Overview of Perspectives on Causal Inference: Campbell and Rubin

Stephen G. West
Arizona State University
Freie Universität Berlin, Germany

Randomized Experiment (RE)

Sir Ronald Fisher

$E(X_{Treatment}) = E(X_{Control})$

Covariate Balance at pretest

Unbiased Estimate of Treatment Effect

Often Viewed as Gold Standard

Strongest Design for Causal Inference

BUT

Fisher’s Work’s was in agriculture

This context does not fully characterize research in community settings with human participants

Corn plants do NOT

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

Limitations of Randomized Experiments 1

Randomization May Not Be Possible or Desirable to Implement

- Ethical Issues (secondhand smoke)
- Policy Interventions
- Unplanned Events (e.g., Hurricane Katrina)
- Doubts About Equipoise
- Participants refuse to be randomized (or only atypical participants agree, e.g. faith-based).

Limitations of Randomized Experiments 2

Broken Randomized Experiments

(not infrequent). Some issues

Treatment Noncompliance

e.g. refuse treatment, actively seek out alternative treatment, receive partial treatment

Attrition

Participants lost to posttest measurement

Need Alterative Perspectives,

Alternative Designs

Two Complementary Perspectives on Causal Inference That Provide Strong Alternative Approaches

Campbell's "Working Scientist" Approach

Rubin's Potential Outcomes Approach

Campbell's "Working Scientist" Approach

- PhD UC Berkeley Psychology
- Psychology, Education
- Applied, Basic
- Manipulated,Non-Manipulated Variables
- *Threats to Validity*
- *Emphasis on Design Elements, Pattern Matching*

Concern with construct validity

Concern with generalization

Campbell (1957); Campbell & Stanley (1966); Cook & Campbell (1979); Shadish, Cook, & Campbell (2002); Shadish & Cook (2009)

Donald Campbell
Key Concept for Causal Inference

Plausible Threats to Internal Validity

-Specific reasons why we can be partly or completely wrong in our causal inference

-Depend on

Design

Obtained pattern of results

Prior results and theory

-Extensive (exhaustive?) list representing “an accumulation of our field’s criticisms of each others research”

Campbell on Plausible Threats

"We took the position that there could be lots of threats to validity that were logically uncontrolled but that one should not worry about unless they were plausible. The general spirit was that any interpretation of a body of data or research should be regarded as innocent until judged guilty for plausible reasons, as determined through the scientific method of mutual criticism."

Campbell (1988, p. 317).

Design Elements and Pattern Matching

Identify specific confounding factors that may produce observed results—threats to internal validity

Add design elements to address threats

Assess Match of Pattern of Results to

Scientific Hypothesis, Potential Threats

Working Scientist Approach—Logically rule out threat

Illustration: Observational Study

Baseline	Treatment	Outcome
X	T	Y
X	C	Y

Threats to Internal Validity

- Selection x Maturation
- Selection x History
- Selection x Instrumentation
- Selection x Statistical Regression
- Differential Attrition

Unknown assignment rule (presumed nonrandom)
Other covariates may be measured at baseline.
Causal Inference: One of weaker quasi-experimental designs

Reynolds & West (1987). Basic Observational Findings
Design Element 1. Matching

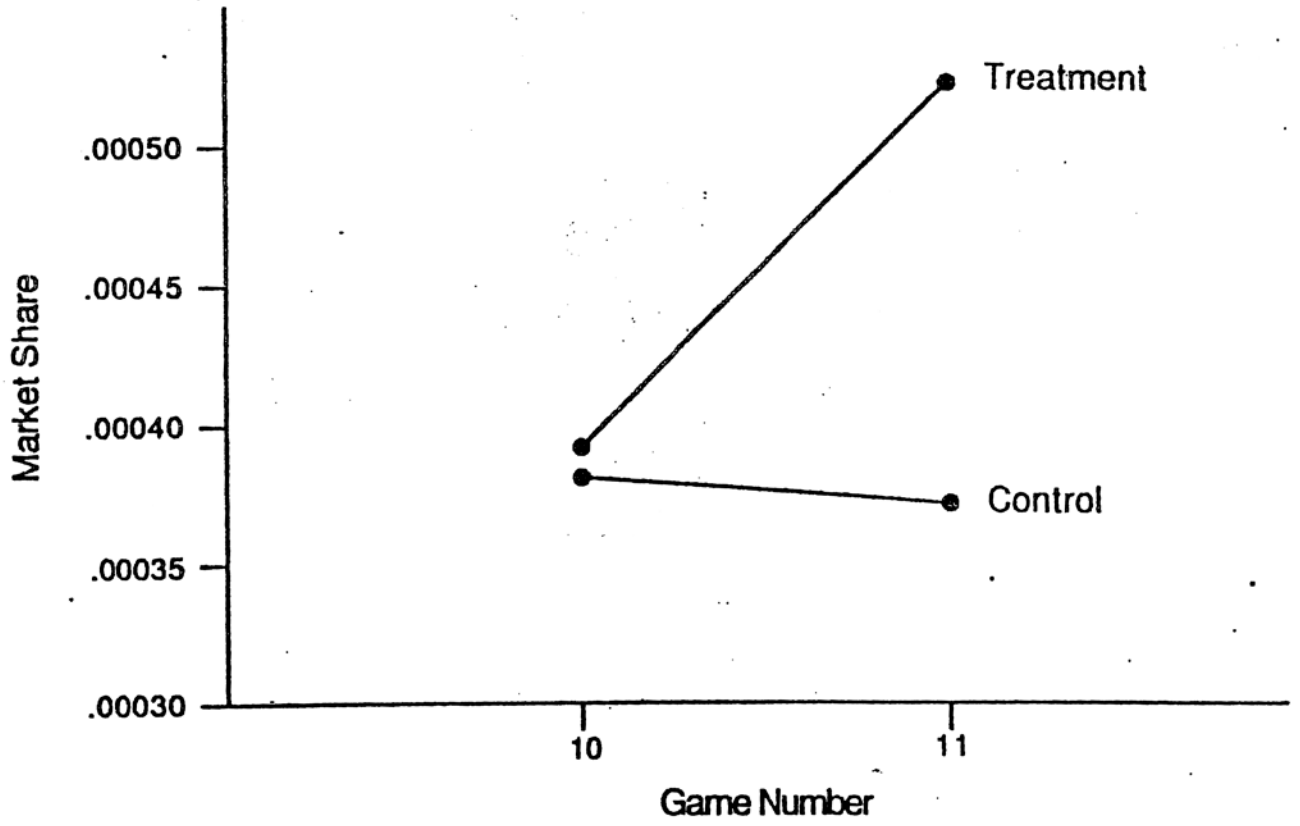


Figure 1: Mean Marketshare for Treatment and Control Stores in Chain A
Note: The program was implemented in treatment stores during game 11.

Reynolds & West (1987. Design Element 2. Nonequivalent DVs

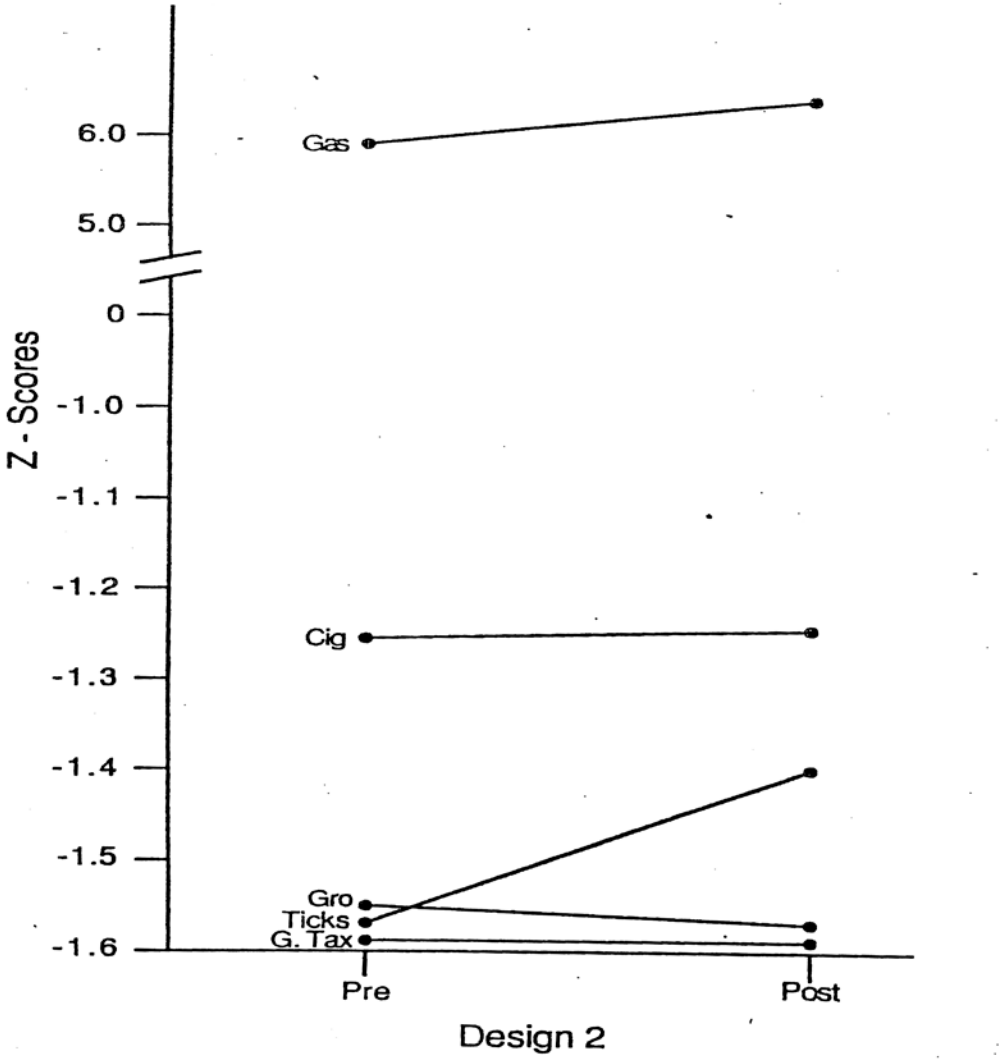
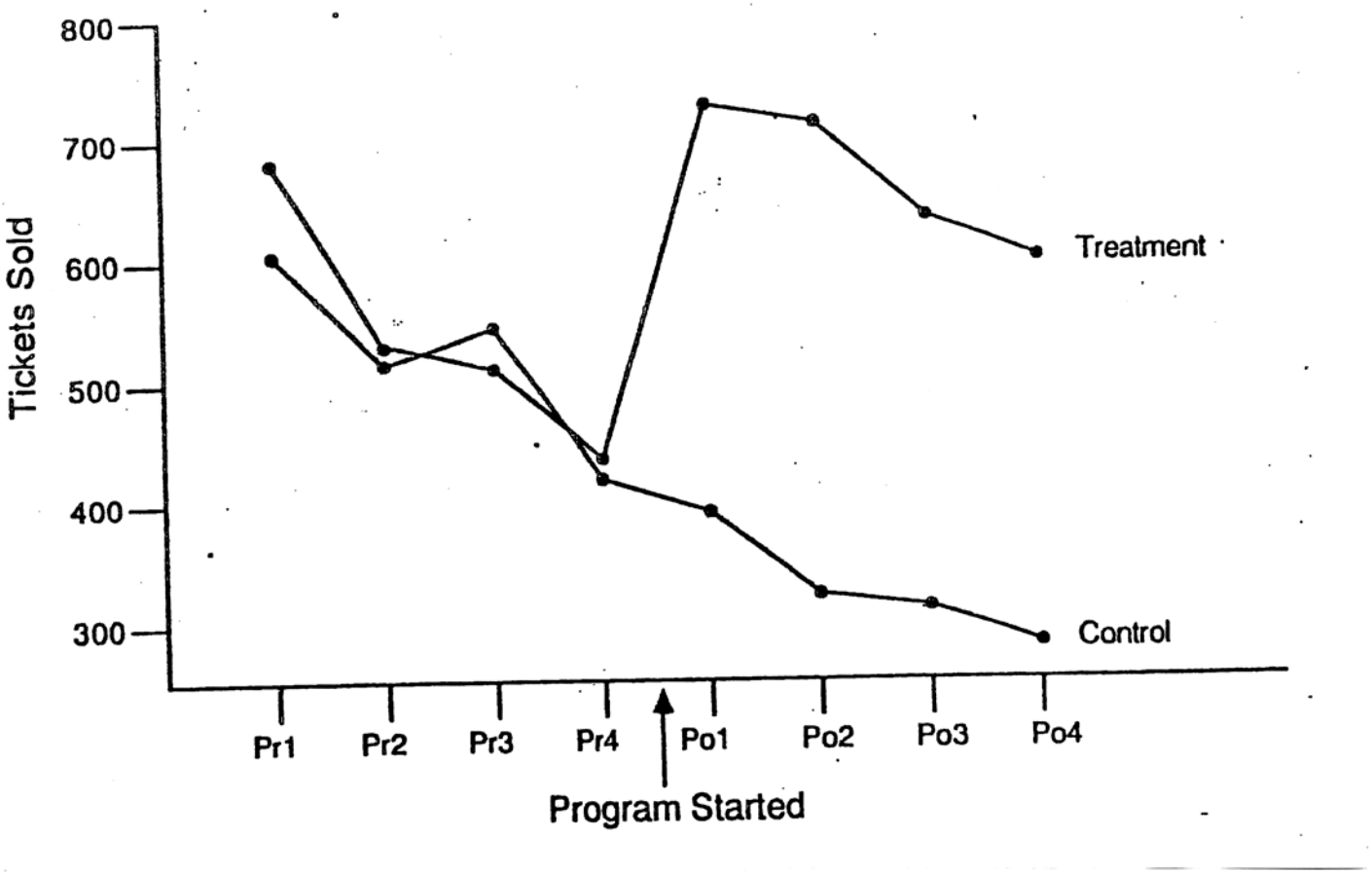


Figure 2 : Pre-Post Change in Sales for Each Dependent Variable in Design 2
Note: Department variables are gasoline sales (Gas), cigarette sales (Cig), non-taxable groceries (Gro), lottery ticket sales (Ticks), and taxable groceries (G. Tax).

Reynolds & West (1987). Design Element 3. Multiple pretests, posttests



Experimental and Quasi-Experimental Designs

Campbell's Perspective Permits Relatively Strong Inferences from:

Randomized Experiments

Broken Randomized Experiments

Regression Discontinuity Design—quantitative assignment

Interrupted Time Series—assignment on time

Observational Studies—unknown assignment rule

Design elements need to be added to strengthen causal inferences in weaker designs

Some Challenges for Campbell's Perspective

(a) Have all threats to internal validity been identified?

(b) All or none ruling out of threats.

(c) Pattern Matching: How can we assess the degree to which the obtained results match the hypothesis vs. threats to validity (confounders)?

Rubin's Potential Outcomes Approach

- PhD Harvard Statistics
- Medicine, Public Health, (Economics)
- Applied
- Manipulated Treatments
- Precise Assumptions—Deductive, Mathematical
- Emphasis on Potential Outcomes Logic
- Strong Basis for Causal Inference When
  - Broken Randomized Experiments, Missing Data
  - Cannot Randomly Assign

- Holland (1986); Rubin (1974, 1978; 2005);
- Imbens & Rubin text (stubbornly unpublished)

Rubin's Potential Outcomes Approach

Starts With Platonic Ideal:

Compare response of unit [person] to T with response of unit to C under identical conditions and at identical time.

Fundamental Problem of Causal Inference

"It is impossible to observe the value of $Y_{T(u)}$ and $Y_{C(u)}$ on the same unit [person] and, therefore, it is impossive to observe the effect of T on u." Holland (1986, p. 947)

Potential Outcomes

For each unit, there are two potential outcomes:

--One that would be observed under T

--One that would be observed under C

Only one outcome may be observed

Need to define approximations to Platonic ideal and the necessary assumptions for the approximation to be correct.

One approximation to Ideal

RANDOMIZATION

Given additional assumptions—

- Clear definition of T and C conditions
- full treatment compliance
- no attrition
- Stable Unit Treatment Value Assumption

→Groups balanced on all covariates

Unbiased Estimate of Average Treatment Effect

What If Randomization Is Not Possible or Fails

Additional Assumptions—permit unbiased estimates of Average Treatment Effect for Alternative Designs

- **Broken Randomized Experiments**
  -Treatment noncompliance
  -Missing Data

- **Regression Discontinuity Design--**quantitative assignment

- **Observational Studies--**Unknown assignment rule

- **Interrupted Time Series--**Not Addressed

Illustration: Observational Study

If we can match T and C subjects on **all** covariates that are related to (a) treatment assignment and (b) outcome, then unbiased estimate of causal effect is possible.

Conditional independence replaces independence.

Strong Ignorability assumption

Devise a statistical method of balancing all key **measured** covariates→ mimic RE

Propensity scores (PS)

PS(Z = 1| X)

—probability of T assignment given X

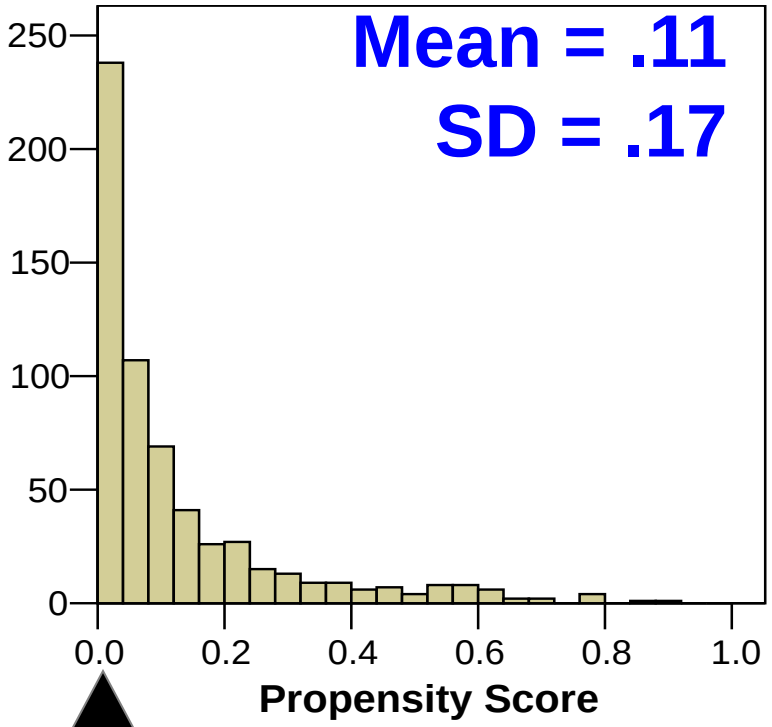
PS: A single number summary based on all available covariates that expresses the likelihood that a given subject is assigned to the treatment condition, based on the values of the set of observed covariates

Propensity scores

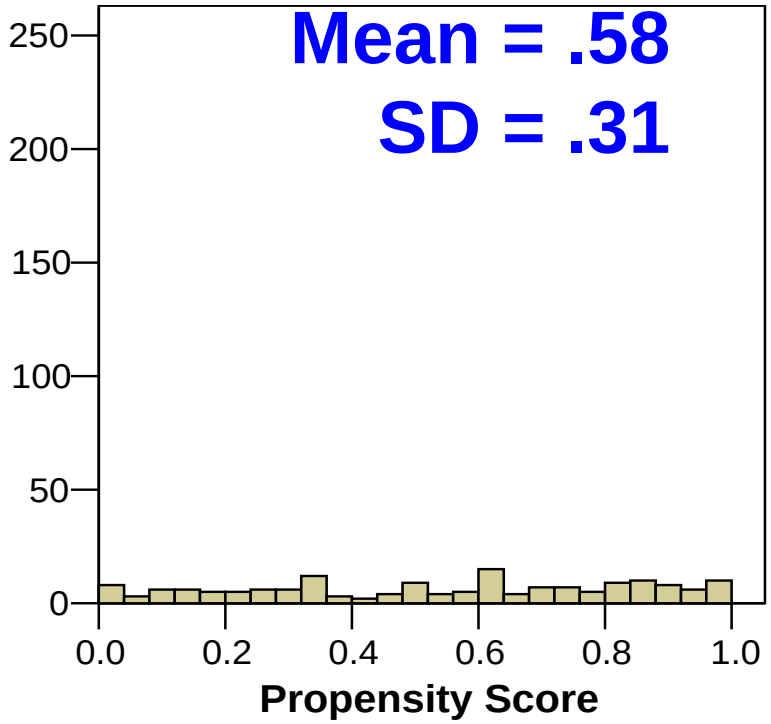
- Balance on the propensity score implies on average balance on all observed covariates
- Two units in the treatment and the control group that have the same propensity score are similar on all covariates. *They only differ in terms of treatment received*
- Balance on observed covariates can be checked

Before Matching:
For 50% of children at highest risk

Promoted (N=604)



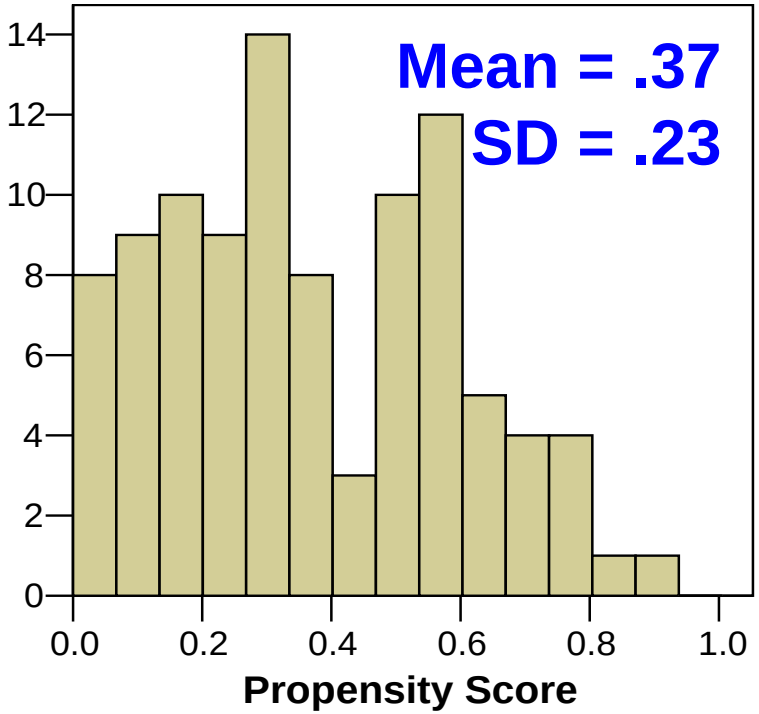
Retained (N=165)



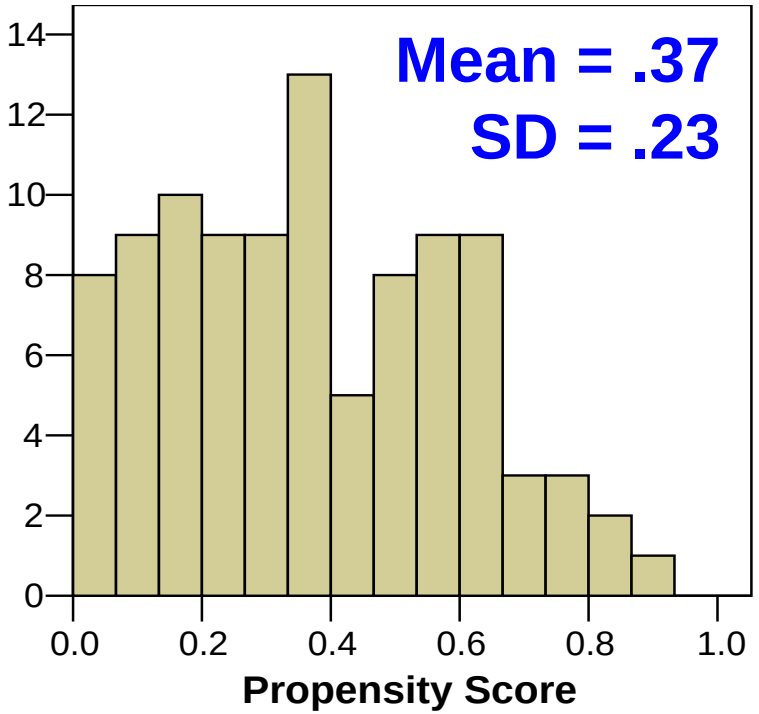
Wu, West, and Hughes (2008)

After Matching

Promoted (N=98)



Retained (N=98)



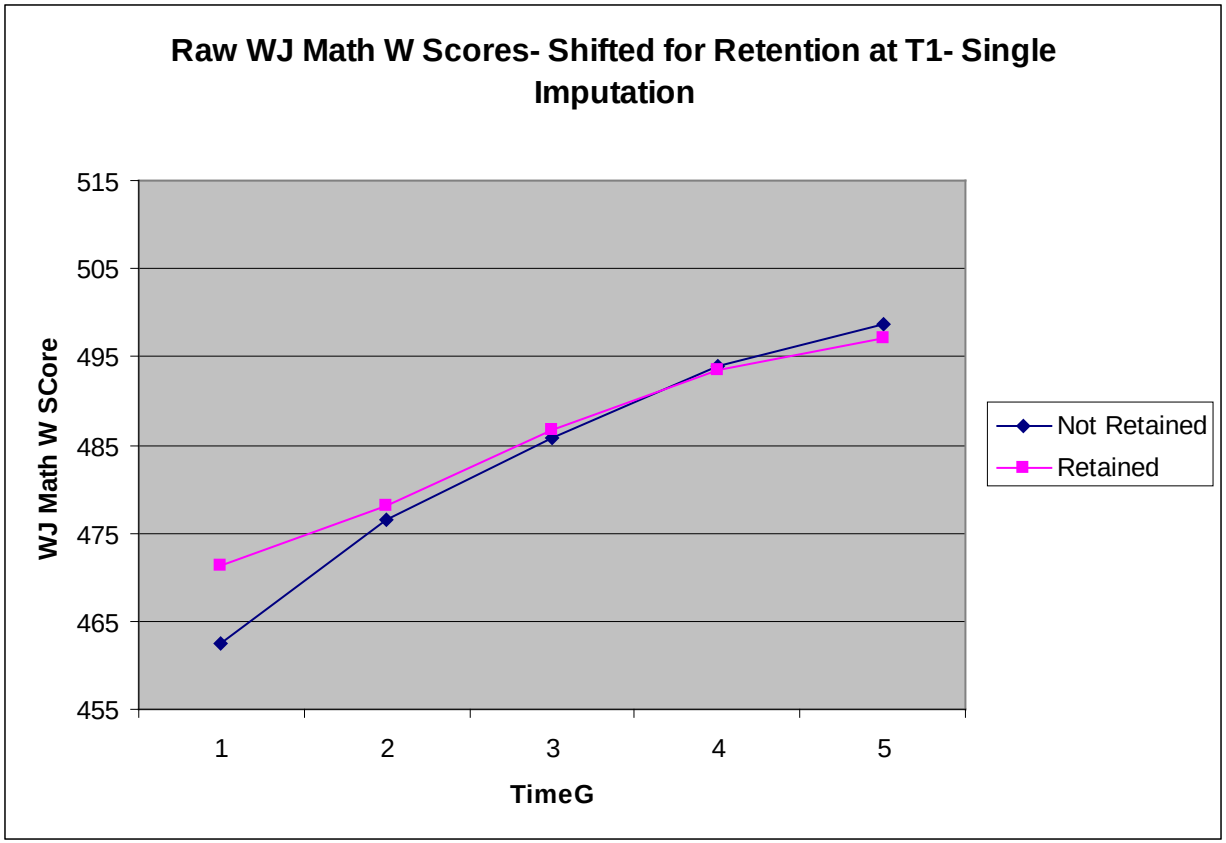
Pretest Balance of *Covariates* in T and C

(a) Conducted 5 (strata) x 2 (T vs. C) ANOVAs—
Fewer than 5% significant at $\alpha = .05$,
maximum d = 0.30

(b) Boxplots showed few differences

Reasonable balance achieved

Results: WJ Math--Grades 1 to 5



Moser, West, and Hughes (2012)

Some Challenges for Propensity Scores 1

(a) Highly dependent on choosing the correct set of covariates.

Subject matter experts: good knowledge of outcomes, poor of selection processes.

(b) Distributions of propensity scores may not overlap—limited region of support

(lower power, limited generalization)

Some Challenges for Propensity Scores 2

c) Estimated PS might not = true PS.

(i) Modeling propensity score

(ii) Selection x maturation—

matched individuals may be growing at different rates. Haviland, Nagin, and Rosenbaum (2007)

(iii) Unreliability/lack of stability of measurement.

Campbell and Rubin: Complementary Approaches

Unique Features, Emphases

Foster different ways of thinking:

Campbell	Rubin
Working Scientist	Mathematical
Validity Threats	Assumptions
Design	Analysis
Inductive/Abductive	Deductive

TABLE 1 Threats to Internal Validity/Key Assumptions and Example Remedies for Randomized Experiments and Alternatives.

TABLE 1

Threats to Internal Validity/Key Assumptions and Example Remedies for Randomized Experiments and Alternatives

Assumption or threat to internal validity	Approaches to mitigating the threat: Design approach	Approaches to mitigating the threat: Statistical analysis strategy
Randomized experiment		
Independent units assumption	Geographical or temporal isolation of units	Multilevel analysis; other statistical adjustment for clustering
Stable unit treatment value assumption (SUTVA); other treatment conditions do not affect participant's outcome; no hidden variations in treatments	Geographical or temporal isolation of units	Statistical adjustment for measured exposure to other treatments
Full treatment adherence assumption	Incentives for adherence	Instrumental variable analysis (exclusion restriction is assumed)
No attrition assumption (measurement of all randomized participants on outcome measure)	Sample retention procedures	Modern missing data analysis (outcome measure assumed to be missing at random)
Regression discontinuity design		
Functional form of the relationship between assignment variable and outcome is properly modeled	Replication with different cutpoint; nonequivalent dependent variable	Sensitivity analysis; nonparametric regression
Interrupted time series design		
Another historical event, a change in population (selection), or change in measures coincides with the introduction of the intervention; functional form of the relationship for the time series is properly specified	Nonequivalent dependent measure; nonequivalent control series in which intervention is not introduced; switching replication in which intervention is introduced at another time point in another group	Sensitivity analysis; diagnostic plots (autocorrelogram; spectral density)
Observational study		
Measured baseline variables equated; unmeasured baseline variables equated; differential maturation	Multiple control groups; nonequivalent dependent measures; additional pre- and postintervention measurements	Propensity score analysis; sensitivity analysis; subgroup analysis

Note. **The list of assumptions/threats to internal validity identifies issues that commonly occur in each of the designs. The alternative designs may be subject to each of the issues listed for the randomized experiment in addition to the issues listed for the specific design. The examples of statistical and design approaches for mitigating the threat to internal validity illustrate some of the commonly used approaches and are not exhaustive. For the observational study design, Rubin's (2005; Imbens & Rubin, in press) and Campbell's (Shadish, Cook, & Campbell, 2002) perspectives differ so that the statistical and design approaches do not map 1:1 onto the assumptions/threats to internal validity that are listed.**

West S G Current Directions in Psychological Science 2009;18:299-304

Selected Recent References

Berzuini, D., Dawid, P., Benardinelli, L. (Eds.). *Causality: Statistical perspectives and applications*. West Sussex, UK: Wiley.

Morgan, S.L. (in press). *Handbook of causal analysis for social research*. New York: Springer.

Morgan, S.L., Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. New York: Cambridge.

Pearl, J. (2009). *Causality: Models, reasoning, inference*. New York: Cambridge.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *JASA*, *100*, 322-331.

Rosenbaum, P. R. (2010). *Design of observational studies*. New York: Springer.

Shadish, W. R., Cook, T. D. (2009). The renaissance of field experimenatation in evaluating interventions. *Annual Review of Psychology*, *60*, 607-629.

Shadish, W. R., Cook, T. D., Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, Houghton Mifflin.

Shadish, W. R., Sullivan, K. J. (in press). Theories of causation in psychological science. In H. Cooper, *APA Handbook of research methods in psychology*. Washington, American Psychological Association.

West, S. G., Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, *15*, 18-37.

West, S. G., Cham, H., Liu, Y. (in press). Causal inference and generalization in field settings. In H. T. Reis, C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd Ed.). New York: Cambridge.