help us to be clear about what we mean by saying that something was causal

It's therefore important to be clear about what counterfactual we are considering

eg "if no-one was obese", or "if everyone had a BMI of 25" are different counterfactuals when we are thinking about the causal effect of abnormally high weight

Summing up

We can’t observe counterfactual outcomes

However if we assume that groups are exchangeable, then we can use the outcome in the other group to estimate the unobserved counterfactual

Randomisation gives us some assurance that the groups are exchangeable

Summing up

In observational research to draw causal conclusions we have to assume that groups are either exchangeable or at least conditionally exchangeable (conditional on some third factor)

We can deal with this by standardising on that third factor or using other similar approaches to adjust for other factors

Summing up

However when randomisation isn't involved, we always have to consider whether there might be other factors that threaten exchangeability

In traditional terms, could our group comparison be confounded?

Conclusion

the counterfactual framework is part of everyday speech and being clear about counterfactuals adds rigour and clarity to discussions about causality

the notation may seem complicated

however it gives us a useful mathematical language to discuss causation

extra material: standardisation

Calculation for treated

	X=0 (not so sick)			X=1 (very sick)		
	Y=1 (died)	Y=0 (lived)	total	Y=1 (died)	Y=0 (lived)	total
A=1 (medicine)	1	9	10	10	40	50
A=0 (no medicine)	10	40	50	4	6	10

Weighted average of conditional risks in the **treated** (A=1)

Weights are Pr[X=1] = Pr[X=0] = 60/120 = 0.5

Pr[Y=1|X=1,A=1] Pr[X=1] + Pr[Y=1|X=0,A=1] Pr[X=0]

(10/50) x (60/120) + (1/10) x (60/120) = 0.15

Calculation for untreated

	X=0 (not so sick)			X=1 (very sick)		
	Y=1 (died)	Y=0 (lived)	total	Y=1 (died)	Y=0 (lived)	total
A=1 (medicine)	1	9	10	10	40	50
A=0 (no medicine)	10	40	50	4	6	10

Weighted average of conditional risks in the **untreated**

Pr[Y=1|X=1,A=0] Pr[X=1] + Pr[Y=1|X=0,A=0] Pr[X=0]

(4/10) x (60/120) + (10/50) x (60/120) = 0.30

Standardised ratio = 0.15/0.30 = **0.5**

Unstandardised ratio was 0.79