Implications of and solutions for covariate measurement error and differential covariate measurement across treatment groups

Elizabeth A. Stuart
Johns Hopkins Bloomberg School of Public Health
Departments of Mental Health and Biostatistics

estuart@jhsph.edu
www.biostat.jhsph.edu/∼estuart
Funding thanks to NIMH 1R01MH099010

Thanks to Yenny Webb-Vargas, Hwanhee Hong, Kara Rudolph, and David Lenis

May 4, 2015

Covariate measurement error

- Lots of attention in causal inference about unmeasured confounders and ensuring that we measure as many confounders as possible
- But what about mis-measured confounders?
- Covariate measurement error can cause some of the same problems as fully unobserved confounders
  - In fact can be thought of as a type of unmeasured confounding, where the "true" covariate is not observed

What sorts of covariate measurement error might exist?

- Weight: May have self-report rather than actually measured
- Blood pressure: Noisy measure of biological state
- Depression: Can't be directly observed, instead use scales to get at underlying construct
- Clinical status: May only have claims data on services received, not on actual clinical condition
- Sometimes things may even be measured differentially across treatment groups
  - e.g., combining two datasets (treatment and comparison), with different depression scales in each

What are the consequences of this measurement error?

What can we do about it?

Outline

Big picture

- Goal: Estimate the causal effect of receiving one treatment relative to a comparison condition
- Non-experimental studies use naturally occurring groups of individuals, some who got the treatment and some who got the comparison condition
- Problem is potential "selection bias"
- Approaches, such as propensity scores, try to limit selection bias by adjusting for (or matching on) covariates before estimating effects
  - Will focus on propensity score weighting today
  - Separation of "design" and "analysis": Outcomes not (typically) used in the propensity score process

The standard assumption underlying propensity score analyses

- Most propensity score analyses rely on assumption of unconfounded treatment assignment:
  - $T \perp (Y(0), Y(1))|X$
  - Given the observed covariates, no unobserved variables related to treatment assignment and outcomes
- What if treatment assignment actually depends on true X but all we observe is a mis-measured version of it, W?
  - e.g., decision to take a new treatment depends on true underlying health status, but all we have are proxies for it
  - e.g., decision to take a new treatment depends on blood sugar levels, but all we have are claims data

Outline

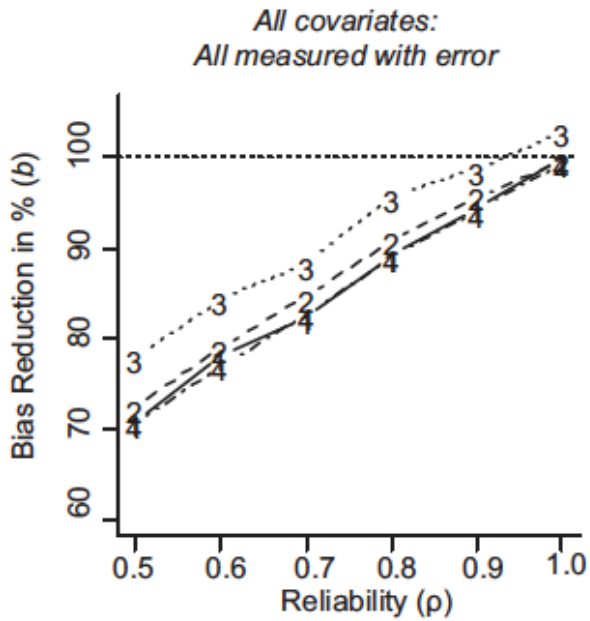
Reduced control for covariates

- Intuitively, if covariates measured with error, then aren't getting as much control for (true) covariates as expected
  - Residual confounding
  - Think you're adjusting for confounder (say blood pressure) but in fact only adjusting for a noisy version of it
- Relatively little investigation of how much trouble this causes
- Analytic results available for simple case with a single covariate measured with error
- Results in more general settings more difficult
  - Depends on reliability of the covariates, bivariate correlations of covariates with treatment, bivariate correlations between covariates, bivariate correlations with outcomes

Consequences of covariate measurement error

Steiner et al. (2010) used simulation to see how much LESS bias reduction is attained when covariates measured with error

- If covariates related to treatment and outcome, then measurement error can lead to a lot of residual bias
- If covariates measured with error unrelated to treatment and outcome, then no problems created



Outline

Some potential solutions for covariate measurement error

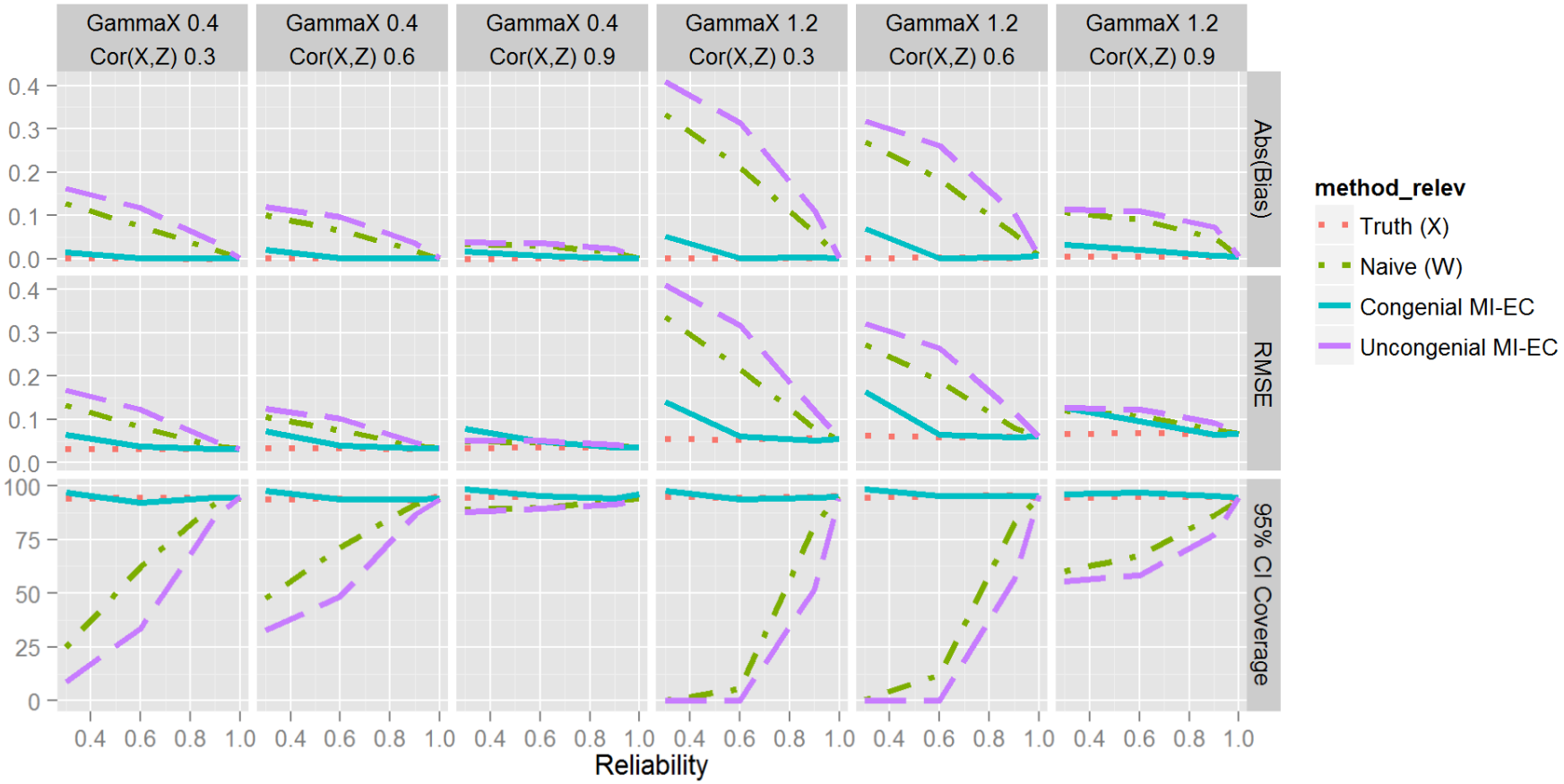
- Latent variable approach (Raykov, 2012)
- Corrected propensity score weighting strategy (McCaffrey et al., 2011)
- Propensity score calibration (Sturmer et al., 2005)
- Empirical expressions for resulting bias (Ogburn and VanderWeele, 2013)
- Simulation-Extrapolation (SIM-EX; McCaffrey and Lockwood, 2014; Lenis et al., 2015)
- Bayesian model for differential measurement error (Hong et al., 2015)
- Multiple imputation approach (Webb-Vargas et al., 2014)

The multiple imputation approach

- Main idea: Use a source of information on the relationship between W and X to multiply impute values of X
  - Intuitively, should account for uncertainty in imputations of X
- For now, will assume that we have some external validation sample with data on X and W (and possibly other common variables Z)
- Using "multiple imputation-external calibration," generate imputations of X given the other covariates, treatment and outcome
- Once imputations generated, run propensity score approach within each imputed dataset, combine effect estimates across imputed datasets

(Note: X=true covariate, W=mis-measured version)

Simulation results



Summary of simulation results

- Ignoring the measurement error leads to bias
- More bias if mis-measured covariate strongly related to treatment assignment
- Less bias if covariates strongly correlated
- Less bias if not a lot of measurement error (high reliability)
- Using MI-EC can correct for most of the bias
- But using an uncongenial MI-EC (with only Z [no outcome or treatment]) worse than naive approach
    - Need to include treatment and outcome in the imputation process for good performance

Outline

Extensions

- These methods may help us make better use of existing data
- Strategies to handle differential measurement error across datasets
  - e.g., when combining two data sources
- Utilize as many measures as possible, even if not measured the same
  - Previously no tools available to help do this combining
  - (Connections with "integrative data analysis" and "measure harmonization")

Considerations for study design

- Researchers used to thinking about which variables are completely missing for analyses (unobserved confounders)
- But measurement error can cause some of the same problems
- Need to better understand when and how much to worry
  - May not be a big deal if covariates correlated and one measured with error not a strong confounder
- Statistical methods can be used to help correct for the measurement error, especially if a validation sample available
  - Multiple imputation and SIMEX-based approaches both seem promising
- Do need to think carefully about the need to include the outcome in the imputation; violates the separation of "design" and "analysis"

References

- http://www.biostat.jhsph.edu/~estuart/propensityscoresoftware.html
- Two-day short course on propensity scores in JHSPH summer institute (2016): http://www.jhsph.edu/departments/mental-health/summer-institute/courses.html
- Stuart, E.A. (2010). Matching Methods for Causal Inference: A review and a look forward. *Statistical Science* 25(1): 1-21
- Liu, W., Kuramoto, S.K., and Stuart, E.A. (2013). An Introduction to Sensitivity Analysis for Unobserved Confounding in Non-Experimental Prevention Research. *Prevention Science* 14(6): 570-580. PMCID: 3800481.
- Webb-Vargas, Y., Rudolph, K.E., Lenis, D., Murakami, P., and Stuart, E.A. (2014). Applying multiple imputation for external calibration to propensity score analysis. Johns Hopkins University, Department of Biostatistics Working Papers, Working Paper 269. http://biostats.bepress.com/jhubiostat/paper269