Comparative Effectiveness Evaluation and Monitoring: An American Perspective

Jeffrey Smith
Paul T. Heyne Distinguished Chair in Economics
University of Wisconsin-Madison

Austrian Embassy
Washington, D.C.
November 13, 2018

Outline of talk

What sites do

What is the point of the measures?

Remarks on choosing the set of conditioning variables

The technocratic ideal

Applied econometric issues

Opportunity knocks!

What do sites do?

Recruiting
Selecting
Testing
Informing
Assigning
Motivating
Following up
Monitoring

What is the point of the adjusted outcome measures?

Monitoring?
Description?
Performance management?
Evaluation?
Identifying outlier sites (on the high side or the low side)

More than one of these?

The goal matters for the design and use of the system that creates and interprets the adjusted outcome measures

An aside on comparative effectiveness

What does effectiveness mean?

Long-run impacts?
Short-run impacts?
Something else?

What do the adjusted outcome measures compare?

Not different programs, but different versions of the same program operated in different locations

Conditioning variables: participants or eligibles?

Existing systems condition on participant characteristics

But sites choose these! In particular, they choose:
    1) Where and how to market their programs
    2) How to triage potential participants who apply

To really "hold sites harmless" the models should adjust for the characteristics of the eligible population rather than of participants

Conditioning variables: training type

Is this wrong? Depends what you want to do with the system.

Training type represents a "mediator" in program evaluation jargon

Conditioning on it removes that aspect of the sites' performance

Conditioning variables: providers

Even within a state, sites differ in the nature, number and quality of available service providers

Not all community colleges are created equal!

Particularly but not entirely an urban / rural issue

Should find a way to quantify this and include it in the model

Conditioning variables: conditional independence

Do the US administrative data contain variables that would suffice to make conditional independence plausible?

Technically, conditioning on some but not all of the variables required for conditional independence need not reduce the bias relative to no conditioning

Does some but not full conditioning suffice to get the ranks right?
Are ranks enough, given the purpose of the system?

Should program staff understand the adjustments?
    SRI study of the JTPA adjustments
    How much can program staff control participant characteristics?

Other substantive issues

Estimate separate models for each state?
    What do the data have to say?
    Perhaps groups of similar / nearby states as an intermediate position

How many years of data to use?
    How stable are the coefficients? Again, let the data speak.
    How many sites have low numbers of participants per year?
    Should the analysis combine small sites?

What about services received outside the program?
    Not an issue in Austria, very much an issue in the US
    Use WIA experiment survey data to get a handle on this

Someone mislaid the costs … finding them means measuring them

The technocratic ideal

Do on-going random assignment to "core" services only

Similar to the set-up in the WIA experiment

No need to consent if part of on-going program?

Still only provides short-run impacts (in the short-run)

Applied econometric issues

"Sample (really the population) size" and shrinkage

No reason to worry overmuch about logit versus probit versus LPM

Opportunity knocks!

The point of the system matters for its design

Think hard about what to condition on and why

Worry about how local program stuff will understand the system:
　　Conditioning, causality, and cost-benefit

Many substantive and applied econometric challenges

Combine with systematic institutional variation among sites

A real opportunity here to improve on both the current system and the
JTPA system that included conditioning