World Vision

Use of Propensity Score Matching
A thought exercise

Annette E. Ghee, PhD, MPH

American Evaluation Association, Denver, CO
16 October 2014

World Vision

Our agenda today

- What issue was PSM designed to address?
- What are some pitfalls to keep in mind?
- How do you choose covariates that would ideally be used to create the Propensity Score?
- For those with experience, what have you learned from using it?

World Vision

Acknowledgement

Propensity Score Matching:
Theories and Applications

Haiyan Bai, Ph.D.
University of Central Florida
M. H. Clark, Ph.D.
University of Central Florida

October 23, 2012
American Evaluation Association Annual Conference
Minneapolis, MN

World Vision

What is PSM designed for?

Lack of

Asking Causal Questions without Randomized Experiments

- Random assignment does prevent us from asking causal questions, it just makes it difficult to make those inferences
  - Do patients in a residential rehabilitation center benefit from having a room with view? [1]
  - Do children who are instructed by teachers who specialize in PE have better physical fitness development than those taught by generalist teachers? [2]
  - How is sleep quality affected by stress among high school students? [3]
  - How did the September 11th attacks affect the mental health of British citizens? [4]

AEA 2012 10/23/2012 6

World Vision

Making Causal Inferences

- Randomized experiments allow evaluators to control temporal precedence and reduce extraneous errors.
  - Therefore, these designs often provide unbiased estimates.
- Random assignment is not always practical in field settings
  - Ethical reasons (denying access to those who need Tx)
  - Logistical limitations (Can't separate participants)
  - Lack of cooperation from stakeholders (controlling Tx may be disruptive in settings)

AEA 2012 10/23/2012 5

World Vision

What is a propensity score?

- A propensity score is the conditional probability of a unit being assigned to a particular study condition given a set of observed covariates [6].
- pr(z = 1 | x) is the probability of being in the treatment condition
  - In a randomized experiment pr(z = 1 | x) is known
    - It equals .5 in designs with two groups and where each unit has an equal chance of receiving treatment
  - In non-randomized experiments the pr(z = 1 | x) is unknown and has to be estimated.

How can it be used?

- These scores can be used to equate groups on observed covariates through
  - Matching
  - Stratification (subclassification or blocking)
  - Weighting
  - Covariate adjustment (analysis of covariance or regression)
- Propensity score adjustments should reduce the bias created by nonrandom assignment, making those adjusted estimates closer to those effects from a randomized experiment [7].

When should it be used?

- With quasi-experiments
  - When the independent variable was manipulated
  - When the intervention was presented before the outcome
- When the selection method is unknown
  - If assignment is based on a criterion, consider using a regression discontinuity design instead
- When there are several covariates related to selection
  - These can be continuous or categorical
  - The goal is to balance groups, so covariates should be related to selection
- The distribution of the dependent variable is unlimited
  - The adjustment analysis will vary based on the distribution

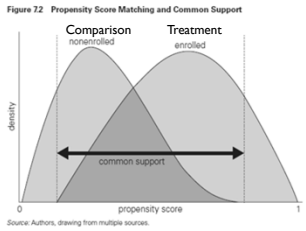
How do you implement?

- Trying to mimic a randomized design
- We architect the analysis so that members of the comparison group have similar propensities to members of the treatment group
- The matched nature of the intent must be followed through in the analytic methods used. The difference in outcomes (Y) between treatment units and their matched comparison unit is the estimate of program impact.
- Covariates should be "balanced" after matching.

Potential challenge

- What if there is not a good match between a treatment and a comparison unit?

Lack of "overlap"



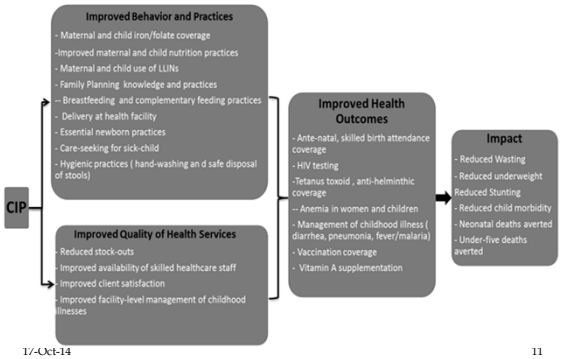
Important limitations to consider

- Requires large data sets; matching requires cases to be dropped from analysis.
- Still subject to participation bias associated with unmeasured/able characteristics; must assume there is none but this is a strong assumption
- Assumes missing information is missing at random
- How does it affect external validity?

World Vision Case Study

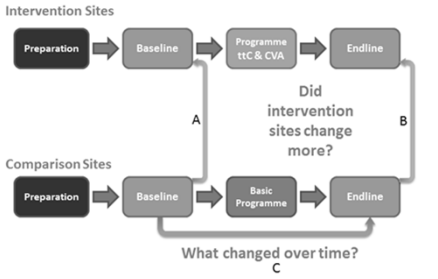
- Child Health Targets Impact Study (chTIS)
- Primary evaluation question
  Does the World Vision package of community-based health programming have a beneficial effect (effectiveness) on a range of child health and nutritional outcomes?
- 4 countries: Cambodia, Guatemala, Kenya & Zambia

Conceptual framework (Theory of Change)



World Vision

Study Design



Intervention period of ~24 months

Difference of Differences Analysis

Impact of the bundled programme = C

C = B – A

World Vision

Discussion

Given that the covariates we select to compute the propensity score must be collected at baseline and that they should mimic criteria for random assignment....

- What covariates would you ideally like to use for this case?
- Ideally would you want them to be measured at the individual, household or community level?
- When you inspect a potential covariate you are considering including, what properties should you consider?

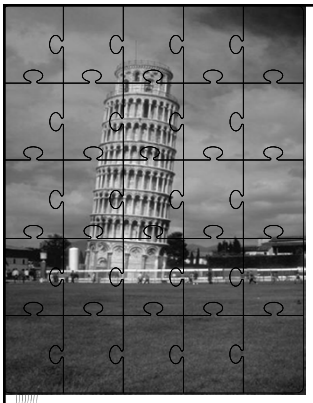
World Vision

Site selection

- Two pairs of Area Development Programs (ADPs) per country
- Matched and prioritized considering various criteria:
  - "Phase" of ADP maturity and availability of health budget
  - Prevalence of key socio-demographic and MCH indicators
  - Ethic/cultural context
  - Risk of contamination
  - Feasible for WV to deploy interventions with quality

Example from Zambia

	Pair 1	Pair 2
Comparison	Chongoo ADP	Luampa ADP
Intervention	Magoye ADP	Nyimba ADP



World Vision

"Eighty percent (80%) of projects fail *due* to poor objectives and misleading or insufficient assumptions . . . *not* because of poor implementation."

World Bank, Butler Cox

World Vision

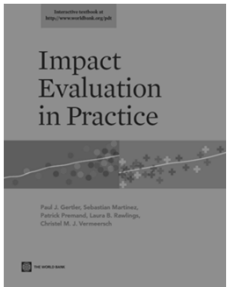
Important design considerations (?)

- Is this really a treatment vs absence of treatment comparison?
- Will all eligible "beneficiaries" receive the World Vision program?
- Will all "beneficiaries" receive the same "dose" when exposed to the program?
- Will direct beneficiaries be identifiable/ quantifiable?

World Vision

Key references

- World Bank document, especially chapters 4 & 7
- Austin, P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine*, 27(12), 2037-2049.



For any follow up discussion:
annette_ghee@wvi.org

Link to download World Bank document