# Introductory causal inference journal club

Joanna Diong

# Paper 1. Hernán MA. The C-Word: scientific euphemisms do not improve causal inference from observational data.

"The proscription against the C[ausal]-word is harmful to science because causal inference is a core task of science [in many instances], regardless of whether the study is randomized or non-randomized."

How a research question is asked reflects its aim:

1) Do the sorts of people who drink a glass of red wine daily have lower risk of heart disease?
2) Does drinking a glass of red wine daily lower the risk of heart disease?

Q1 is interested in association.
Q2 is interested in causation.

# Paper 1.

Being explicit about scientific aims, even when using observational data, improves study design by helping to

- specify the exact causal effect of a research question (e.g. exposure, dose, frequency, population, etc.), which guides the analysis
- better adjust for confounding, the distorted association between variables, using subject-matter knowledge and a plausible causal structure

Even if there's no guarantee a causal model includes all confounders, "we can [only] have an informed scientific discussion ... if we first acknowledge the causal goal of the analysis."

# Discussion

Comments or clarifications?

Does anyone disagree with Hernán (e.g. do you think using 'association'/'correlation'/'link' words are better)? Why?

How are RCTs imperfect, and how are these limitations overcome?

How else are observational data limited at allowing causal inference? Do these problems only affect observational studies?

# Paper 2. Lipsky AM, Greenland S. Causal directed acyclic graphs.

A causal directed acyclic graph

- shows direction of plausible causal effects between exposure, outcome and other variables
- has directed (e.g. E -> O, E -> M -> O) and non-directed (e.g. E <- C -> O, E -> S <- O) paths
- is complete when it includes any variable that has arrows into each possible pair of variables along paths from cause to effect

# Paper 2.

- Association flows along directed and non-directed paths
- Causal association only flows along directed paths
- To identify (then estimate) causal associations, block all non-directed (backdoor) paths
- Do this by
  - adjusting/conditioning on confounders
  - leaving colliders alone

# Discussion

Any comments or clarifications? Terms:

- exposure/treatment
- assumption
- (statistical) adjustment/conditioning/control

Sometimes, a variable can be both a confounder and a collider. Special methods are needed to deal with this.

Sometimes, conditioning on colliders can happen inadvertently (e.g. loss of participants before study completion forces analysis to condition on selection of participants). Methods are needed to deal with selection bias.

1. In practice, how do you build a DAG?
2. What have reviewers challenged when reviewing DAGs?
3. How do you know a DAG is correct?