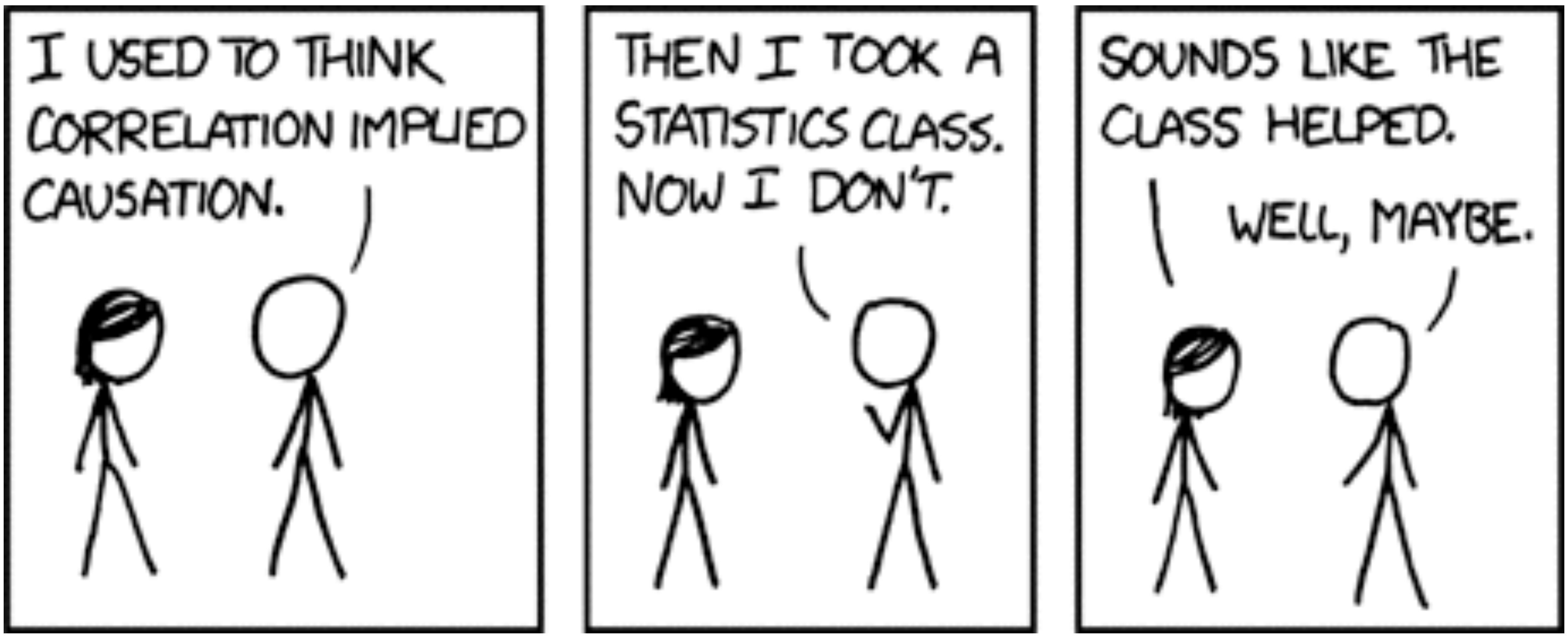
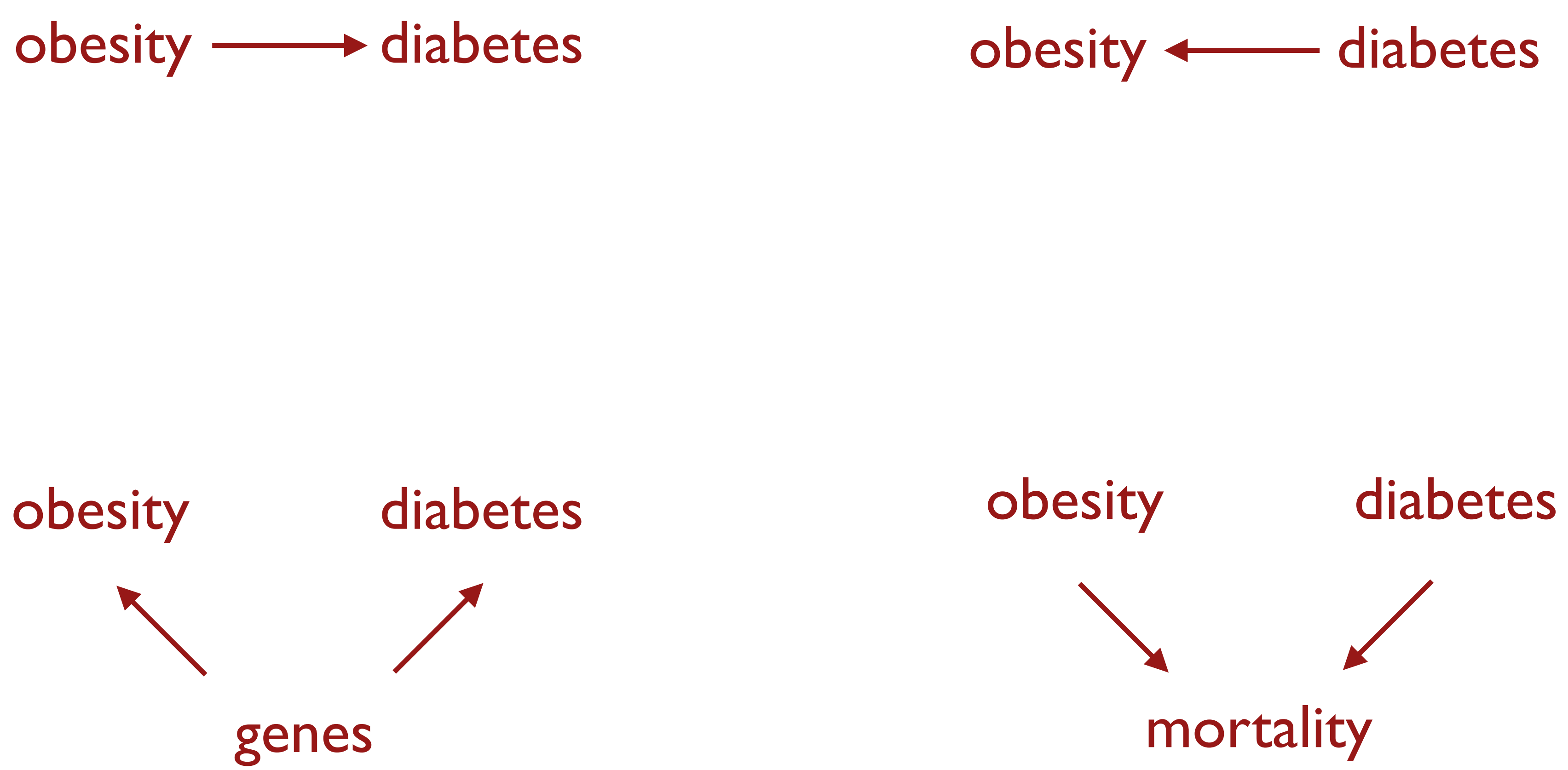
Inferring causal relationships from observational data

Elizabeth L. Ogburn
Department of Biostatistics
Johns Hopkins University



https://xkcd.com/552/

association vs causation



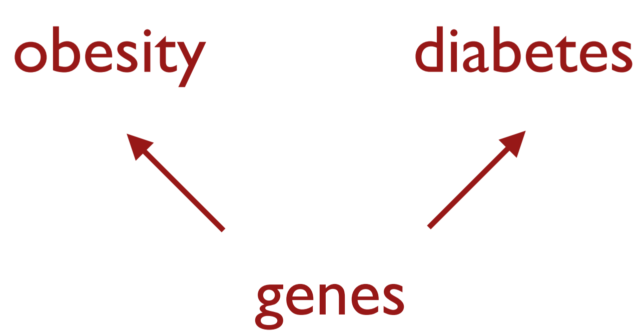
Causal inference articulates assumptions needed to move from conclusions about association to conclusions about causation.

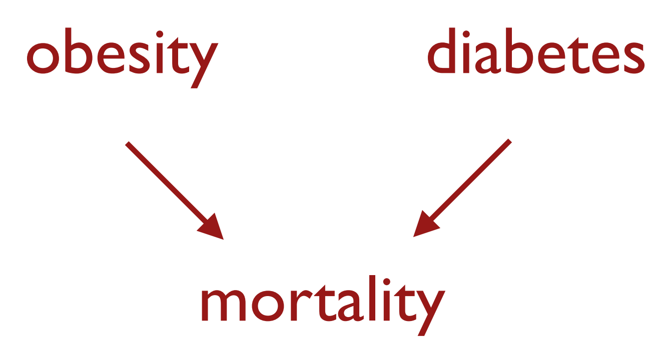
association vs causation

LATEST MEDICAL NEWS

Cardiology

Obesity Paradox for Survival in Diabetes





Causal inference articulates assumptions needed to move from conclusions about association to conclusions about causation.

how statisticians define cause

Critical role of (hypothetical) interventions:

> X causes Y if intervening to change value of X produces change in value of Y.

We can't always perform the intervention we're interested in, hence counterfactuals.

how statisticians define cause

Counterfactuals: what would have happened if, possibly contrary to fact, we had given treatment to everybody? to nobody?

simple questions can be hard

Does hormone replacement therapy cause coronary heart disease?

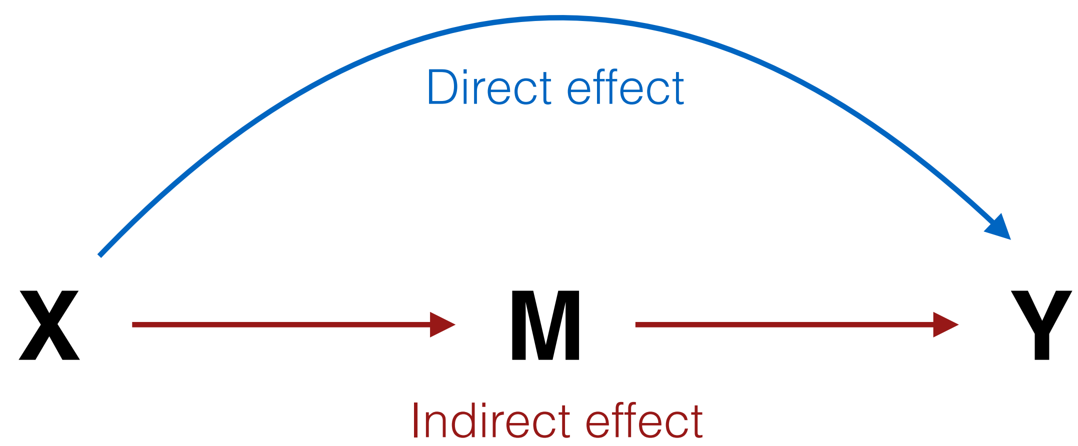
hard questions are harder

- mediation
- longitudinal data with time-varying confounders
- etc
  - networks of interconnected subjects
  - unmeasured confounding
  - generalizability / transportability
  - …

hard questions are harder

- **mediation**
- longitudinal data with time-varying confounders
- etc
  - networks of interconnected subjects
  - unmeasured confounding
  - generalizability / transportability
  - …

mediation



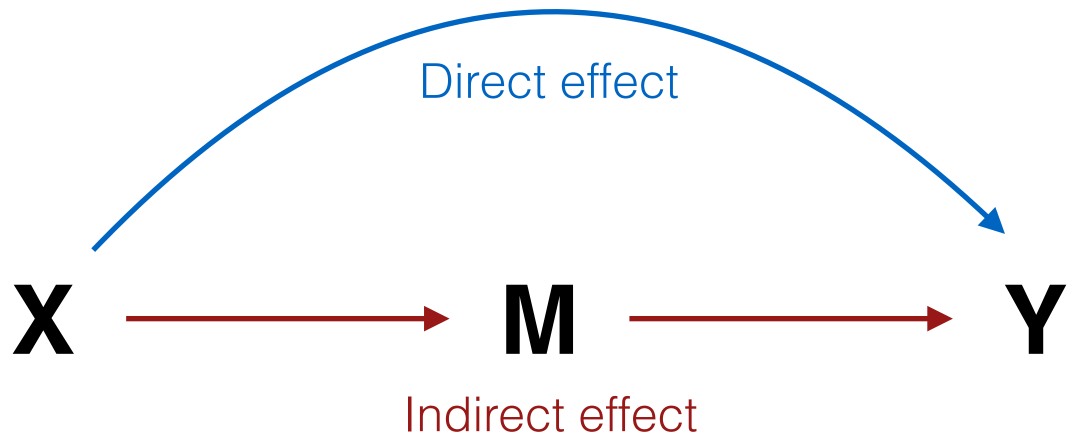
Traditional methods define direct and indirect effects in terms of the coefficients of two linear regression models:

$$Y = aX$$
$$Y = aX + bM$$

But these methods often do not correspond to the causal interpretations they are given, e.g. if

- the relationships among the variables are not linear
- there is an treatment-mediator interaction
- both models do not control for mediator-outcome confounders

mediation

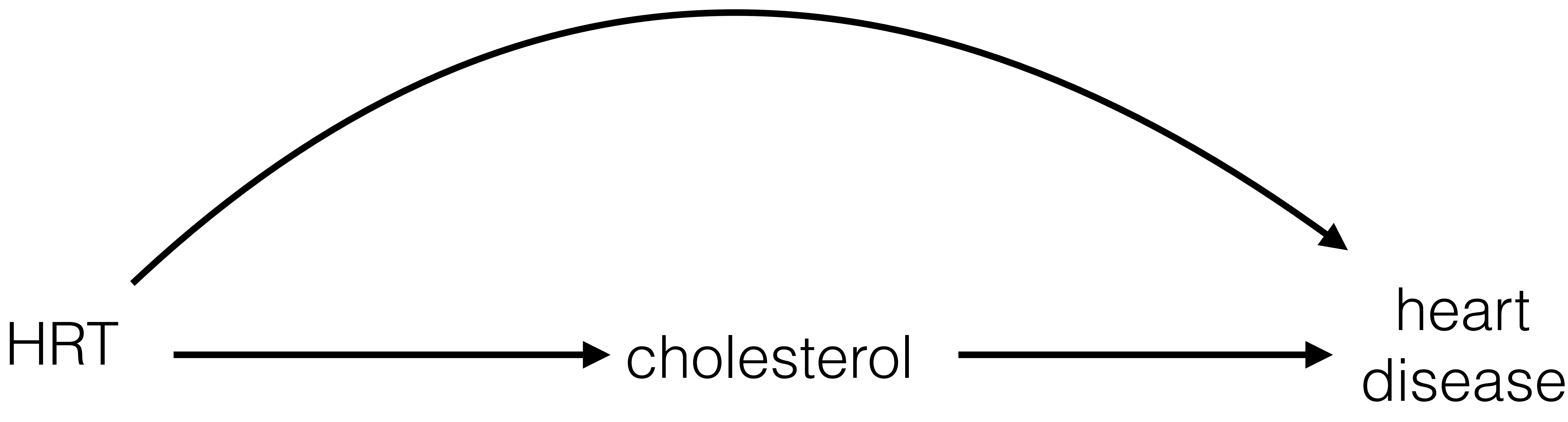


There are many ways to define direct and indirect effects using counterfactuals.

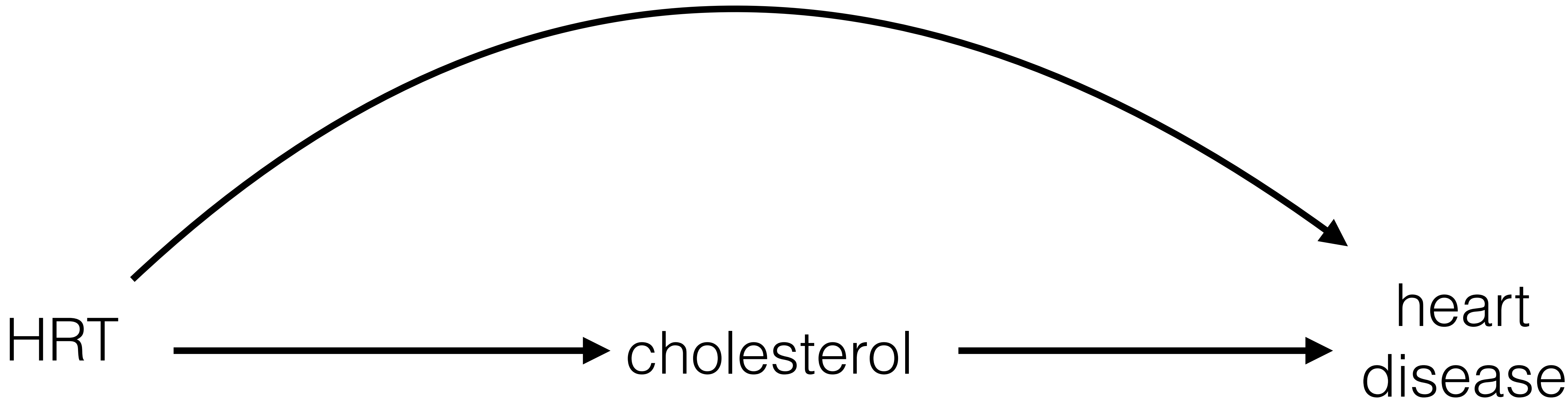
The simplest is to look at a joint intervention on X and M.

But only one set of definitions has the desirable property of decomposing the total effect of X on Y: the “natural” indirect and direct effects.

mediation



mediation



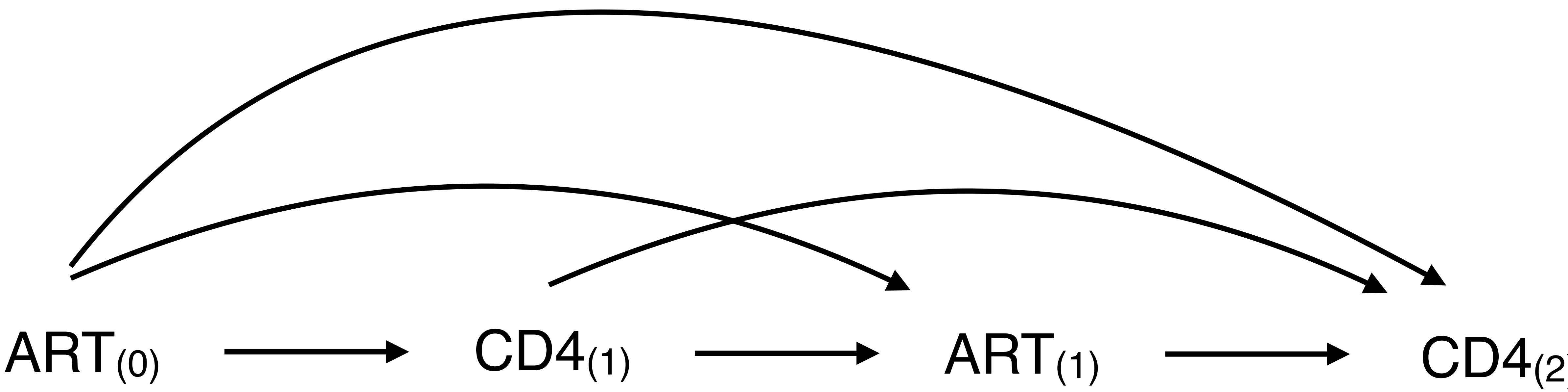
Applications include:

- showing that some polymorphisms that increase risk of lung cancer operate mainly through smoking behavior,
- examining the extent to which the effect of abruption on perinatal mortality is mediated through preterm delivery,
- finding prefrontal-subcortical pathways that mediate successful emotion regulation.

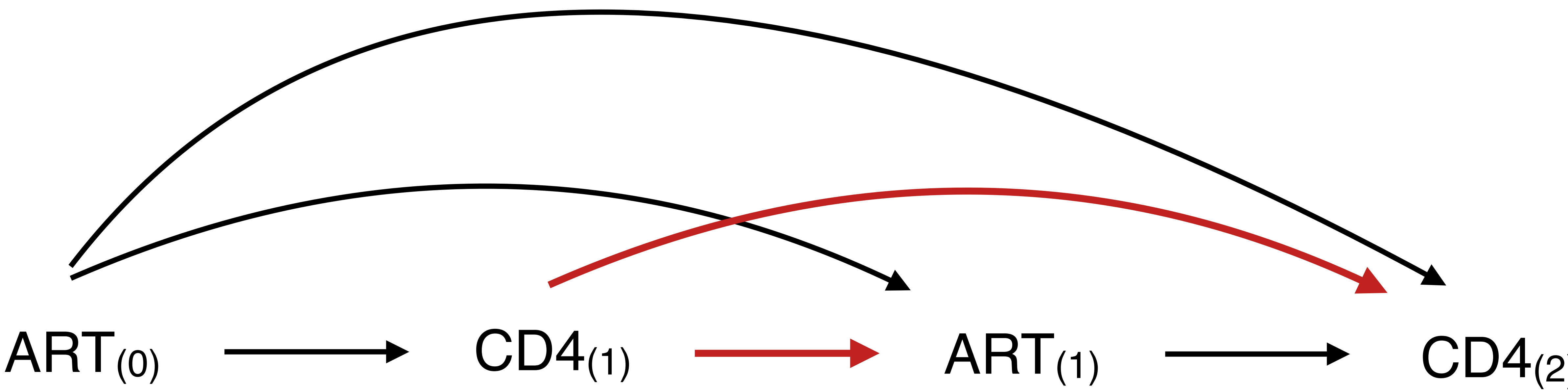
hard questions are harder

- mediation
- **longitudinal data with time-varying confounders**
- etc
    - networks of interconnected subjects
    - unmeasured confounding
    - generalizability / transportability
    - …

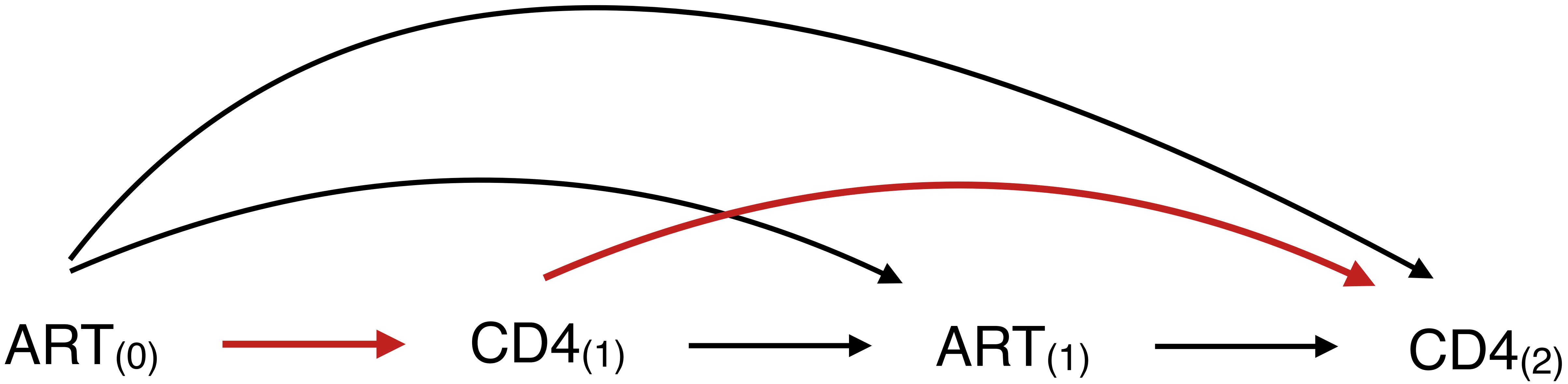
time-varying confounders



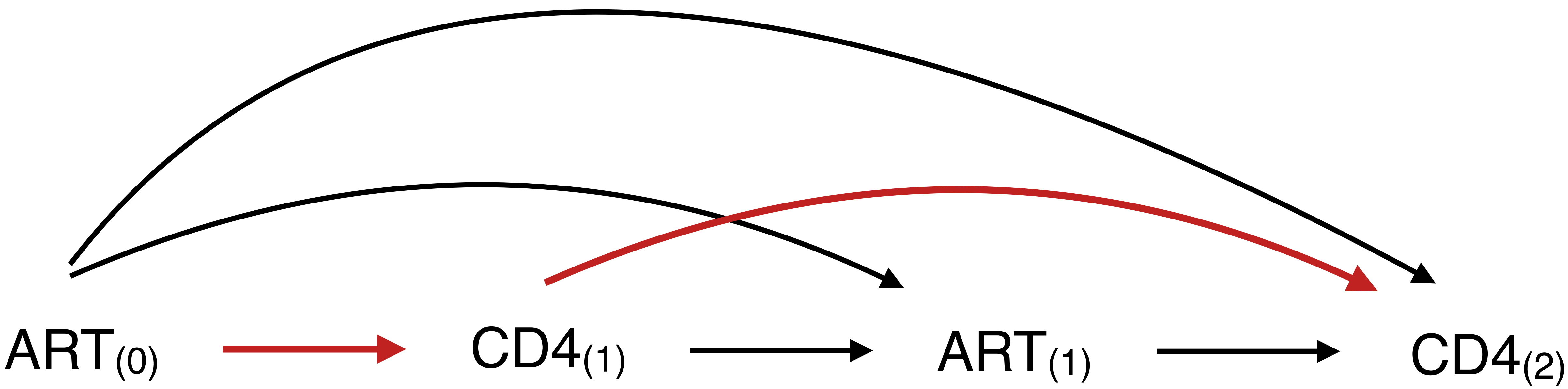
time-varying confounders



time-varying confounders



time-varying confounders



Applications include:

- analyzing large cohort studies like the Nurses Health Study about the effects of different interventions on diet and exercise,
- estimating the effect of aspirin on cardiovascular mortality,
- developing adaptive treatment strategy for adolescent depression,
- personalized medicine.

hard questions are harder

- mediation
- longitudinal treatments with time-varying confounders
- **etc**
  - **networks of interconnected subjects**
  - **unmeasured confounding**
  - **generalizability / transportability**
  - **…**

assessing uncertainty

Two sources of uncertainty can affect confidence in a causal analysis:

- statistical uncertainty, which can be captured by a confidence interval or a predictive interval or a standard error

- uncertainty about whether the fundamental assumptions hold, which can sometimes be assessed with sensitivity analyses

in conclusion

Standard statistical models are really good at estimating relationships, but the relationships they estimate aren’t always causal.

If you want to learn about causal effects, the hard and most important work is thinking about **assumptions**.