A Cautionary Tale: Instrumental Variable Validation and Covariate Balance

Thomas E. Love, Ph. D.

thomaslove@case.edu

NV Dawson, CL Thomas and RD Cebul

Center for Health Care Research & Policy

Case Western Reserve University

Joint Statistical Meetings – 2004 Toronto

Why I'm Here

- I do a lot of causal analyses of observational studies in health services research. I use propensity scores and sensitivity analyses frequently, and think they are useful tools.
- I'm no economist.
- As a statistician, my instinct is to emphasize the importance of assumptions when thinking about causal models.

Right Heart Catheterization and SUPPORT

- SUPPORT: Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments
- 5735 seriously ill adults admitted to an ICU in one of five U.S. teaching hospitals.
- Exposure: right heart catheterization in the first 24 hours after study admission.
- Key Finding: RHC use significantly **increased** 30-day mortality rates, after PS adjustment.

  See Connors AF et al. (1996) *JAMA 276*, 889-897

Instrumental Variables: The Basic Idea

- Designed to estimate causal effects in the presence of unobserved covariate differences.
- The toughest part is identifying the instrument.
- Stage 1: Instrument is used to predict exposure independent of unobserved selection effects.
- Stage 2: Outcome differences compared in terms of predicted exposure, not actual receipt.

Angrist, Imbens & Rubin (1996) *JASA* 434, 444-454.

Choosing the Instrument

- Our goal is to identify variables, called instruments, that are related to treatment but not to outcomes other than through their effect on treatment.
- Our instrument must predict RHC status, with no residual predictive power on 30 day mortality, after controlling for covariates.
- Example: Region of the country? Site?

Cherry-Picked Example

- We have two sites with similar outcomes, but very different rates of exposure for their patients with acute respiratory failure:

Location	Site A	Site B
% receiving RHC [Exposure]	50.2	33.9
% dead at 30 days [Outcome]	30.7	31.0

This is a straw man.

What Do We Estimate?

- Local Average Treatment Effect can be estimated for all "marginal patients" (sometimes called "compliers")
  - would receive RHC if they lived in a high RHC site, but wouldn't receive RHC if they lived in a low RHC site
  - Can't identify compliers from observed data

My Pre-Existing Biases

- IV has long history in econometrics – "bad" data.
- Attractive because they mirror RCT – instrument should adjust for overt AND hidden biases
- Local Average Treatment Effect is sometimes more interesting than PS-based estimates.
- Some specific examples (noncompliance) look well-suited for instruments.
- Awfully difficult to justify key IV assumptions…

IV Assumption 1: SUTVA

- Stable Unit Treatment Value Assumption
  - Potential RHC exposure and mortality for each patient are assumed to be unrelated to the RHC status of all other patients.
  - Certainly practice patterns vary across geographic areas, but still pretty debatable (as usual) – high-RHC sites may also provide increased access to other exposures.

See Landrum and Ayanian (2001) or Angrist, Imbens and Rubin (1996)

IV Assumption 2: Non-zero Causal Effect of Exposure

- The instrument must predict RHC exposure.
  - Can check this out (to some degree) in data.
  - Likelihood of receiving RHC care was indeed significantly associated with site, comparing A to B.
  - However, this **is** a cherry-picked example.

Location	*Site X*	*Site Y*	*Site Z*	**Site A**	**Site B**
% RHC	*~42*	*~29*	*~23*	**50.2**	**33.9**

IV Assumption 3: Exclusion Restriction

- The instrument is assumed to have no effect on outcomes other than through its effect on exposure.
  - Which site you are in doesn't affect outcome?
  - If this is true, why worry about profiling?
  - Looks more plausible comparing A to B, but including X, Y and Z…

Location	*Site X*	*Site Y*	*Site Z*	**Site A**	**Site B**
% Mortality	*~37*	*~25*	*~26*	**30.7**	**31.0**

IV Assumption 4: Monotonic effect of instrument on exposure

- If a patient in site B (low RHC) received RHC, then they would also have received RHC if they had been in site A (high RHC)
- This assumption essentially goes unchecked – it's hard to see a good reason for concern here.

IV Assumption 5: Ignorable Assignment of the Instrument

- Patients from the two sites were similar (in terms of observed and unobserved characteristics) to what they would have been had site been randomly assigned.
  - Also can't verify this directly, but ...
  - If patients stratified by the instrument (site) are similar in terms of observed characteristics, this provides some suggestive evidence…

So Why Am I Here, Really?

- To illustrate the importance of checking residual imbalance for important measured covariates in validating instrumental variables.
- I have a long-standing interest in checking this assumption, because I use related ideas all the time in thinking about propensity score analyses.
- I think checking the balance of propensity for RHC (here I have a swell model, previously validated) across sites is a natural, useful idea.

Checking Baseline Characteristics (Before Site Stratification)

	Sites A & B		
Variable	RHC	No RHC	Stdzd Diff (%)
Age	60.3	58.3	**12.3**
Heart Rate	120	115	**12.7**
2M surv pred	.609	.627	**-10.0**
Male?	.531	.619	**-14.4**

25 of 58 covariates show absolute standardized differences ≥ 10%

Checking Baseline Characteristics (After Site Stratification)

	Site A			**Site B**		
Variable	RHC	No RHC	Stdzd Diff (%)	RHC	No RHC	Stdzd Diff (%)
Age	58.6	57.5	**6.8**	68.4	60.4	**51.4**
Heart Rate	123	116	**17.2**	105	111	**-12.6**
2M surv pred	.624	.631	**-3.7**	.540	.618	**-45.1**
Male?	.539	.602	**-10.3**	.494	.658	**-27.3**

Site A: 23 of 58 show |Stdz D| ≥ 10%; Site B: 31 of 58.
In total, 32 of 58 show |Stdz D| ≥ 10% in at least one site.
24 of 25 initially unbalanced show **worse** problems in at least one site.

Comparing Propensity for RHC: Similar Across Sites?

	Site A			**Site B**		
Variable	RHC	No RHC	Stdzd Diff (%)	RHC	No RHC	Stdzd Diff (%)
Logit of Propensity for RHC exposure based solely on patient characteristics	0.212	-0.979	**101.4**	-0.074	-1.300	**99.1**

Results

- Here, site is clearly *not* a valid instrument for RHC exposure, based on the observed lack of balance within sites for measured covariates.
- If, despite this imbalance, site is used as an instrument, 30-day mortality is estimated to decrease by **1.8 percentage points** for patients receiving RHC among the patients for which the site determined treatment.

Conclusions

- Researchers selecting instruments to estimate causal effects need to carefully demonstrate balance in key covariates related to outcome.
- Bivariate comparisons of instrument to exposure and outcome are insufficient to assess validity.

Three Additional Useful Resources

- Landrum MB, Ayanian JZ (2001) Causal Effect of Ambulatory Special Care on Mortality Following Myocardial Infarction: A Comparison of Propensity Score and Instrumental Variable Analyses. *Health Services & Outcomes Research Methodology*, *2*, 221-245.
- Posner MA et al. (2001) Comparing Standard Regression, Propensity Score Matching, and Instrumental Variables Methods for Determining the Influence of Mammography on Stage of Diagnosis. *HS&ORM*, *2*, 279-290.
- Normand SLT et al. (2001) Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: A matched analysis using propensity scores. *J of Clin Epidemiol*, *54*, 387-398.