Causal Inference III:

Miscellany

Vincent Dorie

Columbia University

January 10, 2019

Instrumental Variable Analysis

When an instrumental variable exists, another form of analysis can be done that avoids problems with confounding

An instrument is a variable that must be:

- Relevant: the IV causes a change in the treatment received
- Effective random assignment: the IV is independent of unmeasured confounding conditional on covariates as if it was randomly assigned conditional on covariates
- Exclusion restriction: the IV does not have a direct effect on outcomes, i.e., it only affects outcomes through the treatment

Instrumental Variable Analysis Continued

If an instrument exists, can get around unmeasured confounding

1. find a valid instrument
2. use this variable to extract variation in the treatment that is free of the unmeasured confounders
3. use this confounder-free variation in the treatment to estimate the causal effect of the treatment

$$E[Y(1) - Y(0) \mid C = \text{complier}] = \frac{E[Y \mid A = 1] - E[Y \mid A = 0]}{E[Z \mid A = 1] - E[Z \mid A = 0]}$$
$$= \frac{\text{Cov}(Y_i, A_i)}{\text{Cov}(Z_i, A_i)}$$

A - IV, Z - treatment, Y - response, C - compliance group

Compliance Categories

An IV estimate only works on people who respond to the instrument

	Instrument $A_i = 1$	Instrument $A_i = 0$
Treatment $Z_i = 1$	Complier/Always-taker	Defier/Always-taker
Treatment $Z_i = 0$	Defier/Never-taker	Complier/Never-taker

- Have to assume there are no defiers (monotonicity)
- Connects to any compliance scenario and "Intent to Treat" (ITT) estimate
- Also known as the Local Average Treatment Effect (LATE) for compliers

IV Pros and Cons

- Helpful if unmeasured confounding is a major concern, but not strong confounding overall
- Need a valid instrument
- Need IV to be strong or else the result is highly variable

In practice, very common to use linear regression to estimate expected values, Huber-White robust cluster standard errors even with no clustering variable

Model only identifies the average treatment effect for compliers and compliers differ substantially from defiers (poor/better health)

Sensitivity to Unmeasured Confounding/Hidden Bias

Imagine the existence of an unmeasured confounder (U) and see how strong its effect on treatment and response would have to be to invalidate a finding/drive the estimate to 0/not significant

- For simple models, can analytically or brute force evaluate values of U by modifying treatment assignment or response
- Requires one or more parameters that control how U influences both functions
- For complicated models, simulation based approaches are common

Rosenbaum's Γ

Rosenbaum's Γ applies to binary outcome, binary treatment models and adjusts the odds of receiving treatment

- Modify $\log \text{odds}(Z = 1) = f(x) + \gamma u$, f is unknown function, γ is a parameter, and u an unobserved 0 or 1 confounder; take modified odds and compute treatment effect estimates
- If i and j are two individuals with the same covariates and $\Gamma = e^{\gamma}$ such that $\Gamma^{-1} \leq \text{odds}(Z_i = 1)/\text{odds}(Z_j = 1) \leq \Gamma$
- $\Gamma = 1$, there is no hidden bias; if $\Gamma = 2$, then of two otherwise identical units one is twice as likely to receive the treatment

In a study of smoking data, in order to explain the higher death rate from lung cancer to U rather than smoking, U would need a $\Gamma = 6$ times increase in the odds of smoking and be a near perfect predictor of lung cancer

Variable Selection

Unbiased estimation of causal effects requires ignorability, however it is rare that all confounders have been measured

What variables should be included?

- Some argue including all pre-treatment variables (Rubin, Rosenbaum)
- Only those that are related to treatment assignment mechanism (DAgostino, Jr, Hernan, Robbins, Pearl)
- Or the outcome of interest (Brookhart, Austin, Hill)

Public back and forth between big names; don't condition on:

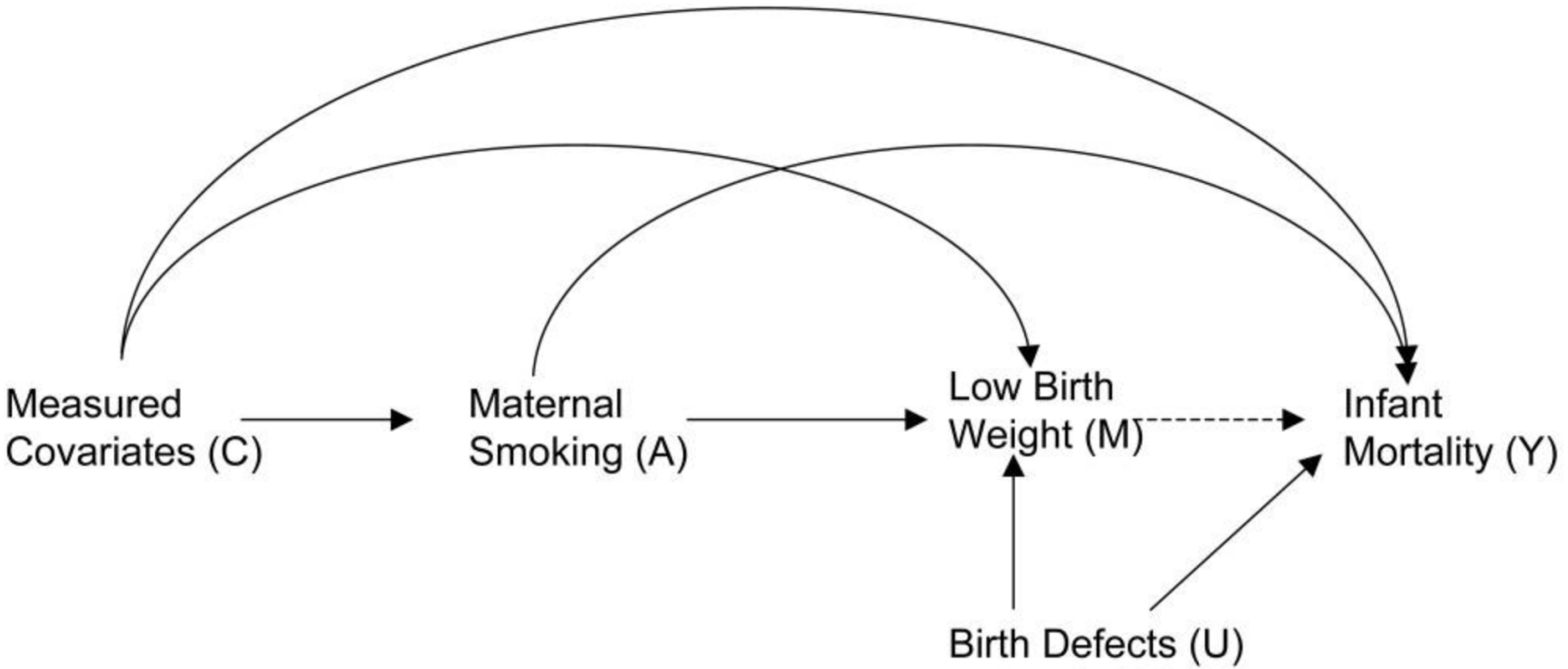
- *bias inducers*, directly add bias, e.g. post-treatment variables/mediators or colliders
- *bias amplifiers*, modify bias due to unobserved confounding, e.g. instruments

Birth-Weight Paradox

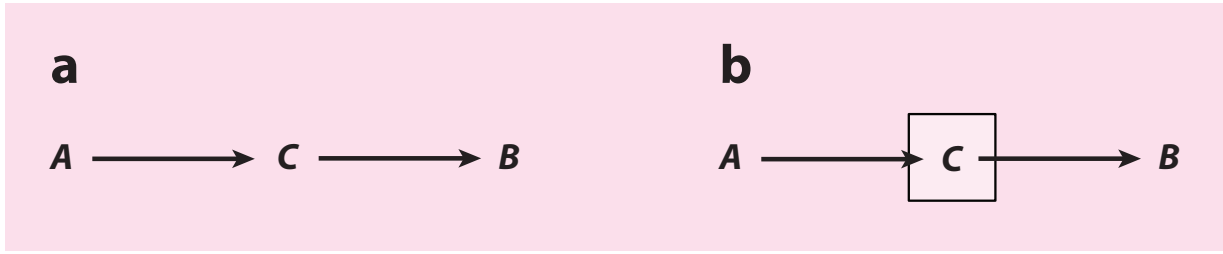
Among low birth-weight infants (intermediate variable), maternal smoking (treatment) is associated with decreased mortality (outcome)

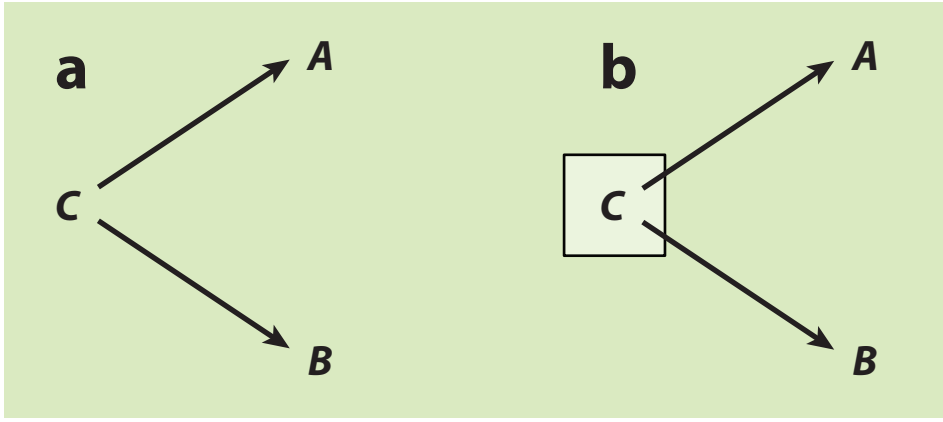
- For smoking, low birth-weight is caused by smoking and birth defects
- For non smoking, low birth-weight is caused only by birth defects
- If birth defects aren't controlled, bias is introduced as in as not smoking and low birth-weight indicates serious complications

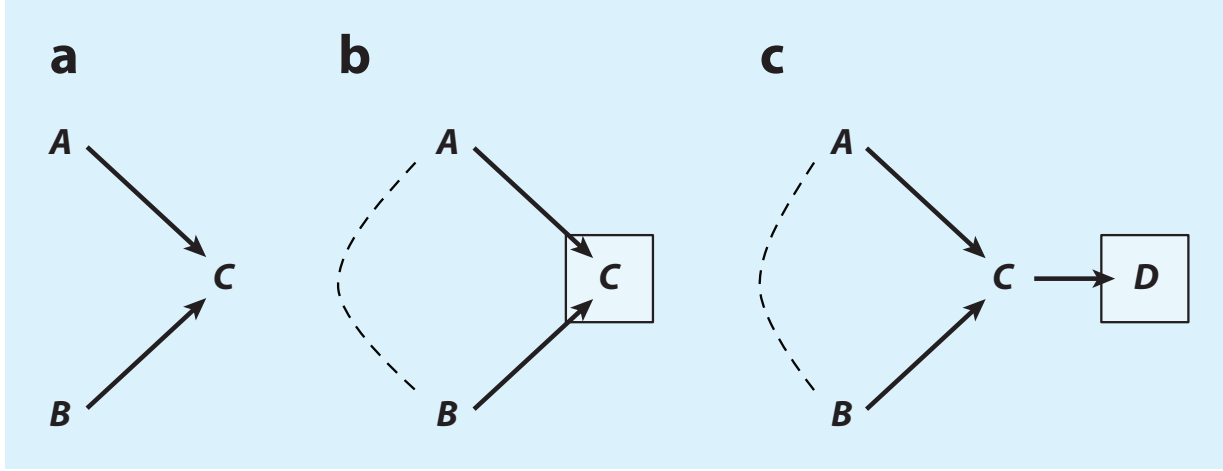
Birth-Weight Paradox Continued



VanderWeele, Mumford, and Schisterman







Elwert & Winship

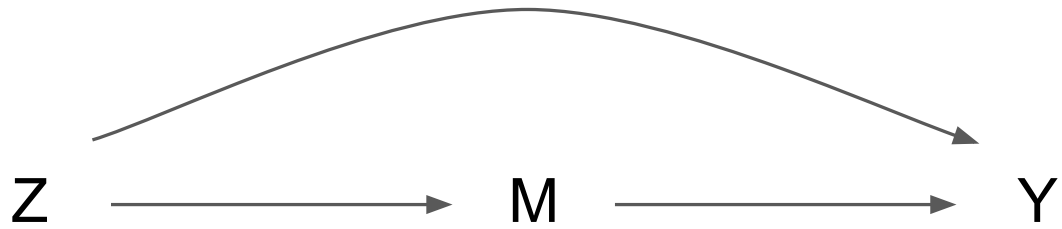
Varible Selection in Practice

- Don't include post-treatment variables
- If you believe that all confounders are measured, condition on everything pre-treatment
- If you don't believe all confounders are measured, don't condition on anything that you know isn't a confounder but try to include as much as possible (the induced bias is generally less than the bias from omission, but not always)

Many practitioners look at the relationships variables have with treatment and response and try to identify the set of true confounders, ignoring everything else

Causal Effect Mediation

What to do if you're interested in a mediator: *model it*, don't condition on it



- Response is $Y(Z, M(Z))$
- Natural direct effect (NDE) for unit i is $Y_i(1, M(0)) - Y_i(0, M(0))$
- Natural indirect effect (NIE) for unit i is $Y_i(1, M(1)) - Y_i(1, M(0))$
- Treatment effect is NDE + NIE

For every effect of interest, requires potential counterfactuals

Causal Effect Mediation Continued

Example with confounders:

$$E[Y \mid Z, M, X] = \beta_0 + \beta_1 Z + \beta_2 M + \beta_3^\top X$$
$$E[M \mid Z, X] = \gamma_0 + \gamma_1 Z + \gamma_3^\top X$$
$$E[Y \mid Z, X] = \beta_0^* + \beta_1^* Z + \beta_3^{*\top} X$$

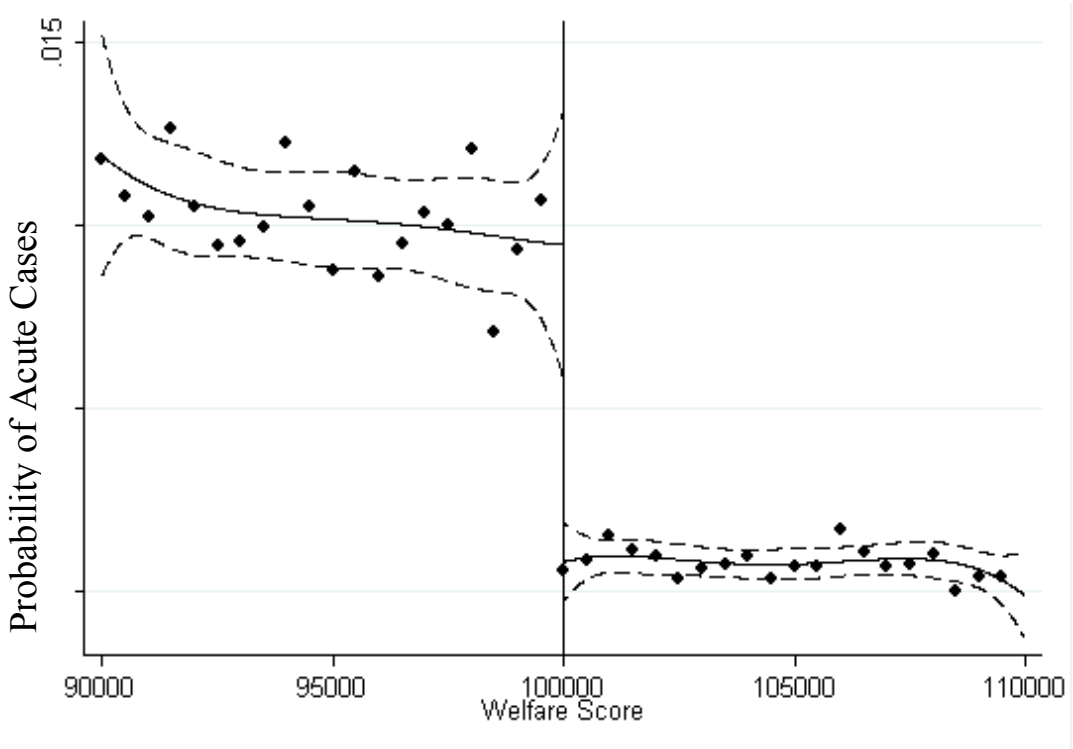
Derive effects from algebra and coefficients, if not linear use estimating equation jointly on (β, β^*) model simultaneously

Donna Spieglman

Regression Discontinuity

Occurs when a treatment is given based on a cutoff

Without extrapolating/assuming, treatment effects can in only be estimated for those directly at the cutoff



Hou and Chao

References

- Rosenbaum - Design of Observational Studies
- Pearl - Causal Inference in Statistics: A Primer
- Imbens and Rubin - Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction
- Hernan and Robins - Causal Inference, forthcoming