measurement error in causal inference

ENAR session 51738 Washington DC 2017/03/13

discussion

by Trang Quynh Nguyen

Department of Mental Health
Johns Hopkins Bloomberg School of Public Health

Bias due to/affected by covariate measurement error

Hong & Stuart: confounders w/ correlated measurement error; mismeasured confounder & correlated non-confounding covariates

Kim & Steiner: measurement error in observed confounders (& near-IVs) affecting omitted variable bias due to unobserved confounders

Bias correction when exposure is mismeasured

Braun, Kioumourtzoglou & Dominici: removing bias when there is measurement error in an aggregate ordinal exposure variable

Hong & Stuart (HS)

- Positively correlated covariates – PS weighting estimation (or PS weighting & outcome regression)
- Situation 1: confounders with same-direction confounding, positively correlated measurement error
  - Higher correlation between confounders -> better balance -> less bias
  - Higher measurement error correlation -> worse balance -> more bias
- Situation 2: confounder X1 plus covariates X2 (influencing outcome) and X3 (influencing treatment)
  - Using X2 helps reduce bias due error in measuring X1. Using W2 also helps, to a lesser extent.
  - When already using W2 (or X2), adding X3 or W3 doesn't seem to help, but actually increases bias slightly. And adding X3 hurts slightly more than adding W3.
- Highlights complexity of how measurement error affects bias and cautions about which auxiliary covariates should (not) be used to help reduce bias

Kim & Steiner (KS)

- Omitted variable bias (OVB), or bias due to unobserved confounding – linear regression estimation
- Adjusting for observed covariates has
  - bias reducing effects: removing their own confounding, and/or reducing unobserved confounding (if correlated with unobserved confounders)
  - bias amplifying effect: amplifying confounding by unobserved confounders due to explaining variance of exposure
- Measurement error in the adjusted covariates attenuates both these effects
- Special cases
  - No unobserved confounder: no bias amplification; measurement error attenuates the bias reducing effect
  - Adjusted covariates are IVs: no bias reduction; measurement error attenuates the bias amplifying effect

KS theory helps explain a result of HS simulation regarding X3 (influencing A but not Y)

- Adding X3
  - has bias reducing effect: due to correlation with X1, reduces bias due to measurement error in W1 – which could be thought of as OVB (Rudolph et al. under review)
  - and bias amplifying effect: due to explaining variance in A, amplifies the same bias
  - here: amplification > reduction
- Adding W3 results in less bias than adding X3
  - measurement error attenuates both bias reducing and bias amplifying effects
  - here: net amplification is reduced

Based on HS, may ask additional questions re X2 (influencing Y, not A) and a related covariate type

- What is behind the bias reduction when X2 (influencing Y only) or W2 is included in addition to W1?
  - Does it only have to do with the correlation between X2 and X1?
  - Or does the effect of X2 on Y matter?
- Should we include X2* variables that are correlated with X1 but are independent of Y conditional on X1?
  - Theoretically, should X2* be included in the PS model only, or both the PS model and the outcome model?
  - Practically can we tell the difference between X2* and X2?
  - If X2* (and thus W2*) independent of all other variables given X1, can treat W1 and X2* as multiple measurements of X1, and use the correction in Nguyen et al. (under review)

KS also point out two important points

- There may be confounding in opposite directions
- There are cases where a “mismeasured” covariate deconfounds the treatment-outcome relationship
    - “mismeasured” is a misnomer, because this variable
        - either is the confounder – it influences both A and Y
        - or is on the back-door path – it influences either A or Y

KS theory

The theory covers continuous variables and linear regression

w/ some extensions: binary exposure, matching/subclassification on an IV

Do you expect the finding of attenuation of bias reducing/amplifying effects to carry over to the case with misclassified binary confounder?

Braun, Kioumourtzoglou & Dominici (BKD)

- Ordinal zip-code PM2.5 X_z^{cat} –> zip-code health outcomes
- Adjust for measurement error using regression calibration (RC)
  - Based on an internal validation sample with monitor locations, recalibrate grid-cell exposure
  - Aggregate recalibrated grid-cell exposure to zip-code average
  - Then categorize into levels: adjusted zip-code level exposure
- Then estimate exposure effect using subclassification, IPTW, and matching based on the GPS (Yang et al. 2016)
- Simulation results: all three effect estimation methods using the calibrated $\hat{X}_z^{\text{cat}}$ remove bias

BKD brings up a very interesting measurement error structure

- Measurement error occurs at a different level than the analysis level, and is adjusted for at that level.
  - relevant to analyses where analysis variables are area-aggregated variables
  - could this be more generally relevant where an analysis variable is a composite variable? e.g., school-average academic performance measured by average test score, SES index with three domains, depressive symptom score with sub-dimensions
- Misclassification of categorical variable due to measurement error in the underlying continuous variable
  - may be relevant for a range of problems, e.g., T-CD4 < 50, < 100, < 200 often thought of as indicating different levels of immune suppression
  - misclassification depends on error structure: e.g., if unrestricted range but retains location and increases spread, leads to increased classification in categories to the two ends; may be very different with bounded range

- Adjustment for means vs. for original data points: the target of measurement error adjustment is X_z, not X_g, even though the correction starts with X_g
    - $\hat{X}_g$ is the predicted mean of X_g, therefore generally different from X_g, even if the calibration model is correct
    - $\hat{X}_z$ (the mean of $\hat{X}_g$ in a zip code) is closer to X_z than $\hat{X}_g$ is to X_g, and if each zip code includes many grid cells, may get close to X_z
- Scale matters for exposure, but not covariate:
    - Recalibration using X|W (but not C) for a covariate does reduce bias (Webb-Vargas et al. 2015)
    - In the current case, recalibration using X|W for an exposure reduces bias. M my guess is it gets the variable back to the scale of X

Thoughts/reactions from the speakers

Comments/questions from the audience