Alternatives to the Randomized Controlled Trial

Stephen G. West
Arizona State University

Based on West, Duan, Pequegnat, Gaist, Des Jarlais, Holtgrave, Szapocznik, Fishbein, Rapkin, Clatts, and Mullen (2007), Manuscript under review. Please do not circulate

Randomized Controlled Trial (RCT)

Random Assignment: Sir Ronald Fisher

At Baseline:

$E(Y_{Treatment}) = E(Y_{Control})$ for any covariate Y

$E(r_{YT}) = 0$

Unbiased Estimate of Average Treatment Effect

Strongest Design for Causal Inference

With Greatest Statistical Power

BUT

Fisher's Work's was in agriculture

This context does not fully characterize prevention trials in the community

Corn plants do NOT

Raise ethical or practical concerns about randomization
Fail to comply with Treatment
Find a better Treatment
Move away—so lost to measurement
Refuse to answer questionnaires

Ethical, Practical Concerns Can Rule Out Randomization of Participants

Gang Membership Increased Aggression?

Hurricane Katrina Increased HIV risk behaviors, violent criminal behaviors?

Faith-based intervention Improved mental health, substance abuse outcomes?

Important questions can't be answered by RCT.

Randomized Trials in Community: *New* Assumptions

- Independent units
- Full treatment adherence
- No attrition
- SUTVA
  - Other treatment conditions do not affect participant's outcome
  - No hidden treatment variations

Randomized Preventive Trial:
Assumptions will almost *never* be met

Barnard, Frangakis, Hill, and Rubin (1998)

“Broken Randomized Experiment”

Good design and analysis solutions now exist.

Addressing Assumption Violations I: Statistical Approach

Rubin's (1974, 2005) Potential Outcomes Model

Identify appropriate counterfactual condition for treatment case
(matched identical case without treatment)

Specify Assumptions
Make appropriate statistical adjustment
Mathematical Approach

Addressing Assumption Violations II: Design Approach

Campbell's Threats to Validity

(Campbell, 1957; Shadish, Cook, & Campbell, 2002)

Identify Specific Confounding Factors that may produce observed results—threats to validity

Add design elements to address threats

Logically rule out threat

Working Scientist Approach

Example: Independent Units

Statistical Approach:

Multilevel Analysis

(other statistical adjustment for clustering)

Design Approach

Temporal or geographical isolation of units

(See Table 2 for other assumptions, solutions)

Beyond The RCT:
Towards General Design Options

Units

1. Subjects (most common)
2. Settings—street corners
3. Times—drug trials; or even
4. Measures—Sesame street (letters)

by

Assignment to T and C conditions: Basis

(a) random, (b) quantitative basis (c) nonrandom

See Table 1 for full 3 x 4 Table of Designs

Alternative I:
Randomized Encouragement Design

Subjects randomly invited to either T or C condition

encouraged to participate in T condition

only some subjects choose to participate

C –typically no or minimal treatment (all comply)

Example: Parents invited to participate in parenting program to improve child's mental health (50% come)

Alternative I:
Randomized Encouragement Design

Key new assumption:

Exclusion restriction--encouragement intervention is mediated solely by treatment

Statistical Approach: Instrumental Variable Analysis + sensitivity analysis

Design Approach: None

Sometimes more ethical, practical than RCT

Prevention Example
Randomized Encouragement Design

Jobs II Program

(Vinokur, Price & Schul, 1995; Little & Yau, 1998)

Recruited from employment offices

Given screening questionnaire

Randomly assigned to T vs. C (No treatment)

T participants—received invitation + financial inducement to participate in Jobs program

Prevention Example II: Randomized Encouragement Design

Program led to decrease in depression at followup for high risk participants

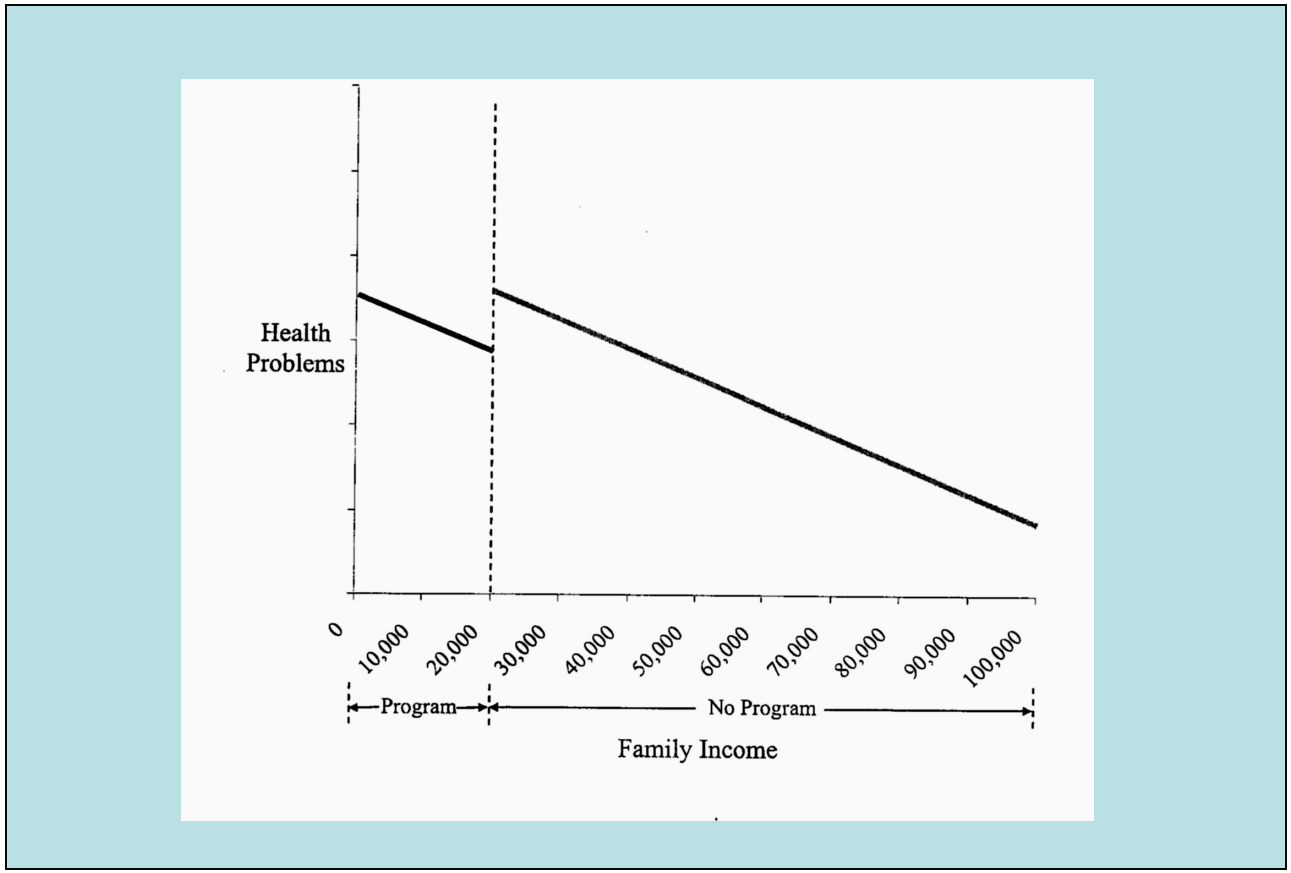
Note. Comparison is Complier Average Causal Effect—Effect of Program on Compliers vs. Effect of Control on People who would have complied if in T condition.

Advantages: Estimate effect of interest;

less biased estimate than ITT if attrition.

Alternative 2: Regression Discontinuity Design

Assignment to T vs. C on basis of **quantitative** measure of need or merit



Prevention Example: Regression Discontinuity Design

Effects of Head Start (Ludwig & Miller, 2007)

300 poorest counties in US received TA (poverty rate above 59.2%)

Rest did not receive technical assistance; most of next tier did not receive Head Start

Prevention Example 2: Regression Discontinuity Design

Head Start Program—education program

also **health services** —

nutrition, screening, immunization

Substantial subsequent improvement in health issues addressed by program

(e.g., children's mortality rate from measles)

Alternative 2: Regression Discontinuity Design

Key Additional Assumption: Functional form of relationship between assignment variable and outcome is properly specified

Statistical Approach: nonparametric regression; sensitivity analysis

Design Approach—replication with different cutoff threshold; nonequivalent dependent variable

Alternative III: Observational Studies

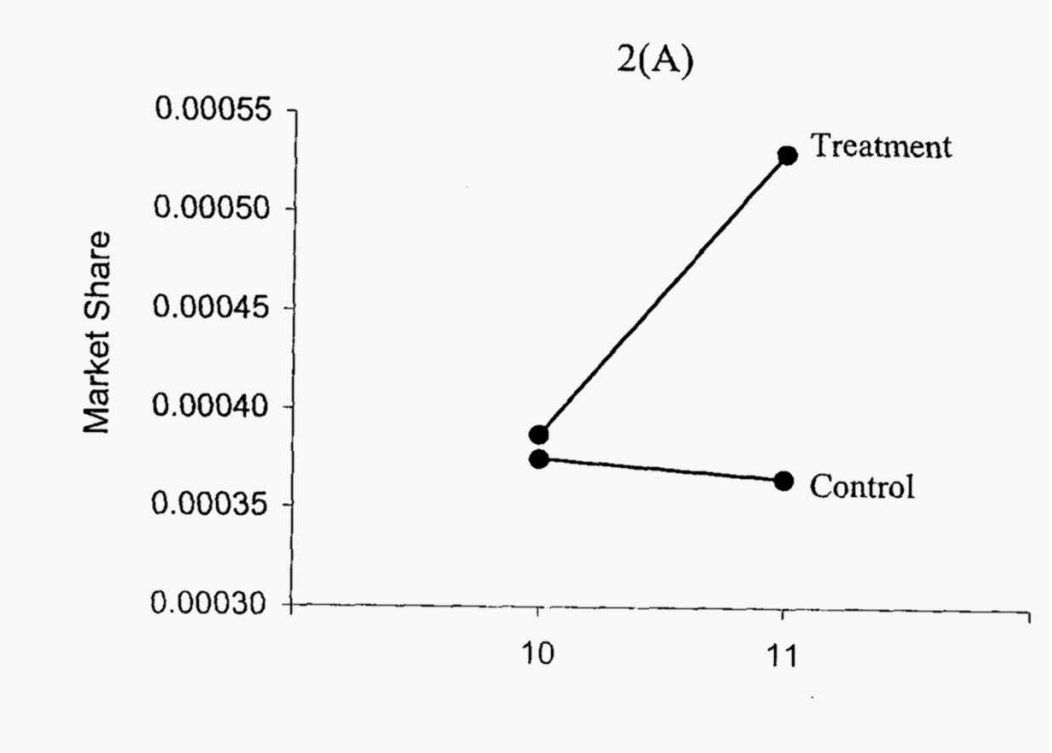
Units are measured at baseline and posttest

Assignment to T, C is unknown and presumed to be nonrandom

Baseline	Treatment	Posttest
x	T	x
x	C	x

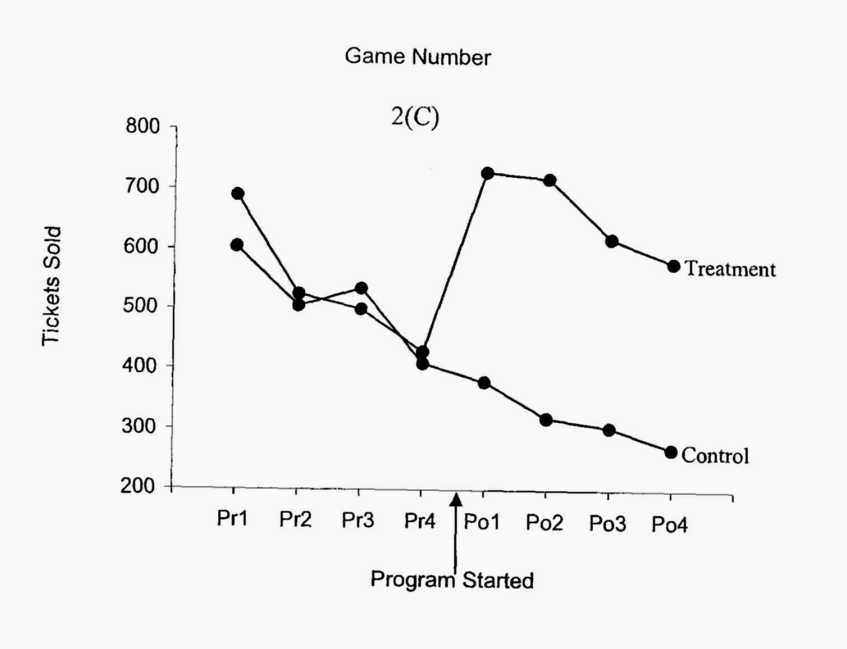
Key Issue: Selection in T and C groups. Many threats to internal validity.

Example: Reynolds and West (1987). AZ Lottery Sales Program

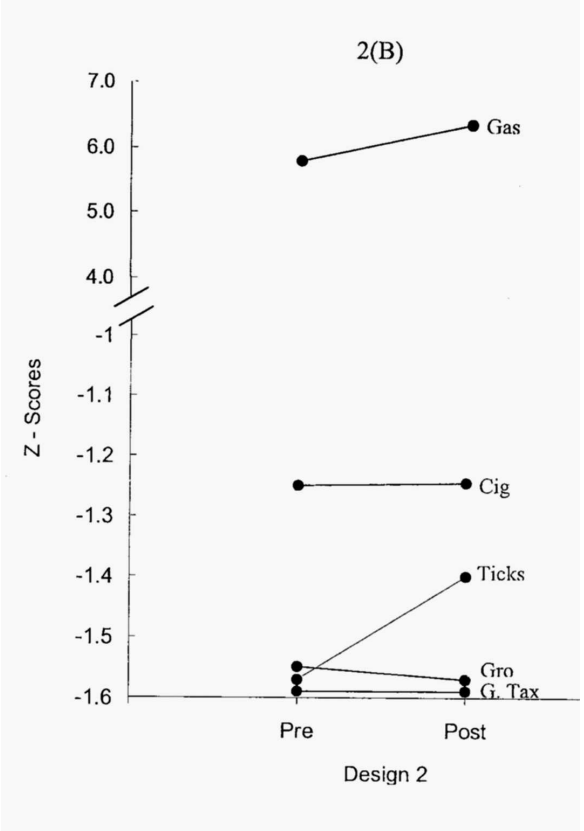


Ruling Out Threats: Design Elements Approach

Repeated Pre- and Post Tests



Nonequivalent DVS



Ruling Out Threats:
Statistical Adjustment--Propensity Scores

Baseline: Measure predictors of outcome and selection; Inclusive

Predict Probability of being in T group

Match (or Statistically Adjust) T and C on propensity

Goal: Achieve Balance between T and C on all Baseline Measures (Researcher must check)

Propensity Score Example
Wu, West, and Hughes (2007)

Children at risk for retention in first grade
<50th percentile on reading scores at entry

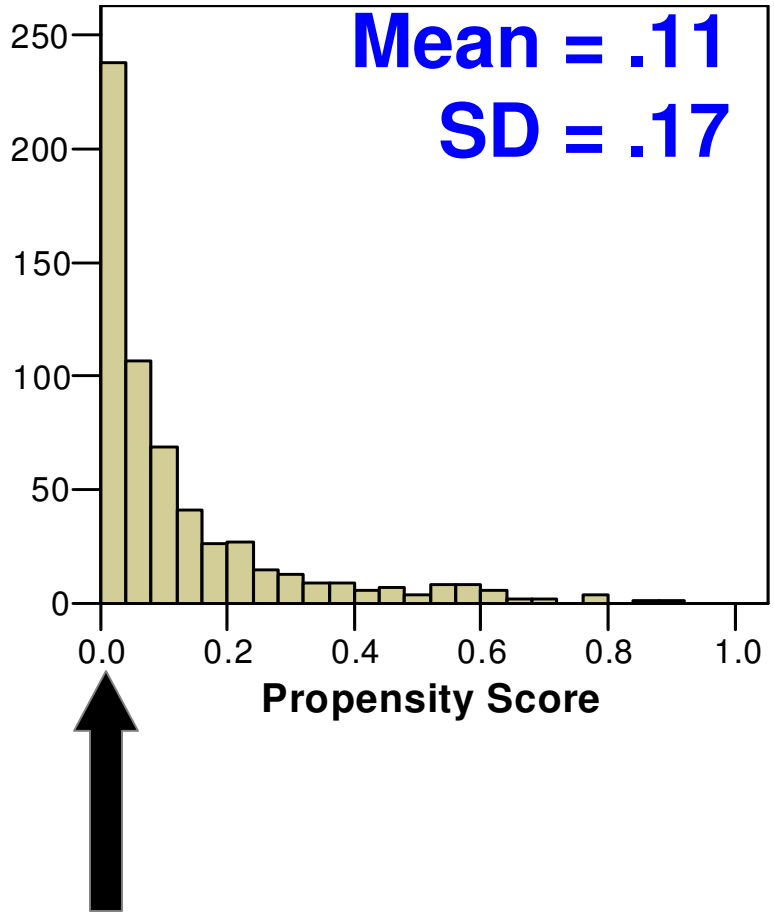
72 Variables measured at pretest
Subject Matter Expert Suggested
--teacher, child, peer, achievement measures

Propensity Score Example II
Wu, West, and Hughes (2007)

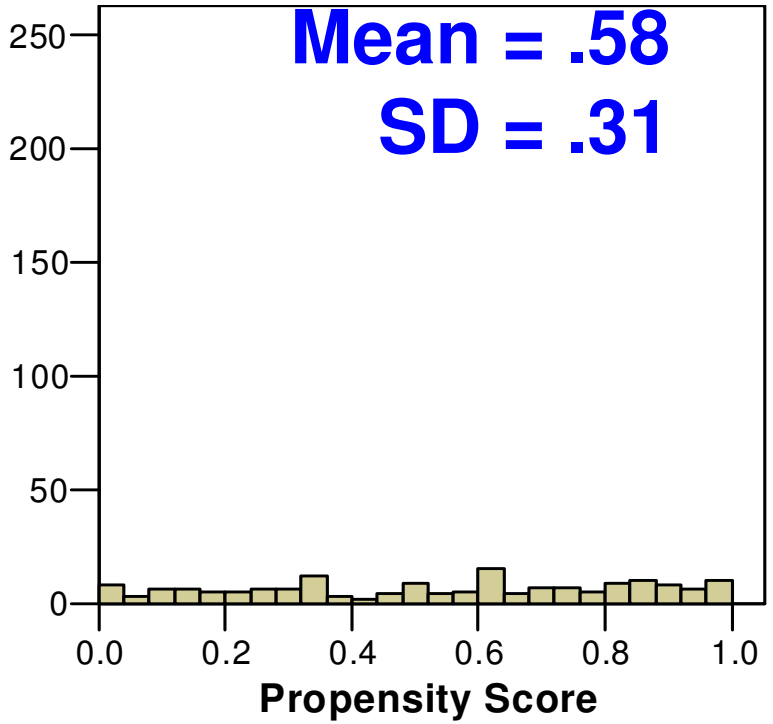
- Probability of Retention
  - Predicted based on 72 variables
  - Logistic Regression to predict
    - probability of retention, .00 to 1.00

Before Matching:
For 50% of children at highest risk

Promoted (N=604)

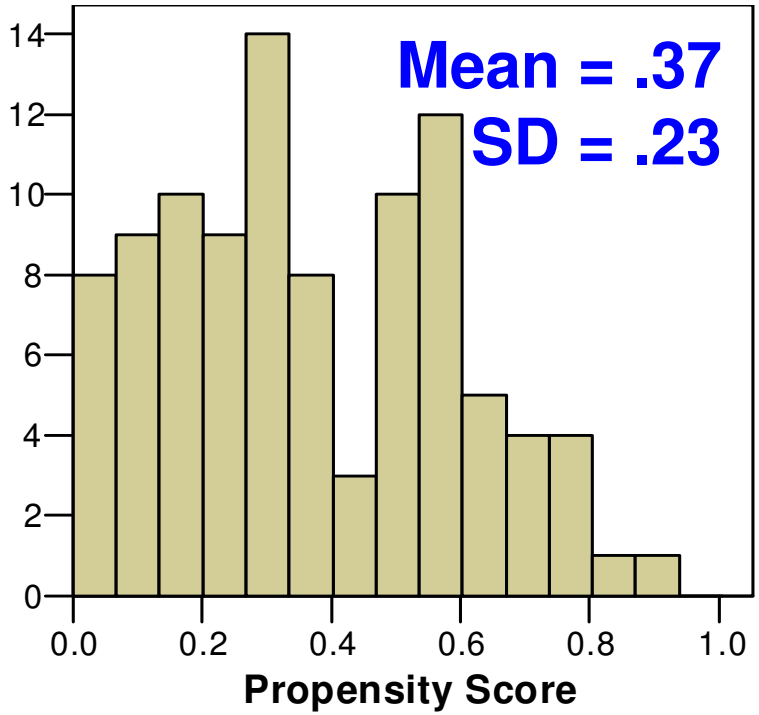


Retained (N=165)

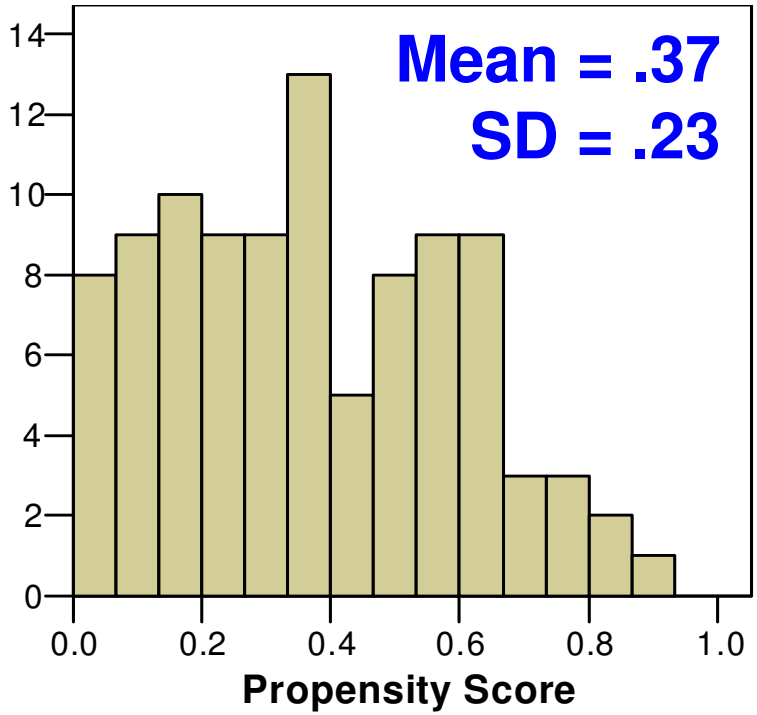


After Matching

Promoted (N=98)



Retained (N=98)



Prevention Example: Observational Study Haviland, Nagin, & Rosenbaum (2007)

Boys measured annually age 11-17

Outcome: violent delinquency

T = join gang at 14; C = not join gang at 14

n =580 boys—not gang members < age 14

Identify three trajectory groups based on delinquency age 11, 12, 13
(high, increasing, low)

12 covariates—violence, hyperactivity, IQ, etc.

Results: Prevention Example

Propensity Scores:

Appropriate Matches found only for low, increasing trajectory groups

No overlap between gang joiners, non joiners in high trajectory group

Outcome: Violence increases at age 14 and age 15 for gang joiners.

Conclusion

Randomized Controlled Trial

- not feasible for all important questions
- can rarely be implemented meeting all assumptions
- design, statistical solutions for violations needed
- (sometimes) may have limited external validity

Alternatives to RCT

- involve additional assumptions
- good design and/or statistical solutions now available to address violations of assumptions

Conclusion II

Cook, Shadish, and Wong (2005)

- Comparisons of RCT with other designs sharing Treatment Conditions
- No difference in effect size estimates for regression discontinuity, time series
- No difference in effect size estimates for Observational Study when careful attention given to selection issues.

Conclusion III

New improved design, statistical solutions for alternatives better estimates of effects

Alternatives to RCT involve more uncertainty

bounds needed on effects to reflect this

(e.g., sensitivity analysis)

Typically lower power than RCT (given = n)

Use of reporting guidelines encouraged

—Consort and Trend transparency

May also help improve research quality.

Three Giants of Causal Inference

Sir Ronald Fisher

Donald Campbell

Donald Rubin