Where are we going?

- What to do if no RCT, RD, ITS or sophisticated matching is possible?
- We describe and analyze principle of pattern matching to improve the basic workhorse design by a feature other than high quality matching
- Illustrate what you might do if there is no pretest at all and so not even a work-horse design is possible

- Before that, a bit of a summary

Yesterday 1

- Do not match from extremes unless forced to
- Then match using a reliable set of measures, moving to propensity score framework, latent variables, or statistical reliability-adjustments to handle unreliability
- But you will still have problem of possible specification error!

Yesterday 2

- In selecting non-equivalent control group:
- Use local focal matching to reduce the degree of initial non-comparability though you cannot expect total non-comparability
- Sometimes the control group so formed will not differ from what would have been achieved with random assignment, at least on observables
- At other times, the initial group non-equivalence will be reduced for when you come to do statistical analysis to “control” initial differences

Yesterday 3

- When there is initial non-equivalence on observables:
- Theory and empiricism about the constructs (covariates) in the “true”selection model helps, as does careful measurement thereof
- Measures in multiple other domains helps too,
- Unclear whether ANCOVA or propensity scores do better--Shadish et al, Glazerman et al.--though demographic variables alone and Heckman IV models do not fare well. Propensity scores preferred on theory alone.
- You can never be sure of the final causal conclusion, though

Implication

- In designing research you do well to avoid the workhorse if you can, though it is modal in current educational practice
- Can you add prior pretest waves?
- Can you add any of the other design elements, some mentioned in our ITS discussion and others we discuss today
- How can you design your way out of reliance on a simple design with non-equivalent groups and pretest and posttest?

When there is no Pretest on the same scale as Outcome

- Do randomized experiment--Abcedarian, Perry-Preschool, Head Start, Early Head Start, CLIO, Even Start and Sesame Street
- If not possible, do everything possible to make control group focally local
- Add design elements to rule out alternative explanations of a possible causal relationship
- Here’s an example

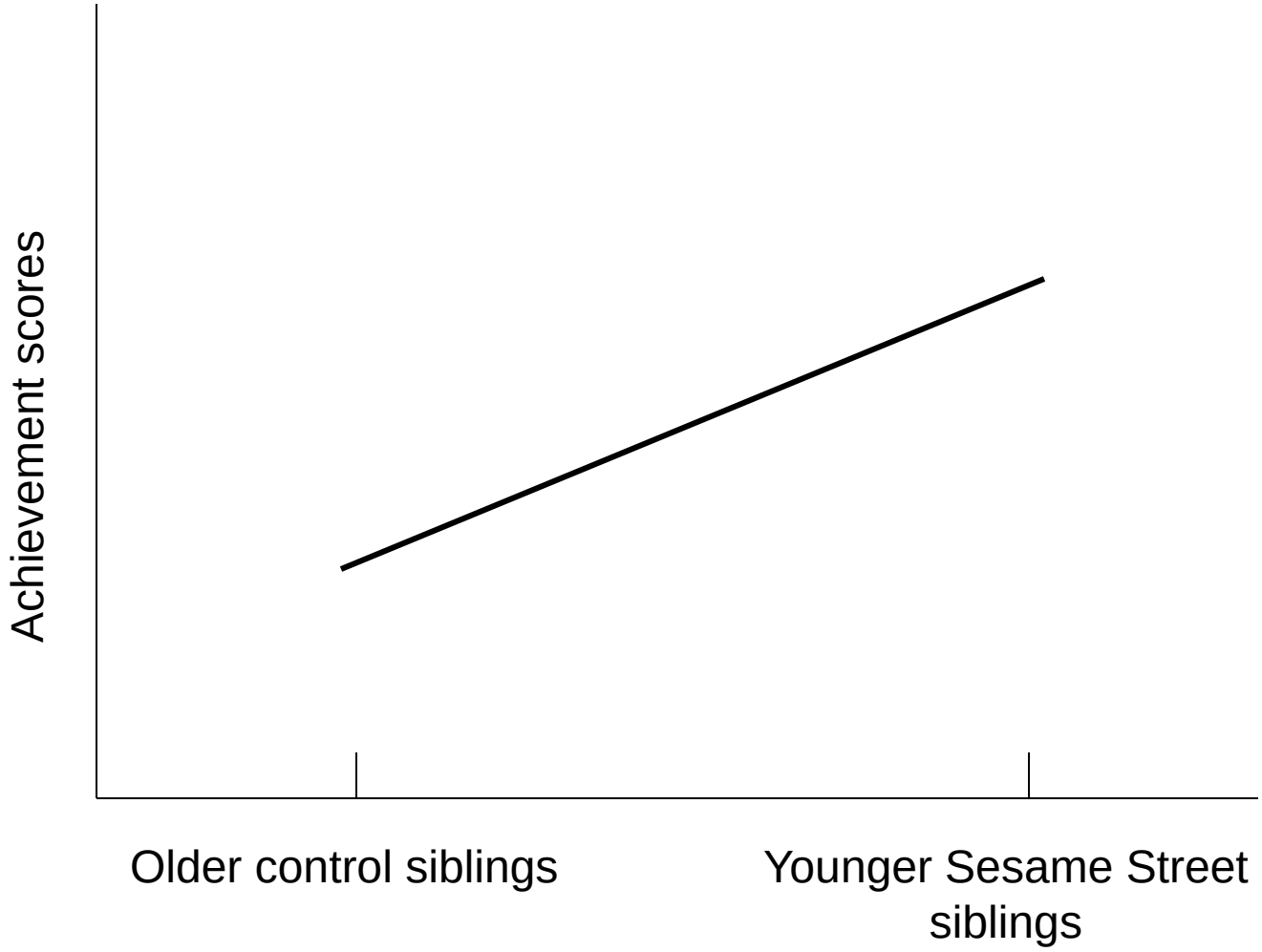
Minton's Dissertation

- Object: Evaluate Sesame St in 1st year
- Problem 1: Program already launched
- Problem 2: No pretest possible
- Problem 3: No money for original data collection
- Setting: One kindergarten in NJ that built SS into its day and that has records on children and their families plus annual PPV assessment

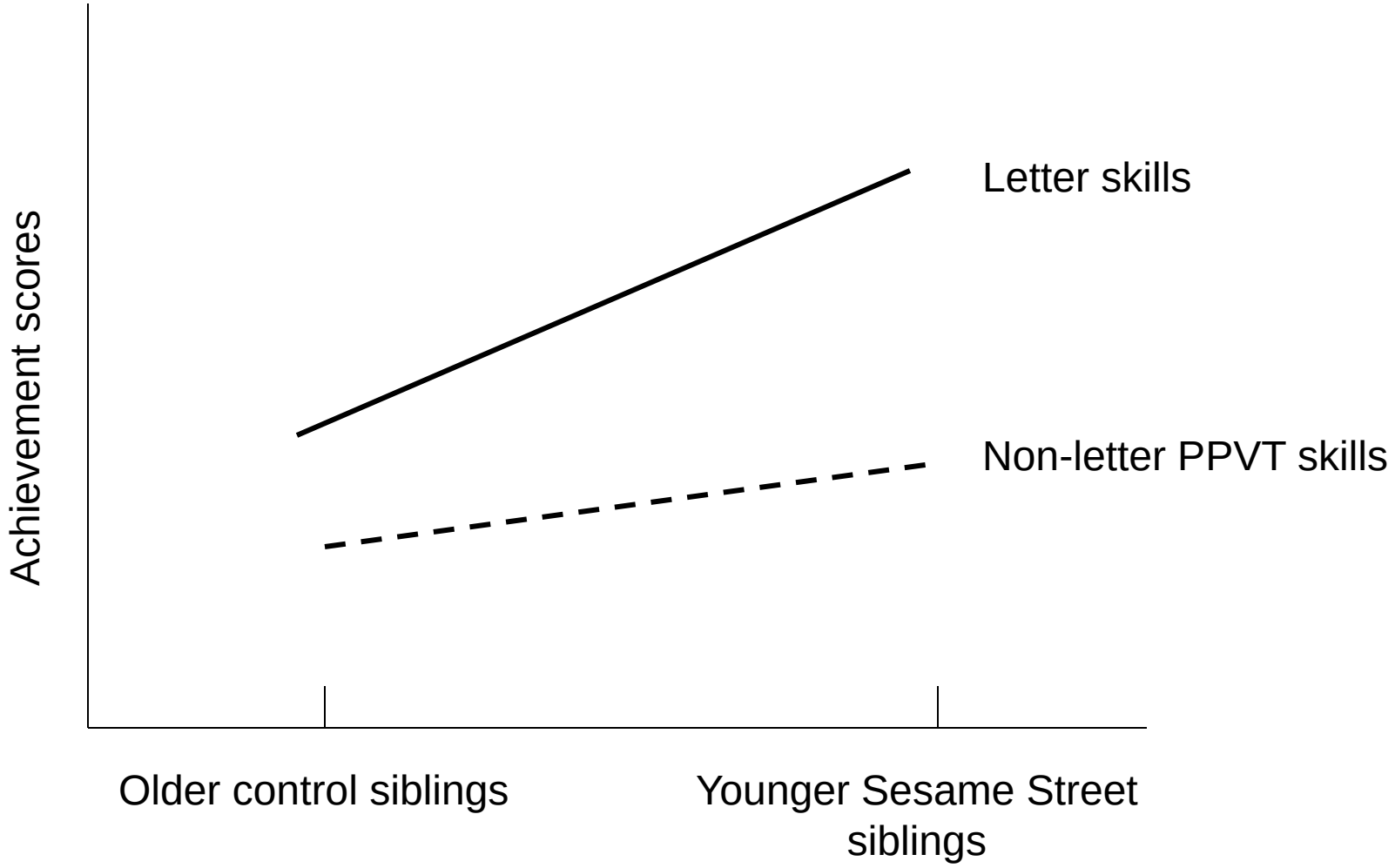
Question 1: What control group is possible?

- What control group to find, given program was very popular in its first year.
- Why is popularity a problem?
- Neighborhood kids who did not view
- Next-door kids of same age who not view?
- Older siblings in general
- Older sibs attending same kindergarten within last N years
- Older sibs attending same kindergarten last 2 years

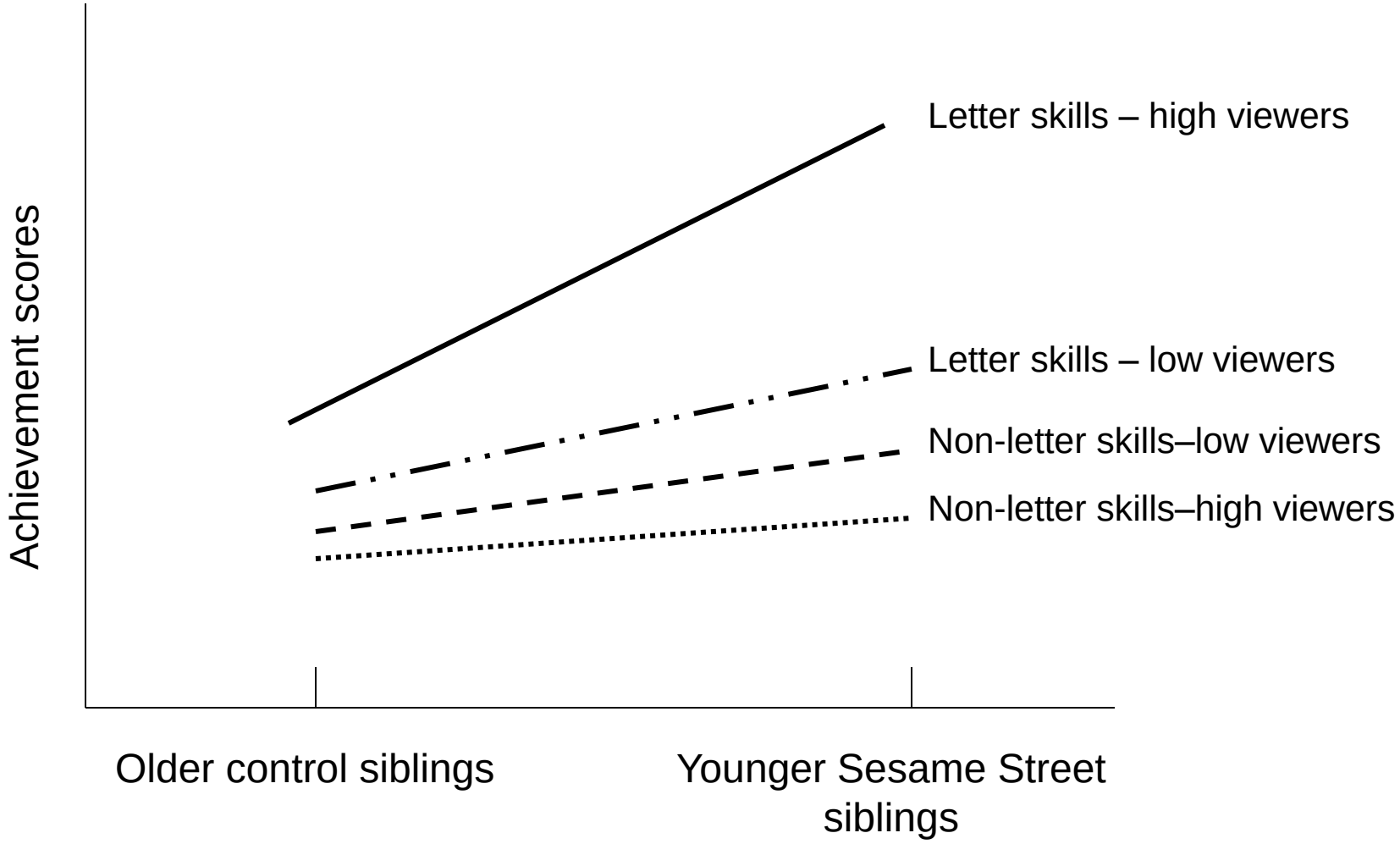
Minton (1975) Sesame Street Study - 1



Minton (1975) Sesame Street Study - 2



Minton (1975) Sesame Street Study - 3



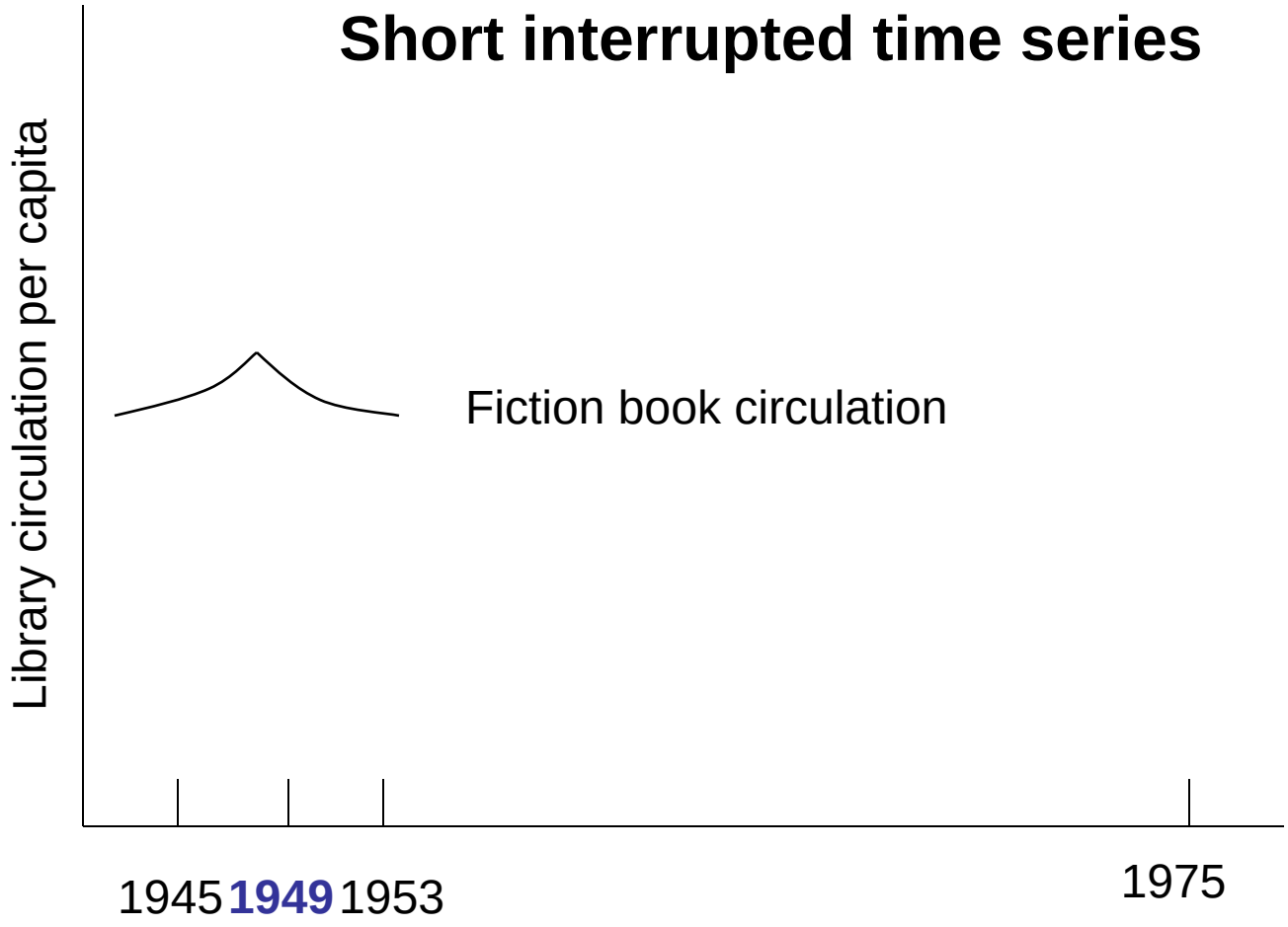
What has happened here?

- Single causal hypothesis of SS effective made to have multiple data implications
- These are meant to rule out alternative hypotheses and not to recreate same bias
- These implications here in the form of a difference in difference in differences
- Collect data and test hypothesis

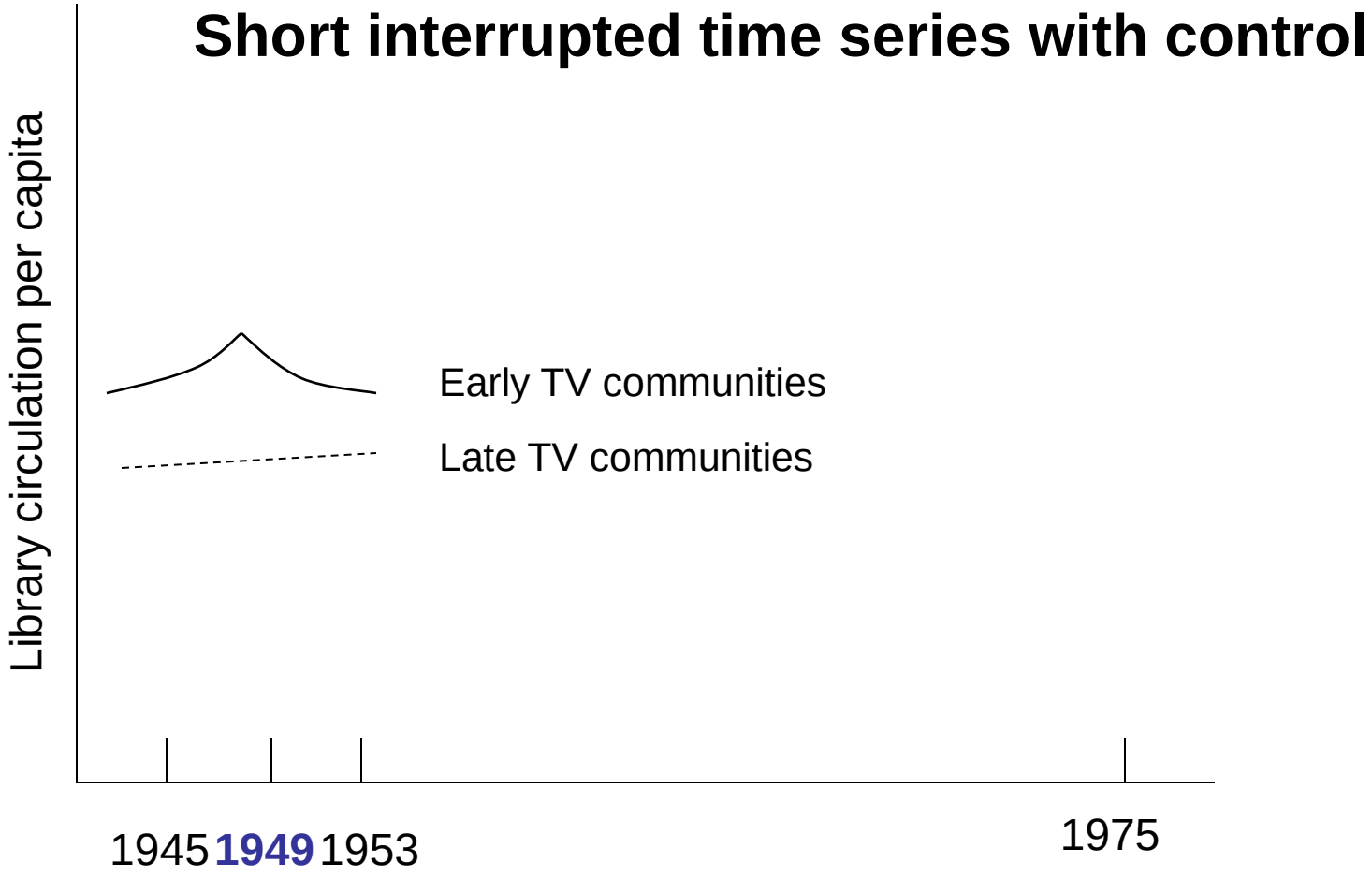
Another Example

- How the Introduction of TV affected Library Circulation

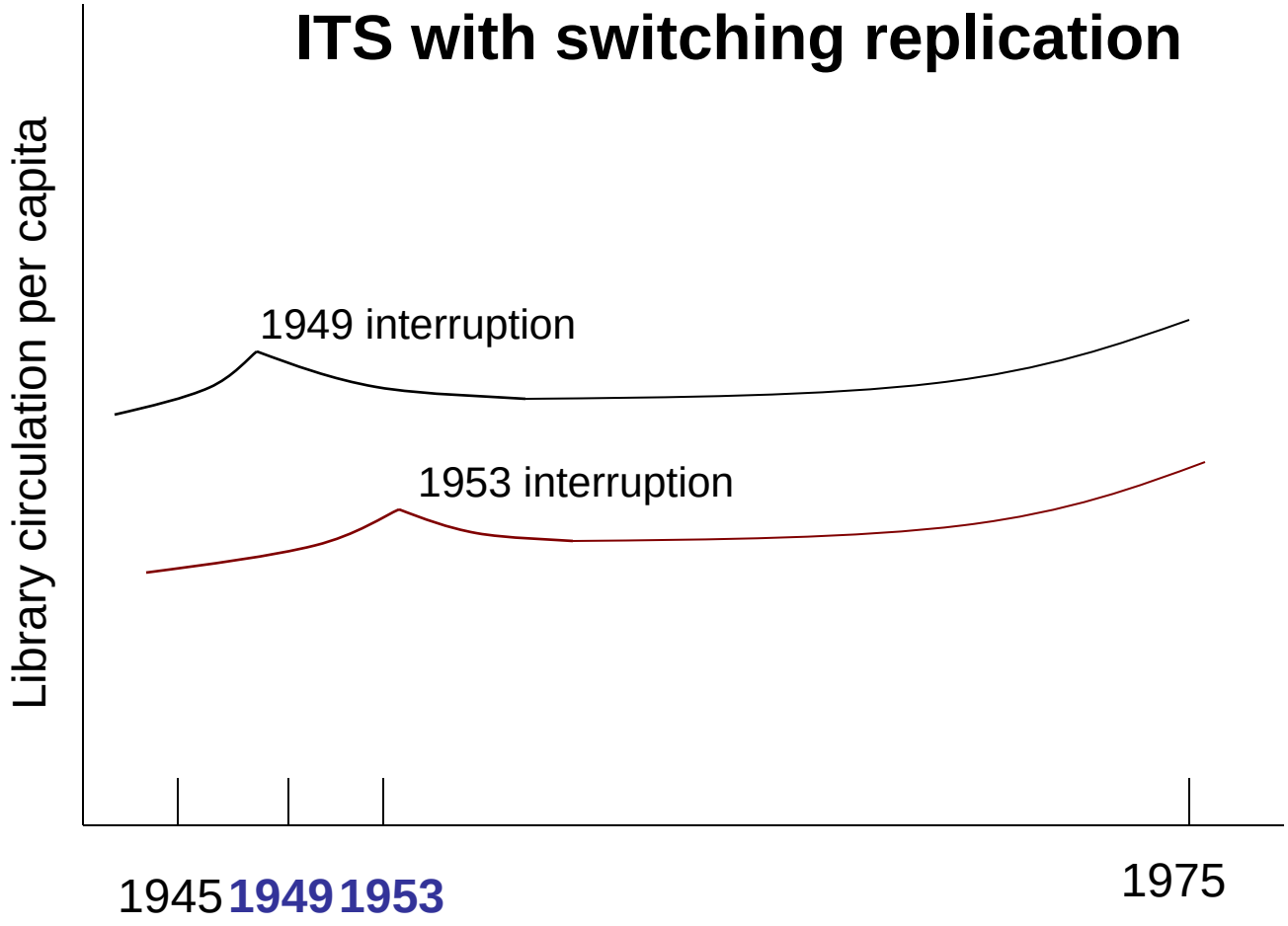
Parker et al. (1966) Effects of TV - 1



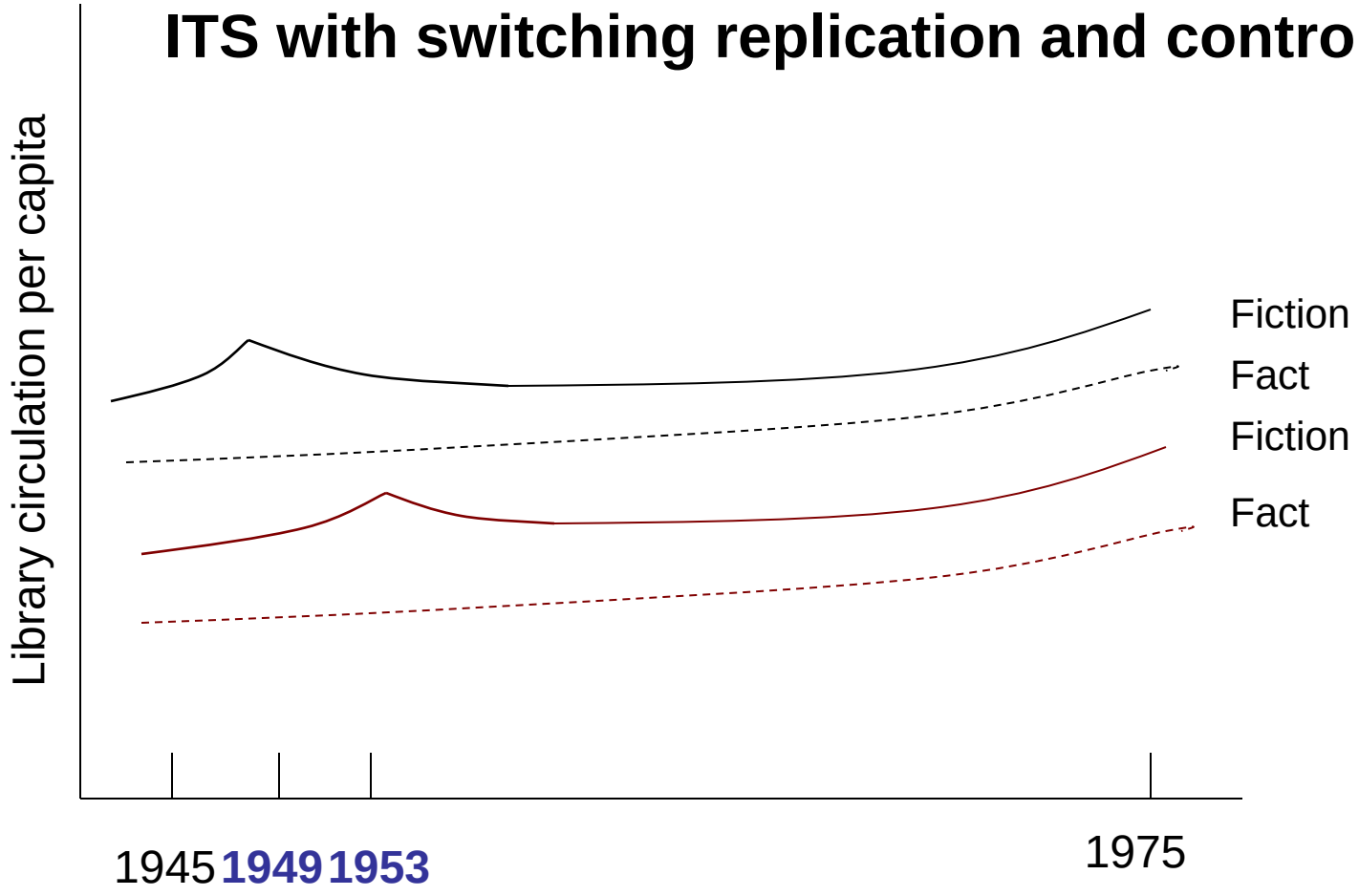
Parker et al. (1966) Effects of TV - 2



Parker et al. (1966) Effects of TV - 3



Parker et al. (1966) Effects of TV - 4



What has happened here?

Combine an ITS with non-equivalent DVs and switching replications

What alternative interpretations can you come up with?

How plausible are these?

- Have we seen this before with RD and ITS?
- One general causal hypothesis has multiple implications in the data
- Predicted hypothesis as multiple differences of differences; as higher order interactions

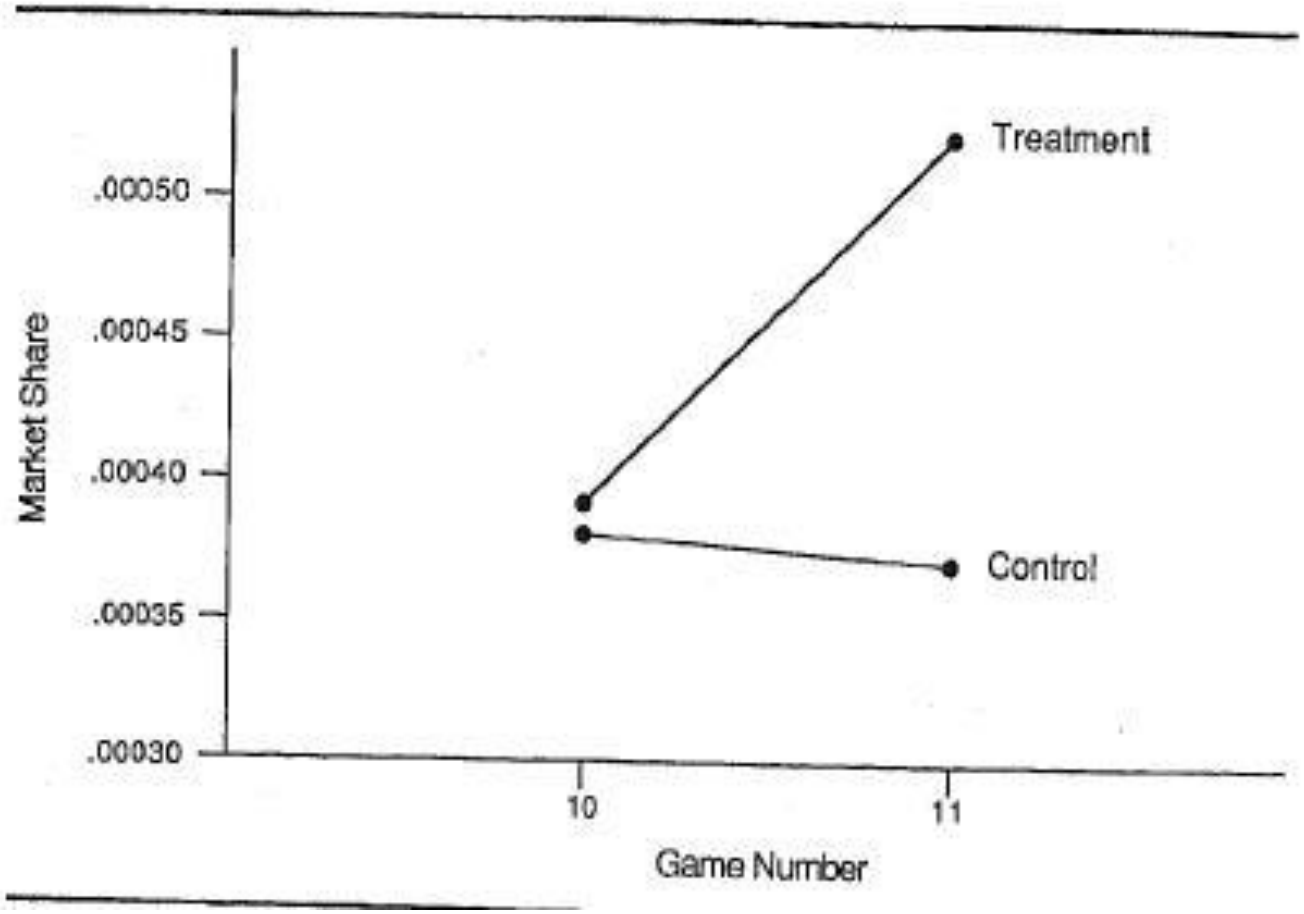
Reynolds and West's (1987) "Ask for the Sale" Experiment

From all stores selling lottery tickets, some stores volunteered (or not) to post a sign reading "Did we ask you if you want a Lottery ticket? If not, you get one free". So this is a basic nonequivalent control group design, with the control matched on zip code, store chain, and pretest ticket sales.

NR O1 X O2

NR O1 O2

The Outcome of the Basic Design



But there might be many reasons besides treatment that caused treatment group sales to rise.

Adding a Nonequivalent DV

- They added three **nonequivalent dependent variables**, showing that the intervention increased ticket sales but not sales of gas, cigarettes, or grocery items.

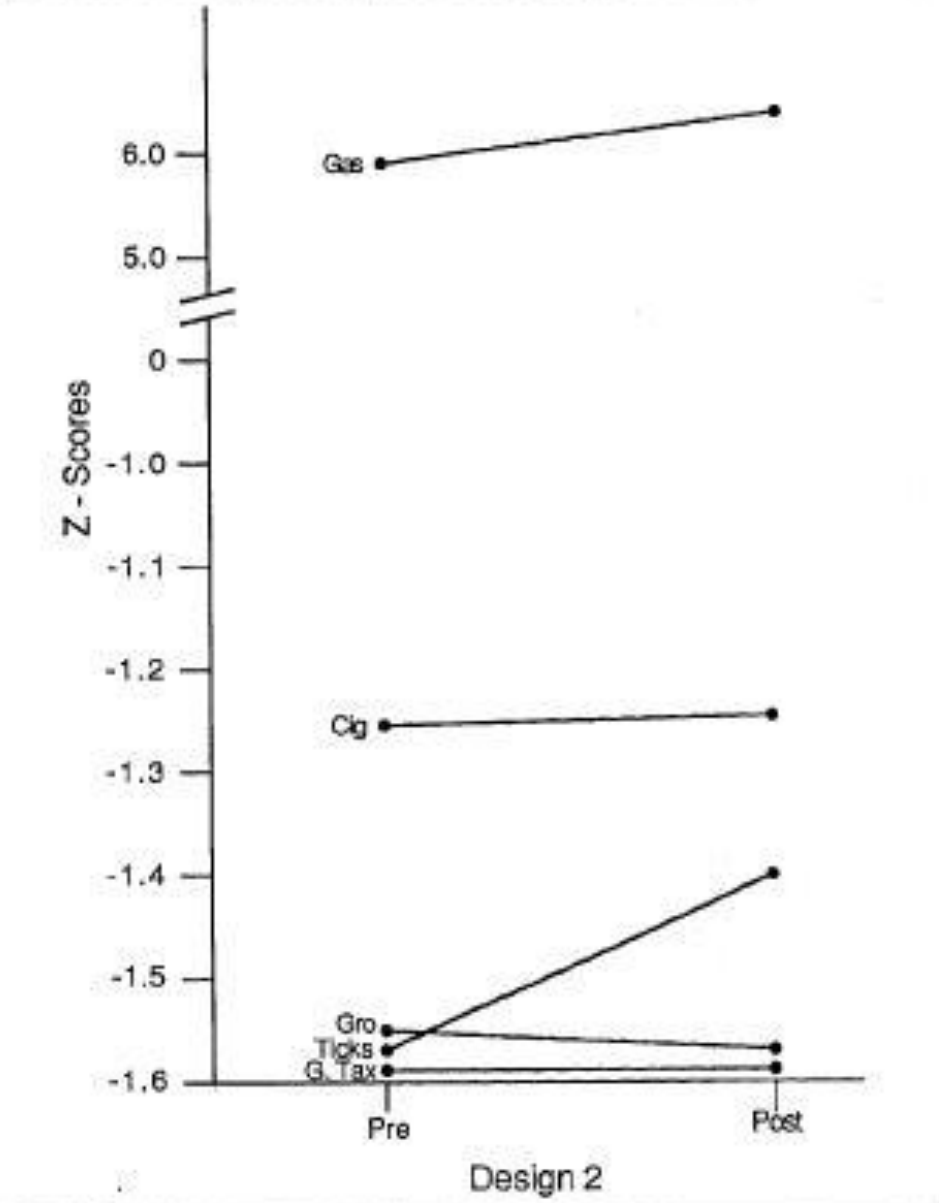


Figure 2: Pre-Post Change in Sales for Each Dependent Variable in Design 2
Note: Department variables are gasoline sales (Gas), cigarette sales (Cig), non-taxable groceries (Gro), lottery ticket sales (Ticks), and taxable groceries (G. Tax).

DESIGN 3

Adding Multiple Baselines - recasting as ITS Design

- They located some stores in which the **treatment was initiated later than in other stores**, **or initiated and then removed,** and found that the outcome tracked the introduction of treatment over time while sales in the matched controls remained unchanged

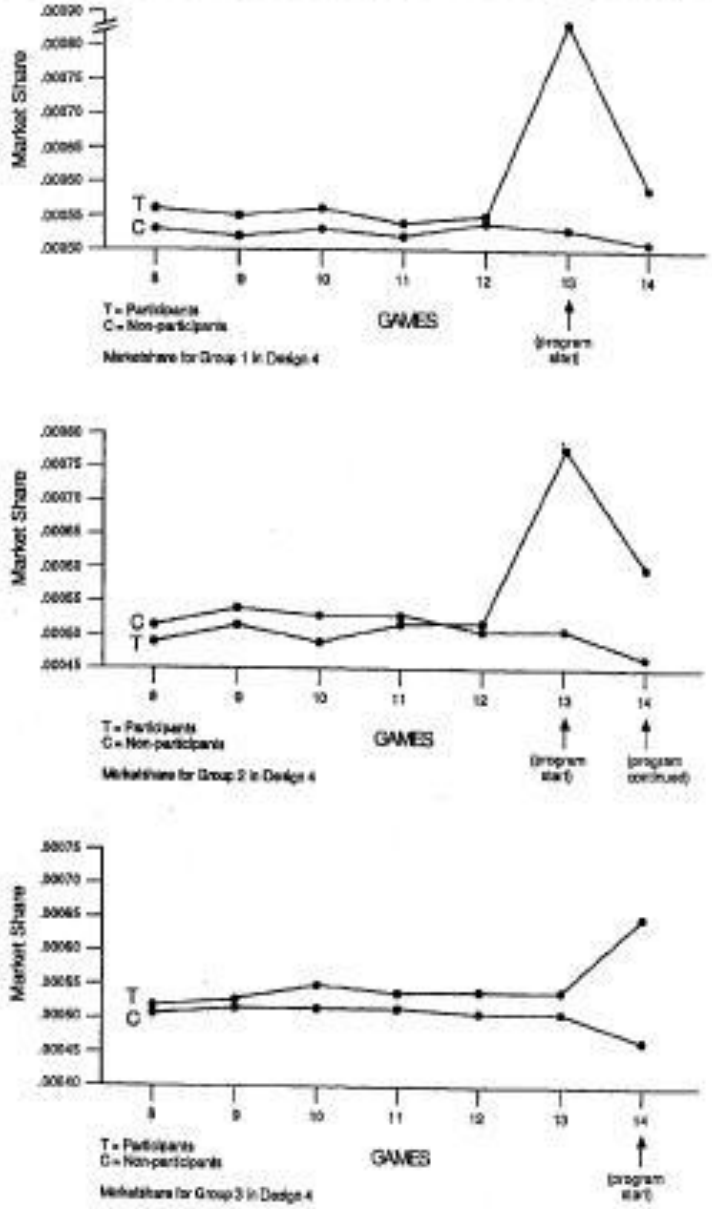


Figure 4: Mean Marketshare for Treatment and Corresponding Control Stores in Group 1, Group 2, and Group 3

Note: Treatment stores in Group 1 began the program during game 13 and discontinued the program during game 14. Treatment stores in Group 2 began the program during game 13 and continued during game 14. Treatment stores in Group 3 began the program during game 14.

Adding Multiple Pretests and Posttests

- They added **multiple pretests and posttests** by examining mean weekly ticket sales for four weeks before and four weeks after the treatment started.

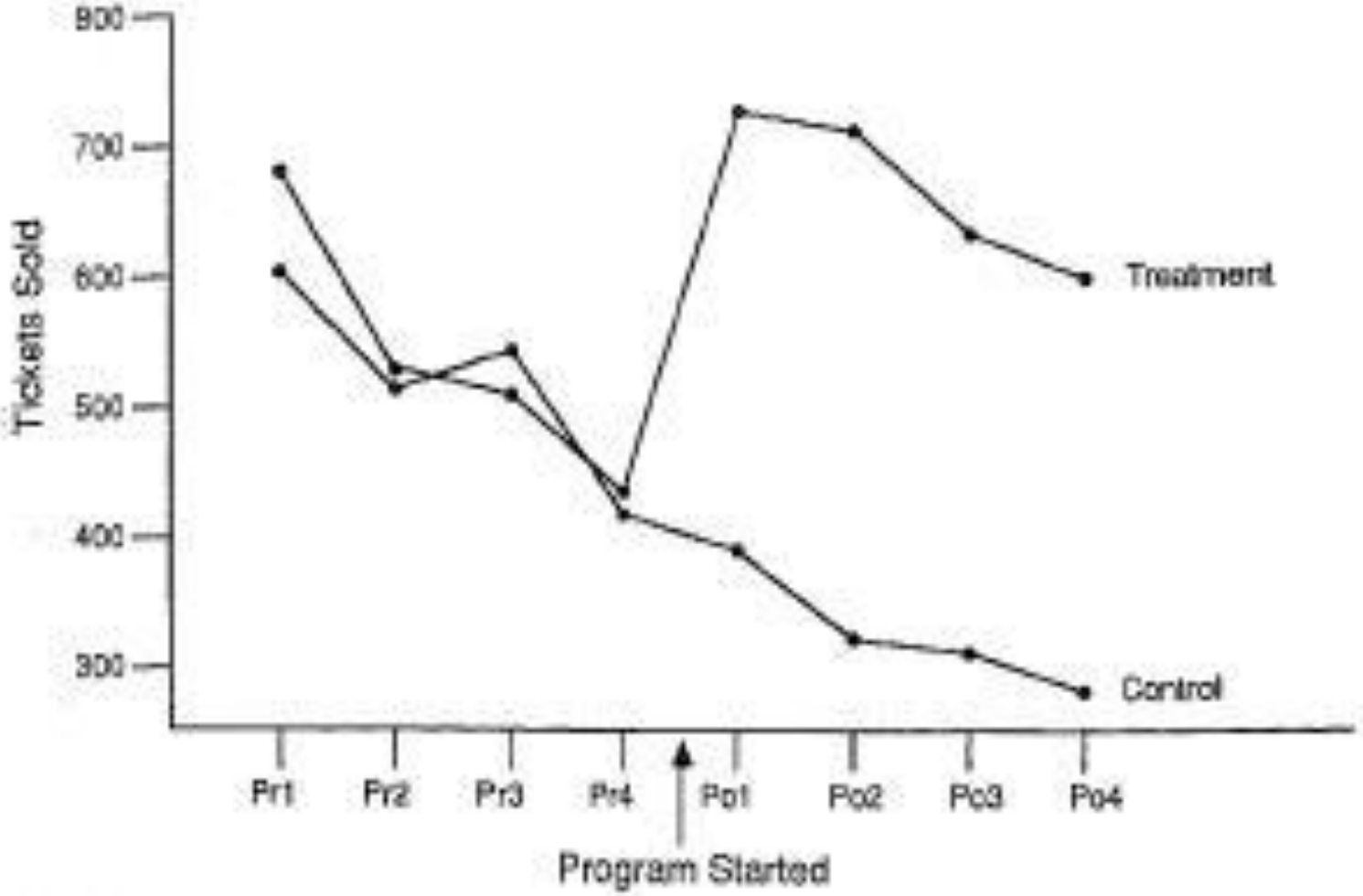


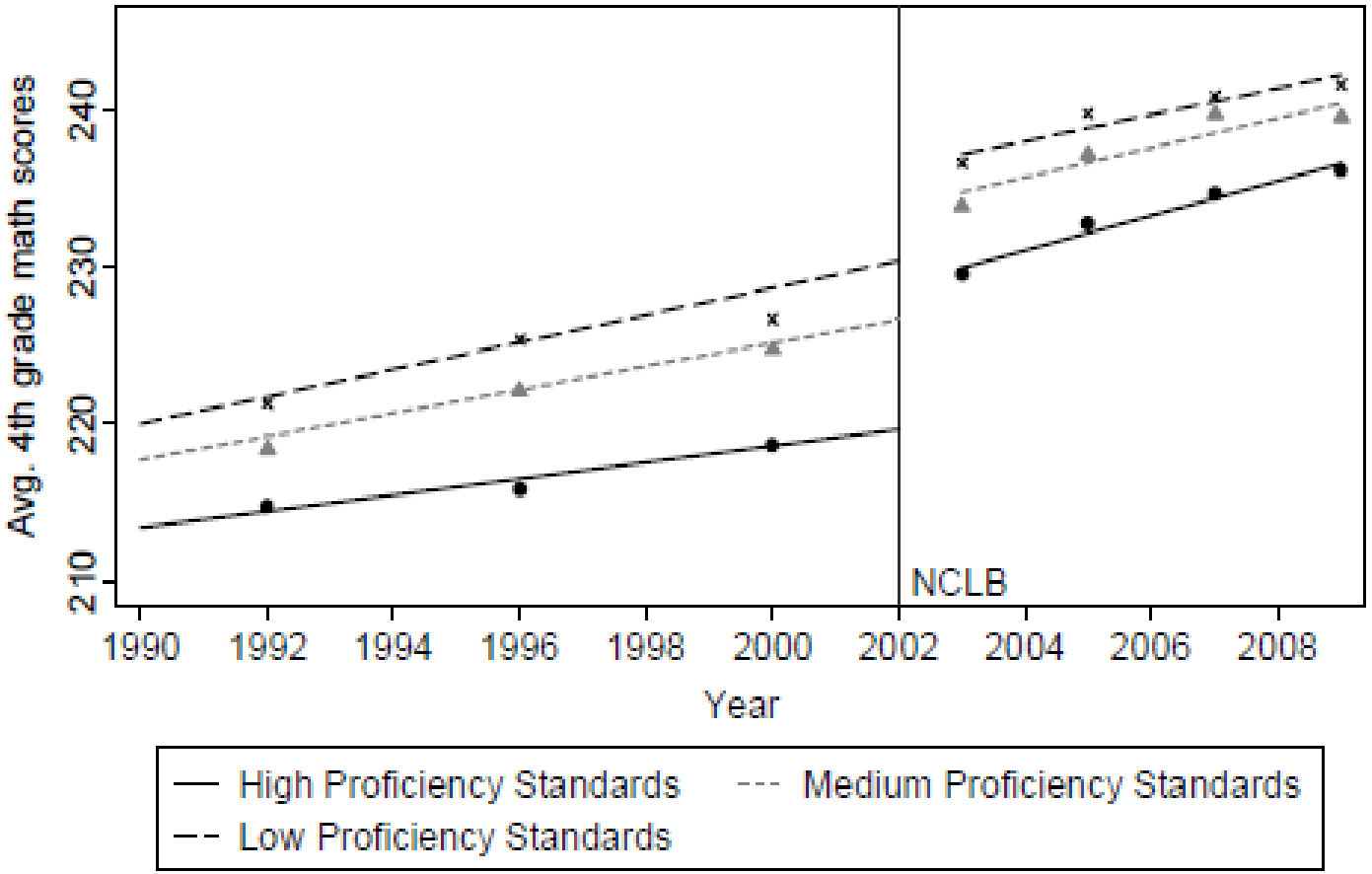
Figure 3: Mean Weekly Ticket Sales Pre- and Posttreatment

Note: Raw mean weekly ticket sales for the 4 weeks prior to treatment (Pr1 – Pr4) and the 4 weeks after treatment (Po1 – Po4) are plotted for treatment (T) and control (C) stores.

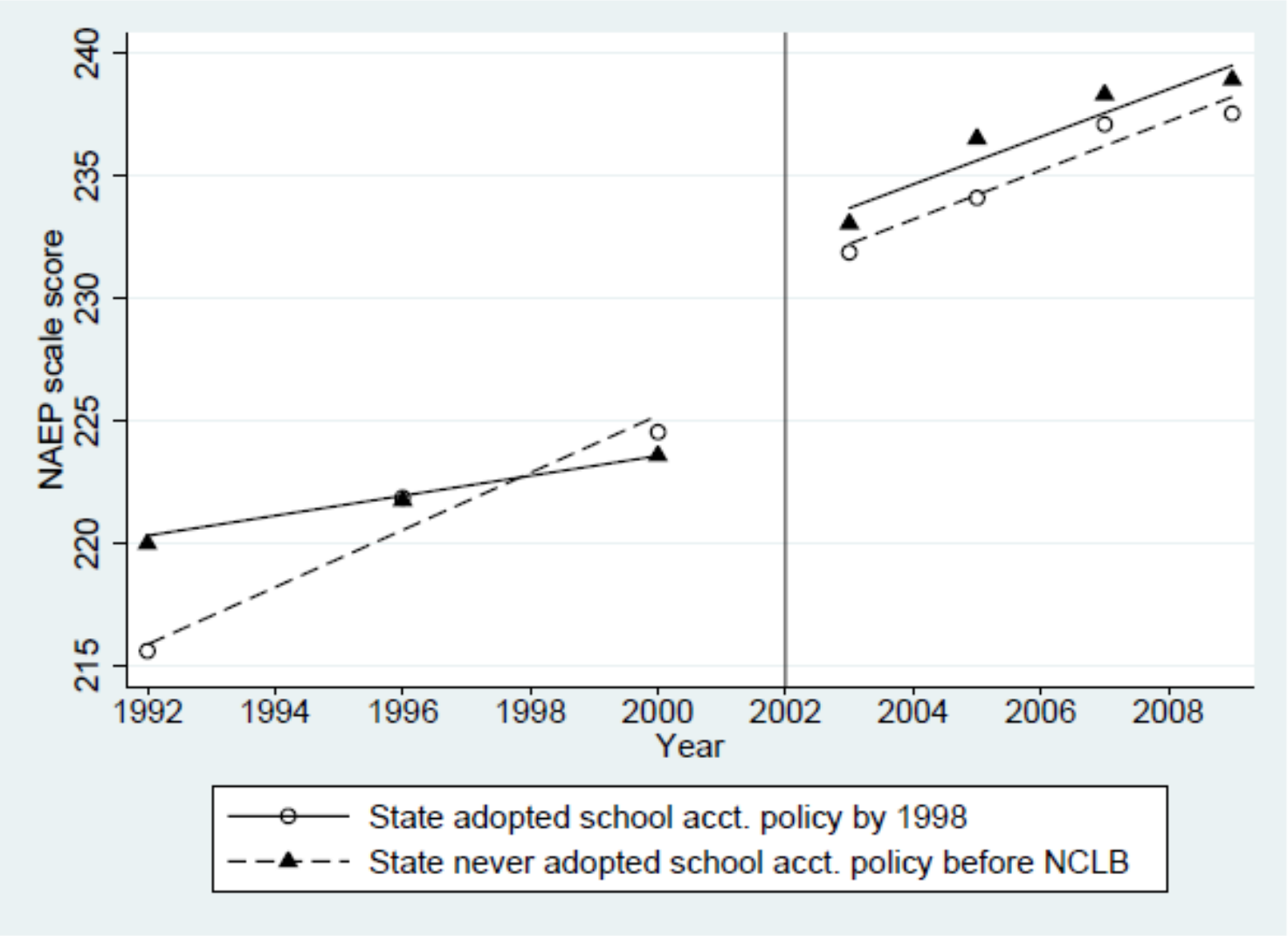
The Point is:

- To use the choice of additional design elements to rule out more alternative interpretations, hopefully all that can be currently identified
- The goal is ruling out plausible alternative interpretations, and it can also be reached via keeping the pattern of results constant but varying the number and type of comparisons involved

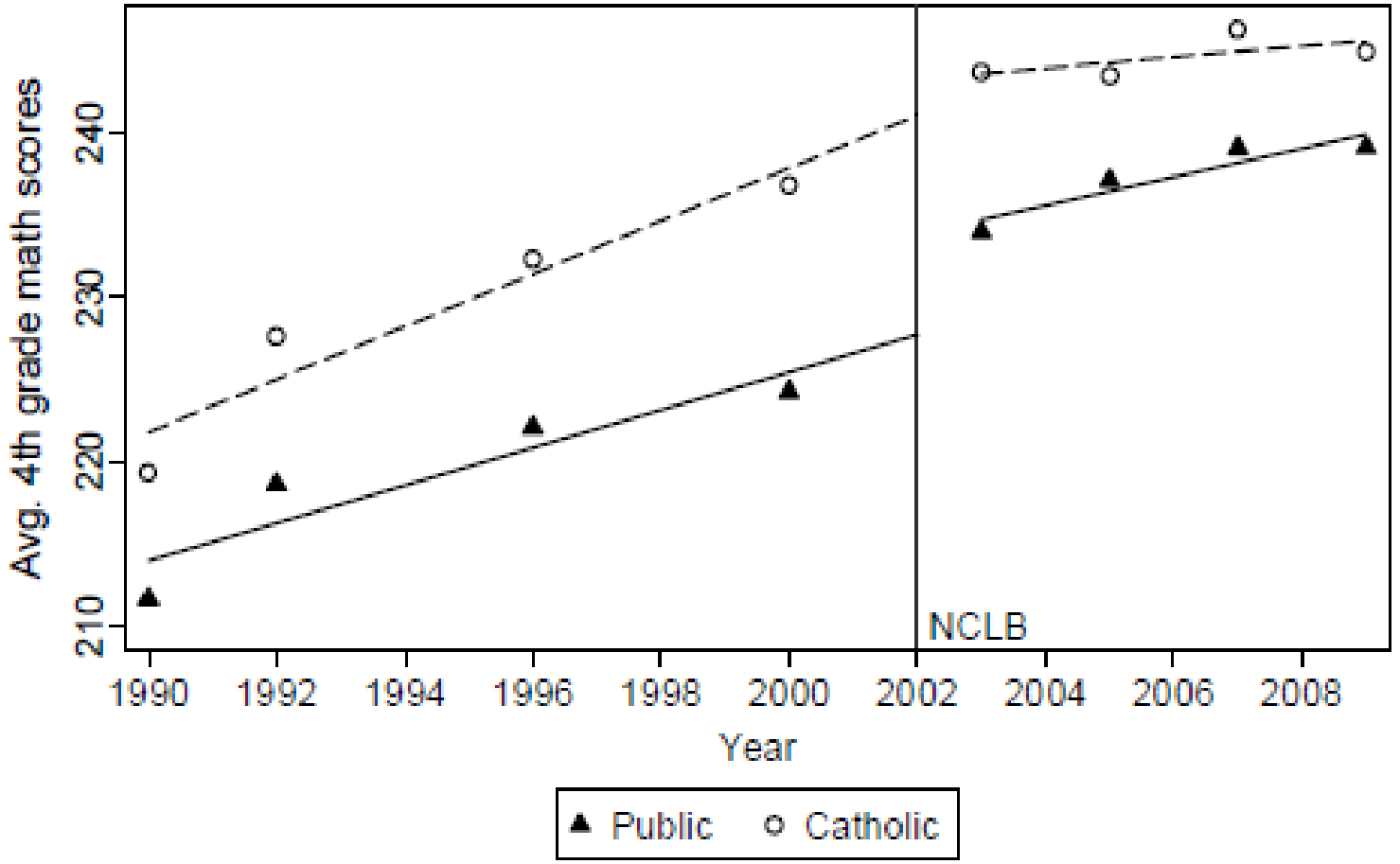
Main NAEP 4th grade math scores by year and proficiency standards



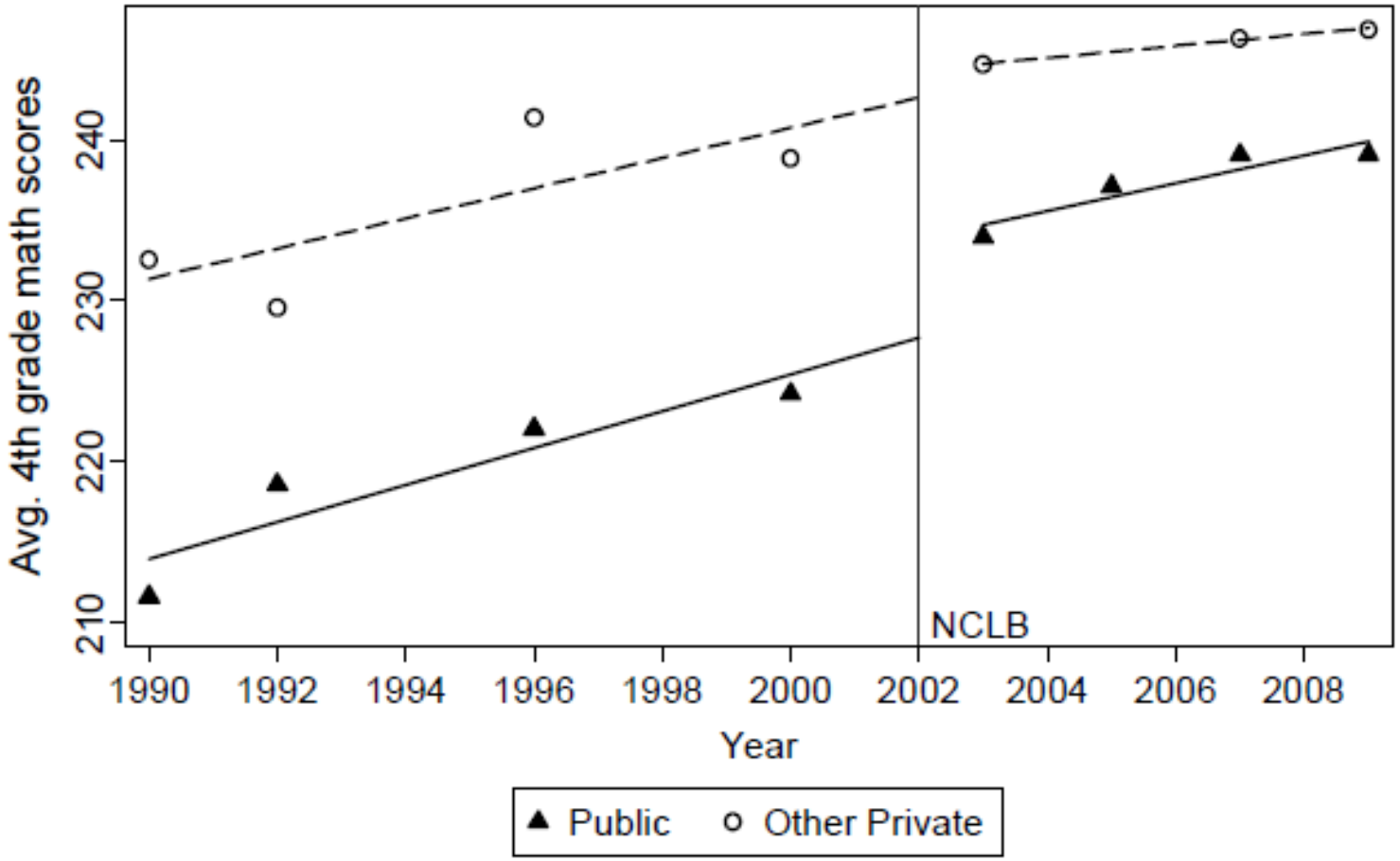
D & J Results: 4th Grade Math



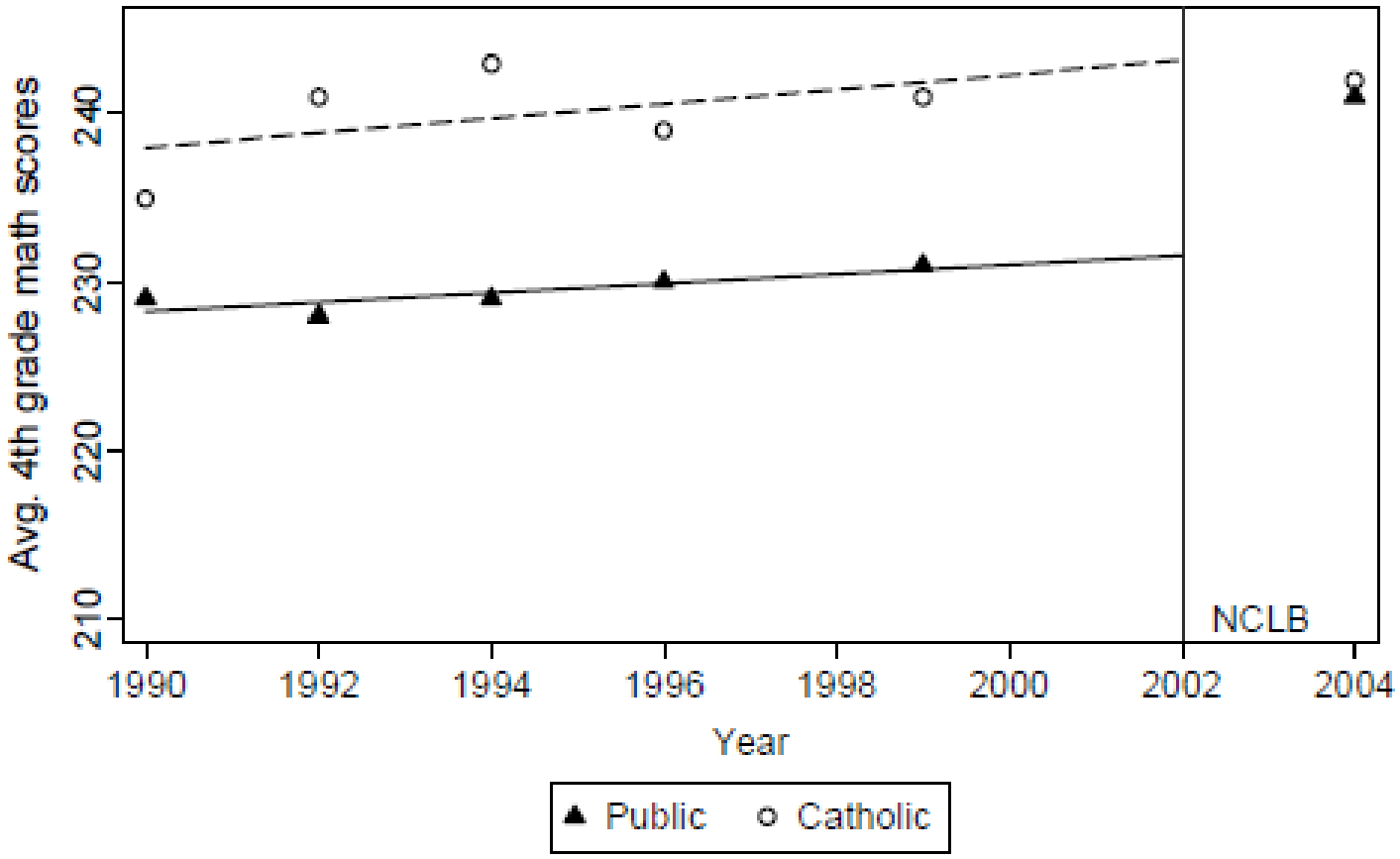
Main NAEP 4th grade math scores by year: Public and Catholic schools



Main NAEP 4th grade math scores by year: Public and Other Private schools



Trend NAEP 4th grade math scores by year: Public and Catholic schools



Student Enrollment

	Catholic	Other Private	Public
1994	5.73	4.72	89.55
1996	5.67	4.74	89.60
1998	5.58	4.87	89.56
2000	5.38	4.81	89.81
2002	5.26	5.13	89.61
2004	4.88	4.93	90.18
2006	4.56	5.07	90.37

Warning!

- This pattern matching strategy requires:
- Clear causal hypothesis - relevance to discontinuity
- Careful measurement - reliability and ceiling or floor effects
- Large samples (or large effects) because hypothesis is of a complex statistical interaction
- How lucky Minton was!

Examples from you of the Basic Work-Horse Design

- Let us take some from you and see if they can be improved by adding design elements.

Design Elements to be combined: Assignment

- Random Assignment
- Cutoff-Based Assignment
- Researcher-controlled Matching -- of many kinds in econometric literature

Design Elements to be combined:Treatments

- Switching Replications
- Reversed Treatments
- Removed Treatments
- Repeated Treatments

Design Elements to be combined: Measurement

- Single Pretest
- Pretest Time Series
- Proxy Pretests
- Retrospective Pretests
- Moderators with predicted Interactions
- Measuring Threats to Validity

Design Elements to be combined: Comparison Groups

- Single Non-Equivalent Groups
- Multiple Non-Equivalent Groups
- Twins/Siblings
- Cohorts
- Other Focal, Local Comparison Groups

Golden Rules (1)

- You can’t put right through statistics what you have done wrong by design
- Statistical adjustments work better the less non-equivalence there is to adjust away in the first place
- Since the work horse is so prevalent but so problematic, how can we complexify the design through adding design elements

Golden Rules (2)

- First, Do an experiment; if not
- Do Regression-discontinuity study. If not,
- Do ITS with some sort of a comparison series. If not
- Do study combining multiple design element, preferably with focal local intact controls, case matching on many covariates, reintroduction of treatment at new time, non-equivalent DVs, etc.

Golden Rules (3)

Don’t be bamboozled by fancy models in Greek clothing. Always translate them into structural design elements before evaluating their likely validity. That will reveal what you have got

- Remember you only control for the reliably measured part of any construct, not the construct itself.

Evaluation, formative

On a scale from 1 to 6, with 6 being high, please rate the following and then indicate how you would improve what we did.

Contact with Valerie about the workshop

Accommodations

Food

Curriculum Content

Curriculum Relevance to your current or anticipated work

Quality of Instruction

Any other Suggestions for Improvements?