Combining RCTs and Observational Studies

Eloise Kaizar
The Ohio State University

Social Science Objectives

- Population distributions
- Comparative effectiveness
  - Well-defined treatment (Hernan, et al)
    - Assign a patient to intervention
    - A patient actually received intervention
    - A patient actually received entire intervention
  - Well-defined outcome
    - Cholesterol vs. death
  - Well-defined population
    - Those who would choose an intervention
    - Those who are most likely to benefit from intervention

Possible Evidence

	Observational	RCT
Large sample size	X	Meta-analysis
Realistic treatment and environment	X	“practical trials” (Tunis et al, 2003)
Probability sample from target population	X	Re-weight and/or extrapolate
“Strong” causal inference	Propensity scores, instrumental variables	X
Customized measurement	sometimes	X
Long-term outcomes	Less rare	Rare

Well-Defined Populations

- What is the target population(s)?
  - RCT Population: Theoretical subset of the target population that consists of all individuals who would be eligible to enroll in the RCT.
  - Obs. Study Population: Population “represented” by the study.
  - Target Population: Population of all individuals for which treatment may be considered for its intended purpose?

Population: Adjusting Obs. Data for causal inference

- Methods to reduce treatment selection bias influence the inferential population
  - Propensity scores (PS) big idea:
    - Essentially re-weight the observational data so that the "controls" are comparable to the "treated".
  - PS population considerations
    - Reweight to those who selected: average treatment effect on the treated (ATT)
    - Poor overlap limits the inference to those who might not choose treatment
  - PS assumptions:
    - All important variables are measured

Population: Adjusting Obs. Data for causal inference

- Methods to reduce treatment selection bias influence the inferential population
  - Instrumental variables (IV) big idea:
    - Essentially scales the effect of the instrument on outcome by the effect of the instrument on treatment selection.
    - Treats treatment as a mediator of the instrument
  - IV effect on the population:
    - The estimand is a “local” average treatment effect (LATE), where “local” describes people who are influenced by the instrument.
  - IV assumptions:
    - The instrument only affects the outcome via the treatment selection

Population: Adjusting RCT Data for participant representativeness

- RCT population might be:
  - Like a SRS from the target population
    - No adjustment needed
  - Like a weighted sample from the target population
    - Use observational data to “standardize” RCT data
      - Survey methods like poststratification; Greenhouse, et al (2008)
      - Propensity-based standardization; Cole and Stuart (2010), Stuart, et al (2011)
  - Like a weighted sample of a **subpopulation** from the target population
    - Sensitivity analyses/Comparisons; Marcus (1997)
    - Use observational data to extrapolate from the RCT data using a model

Overarching models

- Goal: leverage the internal validity of the RCTs and the external validity of the observational data.
- Confidence Profile Method (and other models)
  - Eddy (1986); Eddy et al. (1990 and 1992)
  - Wolpert and Mengersen (2004)
  - Greenland (2005)
- Response Surface Methodology
  - Rubin (1990)
- Cross Design Synthesis
  - General Accounting Office (1992) and (1994)
  - Kaizar (2011)
  - Pressler and Kaizar (2013)

Confidence Profile Method (CPM)

- Approach to handle bias

$$\text{Target parameter} = \theta$$

$$\text{Evidence about } g(\theta, \alpha)$$

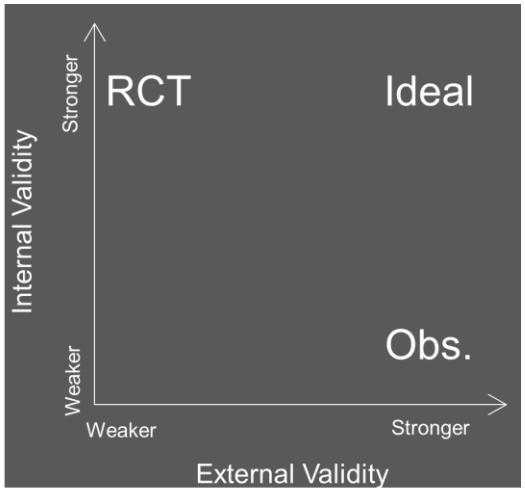
$$\text{Likelihood under target parameter: } f(x \mid \theta)$$

$$\text{Likelihood under CPM: } f(x \mid g(\theta, \alpha))$$

- Many g(θ,α) specified for specific kinds of bias

Response Surface Methodology

- Approximate the response surface as an n^{th} order polynomial of the experimental conditions
- Choose experimental conditions
- Extrapolate to the “ideal” experimental conditions



Cross Design Synthesis

- Examine randomized studies for external validity
- Examine administrative databases for internal validity
- Adjust data to improve validity
- Combine information between and within study types
  - Create a framework (stratify by design, coverage)
  - Combine studies within each design
  - Synthesize information across designs

Simple Case: Linear Bias

Population	Quantity of Interest	Randomized Data Estimator	Obs. Data Estimator
Represented by RCT	$\text{SPATE}_{\text{included}}$	$D^R_{included}$	$D^O_{included}$
Not Represented by RCT	$\text{SPATE}_{\text{excluded}}$	$= D^R_{included} + \delta_S$	$D^O_{excluded}$

Selection bias $= \delta_S \propto D^O_{excluded} - D^O_{included}$

Design bias $= D^O_{included} - D^R_{included}$

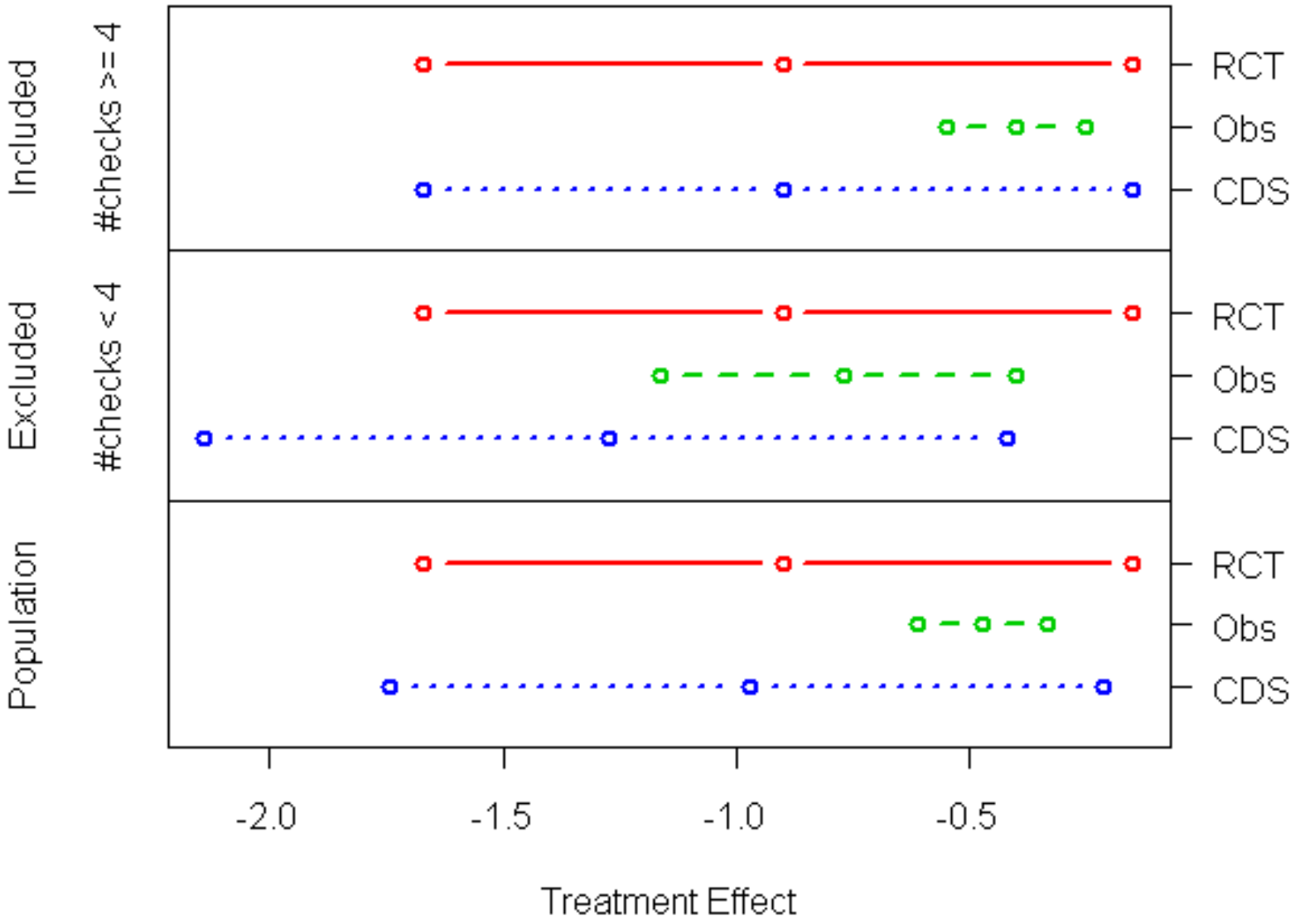
PATE = weighted average of Randomized Data Estimators

Example: Insulin Pump Use

- Problem: Is insulin pump use on average effective in improving metabolic control in the total population of diabetic patients?
  - Outcome: Mean A1C level (lower is better)
  - Control: Insulin injections
- Goal: Estimate the average treatment effect for use in policy decision making
- Data problem: RCTs exclude the noncompliant (<4 checks per day)

Doyle, et al (2004), Paris, et al (2009)

Example


Included
#checks >= 4
Excluded
#checks < 4
Population
RCT
Obs
CDS
RCT
Obs
CDS
RCT
Obs
CDS
-2.0
-1.5
-1.0
-0.5
Treatment Effect

Extensions

- Multiple RCTs and Observational data sets
  - Additional strata used for multiple inclusion criteria
- Multiple treatments (e.g., doses)
  - Additional stratification (multidimensional response “surface”)
- “Fuzzy” group membership

Designing new studies for sequential or simultaneous CDS

- On the RCT side:
  - Clearly identify populations/strata where randomized data is lacking or weak
  - Adapt design: sample size, proxy outcomes
- On the Obs. Side:
  - Check the adjustments for internal validity of the observational data
  - Variable collection

References 1

- Cole SR and Stuart EA. 2010. Generalizing evidence from randomized clinical trials to target population: the ACTG 320 trial. *American Journal of Epidemiology*.172: 107-115,
- Doyle E A, Weinzimer SA, Steffen AT, Ahern JAH, Vincent M, Tamborlane WV. 2004. A randomized, prospective trial comparing the efficacy of continuous subcutaneous insulin infusion with multiple daily injections using insulin glargine. *Diabetes Care*. 27: 1554-1558.
- Eddy DM. 1986. The use of confidence profiles to assess tissue-type plasminogen activator. In G. S.Wagner & R. Califf, eds., *Acute Coronary Care 1987*. Martinus Nihjoff Publishing Company.
- Eddy DM, Hasselblad V, Shachter R. 1990. A Bayesian method for synthesizing evidence: the confidence profile method. *International Journal of Technology Assessment in Health Care.* 6: 31-56.
- Eddy, DM, Hasselblad V, Shachter R. 1992. *Meta-Analysis by the Confidence Profile Method: The Statistical Synthesis of Evidence*. Statistical Modelling and Decision Science. Harcourte Brace Jovanovich.

References 2

- Greenhouse J, Kaizar EE, Seltman H, Kelleher K, Gardner, W. 2008. Generalizing from clinical trial data: A case study. The risk of suicidality among pediatric antidepressant users. *Statistics in Medicine.* 27(11):1801-13.
- Greenland S. Multiple-bias modelling for analysis of observational data. 2005. *Journal of the Royal Statistical Society: Series A.*168: 267-306.
- Kaizar EE. 2011. Estimating treatment effect via simple cross design synthesis. *Statistics in Medicine*. 30(25):2986-3009.
- Marcus SM. 1997. Assessing non-consent bias with parallel randomized and nonrandomized clinical trials. *Journal of Clinical Epidemiology*. 50: 823–828.
- Paris CA, Imperatore G, Klingensmith G, Petitti D, Rodriguez B, Anderson AM, Schwartz ID, Standiford DA, Pihoker C. 2009. Predictors of insulin regimens and impact on outcomes in youth with type 1 diabetes: the SEARCH for Diabetes in Youth study. *Journal of Pediatrics*. 155: 183-189.
- Pressler and Kaizar, 2013. The use of propensity scores and observational data to estimate randomized controlled trial generalizability bias. *Statistics in Medicine*. Early View.
- Prevost TC, Abrams, KR, Jones DR. 2000. Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Statistics in Medicine* 19: 3359-3376.

References 3

- Rubin, D. 1990. A new perspective on meta-analysis. In K. M. Wachter & M. L. Straff, eds., *The Future of Meta-Analysis*. Russel Sage Foundation, 155-165.
- Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. 2011 The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A,* 174: 369–386
- Tunis SR, Stryer DB, Clancy CM. 2003. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 290: 1624-1632.
- Wolpert RL, Mengersen KL. 2004. Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: effects of environmental tobacco smoke. *Statistical Science*. 19: 450-471,
- U.S. General Accounting Office. 1992. Cross-design synthesis: A new strategy for medical effectiveness research (GAO/PEMD-92-18). U.S. General Accounting Office, Washington, DC. http://archive.gao.gov/d31t10/145906.pdf.
- U. S. General Accounting Office 1994. Breast conservation versus mastectomy: patient survival in day-to-day medical practice and in randomized studies. Report to the Chairman, Subcommitte on Human Resources and Intergovernmental Relations, Committee on Government Operations, House of Representatives. PEMD-95-9.

Additional References 1

- **Papers related to population "standardization"**
  - Frangakis C. 2009. The calibration of treatment effects from clinical trials to target populations. *Clinical Trials*. 6: 136-140.
  - Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Sillett WC, Manson JE, Robins, JM. 2008. Observational Studies Analyzed like randomized experiments: An application to postmenopausal hormone therapy and coronary Heart Disease. Epidemiollogy. 19(6): 766—779.
  - Weisberg HI, Hayden VC, Pontes VP. 2009. Selection criteria and generalizability within the counterfactual framework: explaining the paradox of antidepressant-induced suicidality? *Clinical Trials*. 6: 109-118.

Additional References 2

- **Bayesian methods for combining data with explicit models:**
  - Molitor N-T, Best N, Jackson C, Richardson S. 2009. Using Bayesian graphical models to model biases in observational studies and to combine multiple sources of data: application to low birth weight and water disinfection by-products. *Journal of the Royal Statistical Society, Series A.* 172: 615-637.
  - Turner RM, Spiegelhalter DJ, Smith CS, Thompson SG. 2009. Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society, Series A.* 172: 21-47.