NATIONAL Palliative Care RESEARCH CENTER

Propensity Scores: A Method for Improving Validity in Quasi-Experimental/Observational Studies

R. Sean Morrison, MD

Director, National Palliative Care Research Center

Hermann Merkin Professor of Palliative Care

Professor, Geriatrics and Medicine

Vice-Chair for Research

Brookdale Department of Geriatrics & Adult Development

Mount Sinai School of Medicine

New York, NY

sean.morrison@mssm.edu

www.nprc.org

NATIONAL Palliative Care RESEARCH CENTER

- The Randomized Controlled Clinical Trial
- Quasi-experimental Designs
- Propensity Score Methods

RCTs

NATIONAL Palliative Care RESEARCH CENTER

- Strongest design to test efficacy
  - Is a treatment or intervention better than an alternative or placebo under ideal conditions
- Minimizes potential for confounding
  - Both observed and unobserved factors randomly and equally distributed across both groups (theoretically)
- Excellent internal validity
  - Differences in outcomes can be attribute to intervention or treatment

Nevertheless…

NATIONAL Palliative Care RESEARCH CENTER

In certain situations RCTs may be:

- Inappropriate
- Impossible
- Inadequate

An RCT Is Inappropriate:

NATIONAL Palliative Care RESEARCH CENTER

- To accurately assess infrequent adverse outcomes
  - Increased mortality associated with atypical anti-psychotics
- To determine whether an intervention prevents rare events
  - Supine positioning of infants to prevent SIDS
- When the intervention requires the subjects' active participation and thus depends on individual beliefs and preferences
  - Intervention psychotherapy
  - Palliative care consultation

An RCT May Be Impossible:

NATIONAL Palliative Care RESEARCH CENTER

- Clinicians are reluctant/refuse to participate
- Contamination/cross-over is unavoidable
- Ethical objections exist
- When interventions simply cannot be randomized

An RCT May Be Inadequate (Low External Validity)

NATIONAL Palliative Care RESEARCH CENTER

- Physicians that allow their patients to participate in a trial may not reflect the average treating practitioner
- Patients who agree to be randomized may be atypical
- The trial does not reflect real world situations
  - Patients in both arms may receive overall better care
  - The intervention may be artificial (mandatory geriatric/palliative care consultation)

Summary

NATIONAL Palliative Care RESEARCH CENTER

- RCTs offer an indication of efficacy of an intervention rather than its effectiveness in everyday practice
  - Provide evidence of what can be achieved under most favourable circumstances
- Should always be employed when appropriate, practical, ethical

Quasi-Experimental/ Observational Designs

NATIONAL Palliative Care RESEARCH CENTER

- Applied or field or real-world research
- Appropriate for research questions not amenable to RCTs
- External validity may be better ***BUT***
- Internal validity is weaker
  - Non-random assignment can lead to selection bias
  - Differences may be caused by the intervention, by differences in measured and unmeasured confounders, or both

Methods of Addressing Internal Validity

NATIONAL Palliative Care RESEARCH CENTER

- Traditional Multivariable Regression
  - Rely on functional form specifications, which can generate unreliable model-based extrapolations
  - Investigators have access to the study's outcomes when the main analyses of the intervention's effects are performed.
  - Strength, and even direction, of the intervention's effect on outcomes can be manipulated by the choice of control variables selected as potential confounders and how these baseline variables are used (e.g., squares, interactions) in the regression model.

Methods of Addressing Internal Validity

NATIONAL Palliative Care RESEARCH CENTER

- Matching
  - Control subjects matched to treatment subjects on basis of relevant characteristics
  - Difficult to match subjects on all or even many relevant, important covariates as the number of covariates increases

Addressing Internal Validity: Propensity Scores

NATIONAL Palliative Care RESEARCH CENTER

- Summary variable to control for measured patient characteristics in outcomes analyses
- PS methods mimic RCT process and enable investigators to control for relevant covariates simultaneously by matching on a single scalar variable—the propensity score - **without** knowledge of the study's outcomes

Propensity Score Computation

NATIONAL Palliative Care RESEARCH CENTER

1. Establish a **reference day**
   - Day of treatment assignment if this were an RCT
2. Construct dataset that contains only patient characteristics available on reference day and no outcome data.
3. Estimate the conditional probability that a patient will be in the intervention group
   - Dependent variable: Group assignment
   - Independent variables: All the measured background characteristics that are relevant and available on the reference day
4. Divide the sample to evenly distribute the PS
5. Test for balance of independent variables within each block (t-test, chi-square)

Propensity Score Methods

NATIONAL Palliative Care RESEARCH CENTER

6. Match patients who received treatment to patients who did not receive treatment by PS
   - Several different matching procedures available
7. Dataset contains treated patients and non-treated patients who share similar observable relevant baseline covariates – akin to the subject sample of an RCT
8. Dependent variables now made available to the investigators and the effect of the intervention on outcomes is estimated

An Example: Does A Unit-Based Analgesic Intervention Improve Pain and Function In Older Adults Following Surgery?

NATIONAL Palliative Care RESEARCH CENTER

- Hypotheses: An aggressive pain management program will:
  - Improve pain scores
  - Improve functional recovery
  - Not result in serious side effects (constipation, nausea, sedation)
- Subjects:
  - Patients admitted to an acute rehabilitation hospital following lower extremity orthopedic surgery
- Design
  - Prospective quasi-experiment
  - Patients enrolled on admission to rehab. unit
    - Bed assignment based upon availability
  - One intervention unit
  - Three control units

Patient Characteristics – PS matched

NATIONAL Palliative Care RESEARCH CENTER

	Unmatched		PS Matched	
	Control	Interv	Control	Interv
Women	79%	66%	74%	76%
Mean age (years)	72	70	71	71
White	64%	52%	59%	63%
African American	20%	31%	22%	23.0%
Other	16%	17%	19%	14%
Medicare	84%	75%	79%	77%
Hip fracture	29%	15%	21%	21%
Total hip replacement	23%	28%	24%	24%
Total knee replacement	49%	57%	55%	55%
Rand comorbidity index	1.7	1.7	1.7	1.6
Depression Score	3.1	2.6	2.8	2.6
Severe/V. severe pain at admission	38%	35%	37%	37%

Summary

NATIONAL Palliative Care RESEARCH CENTER

- Propensity scores are useful to more accurately estimate the true causal effect
  - The effectiveness of using these scores varies depending on how they are applied.
    - Least effective—Place in multivariate model
    - More effective—Match on propensity score
    - Most effective—Inverse probability of weighting
  - Effectiveness also varies by the richness of the variables included in the regression models used to determine the propensity of “treatment