Comparative Effectiveness Research (CER): Promises and Pitfalls of Observational Data

HealthTechNet

November 12, 2010

Jack Hadley, Ph.D.

College of Health and Human Services

George Mason University

GEORGE MASON UNIVERSITY

Where Innovation Is Tradition

Comparative Effectiveness Research

The Promise – Electronic health records and health IT have the potential to provide access to a wealth of information on medical care use and health outcomes.

The Pitfalls – Failure to adjust for potential "observational data bias" can lead to seriously incorrect inferences and discredit the whole enterprise.

Outline

- Brief Background
- Definitions
- Statistical Problems and Methods
  - (but not a statistics seminar)
- Two Examples (and a few numbers)
  - Prostate Cancer Treatments
  - Medicare Spending Variations

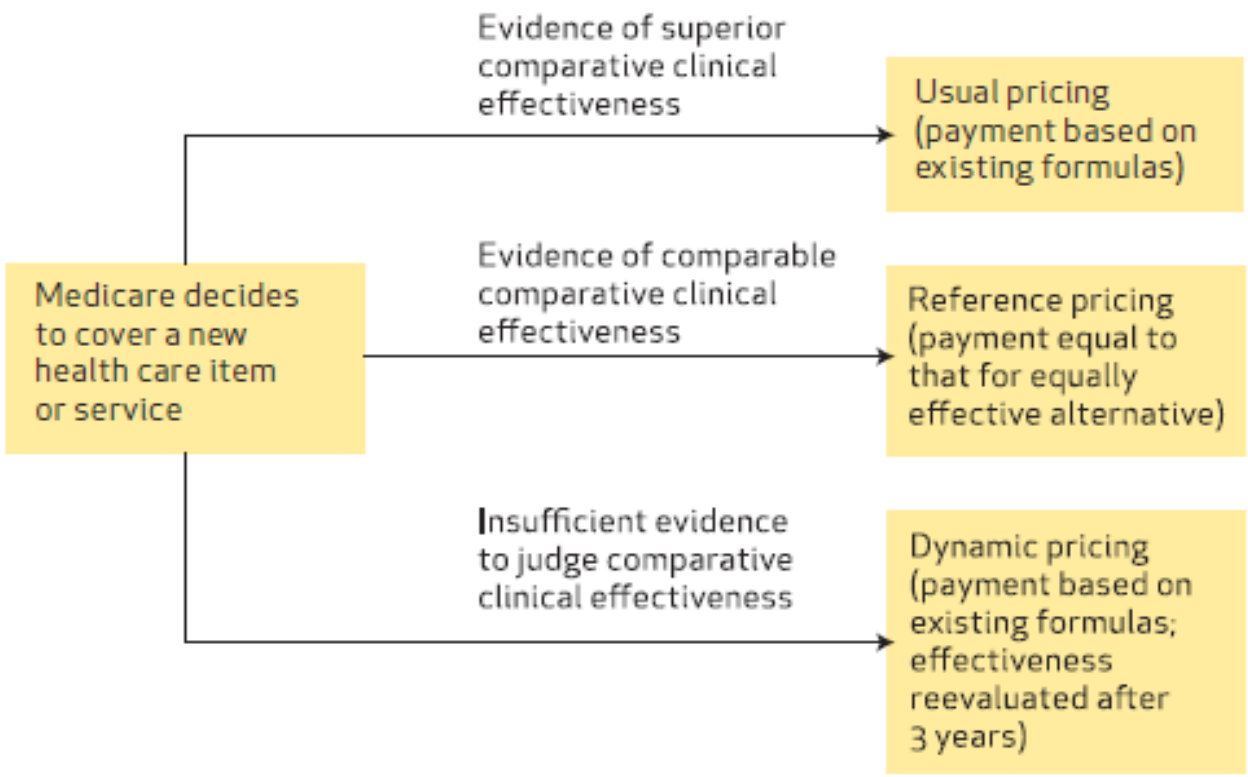
CER Is a High Priority

$1.1 billion in 2009 ARRA legislation

2010 PPACA establishes Patient Centered Outcomes Research Institute (PCORI) to develop methodologies (about $650 million per year by 2014)

Could be an used to vary reimbursements and make coverage decisions

Using Comparative Effectiveness Research To Determine Medicare Coverage And Reimbursement



Source: Pearson and Bach, Health Affairs, Oct. 2010

Definitions

Comparative effectiveness:

What works best?

Which intervention/treatment has the largest positive effect on health (and for which population)?

Observational data:

Data **NOT** generated by a randomized trial--insurance claims, medical records, survey responses, administrative data

Examples of Health Outcomes & Treatments

Health Outcomes

- Mortality
- Survival time
- Quality-adjusted survival
- Health scale or index
- Specific health indicators (blood pressure, weight, t-cell count)

Intervention/treatment

- Specific alternatives (e.g., surgery vs radiation, drug A vs drug B)
- No care
- Standard care
- Amount of care (dose response; more or less medical spending)

Why Not Randomized Trials?

RTs are the gold standard, but have several potential problems

- Can be very costly and time consuming
- Not always feasible (ethical constraints against potential patient harm or withholding of benefits)
- Potentially atypical populations; limited generalizability
- Potential technical problems
  - People self-select into trial
  - Differential drop-out rates across arms
  - Maintaining strict adherence
  - Maintaining comparability across sites in multi-site RTs

Why Observational Data?

- Relatively cheap
- Plentiful – less likely to be underpowered
- Potentially cover full spectrum of clinical settings—real world care situations rather than controlled settings of RTs
- Cover full range of patient characteristics—allows for extensive subgroup analysis

Problems w/ Observational Data

Patients self-select into treatment, presumably based on expected outcome

Reverse causality—poor initial health causes more spending and worse outcomes

Unobservable characteristics (health, economic, cultural) may have differential effects on outcome

Randomization eliminates self-selection and equalizes (usually) both observable and unobservable characteristics

Three Statistical Approaches

Multivariate regression

- controls for observable characteristics
- outcome = f(treatment, measures of observable characteristics)

Propensity score adjustment

- makes populations "look alike" if they are substantially different,
- but does not control for unobservables

Instrumental variable (IV) analysis

- common econometrics tool that has potential to control for unobservables;
- needs to satisfy key conditions;
- sometimes referred to as "pseudo" or "as if" randomization

Instrumental Variable Analysis

"Instrument" is a set of variables that:

- Has a conceptually plausible and significant effect on who gets which treatment (or how much care is received),
- But is uncorrelated with the health outcome
- Note - randomization is a perfect "instrument"

Instrumental Variable Analysis

A two-stage process (similar to Propensity Score analysis)

- Estimate a 1st-stage model to predict probability of treatment (or amount of care) as a function of instrument(s) and all other exogenous variable—**N.B. the presence of the instrument distinguishes IV from PS analysis**

- Estimate 2nd-stage outcome model substituting predicted treatment (or amount of care) for actual treatment—**The predicted treatment represents the pseudo-randomization inherent in IV analysis**

Some Examples of Potential Instruments

- Policy changes (government, health plans, employers)—similar to natural experiments
- Reimbursement rates or methods, e.g., FFS to capitation
- Distance to key providers, e.g., distance to a hospital with a cardiac cath. lab influences probability of heart attack patient receiving bypass surgery
- Local area treatment patterns, e.g., people who live in areas with more surgeons are more likely to get surgery
- Individual physician's treatment propensities, i.e., how physician typically treats other patients with same condition

Comparative Effectiveness of Prostate Cancer Treatments: Evaluating Statistical Adjustments for Confounding in Observational Data

Jack Hadley, K. Robin Yabroff, Michael J. Barrett, David F. Penson, Christopher S. Saigal, Arnold L. Potosky

Source: JNCI J Natl Cancer Inst Advance Access10.1093/jnci/djq393

GEORGE MASON UNIVERSITY

Where Innovation Is Tradition

Study Design

- Compare surgery (radical prostatectomy) and watchful waiting (no treatment)
- Data from SEER-Medicare for 14,302 elderly (66-74) men diagnosed between 1995-2003
- Sample selected to mimic sample from a Randomized Trial—use results as benchmark to compare alternative statistical approaches
- Lagged area (53 hospital referral regions) treatment propensity for watchful waiting as the instrument – average difference between actual and predicted probability of watchful waiting for all cases in geographic region

Selected Hospital Referral Regions	Mean Value of Lagged IV—Difference between Actual and Predicted Pct. Watchful Waiting
Areas with High Propensity for Surgery	
Baton Rouge/Metairie LA	-0.124
Cedar Rapids/Dubuque IA	-0.096
Tacoma WA	-0.078
Orange CA	-0.074
Davenport IA	-0.057
Atlanta GA	-0.047
Provo/ Salt Lake City UT	-0.035
Los Angeles/Ventura CA	-0.033
Areas with a "Neutral" Propensity	
Sacramento CA	-0.019
El Paso/ Lubbock TX	-0.016
Seattle WA	-0.013
Des Moines IA	-0.002
Honolulu HI	0.006
Alameda CA	0.013
Bridgeport CT	0.028
Detroit MI	0.029
Areas with a High Propensity for Watchful Waiting	
Hartford CT	0.033
Albuquerque NM	0.046
San Diego CA	0.051
San Francisco CA	0.058
Newark NJ	0.078
San Jose/San Mateo CA	0.091
Iowa City IA	0.101
New Haven CT	0.194

Patient Age & Comorbidities, by Treatment and Statistical Approach

Variable	Unweighted Observational Data		Propensity Score Reweighted		Below/Above Median Value of Instrumental Variable	
	Surg.	WW	Surg.	WW	Below	Above
Actual Treatment (%WW)	0%	100%	0%	100%	27.1%	35.4%
Age 65-69 (%)	53.2	44.1	50.4	50.6	51.0	49.8
No Comorbidities (%)	75.4	57.8	70.1	69.9	70.6	69.1

GEORGE MASON UNIVERSITY

Where Innovation Is Tradition

Mean Values of Outcome Variables by Treatment and Statistical Approach

Estimation Method and Treatment	Death from Prostate Cancer	Death from Any Cause
	% Died	% Died
All Cases (unweighted)	0.028	0.200
Observational (unweighted)		
Surgery	0.025	0.177
WW – watchful waiting	0.036	0.249
Propensity Score Reweighted		
Surgery	0.026	0.181
WW – watchful waiting	0.035	0.236
Instrumental Variable		
Below median value of instrument	0.027	0.192
Above median value of instrument	0.030	0.208*
RCT (12 years of follow-up)		
Surgery	**0.131**	**0.420**
WW – watchful waiting	**0.132***	**0.393***

* No statistically significant difference

GEORGE MASON UNIVERSITY

Where Innovation Is Tradition

Adjusted Hazard Rates (Watchful Waiting vs Surgery), by Estimation Method

Estimation Method	Adjusted Hazard Rate (p-value)	
	Death from Prostate Cancer	Death from Any Cause
RCT (12 years of follow-up)	**0.87 (.55)**	**1.04 (.81)**
Observational (unweighted)	1.59 (<.01)	1.47 (<.01)
PS Reweighted	1.60 (<.01)	1.54 (<.01)
Instrumental Variable	0.73 (.78)	1.09 (.84)

Medicare Spending, Mortality, and Quality of Care

Data – 1.5 million elderly fee-for-service beneficiaries located in 60 Community Tracking Study sites in 2006

- nationally representative
- 50 MSAs and 10 groups of nonmetro counties
- Analyzed high- and low-cost cases separately

Spending measure - price-adjusted Medicare spending for all services during prior 12 months

Outcome measures

- Mortality
- Likelihood of being hospitalized for an ambulatory care sensitive (ACS) condition

Includes detailed controls for medical conditions (HCC risk adjustment measures)

Instruments

- Area rank (1 to 60) on Medicare spending per beneficiary
- Area medical market structure and supply characteristics

Hypothesis – More Medicare spending reduces the likelihood of an adverse health outcome

Source: Jack Hadley and James Reschovsky (in progress)

Simulated Percentage Changes in Health Outcomes, by Underlying Estimation Method

Outcome Measure (N=1,561,722)	Baseline Rate (% of population with outcome)	Impact of a 10% Increase in Medicare Spending (pct. change)	
		Observational: Multivariate Regression (OLS)	Instrumental Variable (IV)
Died	7.89	1.3%*	-12.7%*
Hospitalized for an ACS (ambulatory care sensitive) Condition	5.11	0.8%*	-6.8%*

GEORGE MASON UNIVERSITY

Where Innovation Is Tradition

* Underlying coefficient has a p-value < 0.10

Conclusions

- Observational data bias can be substantial
- Appropriate statistical adjustment in comparative effectiveness and outcome studies is essential
- IV analysis is a robust approach---if you can find a conceptually plausible and statistically strong instrument
- If you can't, then shouldn't do study using conventional statistical methods

Future Directions

- Explore whether machine learning (artificial intelligence) methods can substitute for or improve parametric statistical analyses; may be especially useful with very rich clinical data
- Investigate issue of patient heterogeneity—results for the "average" patient don't necessarily apply to all patients
- Develop methods (data text mining) for measuring side effects and quality of life dimensions of health outcomes