

QUANTITATIVE METHODOLOGY AND PUBLIC POLICIES: NEW TECHNIQUES AND OLD PROBLEMS

Christopher C. Weiss
Columbia University

8. June. 2010



Overview



- Focus on Estimating Causal Effects of Public Policy
- Randomized Control Trial
- Non-RCT Designs
- Propensity Score Matching
- Example from U.S. Education Policy

Causal Effects of Public Policy

- Evaluating effects of public policy has long, rich history.
- Recent increase in attention to estimating *causal* effects of policy or policy changes.
 - ▣ Not enough to know that outcomes change
 - ▣ Heightened attention to how/why/how much change attributable to policy

Causal Effects of Public Policy

- Has led to a number of randomized control trials in policy evaluation.
- Drawing upon large body of research, theory, and practice in medicine and public health.
- Now in increasing usage in social science and in evaluating social policy.

Randomized Control Trial

- Class of research studies in which participants are randomly assigned to distinct groups (treatment, control).
 - ▣ Random assignment eliminates bias in treatment assignment (selection bias, confounding)
 - ▣ Creates two (or more groups) that, on average, are virtually identical to each other

Randomized Control Trial

- If conditions are met (and RCT is executed well), we can attribute differences in outcomes to the different treatments
 - ▣ rather than any other characteristics of the experimental subjects
- “Gold standard” for making causal inference - and making causal statements about policy’s effects.

Randomized Control Trial

- Conditional Cash Transfer Programs
 - ▣ Progresa/Oportunidades (México)
- Education Reforms
 - ▣ Class size
 - ▣ Vouchers
- Criminology
 - ▣ Recidivism reduction
 - ▣ Policing strategies

Randomized Control Trial

- Strengths of method are many.
 - ▣ Primary one: allows researcher to make statements about causal effect of program/policy.
 - Any differences observed can be attributed to the policy.
 - ▣ Eliminates competing explanations

Randomized Control Trial

- But some weaknesses as well
 - ▣ Can be difficult – even impossible – for some questions.
 - ▣ Expense can be great.
 - ▣ Can be difficult to ensure fidelity to model.
 - ▣ And, in some cases, element of artificiality.

Non-RCT Designs

- Many examples of these kinds of studies:
 - ▣ Quasi-experiments
 - ▣ Case-control studies
 - ▣ Repeated cross-sectional
- Key distinction is that groups compared are not created through random assignment.
 - ▣ To reiterate, random assignment is not always possible.

Non-RCT Designs

- Advantages
 - ▣ Cost much less
 - ▣ Often take less time
 - ▣ May be more generalizable
 - ▣ Larger samples

Non-RCT Designs

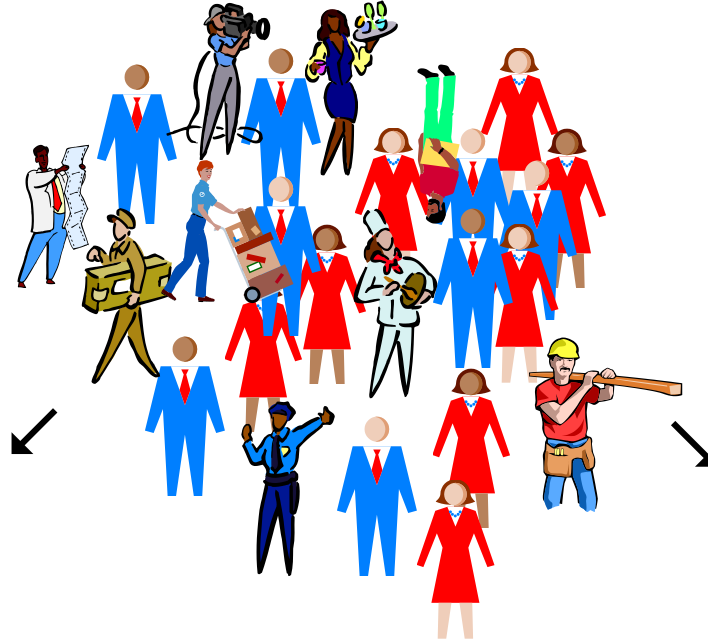
- Primary disadvantage: susceptible to selection bias .
 - ▣ Groups may self-select
 - ▣ Mechanism of selection may be related to outcomes
 - ▣ Baseline characteristics can confound.
- One approach to address the bias in nonrandomized design is through propensity score matching.

Propensity Score Matching

- The logic of propensity score techniques is based in theory of the counterfactual.
 - ▣ Employ language of treatments and control.
 - We can observe outcomes for those who received treatment, those who did not.
 - ▣ Want to estimate what would have happened if people who received treatment hadn't; what would have happened if control group had received treatment.
 - ▣ Cannot truly examine counterfactual.
 - ▣ Propensity score techniques provide a correction strategy that allows estimation of counterfactual.

Propensity Score Matching

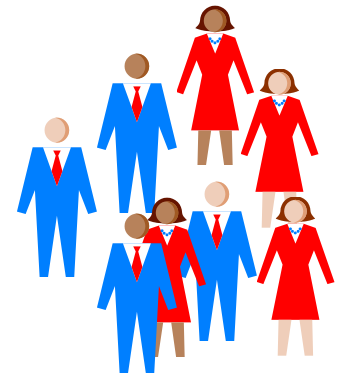
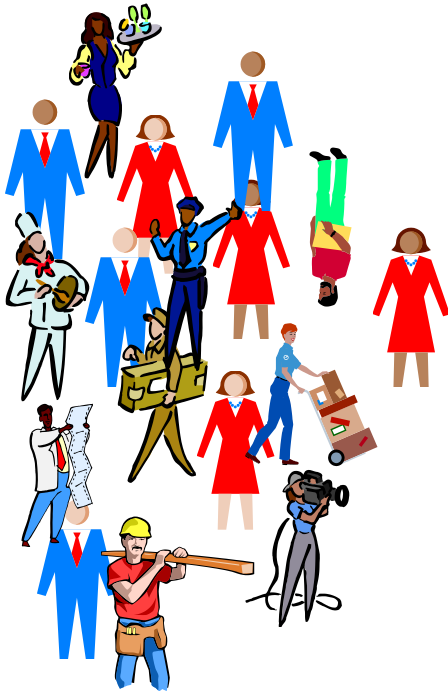
- Propensity score techniques provide a correction strategy that allows estimation of counterfactual.
- Employs a predicted probability of group membership—e.g., treatment vs. control group--based on observed predictors, usually obtained from logistic regression to create a counterfactual group.
- Propensity scores may be used for matching or as covariates—alone or with other matching variables or covariates.

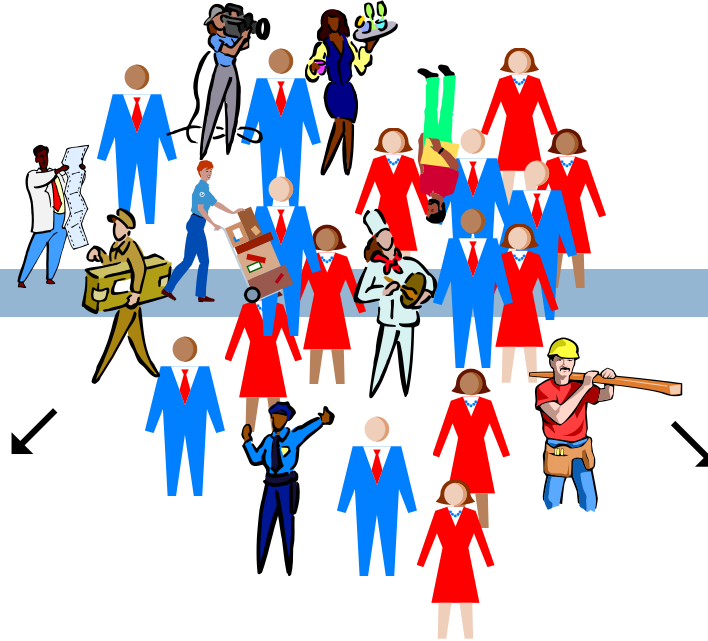


Receives no
training

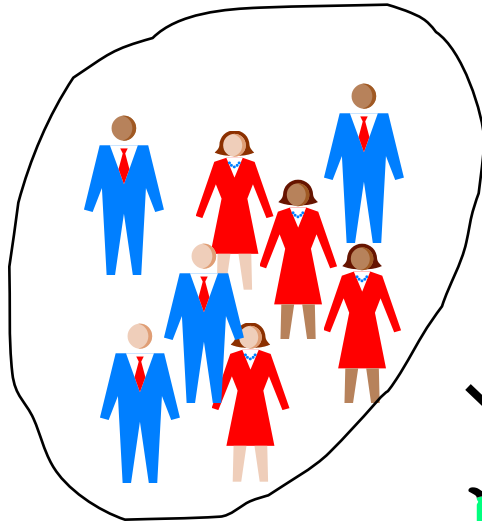
Receives
training

Self-selection into
treatment groups





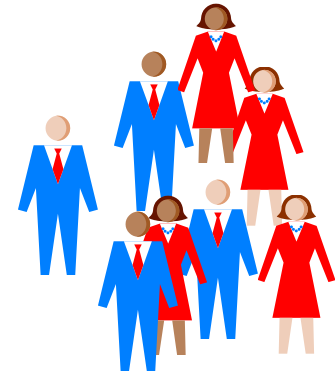
Receives no
training



Propensity score
matching identifies
most similar group
of people



Receives
training



Propensity Score Matching

- Propensity score is the conditional probability of receiving a treatment given a vector of measured (observed) covariates.
- Particularly useful strategy in cases where random assignment cannot be used.
- Talk about procedure through example of educational policy in the United States.
 - Grade retention

Background – Grade Retention Policy

- Great deal of recent attention to policies of retention and social promotion – particularly focused on increasing likelihood of retention.

Background – Grade Retention Policy

- Great deal of recent attention to policies of retention and social promotion – particularly focused on increasing likelihood of retention.
 - ▣ Proponents: children should not progress to the next grade if they are not adequately prepared, it's bad for them and their classmates

Background – Grade Retention Policy

- Great deal of recent attention to policies of retention and social promotion – particularly focused on increasing likelihood of retention.
 - Proponents: children should not progress to the next grade if they are not adequately prepared, it's bad for them and their classmates
 - Opponents: not much evidence that grade retention is effective academically (particularly in the long-term), may cause problems emotionally/behaviorally, and it is expensive

Background – Grade Retention Policy

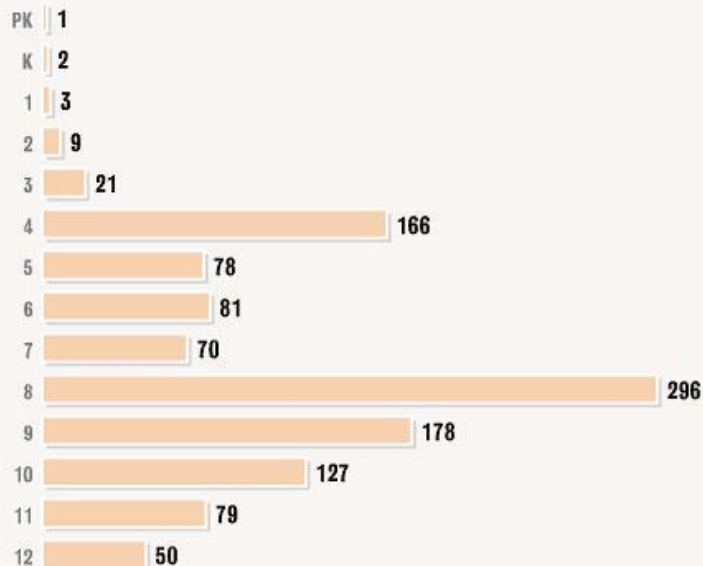
- Great deal of recent attention to policies of retention and social promotion – particularly focused on increasing likelihood of retention.
 - ▣ Proponents: children should not progress to the next grade if they are not adequately prepared, it's bad for them and their classmates
 - ▣ Opponents: not much evidence that grade retention is effective academically (particularly in the long-term), may cause problems emotionally/behaviorally, and it is expensive
- But evidence for effectiveness of policy is thin.

Consequences of Retention Are Many

HELD BACK

1,161 students are two or more years older than the typical age for their grade level in the Recovery School District.

NUMBER OF STUDENTS AT LEAST TWO YEARS OVERAGE BY GRADE LEVEL:



Source: Recovery School District; includes special needs students; does not include charter schools

TYPICAL AGE OF STUDENTS FOR EACH GRADE LEVEL AT THE END OF THE SCHOOL YEAR:

Grade	Age	Grade	Age	Grade	Age	Grade	Age
Pre K	5	3	9	7	13	11	17
K	6	4	10	8	14	12	18
1	7	5	11	9	15		
2	8	6	12	10	16		

THE TIMES-PICAYUNE

In one New Orleans district:

- 10% of students are 2 or more years above average age for grade.
- 20% of 8th graders are 2+ years above average age.

“Older Students Pose Unique Challenges for Teachers, Families”
New Orleans *Times-Picayune*
May 17, 2010

Research on Grade Retention

- Much research suggests negative effects of retention
 - ▣ Increased odds of dropping out of high school
 - ▣ Worse emotional and behavioral outcomes
 - ▣ Greater risk of problem behavior
 - ▣ Lower future earnings
- Yet little of this research is adequate to examine causal effects.

Research on Grade Retention



- Cannot randomly assign children to repeat a grade.
- Useful question for propensity score matching techniques.

Research Questions

- What are the effects of retention in first grade on test scores later in elementary school?
- Are these effects different if retention is combined with tutoring?
- Are these effects different for different subpopulations?

Data



- Early Childhood Longitudinal Study, Kindergarten cohort (ECLS-K)
 - ▣ Collected by the U.S. National Center for Education Statistics (NCES)
 - ▣ Nationally representative panel study of approximately 22,000 kindergarteners in 1998-9
 - Also interviews of families, teachers, and administrators
 - ▣ Samples schools and children within schools

Outcomes

- Cognitive assessments of Math and Reading
 - ▣ Designed to measure a child's knowledge at specific time points.
 - ▣ Derived from state and national standards
 - ▣ These tests were not administered by the district, nor used as a school-based assessment.
 - Therefore, we worry less about the effect of teachers “teaching to the test” in these data.
 - Fewer exemptions than has been seen in other standardized testing scenarios.

Background variables

- About 250 predictors (measured in Kindergarten and 1st grade) that fall into several broad categories:
 - ▣ Characteristics of child: demographic characteristics, behavioral measures, disability classification, child care history, etc.
 - ▣ Characteristics of parents and family: race and ethnicity, income/poverty measures, education levels, family composition, employment, etc.
 - ▣ Parents expectations and participation in child's schooling (both at home and school) and assessments of child's ability and effort levels

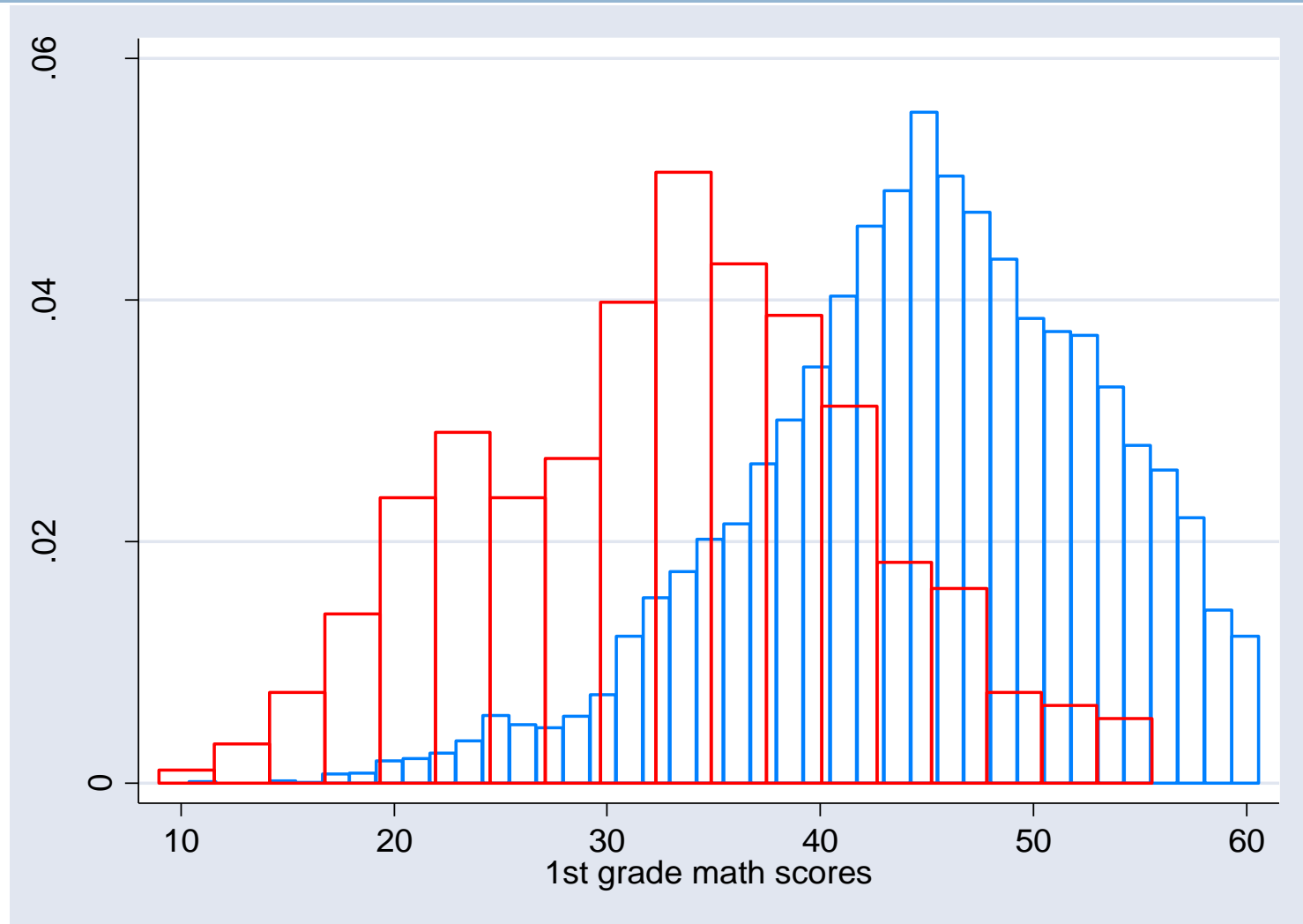
Methods

- Examine students' scores on assessments of math and reading.
- Examine these at two time points:
 - ▣ Expected 3rd grade year
 - ▣ Expected 5th grade year
- Test scores standardized
 - (Mean=100, SD=15)
- Use linear regression and propensity-score weighting controlling for MANY background variables
- (note this leads to same-age comparisons, rather than same-grade comparisons)

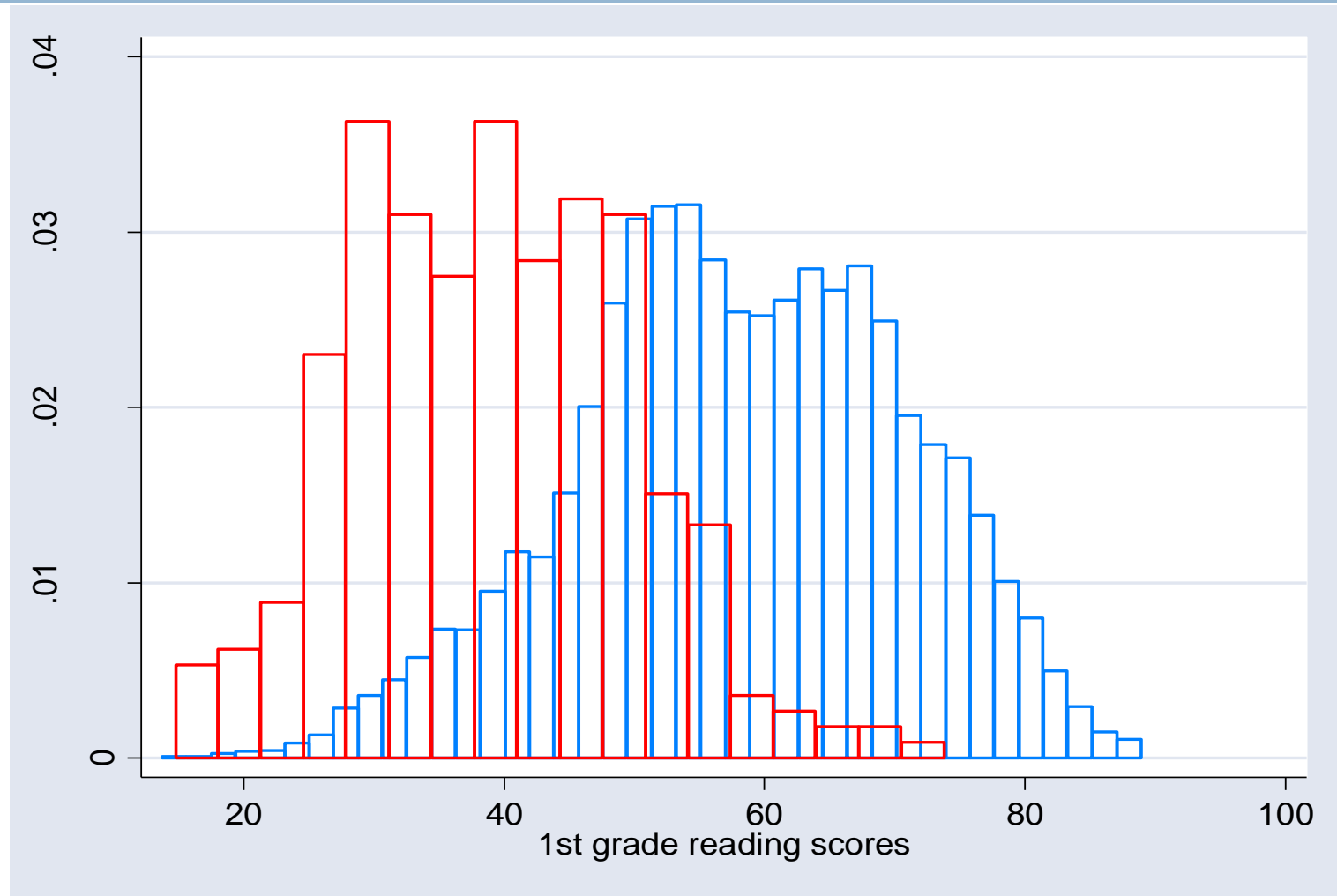
Requirements for Propensity Score Techniques

- In order for propensity score techniques to operate properly, we need:
- Large number of cases (particularly in the non-treated group).
- Large number of covariates used to estimate propensity score.
- Reasonable degree of “overlap”
 - ▣ If no overlap, then the two groups may be just too different to justify comparison

To what extent does grade retention appear to be based on ability? (the “overlap” issue)



To what extent does grade retention appear to be based on ability? (the “overlap” issue)



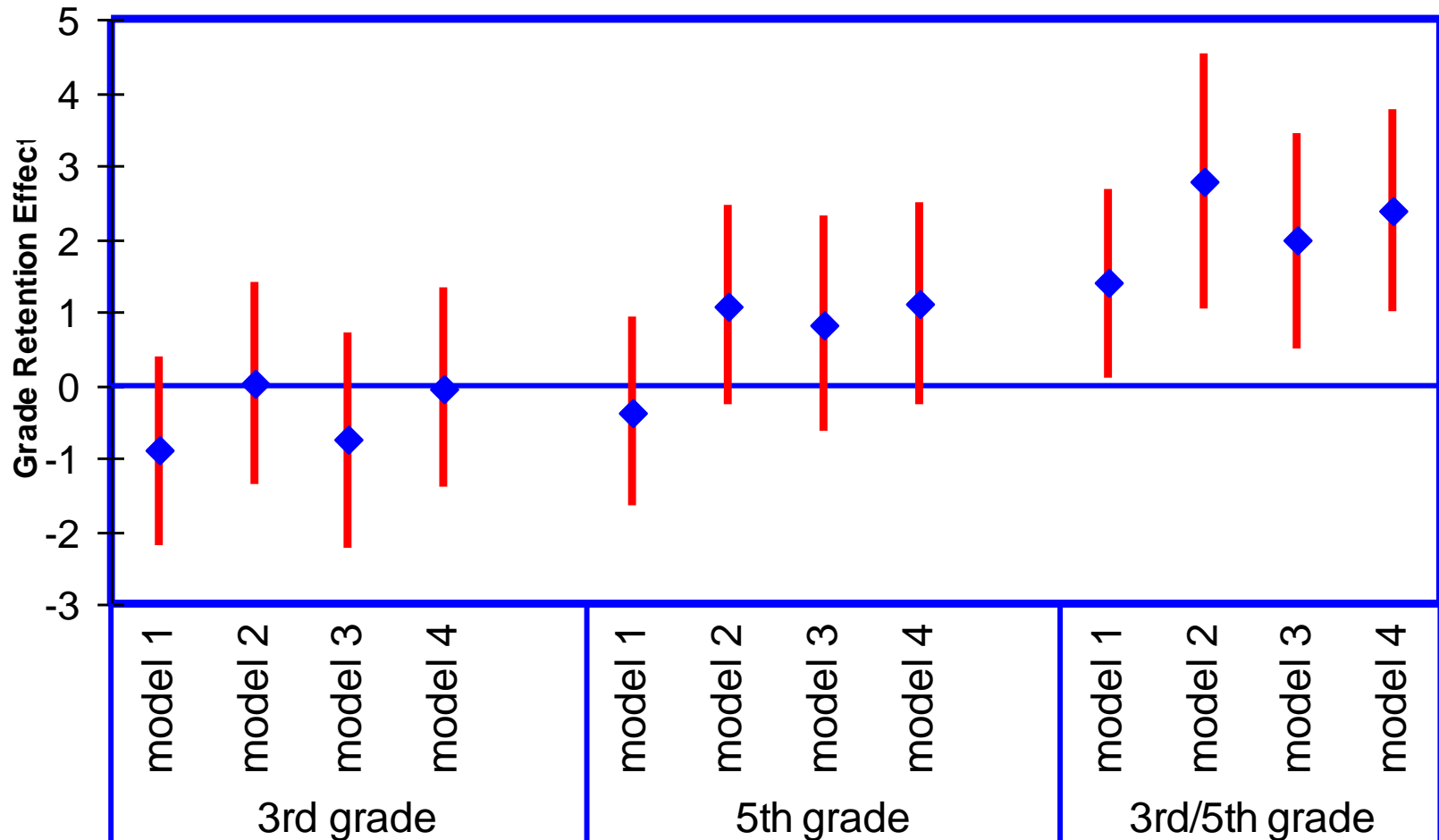
Outcomes: Math and Reading Scores

- Examine students' math and reading scores measured at two time points:
 - Spring 2002 (end of expected 3rd grade)
 - Spring 2004 (end of expected 5th grade)
- Tests standardized with mean=100, SD=15.
- Comparing students of same age, not grade
- We also compare scores for kids retained in first grade.

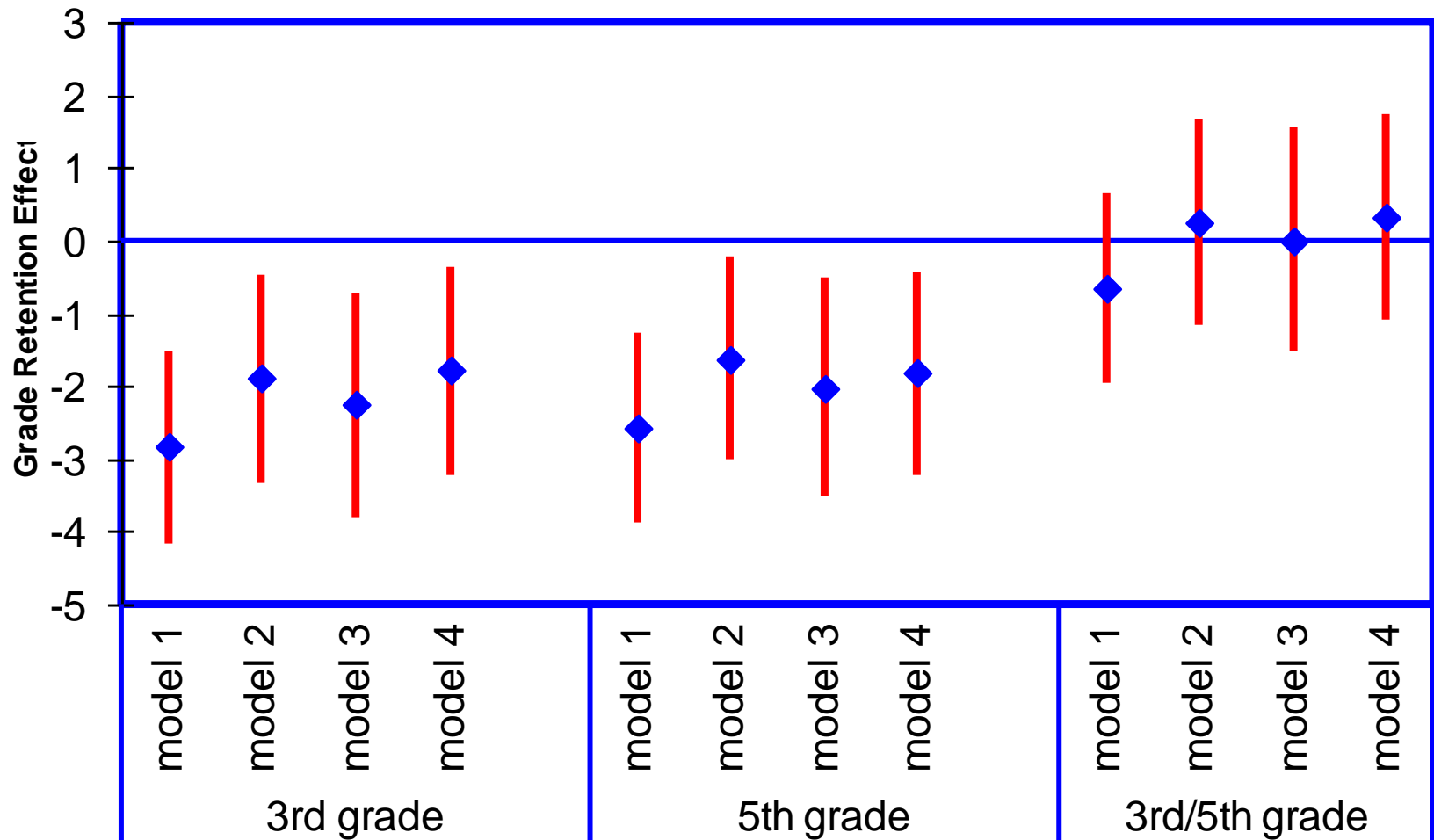
Effects of Retention on Test Scores: Primary Models

- Four models for each outcome:
 - ▣ Model 1 – characteristics of child, parents, family, plus parental involvement (1st grade) and test scores (K and 1st)
 - ▣ Model 2 – Model 1 plus teacher evaluation, program participation, and teacher and school variables.
 - ▣ Model 3 – Model 2 w/ school fixed effects (school variables omitted).
 - ▣ Model 4 – Model 2 w/ county fixed effects

Effects of 1st Grade Retention on Math Scores



Effects of 1st Grade Retention on Reading Scores



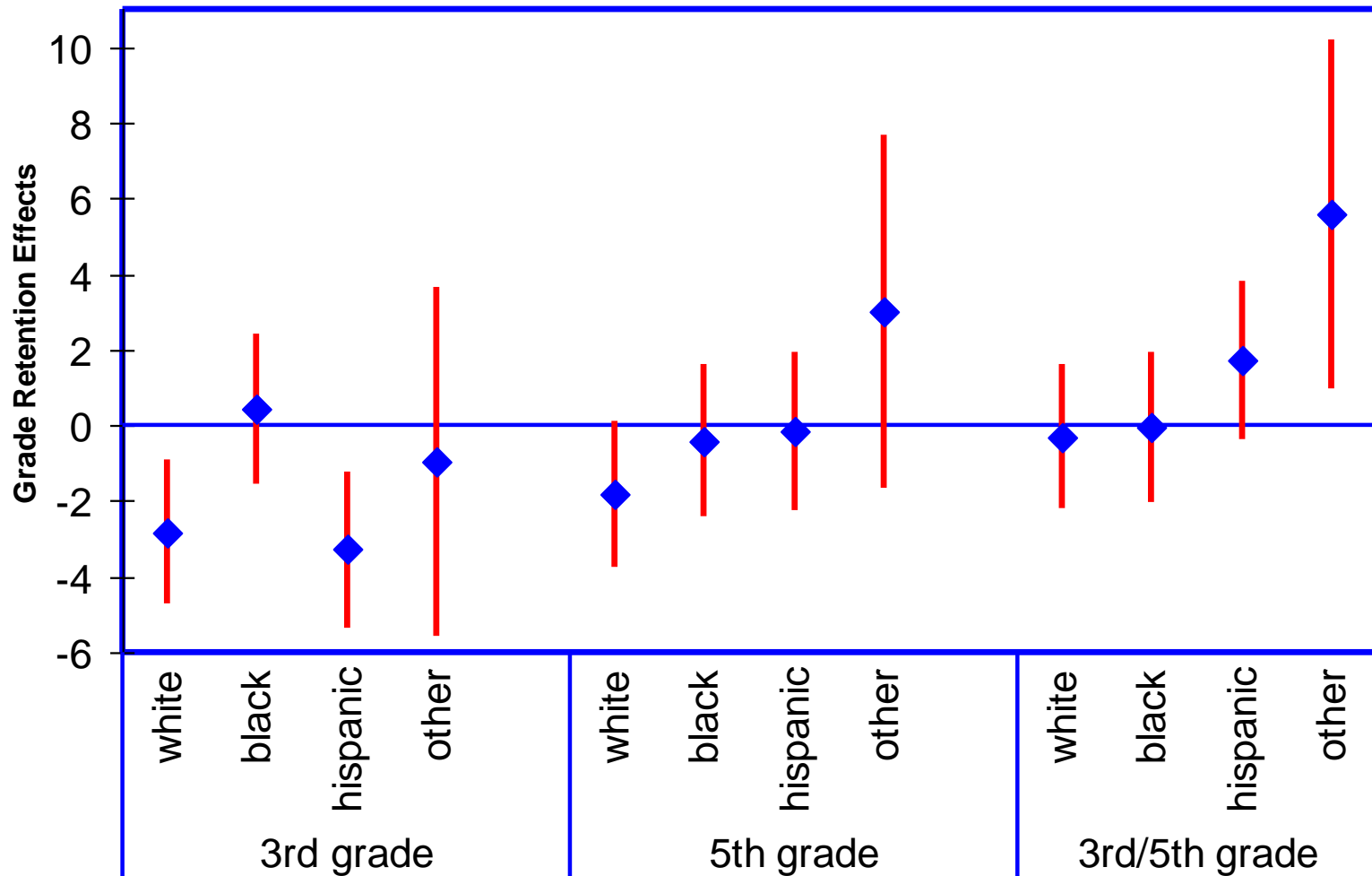
Initial Results: Mixed

- Retention seems to have benefits in math performance for the 3rd/5th group.
- But has negative effects on reading scores.
 - Negative effects that show up in 3rd and 5th grade.
- Overall, not consistent effects.
- Not the clean or clear story we'd hoped for.
 - Perhaps there's a clearer story in examining interactions – different effects for different groups of students.

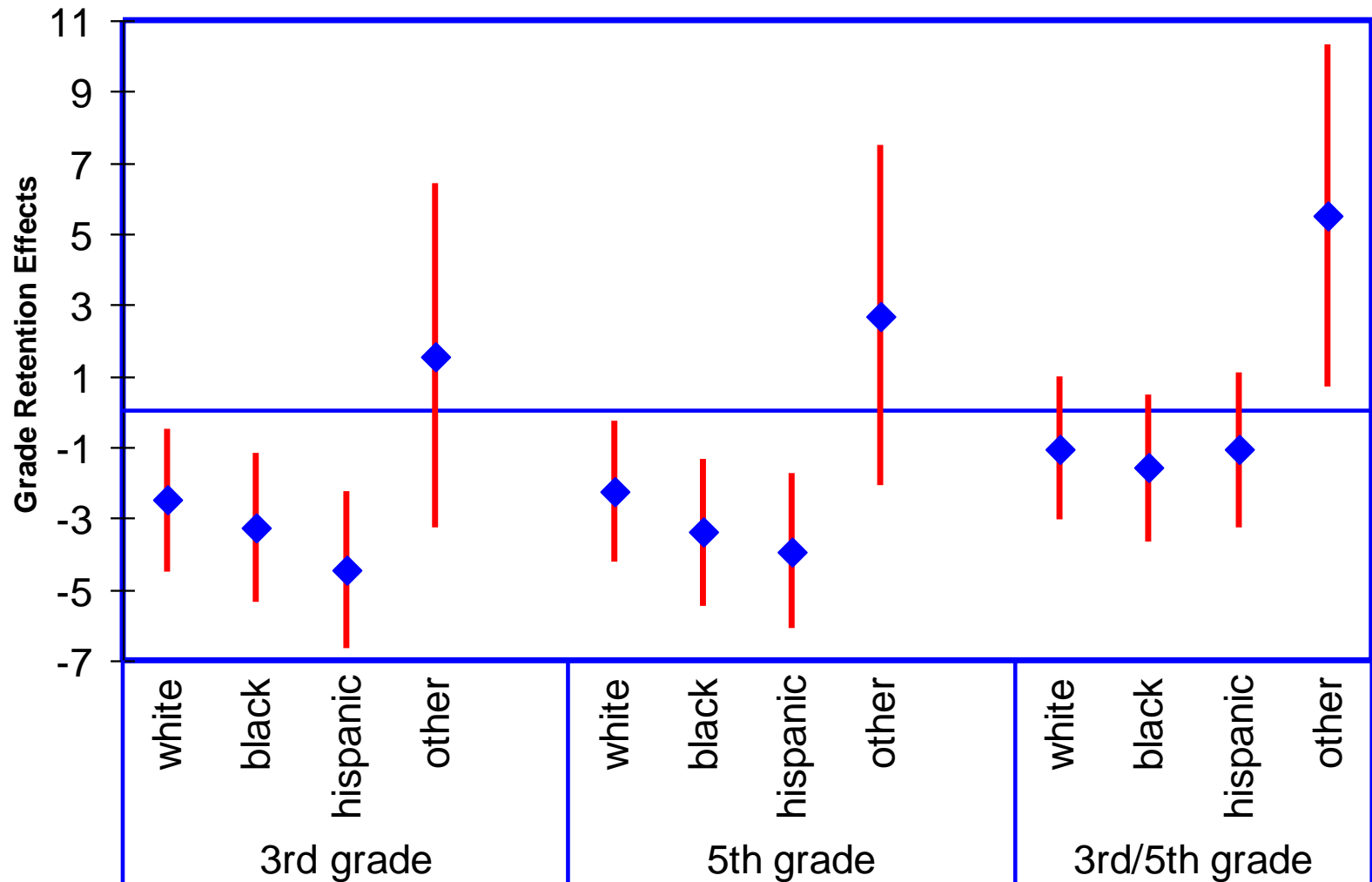
Effects of Retention on Test Scores: Models With Interactions for Race

- In the next stage, we examine whether the effects of retention on school performance are particularly consequential for any specific racial ethnic group.
- Include interaction terms by race/ethnicity to test for this.

Effects of 1st Grade Retention on Math Scores, by Race



Effects of 1st Grade Retention on Reading Scores, by Race



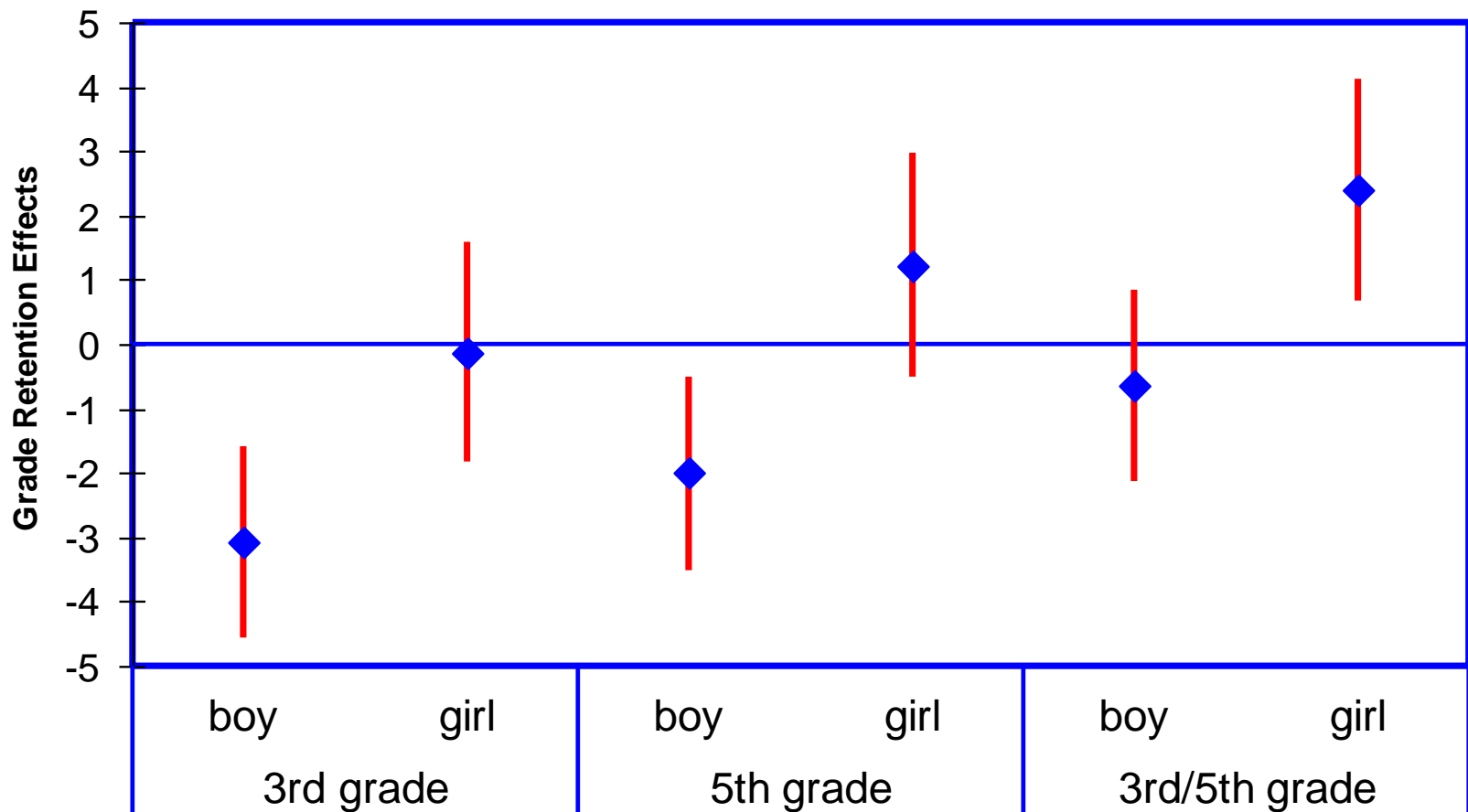
Results of Race Models

- Retention seems to have negative effect on math performance for White and Latino students in 3rd grade.
 - But effects non-significant in 5th grade for these groups
- Negative effects on reading scores in 3rd and 5th grade for all groups except “Other.”
- Did not clarify – if anything, further muddled.

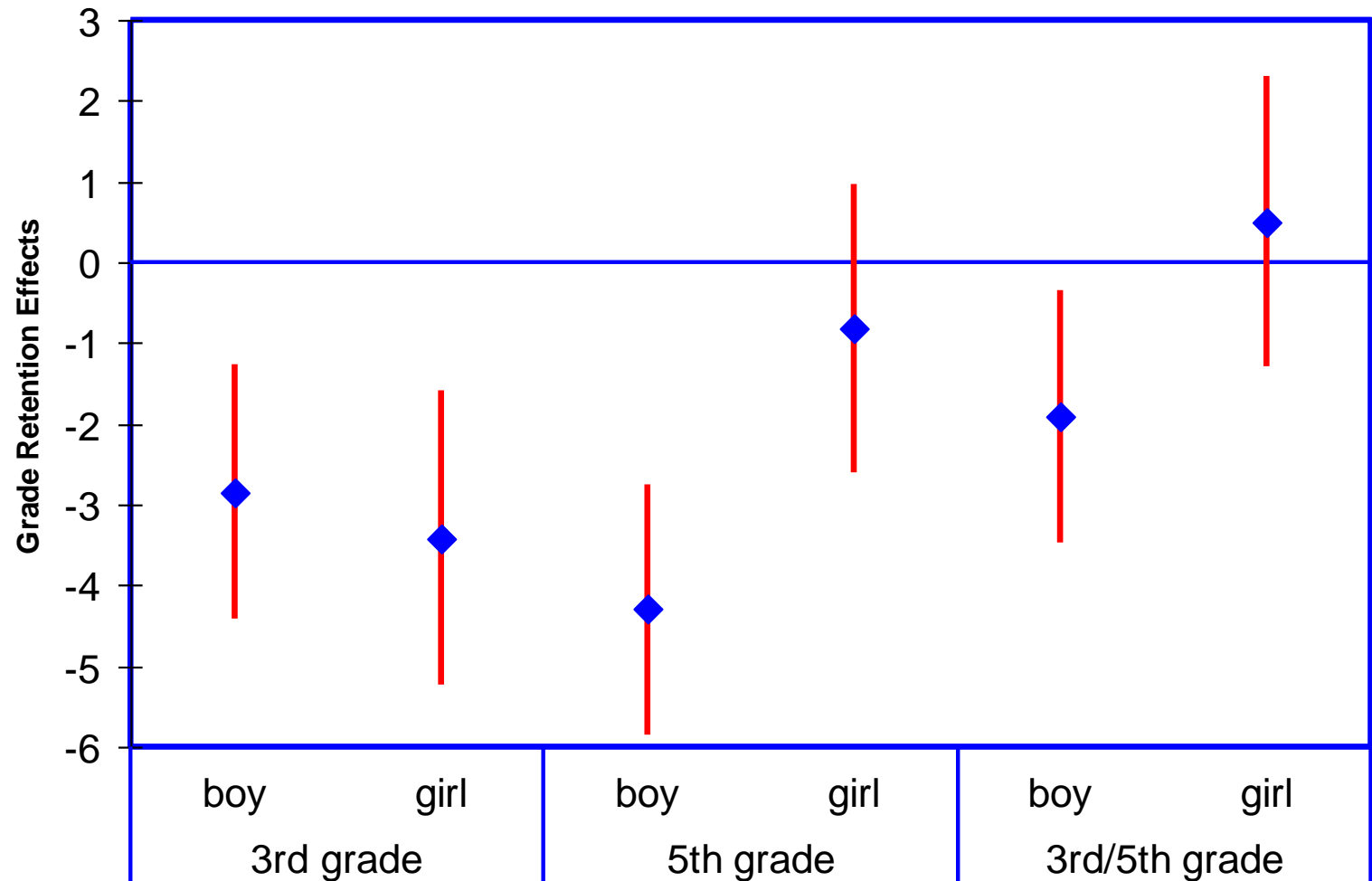
Effects of Retention on Test Scores: Models With Interactions for Gender

- In the next stage, we examine whether the effects of retention on school performance are particularly consequential for males or females.
- Include interaction terms by gender to test

Effects of 1st Grade Retention on Math Scores, by Gender



Effects of 1st Grade Retention on Reading Scores, by Gender



Results of Gender Models

- Marked differences in retention's effects by gender
- For boys, retention has negative impact on reading and math performance in both 3rd and 5th grade.
- For girls, non significant effect on math (and benefit in combined score). And negative effect on reading – but only in 3rd grade.

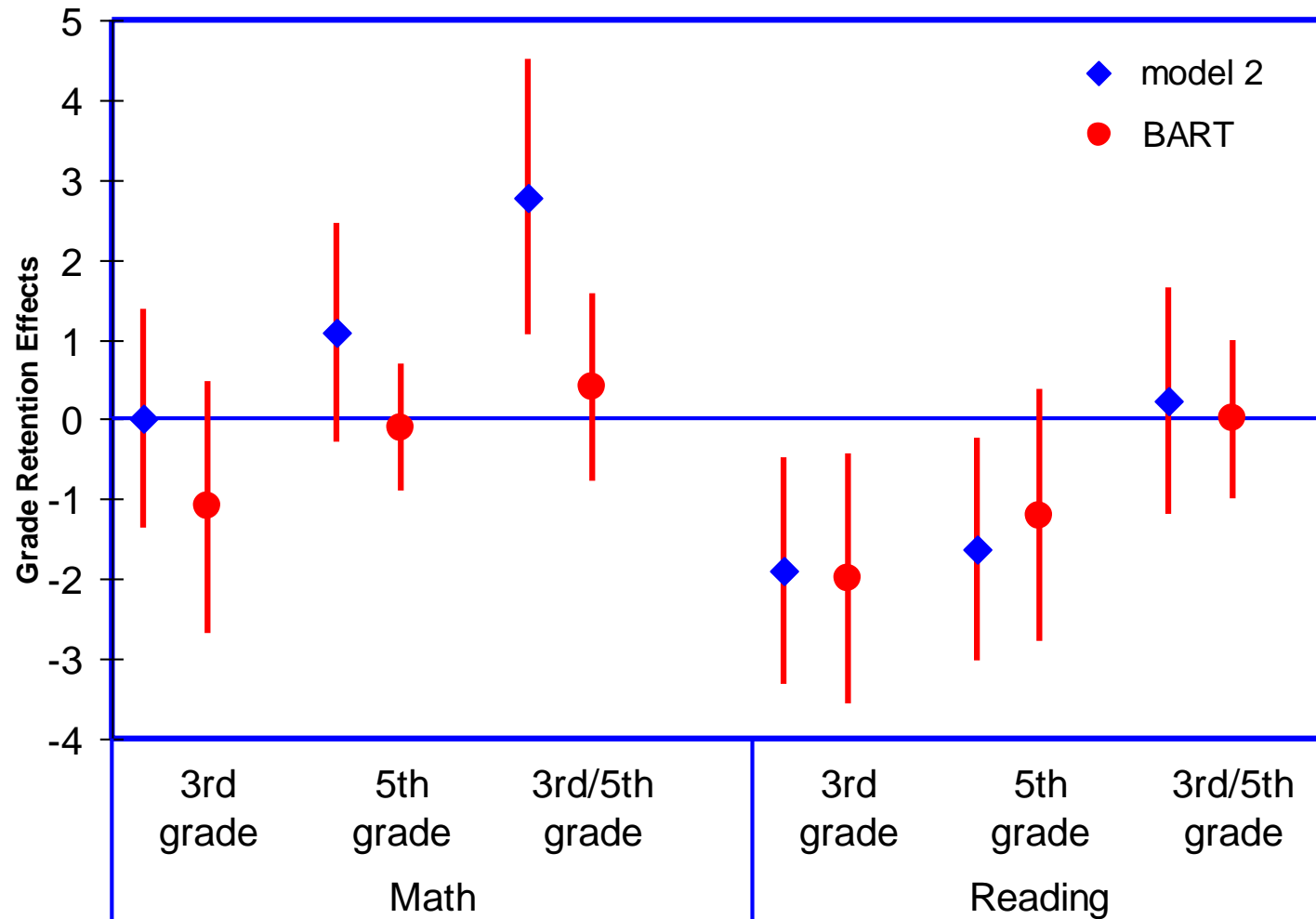
Synthesis: A Mixed Bag of Findings

- Effects of grade retention on students' later outcomes are quite heterogeneous.
- Magnitude and direction of effect depend upon:
 - ▣ Domain of achievement examined
 - ▣ Student race/ethnicity
 - ▣ Student gender
- Before drawing conclusions, subject these results to a series of tests to examine robustness

Sensitivity analyses

- Ran models using propensity score weighting and the answers were consistent
- Also ran models using sophisticated Bayesian non-parametric models and the results again were substantively quite similar
- Ran models that excluded controls (those not retained) most dissimilar from treated (those retained) as modeled using a classification tree -
- point estimates of main effects generally increased a bit in magnitude but no differences in general conclusions

Estimates of Model Robustness



Conclusions

- Effects of grade retention on students' later outcomes are quite heterogeneous.
- Magnitude and direction of effect depend upon:
 - ▣ Domain of achievement examined
 - ▣ Student race/ethnicity
 - ▣ Student gender
- Generally speaking, results suggest that retention has more negative/less beneficial impact on reading than on math.

Conclusions

- So what does this mean for policymakers?
 - ▣ Very equivocal evidence about the benefits and consequences of retention.
- Retention does seem to have generally negative impact on reading scores, esp. for males.
 - ▣ Yet, at the same time, some students benefit.
- Equivocal is the word...

Conclusions

- Important to note that we focus only on academic performance.
 - ▣ Number of studies show impact of retention on other aspects of students' well-being.
 - ▣ Should examine social and emotional outcomes as well.

Conclusions

- Important to note that we focus only on academic performance.
 - ▣ Number of studies show impact of retention on other aspects of students' well-being.
 - ▣ Should examine social and emotional outcomes as well.

- Important also to note that this is only in elementary school.
 - ▣ Most districts allow students to be retained only once before high school. But then...

Future work on this topic with these data

- Explore retention/tutoring interactions more thoroughly
- More coherent approach to missing data
- Examine 3rd grade retention (reconcile possible ECLSK data issues with)
- Group 1st, 2nd and 3rd grade retention to bolster sample size

With thanks...



Chris Weiss
cw2036@columbia.edu