The Implications of (Treatment Effect) Heterogeneity for Internal and External Validity

Jeffrey Smith
Paul T. Heyne Distinguished Chair in Economics
University of Wisconsin-Madison
econjeff@ssc.wisc.edu

Rigorous Impact Evaluation in Europe
A Conference in Honor of Alberto Martini
Torino - May 2018

The treatment effect

The treatment effect for unit *i* is given by $\beta_{Di} = Y_{1i} - Y_{0i}$

In the common effect model that dominates (by presumption) the literature,

$\beta_{Di} = \beta_D$ for all *i*

Substantively, when does this make sense?

Where does treatment effect heterogeneity come from?

In many contexts, the “treatment” is itself heterogeneous.

Example: active labor market programs

Aside: optimal treatment differentiation

In other contexts, the treatment is homogeneous but the responses are heterogeneous

Example: budget set treatment with heterogeneous opportunity costs of work

Differential take-up / dosage

Systematic versus idiosyncratic treatment effect heterogeneity

Djebbari and Smith (2008) *Journal of Econometrics*

Systematic: varies with observed characteristics
Idiosyncratic: the remainder

Link back to opportunity cost of work example

The division between systematic and idiosyncratic depends on the set of moderators available in the data

Treatment effect heterogeneity and internal validity

Generalizing within the population served or evaluated.

Even a compelling causal estimate may be a poor guide to program impacts on particular sub-populations among the treated.

Beware discussion of “the” treatment effect of a program or policy.

Tests and bounds

The literature – e.g. Heckman, Smith and Clements (1997) and Djebbari and Smith (2008) shows how to test the null of the common effect model.

A simple version of the test looks for equal quantile treatment effects.

The literature also shows how to estimate a lower bound on the variance of the treatment effects.

Both should be routine in experimental evaluations.

Heterogeneity and external validity

Why care about treatment effect heterogeneity?

1) Understanding how programs work (i.e. learn about mechanisms)

2) Effects of programs on inequality

3) Understanding program participation choices
    Selection on impacts by agents and/or caseworkers
    Do agents and/or caseworkers know about impacts?
    Link to external validity

4) Targeting / statistical treatment rules
    Note that these require systematic heterogeneity
    Link to external validity

Frame discussion in terms of subgroups but same issues arise in regard to program context and program implementation and operation

Statistical treatment rules

Example: US Worker Profiling and Reemployment Services System (profiles on levels)
Example: Response to Intervention (RtI)
Example: SMART (Sequential, Multiple Assignment, Randomized Trial) designs
Example: “Selective incapacitation” (profiles on levels); Bushway and Smith (2008) *JQC*

Basic idea: use a statistical model to assign individuals to treatment with the largest predicted impacts

See Smith and Staghøj (2008) working paper for a survey and various Manski papers, e.g. Manski (2004) *Econometrica*, for the conceptual framework.

Finding subgroups / moderators: theory

Important for understanding mechanisms

Ex: Rosenzweig on male / female differences in the impact of education

Can provide testable predictions

Ex: Bitler, Gelbach and Hoynes (2006) *AER* on Connecticut Jobs First

See also Weiss, Bloom and Block (2014) *JPAM*

Huge opportunities for research and measurement here

Finding subgroups / moderators: the literature

Men, women and training

A pattern without a model?

The “usual suspects” – but where do they come from?

Finding subgroups: machine learning

Using mechanical statistical procedures to identify statistically and substantively meaningful subgroup effects

Recent example: Davis and Heller (2017)

Modern procedures address fishing concerns

Still limited by available set of candidate moderators

General issues in looking for subgroup effects

Common support

Cannot estimate subgroup effects for subgroups not in the data

Confounding

Observed (not causal) subgroup may proxy for unobserved (causal) subgroup

Imagination

Always limited to the variables measured in the available data

See Hotz, Imbens and Mortimer (2004) *Journal of Econometrics* and Muller (2015) *WBER* for more discussion

External validity has important implications for the design of experiments (i.e. for initial site selection) and of non-experimental evaluations.

Are subgroup effects common?

This is often (implicitly) assumed in the literature

What if effects are heterogeneous within subgroups? Consider an example:

Half of men have impact 10 and half have impact 4
Half of women have impact 12 and half have impact 1

Assume that the cost of participation is five, so top half of both groups participate if agents know their impacts

Evaluation finds program “works better for women” so gender-specific subsidies are provided to induce the remaining women to participate ….

Conditional mean impacts on treated in general do not equal impact on marginal untreated person!

Are subgroup effects structural?

Structural = policy invariant

Subgroups effects may be common, or structural, or both, or neither

The estimated subgroup effect may change when the policy changes even if the distribution of treatment effects within groups is structural if the policy changes the program participation process.

Going deeper: is structural always a binary notion?

Summary and conclusions

Treatment effect heterogeneity has important implications for internal and external validity

Testing the common effect assumption and estimating the lower bound on the treatment effect variance should become standard in experimental evaluations

Lots of scope for better theories of subgroup impacts

Lots of scope for better measurement of potential moderators