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?

Paper 3. Butler et al. Upper limb function but not proprioception is impaired in essential tremor.

Causal mediation analysis is a technique to estimate to what extent the causal effect of an exposure on an outcome is mediated through other paths

To do mediation analysis, plausible causal mechanisms are specified in a causal graph, and adjustments are made for potential confounding.

The analysis then partitions the average total effect of the exposure on outcome into an average effect acting through the mediator, and an average direct effect.

Paper 3.

Fig 2 of paper shows separation of the total effect of essential tremor on physical function, into indirect effects through proprioception and tremor amplitude, and a direct effect.

In this causal structure, the 2 mediators (proprioception, tremor amplitude) are causally dependent (i.e. they are connected by an arrow)

- When mediators are causally dependent, the average causally mediated effect (ACME) assumes interactions between the primary mediator and the outcome are independent of the exposure (i.e. causal effects of mediator on outcome don't depend on levels of exposure)
- This assumption can't be verified using data, so
- sensitivity analyses are performed to assess, if this assumption is violated, to what extent the ACME is prone to bias.

This study was the first application of causal mediation analysis in this field, and provided a better alternative to traditional approaches using correlations to infer causation.

Discussion

Comments or clarifications?

- Total effect: Between-group difference of effect of exposure on outcome
- Average causally mediated effect (ACME): Between-group difference of effect that acts through mediator
- Average direct effect (ADE): Between-group difference of effect that does not act through mediator
- Average causally mediated effect in the treated (ACME treated): Between-group difference of effect, only in the treated/exposed
- Average causally mediated effect in the control (ACME control): Between-group difference of effect, only in the controls

What is the value of ACME control?

We used R's `multimed` function in the `mediation` package; it's stable and has good documentation What other approaches or software do people use?