**Alternative matching methods with non-experimental data:
Application to out-of-home behavioral health care**

**John Robst, Ph.D.
Norin Dollard, Ph.D.
Mary Armstrong, Ph.D.
Keren Vergon, Ph.D.
Stephanie Romney, Ph.D.**

Florida Mental Health Institute
University of South Florida

Introduction

- Administrative data provide large samples, but beneficiaries are rarely randomly assigned to the treatment they receive.
- Selection or assignment into treatment can bias comparisons because outcomes might be due to differences in individuals observed in programs; not the programs.

- This paper focuses on two methods that account for observed differences between treatment and control groups.
    - 1 - Propensity score matching
    - 2 - Risk adjustment
- Focus methodological
- Methods illustrated using out-of-home care in Florida for children with mental health needs.

Prior research on TFC and TGC

- Therapeutic foster care and therapeutic group care. Assignment might be based on characteristics of children.
- Berrick, Courtney, and Barth (1993) found behavioral health needs of children in TFC and TGC in California were similar.
- But Armstrong et. al. (2006) found significant pre-treatment differences between children in TFC and TGC in Florida.

Data

- Primary data: Florida Medicaid eligibility and claims files.
- July 2000 through June 2005.
- Therapeutic group care (TGC) n=141 and specialized therapeutic foster care (TFC) n=386.
- Outcomes: Medicaid costs, involuntary commitment (IC), juvenile justice (JJ), and adult law enforcement (LE) encounters.

Means - Outcomes

- Avg. costs:
    TFC: increased from $1,702 to $2,496.
    TGC: fell from $3,297 to $2,856.
- IC fell for both groups approx. same %.
- JJ and LE encounters changed little for both groups.

Propensity score matching

- Propensity score matching approximates a random experiment by matching individuals in the treatment and "control" groups based on observed characteristics.
- Differences in outcomes between matched individuals due to treatment and not preexisting factors.

- The propensity score is the estimated probability of selecting or being assigned to a treatment group.
- Probability computed from a logit model where treatment placement is a function of observed characteristics:

    $TFC_i = X_i \cdot \beta + u_i$

- Independent variables:
    - demographics (gender, race, and age),
        - 11 diagnostic categories,
   - 13 treatment categories in prior year

---

Risk Adjustment

- While risk adjustment refers to several different concepts, for the purposes of this paper "risk adjustment …describes a way of accounting for differences in health status among various study populations" (Greenwald, 2000, p. 1).
- Health status measured by proxy, such as behavioral health costs, service utilization, or the likelihood of mortality (e.g., suicide).

- Many risk adjustment models, but none appropriate for this analysis.
    - overall health status.
    - mental health models focus on adults.
    - designed as part of a managed care payment system and thus exclude information deemed inappropriate for payment (e.g., race, prior period utilization).
- Thus a risk adjustment model was developed specific to this analysis.

- Model estimated using OLS regression.
- Dependent variable:
    - pre-treatment Medicaid behavioral health costs
- Same independent variables as logit regression.

- Risk scores are computed by predicting expenditures for each individual based on the coefficients and individual characteristics, standardized to a mean of 1.0
- Higher risk scores imply the presence of more conditions and characteristics related to higher costs. As such, higher risk scores denote poorer mental health.

Logit results
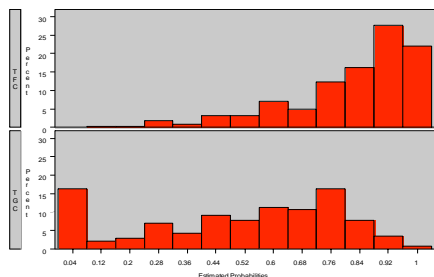
Risk scores
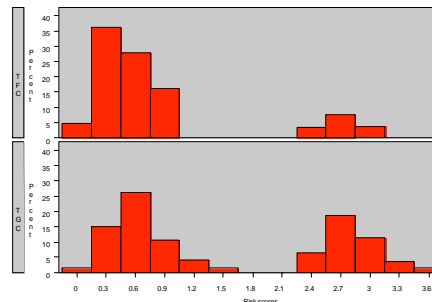
Matching

- Nearest neighbor matching used by matching youth in TFC with the person in TGC with the closest propensity or risk score (Dehejia & Wahba, 2002).
- Matching can be performed either with or without replacement.
- Common support range

Outcomes

- Observations are retained within the common support range and matched without replacement.
- PSM: 108 observations (23% of TGC lost).
- Risk adj: 129 observations (9% lost).

- Results qualitatively similar for propensity score and risk score matching.
- Similar youth in TFC and TGC were not statistically different in the year after treatment.
- Cannot conclude TFC or TGC is more effective.
- Accounting for differences between TFC and TGC children important. A standard pre-post comparison of costs would suggest that TGC was more effective.

Comparing methods

- Propensity scores adjust for the likelihood of being placed into a treatment group, but the score does not provide any indication of the need for services or health status.
- Risk adjustment models are often estimated using costs, utilization, or mortality as a proxy for health status. While such factors are correlated with health status, the proxies are often outcomes of interest.

Comparing methods

- Both approaches rely on observable characteristics to match observations. Consequently, unobservable characteristics potentially problematic.
- Common support problem. The better the model that predicts treatment placement, the harder to match observations. This contradiction within the PSM methodology is less evident with risk models.

Conclusion

- Accounting for differences prior to treatment is important to understanding treatment effects.
- Risk adjustment models provide an alternative method for matching individuals.