BOLD THINKERS DRIVING REAL-WORLD IMPACT

50 YEARS

Using ASPES (Analysis of Symmetrically-Predicted Endogenous Subgroups) to understand variation in program impacts

Presented by:

Laura R. Peck

OPRE Methods Meeting on What Works
Washington, DC | September 3-4, 2014

Today’s Agenda

- Motivating Challenge & Solutions
- Logic & Execution of ASPES
- Illustration: Supporting Healthy Marriage program
- Ideal Conditions for ASPES
- Conclusion

Motivating Challenge

- Policy guidance requires "more than just an estimate of the net effects of a program or policy; it is also necessary to understand…the circumstances under which a program or policy has effects, and how and why it works." (OPRE Meeting Summary)
- Relevant policy questions:
  – What are the effects of participating in the intervention (not just being offered access)? How does variation in dosage affect program impacts?
  – How does exposure to various levels of program quality influence program impacts?
  – What is the effect of participating in component A (or B or C) of a multi-faceted intervention?

Possible Solutions…

… that *use* the experimental design

Analytic Approach
<u>Instrumental Variables</u> works for no shows: assumes the only pathway of randomization’s effect is through participation

Possible Solutions & Limitations

… that *use* the experimental design

Analytic Approach	Limitation
<u>Instrumental Variables</u> works for no shows: assumes the only pathway of randomization's effect is through participation	applicable only where we have two groups, and one experiences no effect of being offered treatment

Possible Solutions & Limitations

… that *use* the experimental design

Analytic Approach	Limitation
<u>Instrumental Variables</u> works for no shows: assumes the only pathway of randomization's effect is through participation	applicable only where we have two groups, and one experiences no effect of being offered treatment
<u>Propensity Score Matching</u> predicts status in one arm to find matched counterparts in other arm	omitted variables in prediction create inconsistent impact estimates

Possible Solutions & Limitations

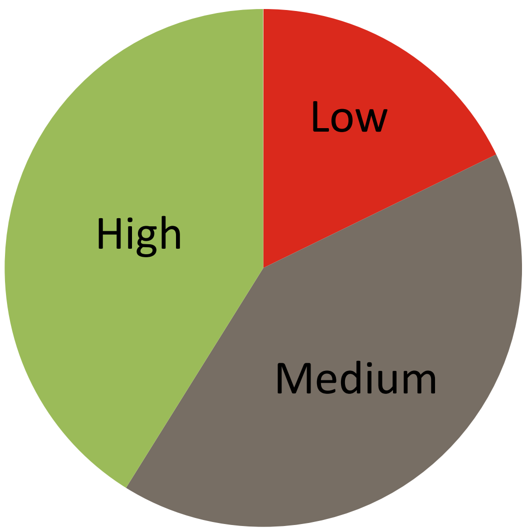
… that *use* the experimental design

Analytic Approach	Limitation
<u>Instrumental Variables</u> works for no shows: assumes the only pathway of randomization's effect is through participation	applicable only where we have two groups, and one experiences no effect of being offered treatment
<u>Propensity Score Matching</u> predicts status in one arm to find matched counterparts in other arm	omitted variables in prediction create inconsistent impact estimates
<u>Principal Stratification</u> uses status in each arm to predict and compare potential outcomes in other arm	applicable when subgroup is observed in both experimental arms

Analysis of Symmetrically-Predicted Endogenous Subgroups (ASPES)

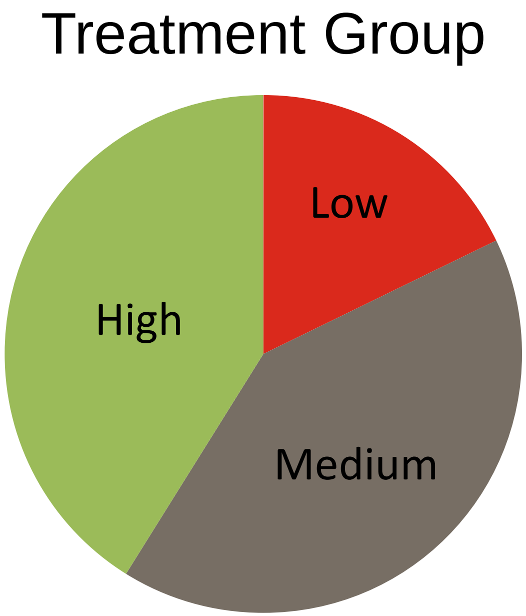
- Logic: any group in one experimental arm has a counterpart in the other arm

Treatment Group



Analysis of Symmetrically-Predicted Endogenous Subgroups (ASPES)

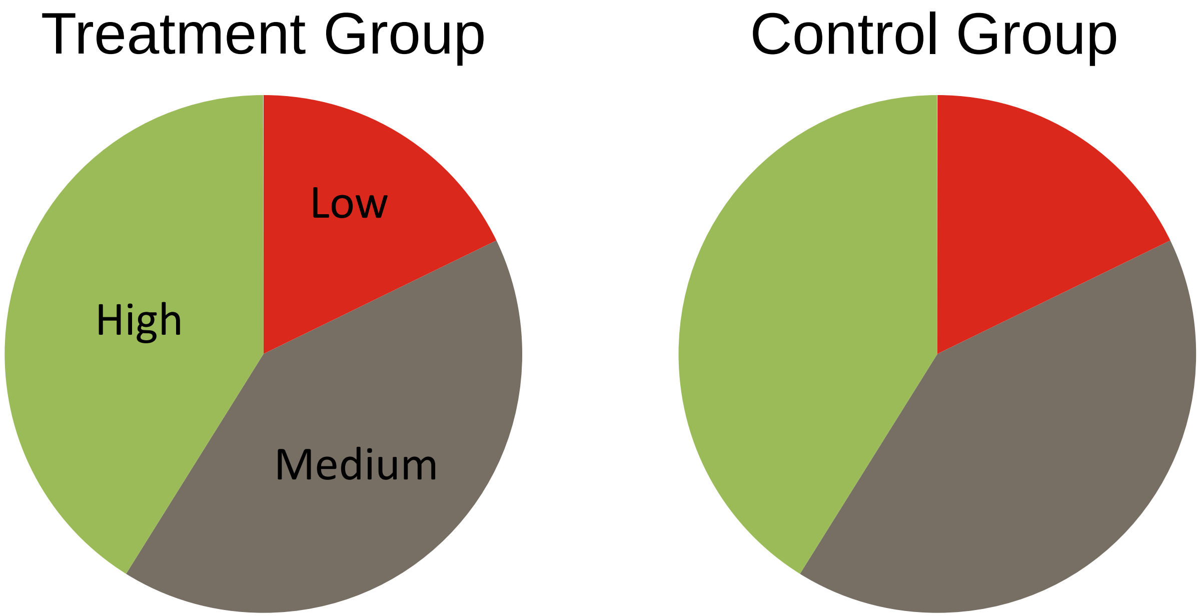
- Logic: any group in one experimental arm has a counterpart in the other arm



- Low, medium or high:
  - Dosage exposure
  - Quality experience
  - Likelihood of program component take-up
  - Risk of drop-out

Analysis of Symmetrically-Predicted Endogenous Subgroups (ASPES)

- Logic: any group in one experimental arm has a counterpart in the other arm

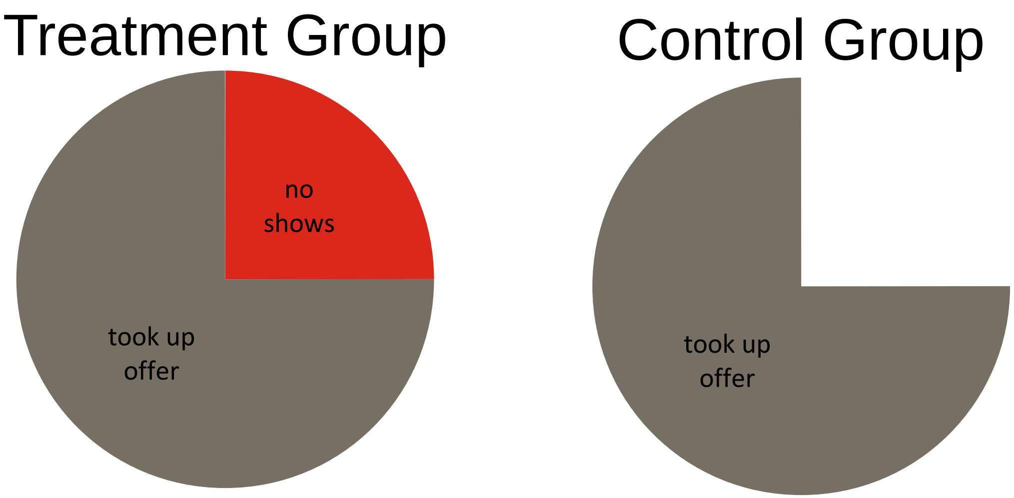


ASPES Defined

- Endogenous subgroups = post-assignment events, experiences
- Symmetrically-predicted = leverages experimental design in identifying subgroups

ASPES Defined & Compared

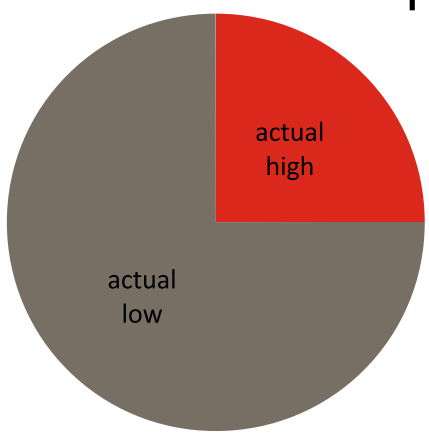
- Endogenous subgroups = post-assignment events, experiences
- Symmetrically-predicted = leverages experimental design in identifying subgroups
- As opposed to IV strategy (all impact among takers):



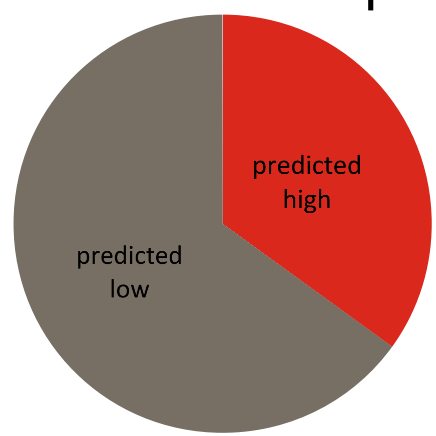
ASPES Defined & Compared

- Endogenous subgroups = post-assignment events, experiences
- Symmetrically-predicted = leverages experimental design in identifying subgroups
- As opposed to (asymmetric) PSM strategy:

Treatment Group



Control Group



Execution of ASPES

- Step 1: Use baseline (exogenous) characteristics to predict subgroup membership

Execution of ASPES

- Step 1: Use baseline (exogenous) characteristics to predict subgroup membership
- *Leveraging the experimental design*: use out-of-sample prediction (to ensure T and C subgroups are *symmetric*)

Execution of ASPES

- Step 1: Use baseline (exogenous) characteristics to predict subgroup membership
- Step 2: Estimate impacts on predicted subgroups

- *Leveraging the experimental design*: use out-of-sample prediction (to ensure T and C subgroups are *symmetric*)

Execution of ASPES

- Step 1: Use baseline (exogenous) characteristics to predict subgroup membership
- *Leveraging the experimental design*: use out-of-sample prediction (to ensure T and C subgroups are *symmetric*)
- Step 2: Estimate impacts on predicted subgroups
- *Achieving internal validity*: the difference between T and C subgroup mean outcomes is unbiased treatment effect (but for a blend of correctly and incorrectly predicted actuals)

Execution of ASPES

- Step 1: Use baseline (exogenous) characteristics to predict subgroup membership
- Step 2: Estimate impacts on predicted subgroups
- Step 3: Convert estimated impacts for predicted subgroups into impacts for actual subgroups

- *Leveraging the experimental design*: use out-of-sample prediction (to ensure T and C subgroups are *symmetric*)
- *Achieving internal validity*: the difference between T and C subgroup mean outcomes is unbiased treatment effect (but for a blend of correctly and incorrectly predicted actuals)

Execution of ASPES

- Step 1: Use baseline (exogenous) characteristics to predict subgroup membership
- *Leveraging the experimental design*: use out-of-sample prediction (to ensure T and C subgroups are *symmetric*)
- Step 2: Estimate impacts on predicted subgroups
- *Achieving internal validity*: the difference between T and C subgroup mean outcomes is unbiased treatment effect (but for a blend of correctly and incorrectly predicted actuals)
- Step 3: Convert estimated impacts for predicted subgroups into impacts for actual subgroups
- *Achieving external validity*: use assumptions to convert the experimental impact estimate

Execution of ASPES (cont.)

- Step 3: Convert estimated impacts for predicted subgroups into impacts for actual subgroups
  – Consider that the impact on predicteds is a weighted sum of the impacts on actuals, where the weights involve correct prediction rates:

$$\hat{I}_1 = p_1 I_1 + (1 - \mathrm{p}_1) I_2$$

$$\hat{I}_2 = p_2 I_2 + (1 - \mathrm{p}_2) I_1$$

Execution of ASPES (cont.)

- Step 3: Convert estimated impacts for predicted subgroups into impacts for actual subgroups
  - Consider that the impact on predicteds is a weighted sum of the impacts on actuals, where the weights involve correct prediction rates:

$$\hat{I}_1 = p_1 I_1 + (1 - \mathrm{p}_1) I_2$$

$$\hat{I}_2 = p_2 I_2 + (1 - \mathrm{p}_2) I_1$$

  - Use "homogeneity assumption" (and algebra) to solve for I

$$I_1 = \frac{p_2 \hat{I}_1 - (1 - p_1)\hat{I}_2}{p_1 + p_2 - 1}$$

$$I_2 = \frac{p_1 \hat{I}_2 - (1 - p_2)\hat{I}_1}{p_1 + p_2 - 1}$$

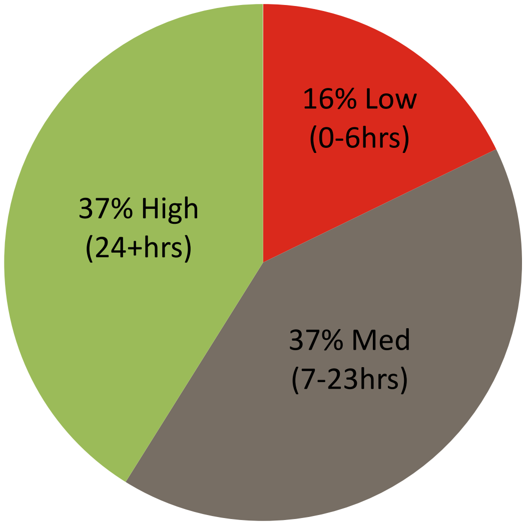
Illustrative Example: SHM

- What is the impact of the number of hours of SHM participation (which is endogenous) on couples' marital stability and relationship happiness?

Illustrative Example: SHM

- What is the impact of the number of hours of SHM participation (which is endogenous) on couples' marital stability and relationship happiness?

We know Ts' Dosage



Illustrative Example: SHM (cont.)

- Select random subsamples of the treatment group from which to predict SHM dosage levels: low (0-6 hrs), medium (7-23 hrs) and high (24+ hrs)
- Using baseline characteristics, predict dosage (used multinomial logit):
  - <u>Personal and Couple Characteristics</u> - earnings, education, age, race, ethnicity, children, psychological measures, marital tenure, communication, satisfaction measures
  - <u>Program Characteristics</u> - site dummies
- Use the resulting predicted dosage variable to symmetrically identify subgroups in the treatment and control groups, then compare groups' outcomes and convert

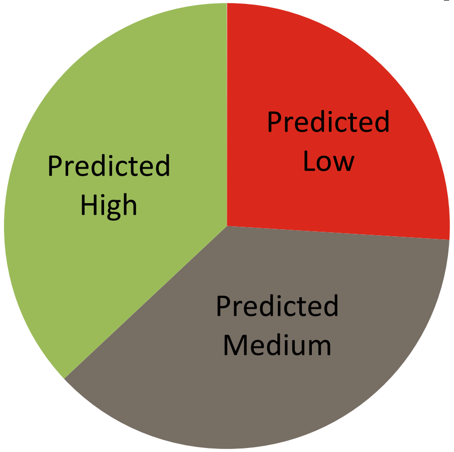
Illustrative Example: SHM (cont.)

Selected Outcomes → Impact Analyzed:	Relationship Status: % married		Average Happiness level	
Overall Study Sample	0.80		0.15	***
Predicted Low-Dosage Group	-0.40		0.12	
Predicted High-Dosage Group	2.20	***	0.15	***
between-group diffs	n.s.		n.s.	

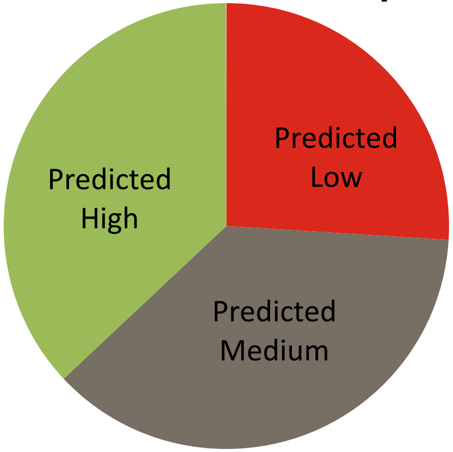
Illustrative Example: SHM (cont.)

Selected Outcomes → Impact Analyzed:	Relationship Status: % married		Average Happiness level	
Overall Study Sample	0.80		0.15	***
Predicted Low-Dosage Group	-0.40		0.12	
Predicted High-Dosage Group	2.20	***	0.15	***

Treatment Group



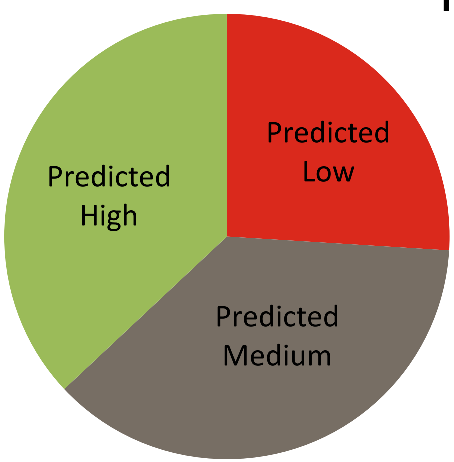
Control Group



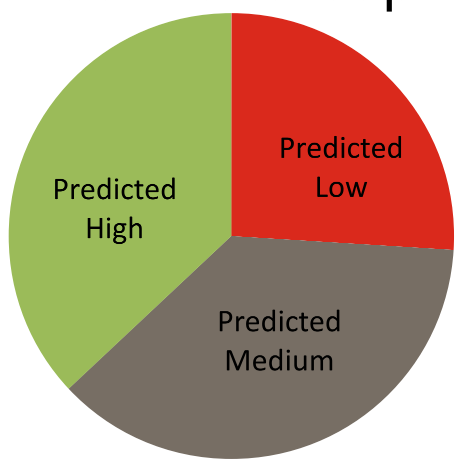
Illustrative Example: SHM (cont.)

Selected Outcomes → Impact Analyzed:	Relationship Status: % married		Average Happiness level	
Overall Study Sample	0.80		0.15	***
Predicted Low-Dosage Group	-0.40		0.12	
Predicted High-Dosage Group	2.20	***	0.15	***

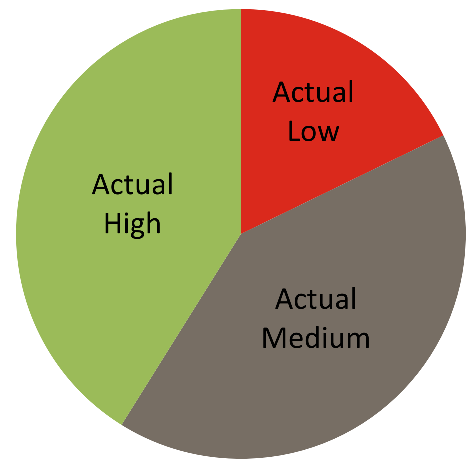
Treatment Group



Control Group



→ Convert to Actuals



Illustrative Example: SHM (cont.)

Selected Outcomes → Impact Analyzed:	Relationship Status: % married		Average Happiness level	
Overall Study Sample	0.80		0.15	***
Predicted Low-Dosage Group	-0.40		0.12	
Predicted High-Dosage Group	2.20	***	0.15	***
between-group diffs	n.s.		n.s.	
Actual Low-Dosage Group	-5.10		-0.07	
Actual High-Dosage Group	6.10		0.16	
between-group diffs	n.s.		n.s.	

Ideal Conditions for ASPES

- Baseline data
  - Standard: demographics, education, works/earnings history
  - “Unobservables”: motivation, behaviors, preferences
- First stage prediction “success”
  - Better than chance?
- Predicted → actual conversion assumption credibility
- Sample size
  - OLS → IV requires greater sample
  - 1,000 for 0.30ES within predicted subgroup → 12,500 for 0.30ES within actual (sample, prediction, noise dependent)

Conclusion

- Assumptions Tradeoffs
  - Exchange IV's exclusion restriction, for example, for ASPES homogeneity assumption

Conclusion

- Assumptions Tradeoffs
  - Exchange IV's exclusion restriction, for example, for ASPES homogeneity assumption
- Research Questions
  - Participation, including potential effects on "no-shows"
    - MTO: what is the effect of using a voucher when offered?
  - Treatment dosage or quality
    - HSIS: what generates greater impacts… two years, rather than one? being in a better quality center?
  - Multi-faceted treatment components/pathways
    - HPOG: *what is it* about the intervention that drives impacts? (experience of boot camp style orientation; participation in facilitated peer support groups; use of emergency assistance…)

References

Angrist, Joshua D., Guido W. Imbens, & Donald B. Rubin. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434), 444-455. DOI: 10.2307/2291629

Bell, Stephen H. & Laura R. Peck. (2013). Using Symmetric Predication of Endogenous Subgroups for Causal Inferences about Program Effects under Robust Assumptions: Part Two of a Method Note in Three Parts. *American Journal of Evaluation*, 34(3): 413-426. DOI: 10.1177/1098214013489338

Bloom, Howard S. (1984). Accounting for No-shows in Experimental Evaluation Designs. *Evaluation Review*, 8(2), 225-246. DOI: 10.1177/0193841X8400800205

Frangakis, Constantine E. & Donald B. Rubin. (2002). Principal Stratification in Causal Inference. *Biometrics*, 58(1), 21-29.

Harvill, Eleanor L., Laura R. Peck & Stephen H. Bell. (2013). On Overfitting in Analysis of Symmetrically Predicted Endogenous Subgroups from Randomized Experimental Samples: Part Three of a Method Note in Three Parts. *American Journal of Evaluation*, 34(4): 545-556. DOI: 10.1177/1098214013503201

Moulton, Shawn, Laura R. Peck, & Keri-Nicole Dillman. (2014). Moving to Opportunity's Impact on Health and Well-being Among High Dosage Participants. *Housing Policy Debate*, 24(2): 415-446. DOI: 10.1080/10511482.2013.875051

Peck, Laura R. (2003). Subgroup Analysis in Social Experiments: Measuring Program Impacts Based on Post-Treatment Choice. *American Journal of Evaluation*, 24(2), 157-187. DOI: 10.1016/S1098-2140(03)00031-6

Peck, Laura R. (2007). What are the Effects of Welfare Sanction Policies? Or, Using Propensity Scores as a Subgroup Indicator to Learn More from Social Experiments. *American Journal of Evaluation*, 28(3), 256-274. DOI: 10.1177/1098214007304129

Peck, Laura R. (2013). On Analysis of Symmetrically-Predicted Endogenous Subgroups: Part One of a Method Note in Three Parts. *American Journal of Evaluation*, 34(2): 225-236. DOI: 10.1177/1098214013481666

Peck, Laura R. and Stephen H. Bell. (2014). <u>The Role of Program Quality in Determining Head Start's Impact on Child U Development</u>. OPRE Report #2014-10, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Schochet, Peter Z., & John Burghardt. (2007). Using Propensity Scoring to Estimate Program-Related Subgroup Impacts in Experimental Program Evaluations. *Evaluation Review*, 31(2), 95-120.

For more information, please contact:

Laura R. Peck

Principal Scientist

Social & Economic Policy Division

T: 301.347.5537

E: Laura_Peck@abtassoc.com