PennState

Y-RISE

Meta-Analysis and External Validity

Michael Gechter

Pennsylvania State University, Y-RISE

June 2019

Outline of this module

1. **Defining external validity**
2. **When do standard meta-analyses deliver external validity?**
3. **A cautionary tale**: site selection bias in action
4. **Meta-analysis and external validity**: some ways forward

Outline of this module

1. **Defining external validity**
2. **When do standard meta-analyses deliver external validity?**
3. **A cautionary tale**: site selection bias in action
4. **Meta-analysis and external validity**: some ways forward

Defining external validity

- The concept dates back to at least Campbell (1957)
- Rachael talked about extrapolation error arising from different values of the same estimand across contexts
  - Ex. the underlying Average Treatment Effects of microcredit expansions differ
- We may want to make adjustments before attempting to generalize
  - Reweight subgroup Average Treatment Effects (CATEs) in reference context to match distribution of subgroups in target context (e.g. Hotz, Imbens, and Mortimer (2005))
  - Similarly, used a mixed model (e.g. Vivalt (2015))
  - Interpret differences in contexts through the lens of an economic model

This module: a “big tent” definition

External validity can be established for:

1. One or more reference contexts
2. One target context

Using:

1. A method for using the reference context to predict a feature of the target context
2. A measure of performance

This is distinct from another meta-analytical goal of quantitatively characterizing data from a set of studies

- “Quantitative literature review” perspective

Outline of this module

1. **Defining external validity**
2. **When do standard meta-analyses deliver external validity?**
3. **A cautionary tale**: site selection bias in action
4. **Meta-analysis and external validity**: some ways forward

Setup

- **Abstracting from sampling**
- C different contexts, indexed by $c \in \{1, \ldots, C\}$
- Individual i belongs to a context c

Treatments

- Binary treatments in each c
  - $T_{ic} = 0$: untreated
  - $T_{ic} = 1$: treated
- Potential outcomes framework
  - Y_{0ic}: untreated outcome
  - Y_{1ic}: treated outcome

$$Y_{ic} = T_{ic} Y_{1ic} + (1 - T_{ic}) Y_{0ic}$$

External validity

- Two sets of contexts, indexed by D_c
- 0: reference contexts included in the meta-analysis
- 1: target context, not included in meta-analysis

Using

1. A method for using the reference context to predict a feature of the target context. Assume:

$$E[Y_{1ic} - Y_{0ic}|D_{ic} = 1] = E[Y_{1ic} - Y_{0ic}|D_{ic} = 0]$$

2. A measure of performance

$$\text{External validity fails if } E[Y_{1ic} - Y_{0ic}|D_{ic} = 1] \neq E[Y_{1ic} - Y_{0ic}|D_{ic} = 0]$$

Using

1. A method for using the reference context to predict a feature of the target context. Assume:

$$E[Y_{1ic} - Y_{0ic}|D_{ic} = 1] = E[Y_{1ic} - Y_{0ic}|D_{ic} = 0]$$

2. A measure of performance

$$\text{External validity fails if } E[Y_{1ic} - Y_{0ic}|D_{ic} = 1] \neq E[Y_{1ic} - Y_{0ic}|D_{ic} = 0]$$

So,

- External validity fails if individual treatment effect $Y_{1ic} - Y_{0ic}$ is not mean-independent of the indicator for inclusion in the meta-analysis.

Using

1. A method for using the reference context to predict a feature of the target context. Assume:

$$E[Y_{1ic} - Y_{0ic}|D_{ic} = 1] = E[Y_{1ic} - Y_{0ic}|D_{ic} = 0]$$

2. A measure of performance

$$\text{External validity fails if } E[Y_{1ic} - Y_{0ic}|D_{ic} = 1] \neq E[Y_{1ic} - Y_{0ic}|D_{ic} = 0]$$

So,

- External validity fails if individual treatment effect $Y_{1ic} - Y_{0ic}$ is not mean-independent of the indicator for inclusion in the meta-analysis.
- Inclusion criteria are an important feature of a meta-analysis if the goal is external validity as defined above

Using

1. A method for using the reference context to predict a feature of the target context. Assume:

$$E[Y_{1ic} - Y_{0ic}|D_{ic} = 1] = E[Y_{1ic} - Y_{0ic}|D_{ic} = 0]$$

2. A measure of performance

$$\text{External validity fails if } E[Y_{1ic} - Y_{0ic}|D_{ic} = 1] \neq E[Y_{1ic} - Y_{0ic}|D_{ic} = 0]$$

So,

- External validity fails if individual treatment effect $Y_{1ic} - Y_{0ic}$ is not mean-independent of the indicator for inclusion in the meta-analysis.
- Inclusion criteria are an important feature of a meta-analysis if the goal is external validity as defined above
- One inclusion/screening criterion: observational studies may not identify $E[Y_{1ic} - Y_{0ic}|c]$ so exclude them

Site Selection Bias

- Now D_{ic} is an indicator for belonging to a site where an RCT of T_{ic} was conducted
- If a context's having a partner organization willing and able to conduct an RCT is related to the context Average Treatment Effect, external validity will fail
- Allcott (2015) calls this Site Selection Bias

Examples of Site Selection Bias

- "Gold plating": organizations agree to conduct an RCT if they have a particularly good program

$$E[Y_{1ic} - Y_{0ic}|D_{ic} = 1] < E[Y_{1ic} - Y_{0ic}|D_{ic} = 0]$$

- "Substitution bias": locations with the capacity to conduct RCTs have many social programs in place, supporting untreated outcomes

$$E[Y_{1ic} - Y_{0ic}|D_{ic} = 1] > E[Y_{1ic} - Y_{0ic}|D_{ic} = 0]$$

Outline of this module

1. **Defining external validity**
2. **When do standard meta-analyses deliver external validity?**
3. **A cautionary tale**: site selection bias in action
4. **Meta-analysis and external validity**: some ways forward

An alternative method

Perhaps our earlier method relied on an uncomfortably strong assumption.

An alternative method

Perhaps our earlier method relied on an uncomfortably strong assumption. Instead, use

1. A method for using the reference context to predict a feature of the target context. Assume:

$$E[Y_{1ic} - Y_{0ic}|D_{ic} = 1, X_{ic}] = E[Y_{1ic} - Y_{0ic}|D_{ic} = 0, X_{ic}]$$

2. A measure of performance

External validity fails if $E[Y_{1ic} - Y_{0ic}|D_{ic} = 1, X_{ic}] \neq E[Y_{1ic} - Y_{0ic}|D_{ic} = 0, X_{ic}]$

Allcott (2015) assesses this performance of this method

- T_{ic} denotes receipt of an Opower Home Energy Report
- Key features
  - Neighbor Comparison Module comparing household's energy use to its 100 geographically nearest neighbors in similar house sizes
  - The Action Steps Module including energy conservation tips targeted to the household based on its historical energy use patterns and observed characteristics.

UtilityCo

Home Energy Report

Account number: 1234567890
Report period: 12/01/12–01/31/13

We are pleased to provide this personalized report to you as part of an energy savings program.

The purpose of this report is to:

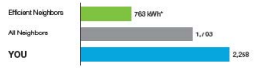
- **Provide information**
- **Track your progress**
- **Share energy efficiency tips**

This information and more available at **www.utilityco.com/reports**

John Doe
1235 Main St.
Bellevue, WA 98006

Last 2 Months Neighbor Comparison

You used **33% more** electricity than your neighbors.



* kWh: A 100-Watt bulb burning for 10 hours uses 1 kilowatt-hour.

How you're doing:

You used more than average

Turn over for ways to save →

All Neighbors: Approximately 100 occupied, nearby homes (avg 0.11 mi away)

Efficient Neighbors: The most efficient 20 percent from the "All Neighbors" group

Are we comparing you correctly?
Tell us more about your home:
www.utilityco.com/reports

Last 12 Months Neighbor Comparison

You used **30% more** electricity than your neighbors.
This costs you about **$246 extra** per year.

Turn over for savings →

Personal Comparison

How you're doing compared to last year:

* kWh: A 100-Watt bulb burning for 10 hours uses 1 kilowatt-hour.

So far this year, you used **8% MORE** electricity than last year.

Looking for ways to save? Visit **www.utilityco.com/reports**

Action Steps | Personalized tips chosen for your home

Smart Purchase
An affordable way to save more

☐ **Program your thermostat**

A programmable thermostat can automatically adjust your heat or air conditioning when you're away, then return to your preferred temperature when you're home to enjoy it.

If you don't already have a programmable thermostat, look for one at your local home improvement store. For comfort and convenience, be sure to program your thermostat with energy-efficient settings.

If you need help installing or programming your thermostat, consult your manual or call the manufacturer for assistance.

SAVE UP TO $80 PER YEAR

Smart Purchase
An affordable way to save more

☐ **Check your air filters every month**

You can improve the energy efficiency of your heating and cooling systems and improve your indoor air quality by checking your filters monthly.

First, remove the filter — it usually slides right out. Next, hold the filter up to a light to see if it is clogged.

You can find an inexpensive replacement for a clogged disposable filter at your local hardware store. Check your manual for cleaning instructions if you have a permanent filter.

SAVE UP TO $45 PER YEAR

Smart Purchase
An affordable way to save more

☐ **Seal air leaks**

Gaps and cracks between the inside and outside of your home can allow heated or cooled air to escape. This forces your heating or cooling system to work harder, increases energy costs, and decreases comfort.

To find leaks, follow drafts to their source. Check where materials meet, like between the foundation and walls, the chimney and siding, and where gas and electricity lines exit your house.

Seal any small cracks you find with caulk and larger ones with polyurethane foam.

SAVE UP TO $215 PER YEAR

This is a best case scenario for assessing this version of external validity

- Almost no treatment heterogeneity
- What there is (frequency of report mailing) can easily be modeled
- Sample sizes are very large

TABLE VI

OPOWER PROGRAM: PREDICTED EFFECTS USING MICRODATA

	(1)	(2)	(3)	(4)	(5)	(6)
	First 10 Sites			Later 101 Sites		
		Nonexperimental estimates			Prediction from first ten sites	
	True ATE	Pre-post	W/ state control	True ATE	Linear	weighted
Frequency-adjusted ATE (percent)	1.67	2.92	2.88	1.26	1.92	1.66
Standard error	0.06	0.08	0.08	0.04	0.08	0.06
Difference from true value (percent)	—	1.25	1.20	—	0.66	0.41
Value of difference in a nationally-scaled program (billion)	—	$1.72	$1.66	—	$0.92	$0.56

- Difference between predicted and measured ATEs for later 101 sites significant with p-values < .0001

Takeaways

- Looks like gold-plating
- We can argue about the importance of the magnitude
- But remember this case is idealized due to low treatment heterogeneity
- So the paper is best thought of as a counterexample

Outline of this module

1. **Defining external validity**
2. **When do standard meta-analyses deliver external validity?**
3. **A cautionary tale**: site selection bias in action
4. **Meta-analysis and external validity**: some ways forward

For external validity: new methods

- Maybe methods based on assuming

$$E[Y_{1ic} - Y_{0ic}|D_{ic} = 1, X_{ic}] = E[Y_{1ic} - Y_{0ic}|D_{ic} = 0, X_{ic}]$$

  don't perform well

- Unobserved heterogeneity relevant for treatment effects remains
- Even when considering individuals with the same x

For external validity: new methods

- Gechter (2016): later today, use differences in $F(Y_{0ic}|X_{ic})$ across reference and target to bound ATE in target context
- Andrews and Oster (2017): sensitivity analysis benchmarked to observed differences between reference and target
- Kowalski (2016) and related papers: linear marginal treatment effect-based extrapolation in settings with imperfect compliance
- Gechter and Meager (2018) (in progress): estimate site selection bias, use estimates to correct for internal selection bias in observational studies
- Attanasio, Meghir, and Szekely (2003): incorporate economic theory by way of a structural model, allows for modeling of treatment differences across contexts

For meta-analysis and external validity: more examples

- Dehejia, Pop-Eleches, and Samii (2017):
  - Examine Angrist and Evans (1998) effect of having two children of the same sex on subsequent fertility across $\approx$ 100 censuses in different countries
  - Country-level covariates are the most important for acheiving external validity
  - Bad news for predictions based on a more typical number of reference contexts (< 10)
- Many others (apologies for omissions): Hotz et al. (2005); Attanasio et al. (2003); Angrist and Fernández-Val (2013); Angrist and Rokkanen (2015); Cole and Stuart (2010); Stuart, Cole, Bradshaw, and Leaf (2011); Pearl and Bareinboim (2014); Vivalt (2017); Meager (2016)

For meta-analysis and external validity: evaluating methods (Gechter, Samii, Dehejia, and Pop-Eleches (2019))

- Consider a more policy-relevant measure of performance
- The ATE that would be acheived if adopting each method's policy recommendations
- Policy recommendation: treat individuals with $X_{ic} = x$ if method-predicted CATE $>$ cost-effectiveness threshold
- Framework can asses all the methods discussed so far
- Application to conditional cash transfer programs

Allcott, H. (2015). Site Selection Bias in Program Evaluation. *Quarterly Journal of Economics 130*(3), 1117–1165.

Andrews, I. and E. Oster (2017). WEIGHTING FOR EXTERNAL VALIDITY.

Angrist, J. and I. Fernández-Val (2013). ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework. In *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress, Volume III: Econometrics*. Econometric Society Monographs.

Angrist, J. and M. Rokkanen (2015). Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff. *Journal of the American Statistical Association 110*(512), 1331–1344.

Angrist, J. D. and W. Evans (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review 88*(3), 450–477.

Attanasio, O., C. Meghir, and M. Szekely (2003). Using Randomised Experiments and Structural Models for 'Scaling Up': Evidence from the PROGRESA Evaluation. *Mimeo*.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin 54*(4).

Cole, S. and E. Stuart (2010). Generalizing Evidence from Randomized Clinical Trials to Target Populations: The ACTG 320 Trial. *American Journal of Epidemiology 172*(1), 107–15.

Dehejia, R., C. Pop-Eleches, and C. Samii (2017). From Local to Global: External Validity in a Fertility Natural Experiment. *NBER Working Paper 21459*.

Gechter, M. (2016). Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India. *Working Paper*.

Gechter, M. and R. Meager (2018). Incorporating Experimental and Observational Studies in Meta-Analysis. *Working Paper*.

Gechter, M., C. Samii, R. Dehejia, and C. Pop-Eleches (2019). Evaluating Ex Ante Counterfactual Predictions Using Ex Post Causal Inference. *Working Paper*, 1–32.

Hotz, V. J., G. Imbens, and J. Mortimer (2005). Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations. *Journal of Econometrics 125*, 241–270.

Kowalski, A. E. (2016). How to Examine External Validity Within an Experiment. *Journal of Economic Perspectives*, 1–16.

Meager, R. (2016). Aggregating Distributional Treatment Effects : A Bayesian Hierarchical Analysis of the Microcredit Literature.

Pearl, J. and E. Bareinboim (2014). External Validity: From do-calculus to Transportability across Populations. *Statistical Science 29*(4), 579–595.

Stuart, E., S. Cole, C. Bradshaw, and P. Leaf (2011). The Use of Propensity Scores to Assess the Generalizability of Results from Randomized Trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 174*(2), 369–386.

Vivalt, E. (2015). How Concerned Should We Be About Selection Bias , Hawthorne Effects and Retrospective Evaluations ? pp. 1–40.

Vivalt, E. (2017). How Much Can We Generalize From Impact Evaluations? *Mimeo*.