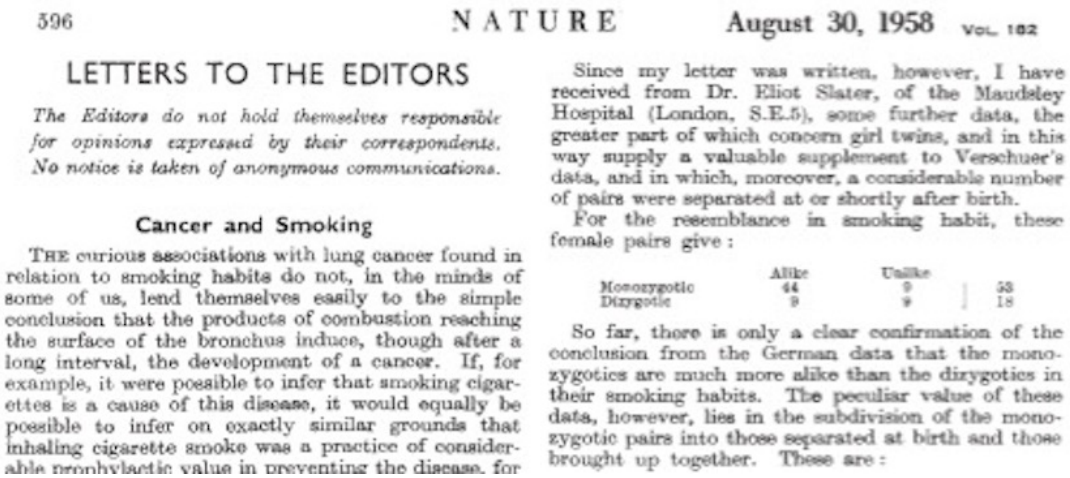
Lecture 17: Intro to causal inference

Criminology 250

Prof Maria Cuellar

University of Pennslyvania

RA Fisher on smoking



596 NATURE August 30, 1958 VOL. 182

LETTERS TO THE EDITORS

The Editors do not hold themselves responsible for opinions expressed by their correspondents. No notice is taken of anonymous communications.

Cancer and Smoking

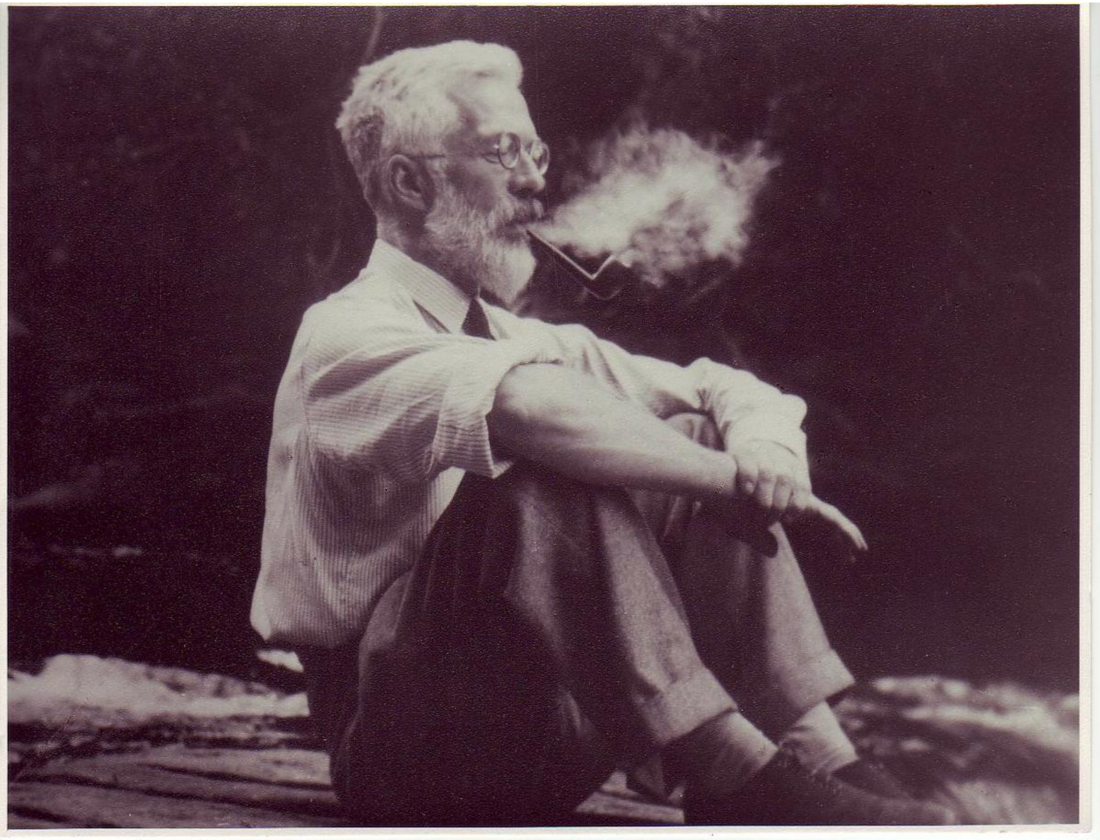
THE curious associations with lung cancer found in relation to smoking habits do not, in the minds of some of us, lend themselves easily to the simple conclusion that the products of combustion reaching the surface of the bronchus induce, though after a long interval, the development of a cancer. If, for example, it were possible to infer that smoking cigarettes is a cause of this disease, it would equally be possible to infer on exactly similar grounds that inhaling cigarette smoke was a practice of considerable prophylactic value in preventing the disease, for

Since my letter was written, however, I have received from Dr. Eliot Slater, of the Maudsley Hospital (London, S.E.5), some further data, the greater part of which concern girl twins, and in this way supply a valuable supplement to Verschuer's data, and in which, moreover, a considerable number of pairs were separated at or shortly after birth.

For the resemblance in smoking habit, these female pairs give:

	Alike	Unlike	
Monozygotic	44	9	53
Dizygotic	9	9	18

So far, there is only a clear confirmation of the conclusion from the German data that the monozygotics are much more alike than the dizygotics in their smoking habits. The peculiar value of these data, however, lies in the subdivision of the monozygotic pairs into those separated at birth and those brought up together. These are:



Individual causal effects

- Zeus gets a heart transplant, and five days later he dies. Suppose you knew somehow (e.g. via an oracle) that had Zeus not gotten the transplant, he would have survived.
- Then you must agree that it was the transplant that caused Zeus's death.

- Hera gets a heart transplant, and five days later she does not die. Suppose you knew that had she not received the transplant, she'd still be alive.
- Then you must agree that the transplant did not have a causal effect on Hera's survival.

The fundamental problem of causal inference

- We cannot observe both situations for an individual: with heart transplant and without.
- The person will either get the transplant or not, but we will only get to see what happens when one of these is actualized.

Notation

Random variable: Informally, it is a variable whose values depend on outcomes of a random phenomenon.

Let Y and A denote the random variables for the outcome and treatment or exposure.

Y: outcome (e.g. 1=death, 0=no death)

A: treatment (e.g. 1=received transplant, 0=did not receive it)

Counterfactual outcomes

- Let $Y^{a=1}$ (read Y under treatment $a = 1$) be the outcome variable that would have been observed under the treatment value $a = 1$.
- Let $Y^{a=0}$ (read Y under treatment $a = 0$) be the outcome variable that would have been observed under the treatment value $a = 0$.
- What values can $Y^{a=1}$ and $Y^{a=0}$ have?
- Y^{a} are called *potential outcomes* or *counterfactual outcomes*. These are also random variables.

Only one potential outcome gets actualized

- For each individual, only one of the counterfactual outcomes–the one that corresponds to the treatment value that the individual did receive–is actually factual.
- For example, because Zeus was actually treated ($A = 1$), his counterfactual outcome under treatment $Y^{a=1} = 1$ is equal to his observed (actual) outcome $Y = 1$.

Only one of those outcomes is observed for each individual – the one corresponding to the treatment value actually experienced by the individual. All other counterfactual outcomes remain unobserved. Because of missing data, individual effects cannot be identified, that is, they cannot be expressed as a function of the observed data.

Example

- Suppose there is a disease going around and researchers are testing whether a vaccine works.
- Let:

Y: whether you get the disease (1=yes, 0=no)

A: whether you get the vaccine (1=yes, 0=no)

$Y^{a=0} = 1$: whether you get the disease if no vaccine (1=yes, 0=no)

$Y^{a=1} = 0$: whether you don't get the disease if you are vaccinated (1=yes, 0=no)

Consistency

- Once an exposure (or no exposure) happens, then the outcome takes on the value of the potential outcome under exposure (or no exposure). This can be expressed by $Y = Y^A$, and it is called consistency.
- This is often a causal assumption. When would it be violated?
- Think about interference in a vaccine trial. If your friends are all vaccinated, but you are not, your outcome might still be your potential outcome under vaccination.
- i.e. Your potential outcomes could be $Y^{a=1} = 0$ and $Y^{a=0} = 1$, but it turns out that you have no vaccine ($A = 0$), yet you still get no disease $Y = 0$! So even without exposure, your outcome took the value of your potential outcome under exposure.
- Also known as: herd immunity.

Causal effect for an individual

The treatment A has a causal effect on an individual's outcome Y if $Y^{a=1} \neq Y^{a=0}$ for the individual.

But we can't usually observe that (by definition). How do we evaluate it then?

Average causal effects

Table 1.1

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Cyclope	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

- With this table we can calculate the average causal effect for the population: An average causal effect is present if $P(Y^{a=1} = 1) \neq P(Y^{a=0} = 1)$.
- Here, $P(Y^{a=1} = 1) = 10/20 = 0.5$ and $P(Y^{a=0} = 1) = 10/20 = 0.5$. So there is no *average causal effect*.
- You can still have *individual effects* even if there is no average causal effect (here 12 have an effect, 6 were harmed, $(Y^{a=1} - Y^{a=0} = 1)$, and 6 were helped, $(Y^{a=1} - Y^{a=0} = -1)$.

Measures of causal effect

We can represent the *causal null* in different ways:

Causal risk difference:

$$i)\ \ P(Y^{a=1}=1)-P(Y^{a=0}=1)=0$$

Causal risk ratio:

$$ii)\ \ \frac{P(Y^{a=1}=1)}{P(Y^{a=0}=1)}=1$$

Causal odds ratio:

$$iii)\ \ \frac{P(Y^{a=1}=1)/P(Y^{a=1}=0)}{P(Y^{a=0}=1)/P(Y^{a=0}=0)}=1$$

- But we'll never observe the information in Table 1. What do we do in real life?

Reality

What we actually observe is the exposure and the outcome for each individual.

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Cyclope	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

Associational risk difference:

$$i)\ \ P(Y=1|A=1) - P(Y=1|A=0) = 0$$

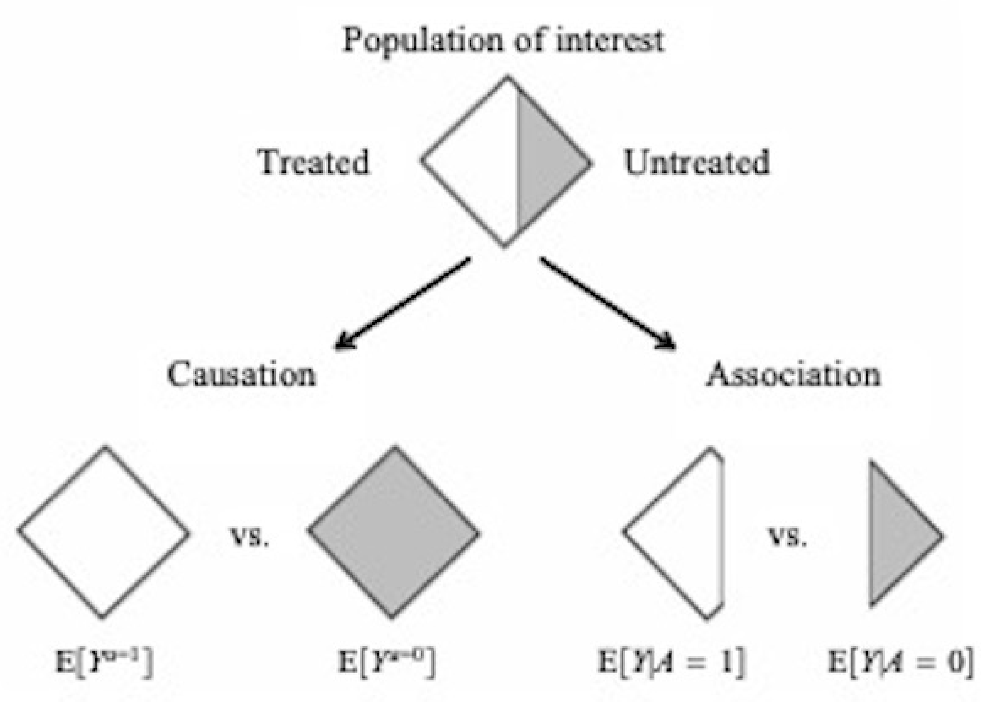
Associational risk ratio:

$$ii)\ \ \frac{P(Y=1|AZ)}{P(Y=1|A=0)} = 1$$

Associational odds ratio:

$$iii)\ \ \frac{P(Y=1|A=1)/P(Y=0|A=1)}{P(Y=1|A=0)/P(Y=0|A=0)} = 1$$

Cause vs. association



The definition of causation implies a contrast between the whole white diamond (all individuals treated) and the whole grey diamond (all individu- als untreated), whereas association implies a contrast between the white (the treated) and the grey (the untreated) areas of the original diamond.

When can you use observational data to answer causal questions?

Table 2.1

	A	Y	Y^0	Y^1
Rheia	0	0	0	?
Kronos	0	1	1	?
Demeter	0	0	0	?
Hades	0	0	0	?
Hestia	1	0	?	0
Poseidon	1	0	?	0
Hera	1	0	?	0
Zeus	1	1	?	1
Artemis	0	1	1	?
Apollo	0	1	1	?
Leto	0	0	0	?
Ares	1	1	?	1
Athena	1	1	?	1
Hephaestus	1	1	?	1
Aphrodite	1	1	?	1
Cyclope	1	1	?	1
Persephone	1	1	?	1
Hermes	1	0	?	0
Hebe	1	0	?	0
Dionysus	1	0	?	0

- Causal inference requires data like the hypothetical data in Table 1.1, but all we can ever expect to have is real world data like those in Table 1.2. The question is then under which conditions real world data can be used for causal inference.

--

- Next: Randomized experiments.