A National Study Comparing Charter and Traditional Public Schools Using Propensity Score Analysis

Dissertation Defense

Jason M. Bryer

School of Education
Department of Educational & Counseling Psychology
Division of Educational Psychology & Methodology
University at Albany

September 8, 2014

It takes a village...
to complete a dissertation!

Outline

Purpose

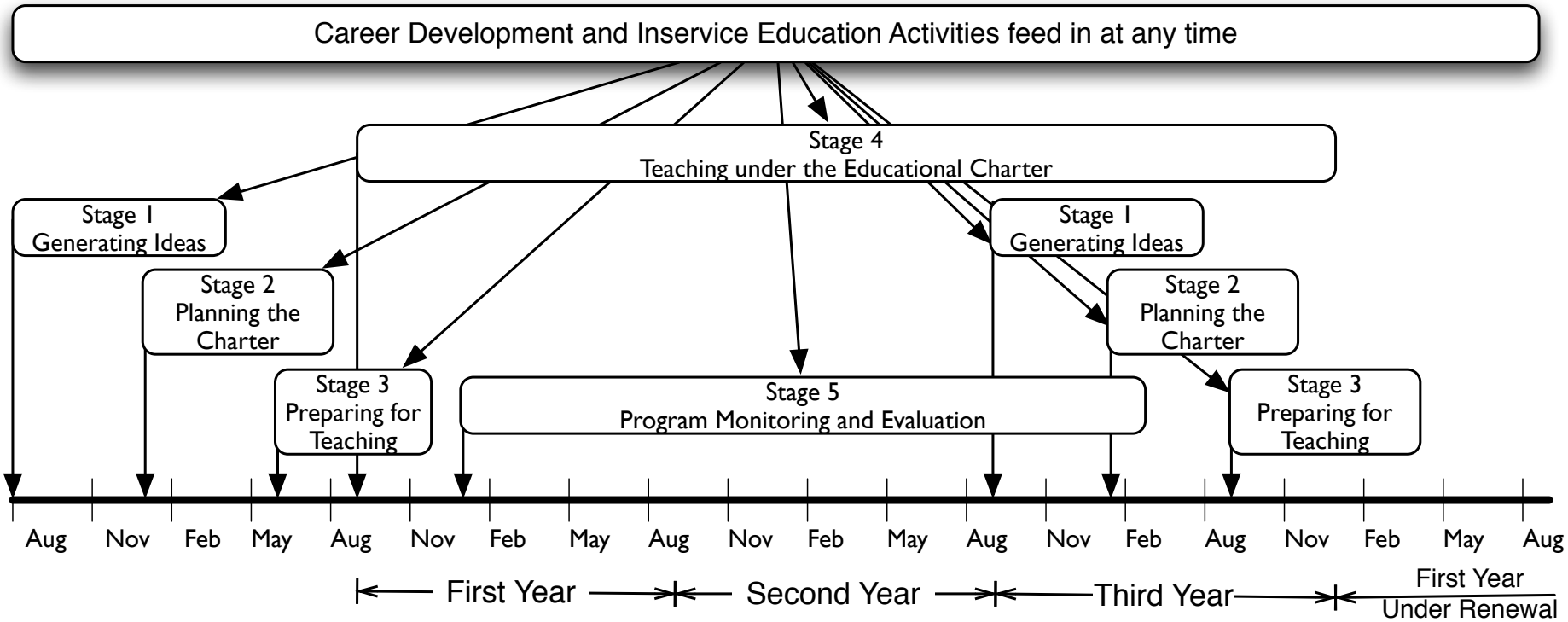
There are two primary purposes of this study:

1. Explore the differences in performance between charter and traditional public schools. Additionally, examine the differences between state performances on NAEP in terms of the quality of state charter laws as measured by the National Alliance for Public Charter Schools (2010).

2. Develop new methods for analyzing and visualizing propensity score analysis for multilevel, or clustered, data.

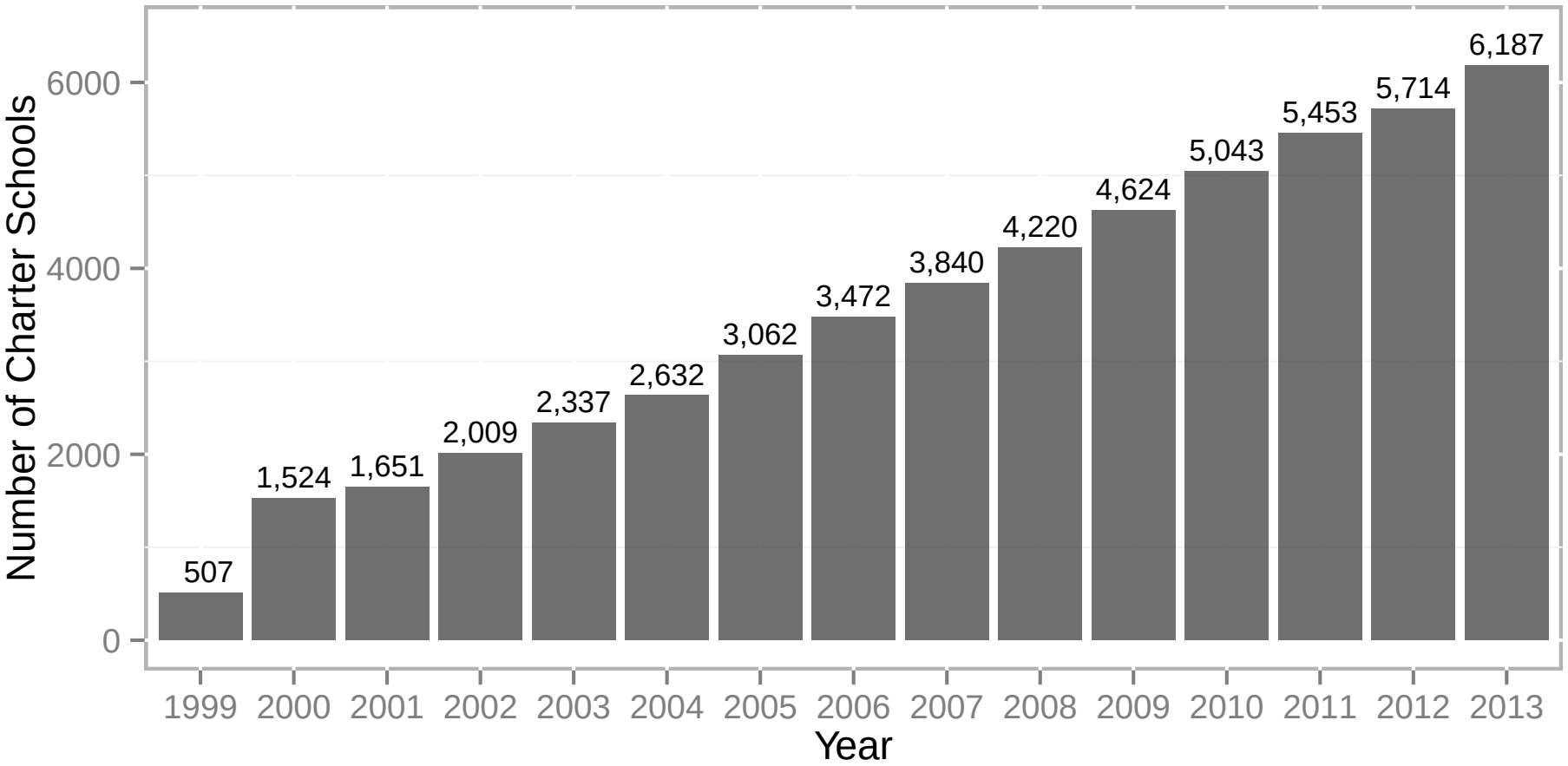
What are Charter Schools?

The first charter school opened in Minnesota in 1991 In principle, charter schools have opted out of bureaucratic rules and union contracts in order to gain academic autonomy. The standard argument has become that this autonomy will lead to higher student test scores and better academic environments (Wells, 2002). The idea is that, under the charter framework, teachers, administrators, students and the community that comprise the charter school would be free to innovate



Adapted from Budde (1988).

Growth in Number of Charter Schools



National Studies Examining Charter and Traditional Public Schools

The number of studies that consider charter school effectiveness from a national perspective are limited.

1. Braun, Jenkins, and Grigg (2006) examined charter and traditional public schools using hierarchal linear modeling (HLM) with NAEP 2005. They found that there was often no difference, and in some instances charter school students performed worse.

2. *The CREDO Study* - Center for Research on Education Outcomes (2009) conducted a study of more than 1.7 million records from 2400 charter within 16 states. The methodology involves creating a Virtual Control Record (VCR) for each charter school student which is used to find matching student from an eligible traditional public school. Overall results show that charter school students performed, on average, 0.01 and 0.03 standard deviations below public school students for reading and math, respectively.

Issues with Charter Schools Research

Betts and Hill (2006) outline three major issues with charter school research:

1. The issue of selection bias
2. The variation in types or kinds of charter schools.
3. The nature of student achievement. Research has shown there are numerous factors that contribute to student success including, but not limited to, social economic status, parental education, motivation, etc. The ability to decipher how school choice contributes to student learning in the context of all the other factors cannot help but be difficult.

A fourth issue not cited by Betts and Hill (2006) is the role of charter school laws.

Research Questions

1. Given appropriate adjustments based on available student data, is there a discernible difference between charter and traditional public schools with regard to math and reading scores on the NAEP scores at grades 4 and 8?

2. If so, what is the nature and magnitude of this difference for the two outcomes, reading and mathematics scores?

3. What is the relationship, if any, of different charter school laws on charter school student performance in math and reading at grades 4 and 8?

Propensity Scores

Propensity scores are the "conditional probability of being in the treatment" (Rosenbaum & Rubin, 1983), defined as:

$$\pi(X_i) \equiv Pr(T_i = 1|X_i) \tag{1}$$

Let X is a matrix of observed covariates. The balancing property under exogeneity states that,

$$T_i \perp\!\!\!\perp X_i \mid \pi(X_i) \tag{2}$$

For randomized experiments, the strong ignobility assumption states,

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i|X_i \tag{3}$$

for all X_i. That is, treatment is independent of all covariates, observed or otherwise.
However, the strong ignobility assumption can be restated with the propensity score as,

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i \mid \pi(X_i) \tag{4}$$

Propensity Score Analysis in Two Phases

PSA is typically conducted in two phases:

Phase 1 Estimate propensity scores (model the treatment indicator).

Phase 2 Estimate effects (model the outcome of interest).

Propensity Score Methods for this Study

1. *Propensity score analysis using stratification.*
   1. Full logistic regression.
   2. Logistic regression with step AIC.
   3. Conditional inference trees (Hothorn, Hornik, & Zeileis, 2006).

2. *Propensity score matching.*
   1. One-to-one
   2. One-to-five
   3. One-to-ten.

   A dependent sample analysis will be performed on the resulting matched pairs (Austin, 2011).

3. *Multilevel propensity score analysis.*
   1. Full logistic regression.
   2. Logistic regression with step AIC.
   3. Conditional inference trees.

The `multilevelPSA` Package

- The `multilevelPSA` R package was written to perform the analysis and generate the figures used for this dissertation.
- Currently available from the Comprehensive R Archive Network (CRAN) at `http://cran.r-project.org/web/packages/multilevelPSA/index.html`
- Contains 16 functions, 4 demos, and a vignette. Examples utilize the Programme of International Student Assessment (PISA) examining the differences between private and public schools internationally.

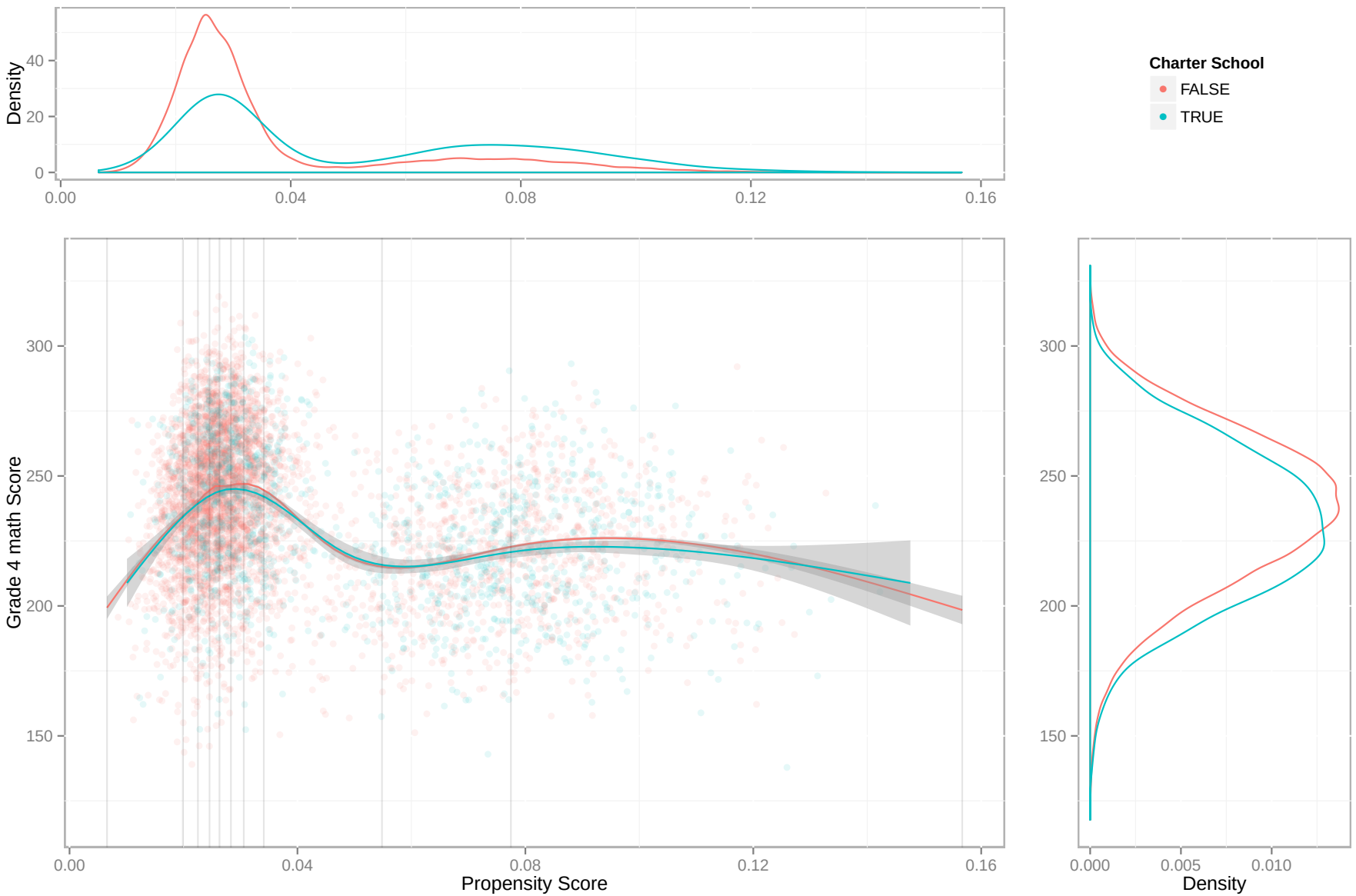
National Assessment of Educational Progress (NAEP)

- Congressionally mandated.
- Started in 1971.
- Provides national measures of student achievement in many subjects including mathematics, reading, science, writing, history, civics, and the arts.
- Used the 2009 implementation of grades 4 and 8 in math and reading.
- The 2009 assessment included over 6,000 public schools and over 200 charter schools comprising of over 145,000 and 3,000 students, respectively.
- The samples is representative of the entire United States, although only states that have charter laws were used ($n \approx 40$, varies slightly for grade and subject.).

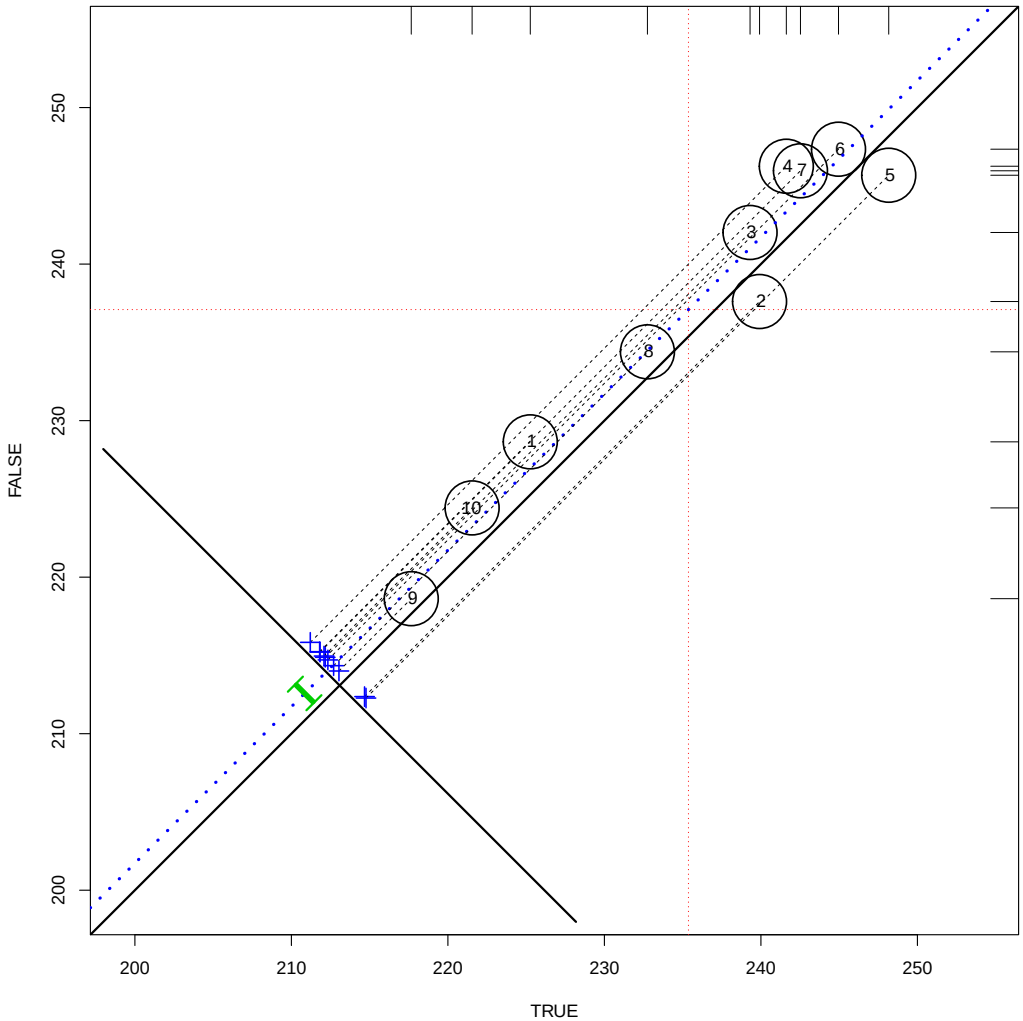
Covariates Available from NAEP

1. Are you Hispanic or Latino? [No, I am not Hispanic or Latino; Yes, I am Mexican, Mexican American, or Chicano; Yes, I am Puerto Rican or Puerto Rican American; Yes, I am Cuban or Cuban American; Yes, I am from some other Hispanic or Latino background]
2. Which of the following best describes you? [White; Black or African American; Asian; American Indian or Alaska Native; Native Hawaiian or other Pacific Islander]
3. Does your family get a newspaper at least four times a week?
4. Does your family get any magazines regularly?
5. About how many books are there in your home?
6. Is there a computer at home that you use?
7. Is there an encyclopedia in your home? It could be a set of books, or it could be on the computer.
8. About how many pages a day do you have to read in school and for homework?
9. How often do you talk about thinks you have studied in school with someone in your family?
10. How many days were you absent from school in the last month?
11. How far in school did your mother go? [Grade 8 Only]
12. How far in school did your father go? [Grade 8 Only]
13. How often do people in your home talk to each other in language other than English?

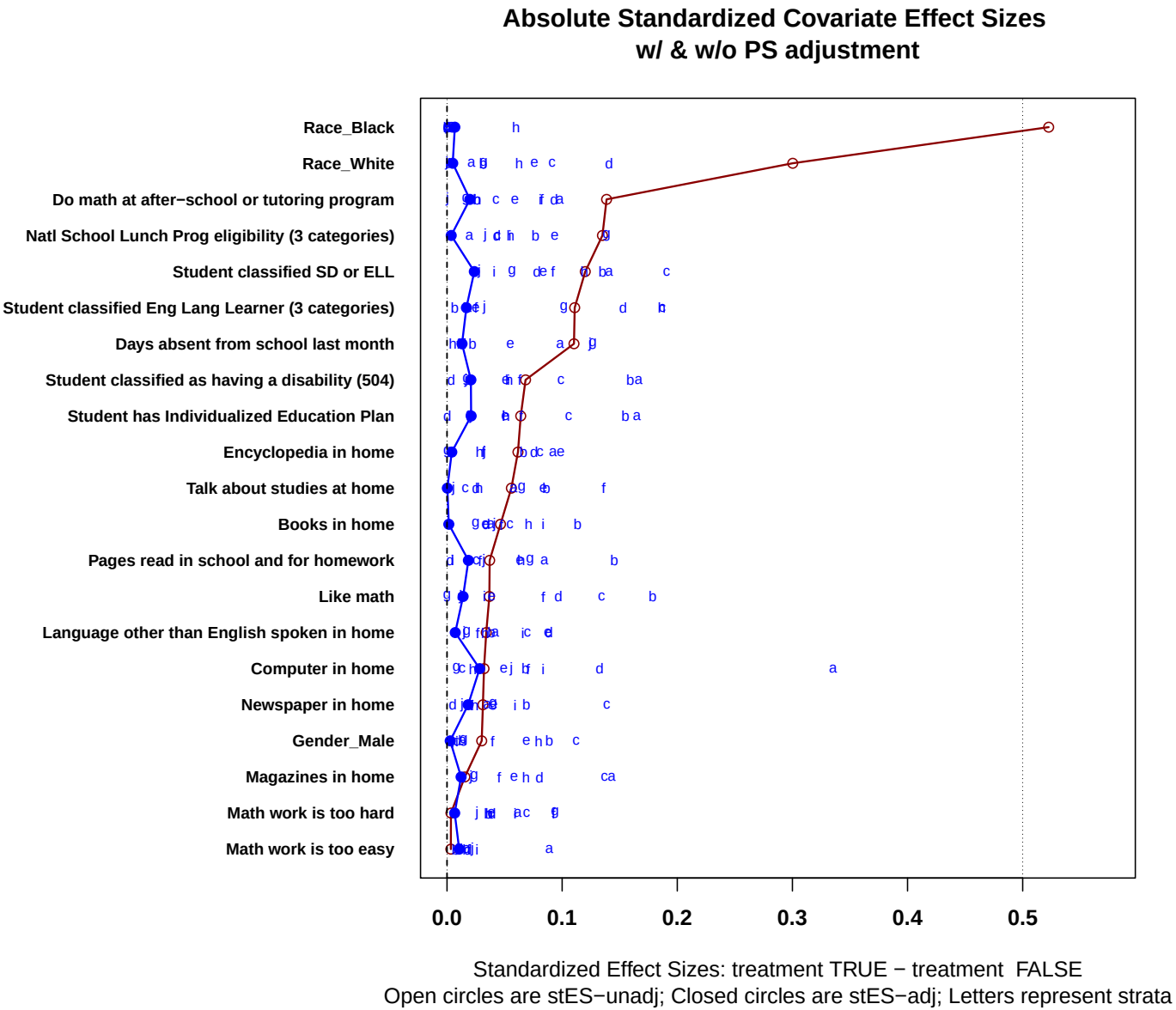
Loess Plot of Propensity Scores vs. Grade 4 Math



PSA Assessment Plot: Grade 4 Math (Stratification)



Balance Plot: Grade 4 Math



Propensity Score Matching

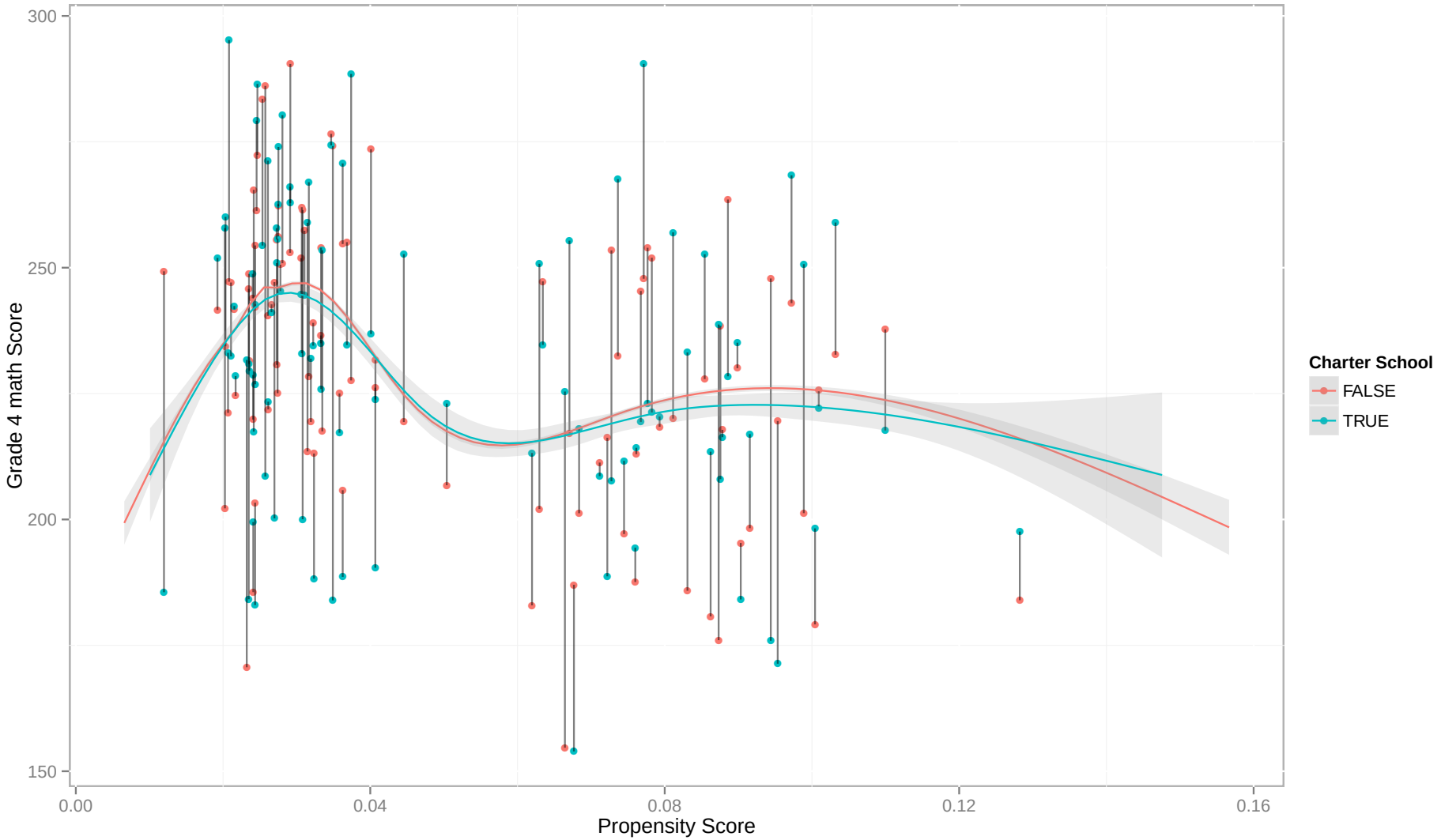
Partial exact matching using nearest neighbor was used using the `Matching` (Sekhon, 2011) package. For each charter school student:

- Traditional public school students in the same state with the same gender and ethnicity are located.
- Of those students, the student with the smallest difference in propensity scores is selected and paired with that charter school student.

Three match sets were estimated: one-to-one, one-to-five, and one-to-ten.

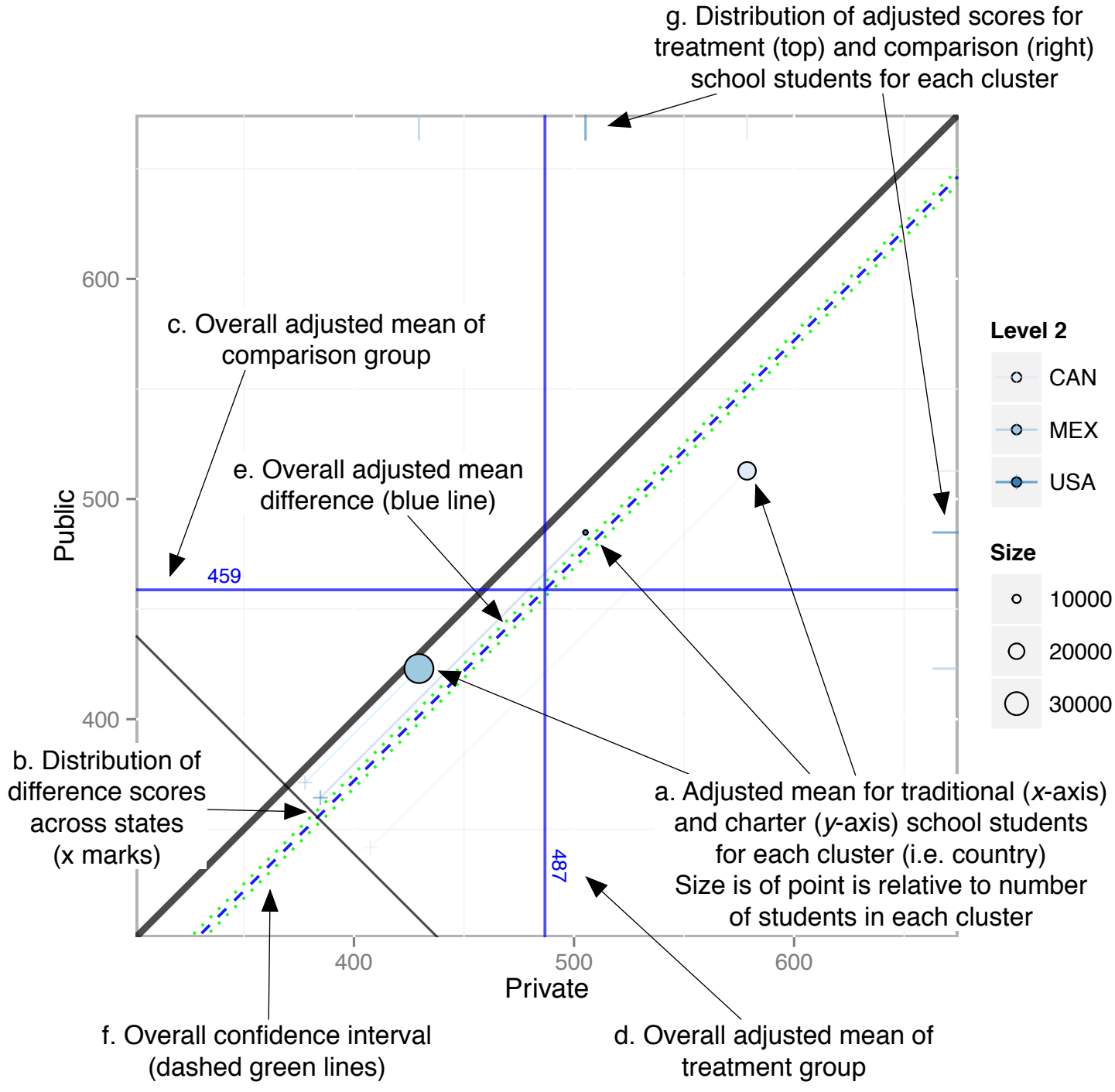
For each match set, a dependent sample t-test were used to estimate ATE (Austin, 2011).

The Relationship of Matched Pairs and Propensity Scores

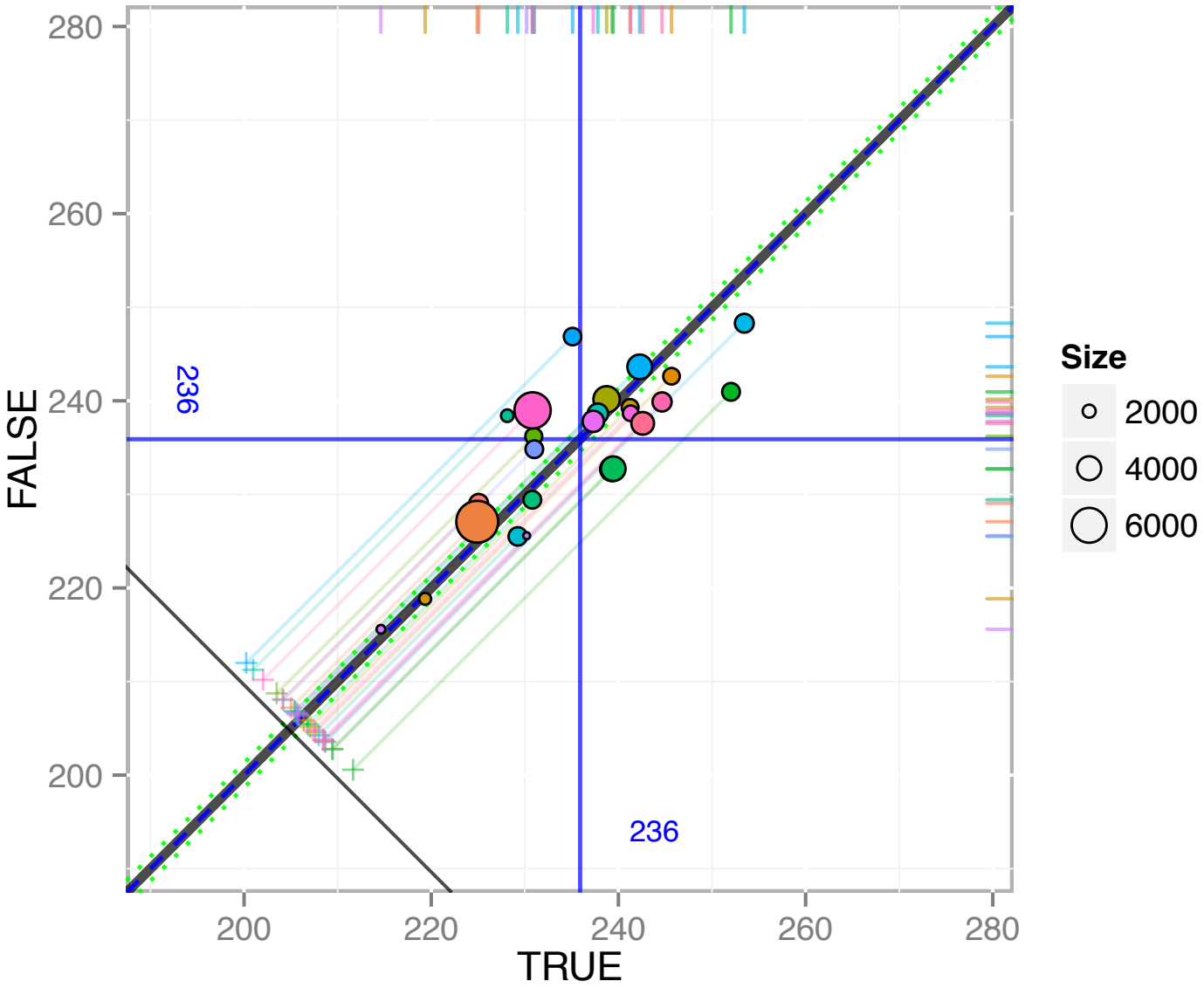


Propensity Score Matching Results

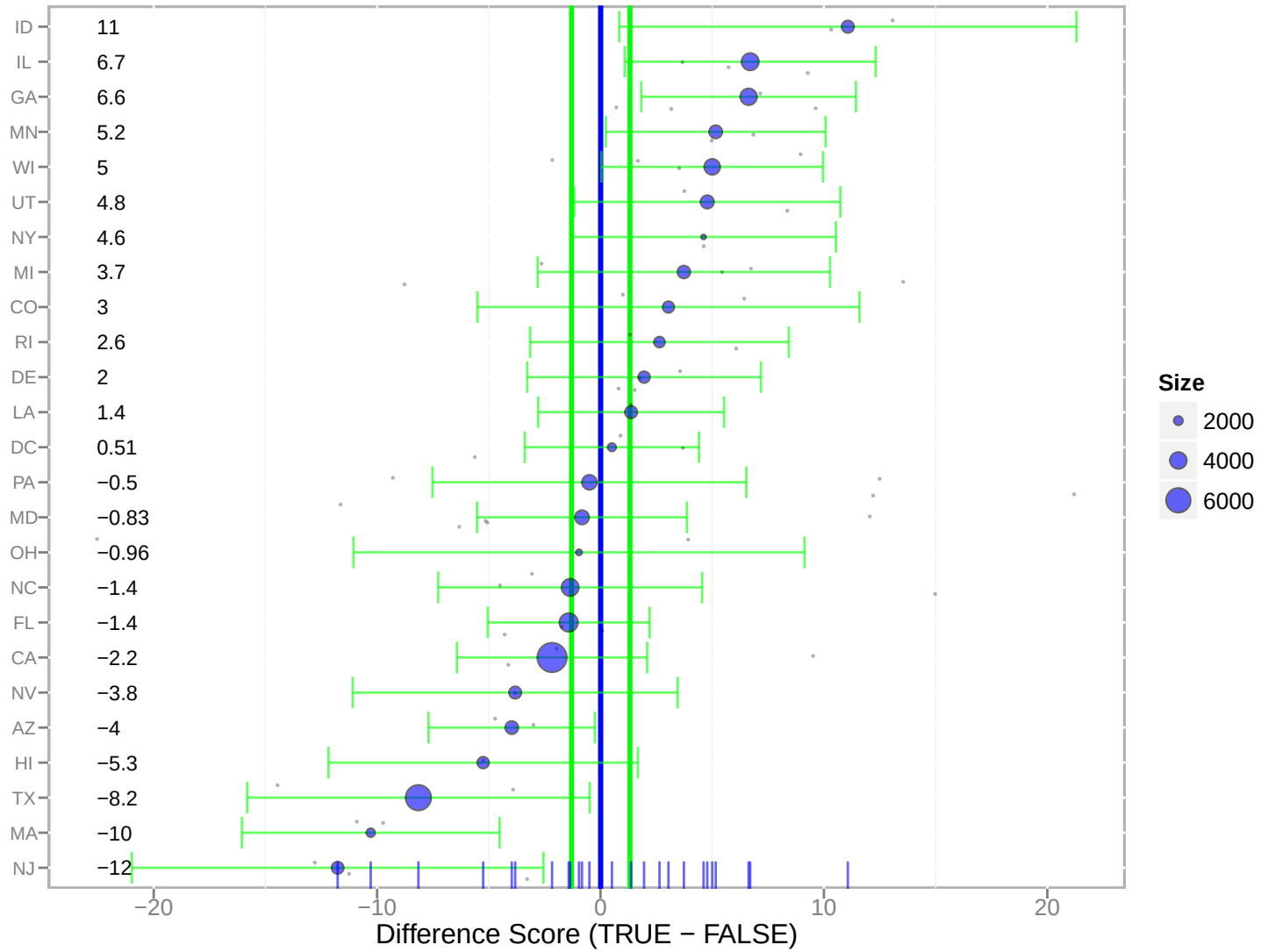
Method	Charter	Public	ATE	95% CI	
			Grade 4 Math		
Matching One-to-One	231.22	231.18	0.00	-1.12	1.20
Matching One-to-Five	232.67	232.14	0.02	-0.01	1.07
Matching One-to-Ten	234.02	233.33	0.02	0.29	1.09
			Grade 4 Reading		
Matching One-to-One	212.88	211.82	0.03	-0.30	2.41
Matching One-to-Five	214.51	213.54	0.03	0.33	1.60
Matching One-to-Ten	215.63	214.75	0.03	0.42	1.35
			Grade 8 Math		
Matching One-to-One	272.05	269.16	0.08	1.51	4.28
Matching One-to-Five	274.85	273.46	0.04	0.72	2.06
Matching One-to-Ten	276.21	274.08	0.06	1.63	2.62
			Grade 8 Reading		
Matching One-to-One	256.29	253.48	0.09	1.51	4.12
Matching One-to-Five	258.19	256.24	0.06	1.33	2.58
Matching One-to-Ten	258.93	256.82	0.06	1.65	2.57


g. Distribution of adjusted scores for treatment (top) and comparison (right) school students for each cluster
c. Overall adjusted mean of comparison group
e. Overall adjusted mean difference (blue line)
b. Distribution of difference scores across states (x marks)
a. Adjusted mean for traditional (x-axis) and charter (y-axis) school students for each cluster (i.e. country) Size is of point is relative to number of students in each cluster
f. Overall confidence interval (dashed green lines)
d. Overall adjusted mean of treatment group
Level 2
CAN
MEX
USA
Size
10000
20000
30000
Public
Private
600
500
400
459
487

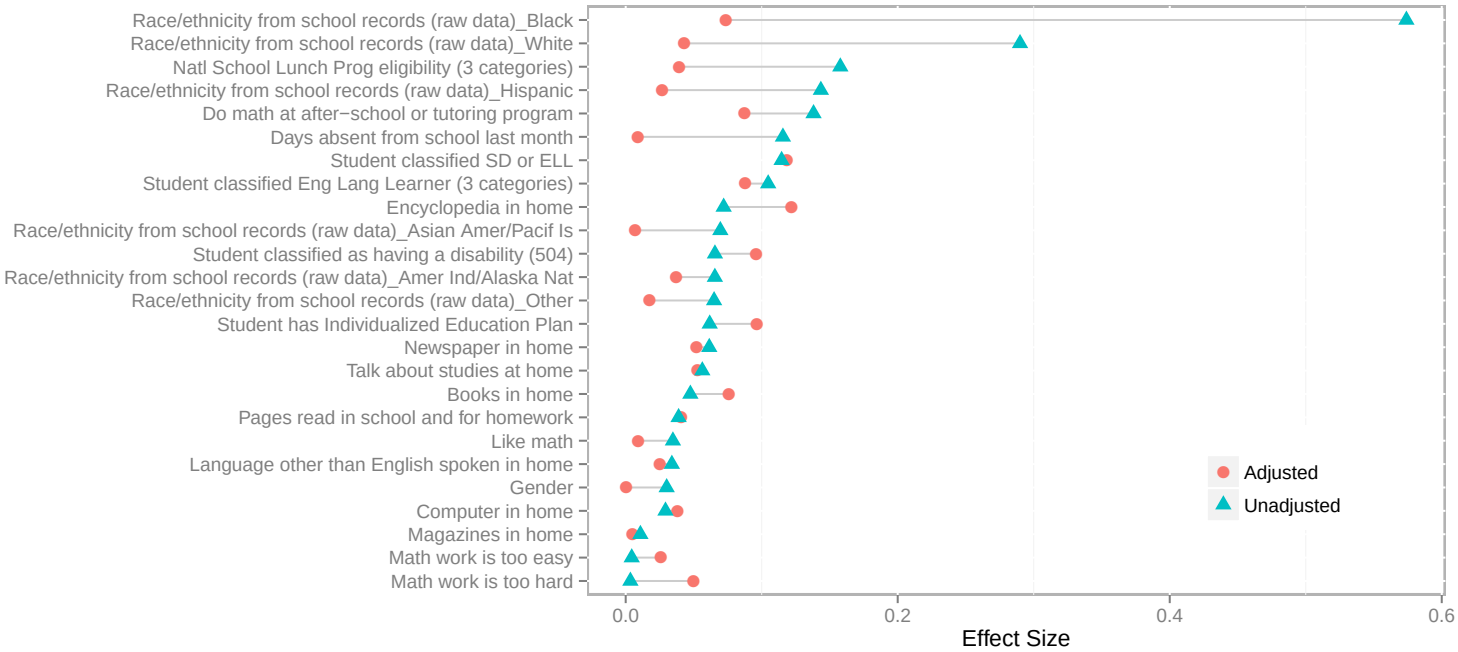
Multilevel PSA Assessment Plot: Grade 4 Math



Multilevel PSA Difference Plot: Grade 4 Math

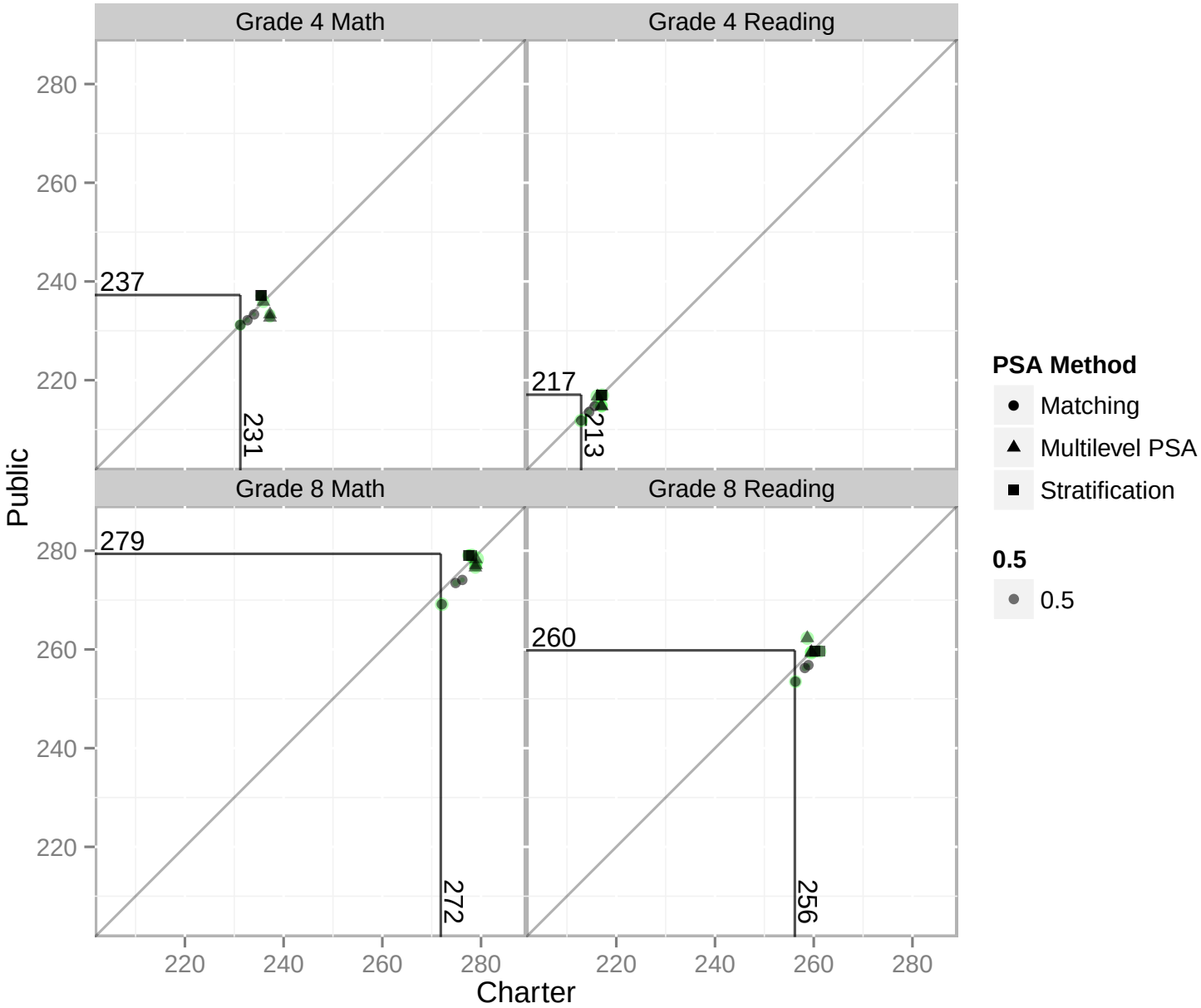


Multilevel PSA Balance Plot: Grade 4 Math

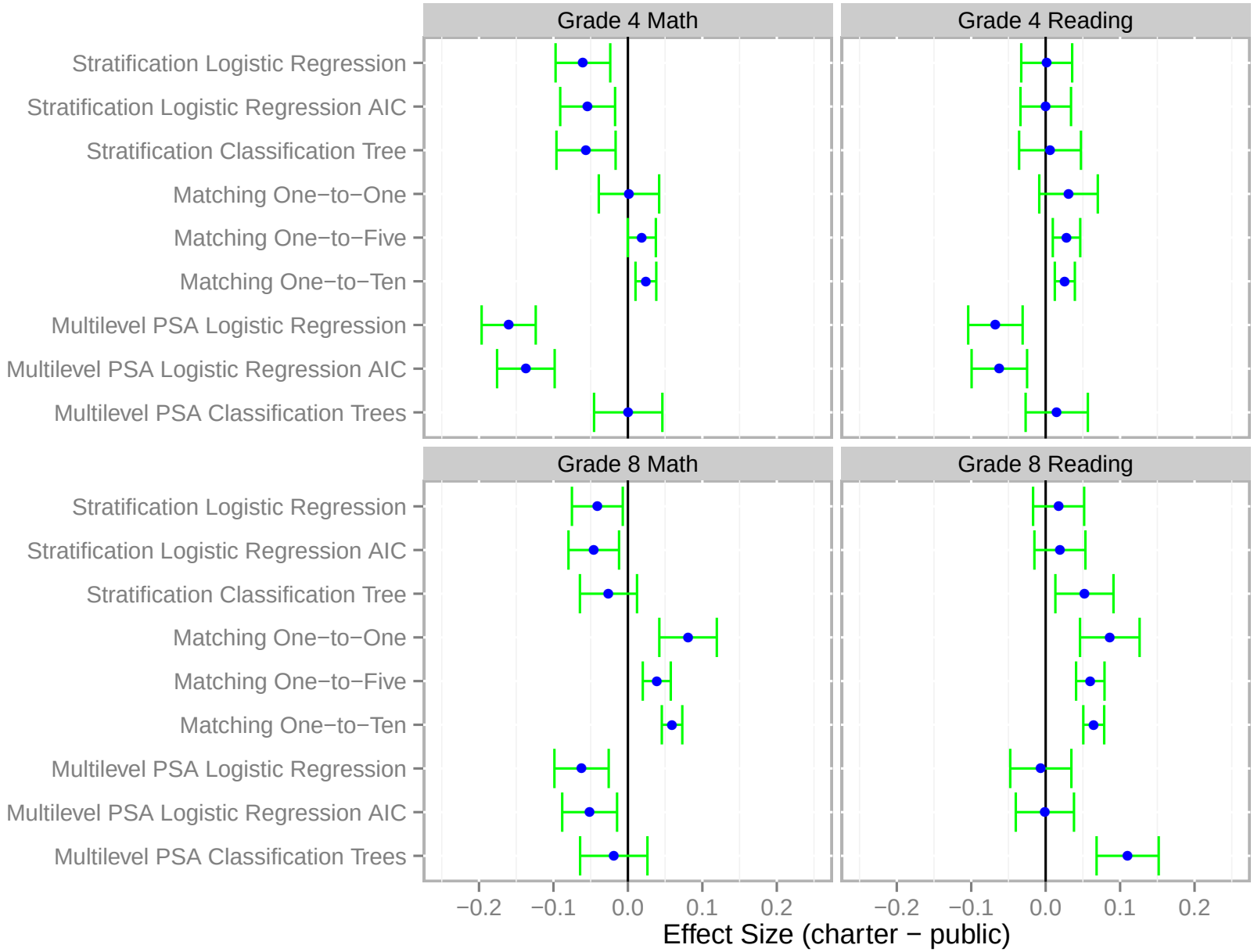


Lather, rinse, and repeat for grade 4 reading,
grade 8 math, and grade 8 reading.

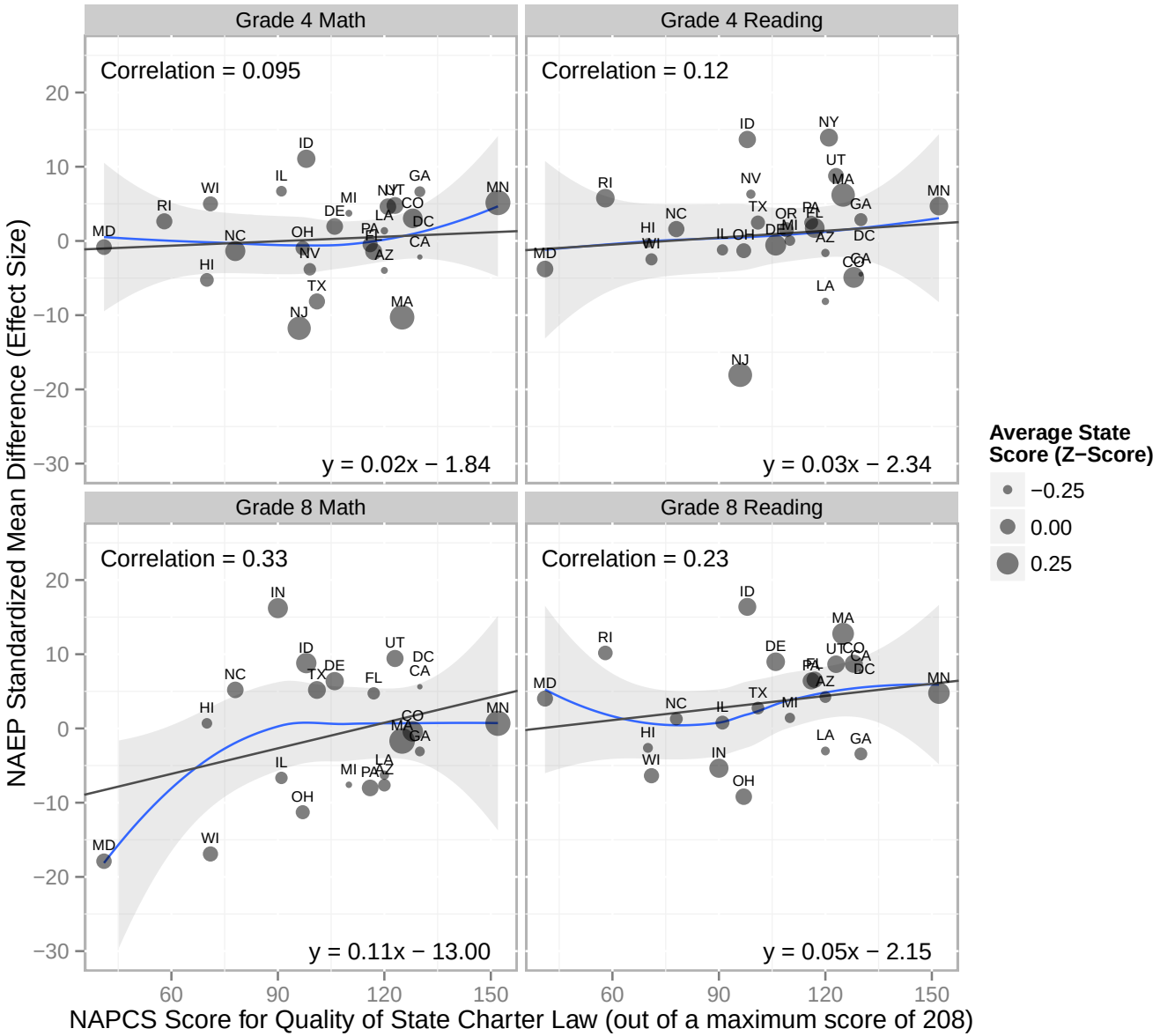
Scatterplot of Overall Adjusted Means (RQs 1 & 2)



Overall Effect Sizes (RQs 1 & 2)



NAPCS Charter Law Scores vs. NAEP Effect Sizes (RQ 3)



Conclusions

- There very little difference between the performance of charter and traditional public school in math and reading in grades 4 and 8 using NAEP as a measure of academic performance.
- The use of visualizations can/is be an important way to present results.
- The `multilevelPSA` R package is an important contribution to the PSA literature and is particularly useful in education research where many studies use data that is naturally multilevel (e.g. students within schools, within districts, within states).
- There is a small to moderate correlation between the quality of charter laws (as identified by NAPCS) and the outcomes from NAEP.

Limitations

- This study only examines reading and math and does not consider other important subjects (e.g. history, science, the arts, etc.).
- There is variation in the quality and types of charter and traditional public schools, but they are considered together (see issue two of charter school research identified by Betts and Hill (2006)).
- Achieving balance using the multilevel PSA methods requires large sample sizes, especially for logistic regression stratification methods.
- The analysis of the quality of charter laws is correlational. Further studies should look at how specific components that NAPCS has identified relate specifically to the performance of specific charter schools.

Resources

- All of the R and LaTeX code is available on Github (`http://github.com/jbryer/Dissertation`).
- `multilevelPSA` - An R package for the methods outlined here. Hosted at `http://github.com/jbryer/multilevelPSA` and documented at `http://jason.bryer.org/multilevelPSA/`.
- `naep` - An R package for accessing and analyzing NAEP data. This work was supported by Educational Testing Services (ETS) and the National Center for Education Statistics (NCES). Documentation is available at `http://jason.bryer.org/naep/`.

Thank You

Jason Bryer (jason@bryer.org)

`http://bryer.org`

References

Austin, P. C. (2011). Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples. *Statistical in Medicine*, *30*, 1292-1301.

Betts, J. R., & Hill, P. T. (2006). *Key issues in studying charter schools and achievement: A review and suggestions for national guidelines*. National Charter School Research Project, Center on Reinventing Public Education, University of Washington Bothell.

Braun, H., Jenkins, F., & Grigg, W. (2006). *A closer look at charter schools using hierarchical linear modeling*. U.S. Government Printing Office.

Center for Research on Education Outcomes. (2009). *Multiple choice: Charter school performance in 16 states*. Stanford University.

Helmreich, J. E., & Pruzek, R. M. (2009). PSAgraphics: An R package to support propensity score analysis. *Journal of Statistical Software*, *29*(6).

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*, 651-674.

National Alliance for Public Charter Schools. (2010). *How state charter laws rank against the new model public charter school law*. Washington D.C.: National Alliance for Public Charter Schools.

Pruzek, R. M., & Helmreich, J. E. (2009). Enhancing dependent sample analysis with graphics. *Journal of Statistical Education*, *17*(1).

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41-55.

Sekhon, J. S. (2011, 6 14). Multivariate and propensity score matching software with automated balance optimization: The matching package for r. *Journal of Statistical Software*, *42*(7), 1–52. Retrieved from `http://www.jstatsoft.org/v42/i07`

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*(3), 1–67. Retrieved from `http://www.jstatsoft.org/v45/i03/`

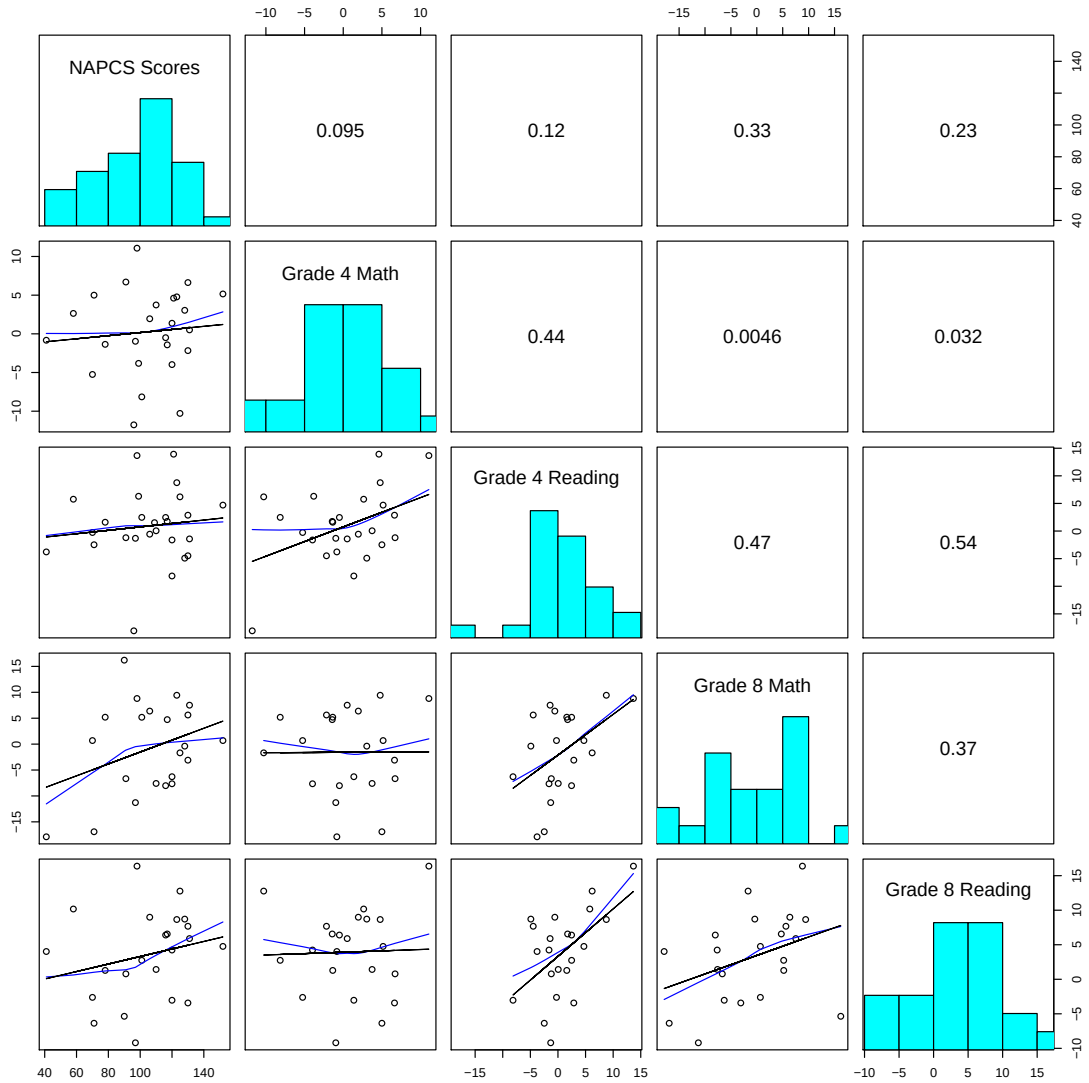
Wells, A. S. (Ed.). (2002). *Where charter school policy fails: The problems of accountability and equity*. New York, NY: Teachers College Press.

20 Components of a Quality Charter Law

1. No caps.
2. A variety of public charter schools allowed.
3. Multiple authorizers available.
4. Authorizer and overall program accountability system required.
5. Adequate authorizer funding.
6. Transparent charter application, review, and decision-making processes.
7. Performance-based charter contracts required.
8. Comprehensive public charter school monitoring and data collection processes.
9. Clear processes for renewal, nonrenewal, and revocation decisions.
10. Educational service providers allowed.
11. Fiscally and legally autonomous schools.
12. Clear student recruitment, enrollment and lottery procedures.
13. Automatic exemptions from many state and district laws and regulations.
14. Automatic collective bargaining exemption.
15. Multi-school charter contracts and/or multi-charter contract boards allowed.
16. Extra-curricular and interscholastic activities eligibility and access.
17. Clear identification of special education responsibilities.
18. Equitable operational funding and equal access to all state and federal categorical funding.
19. Equitable access to capital funding and facilities.
20. Access to relevant employee retirement systems.

(National Alliance for Public Charter Schools, 2010)

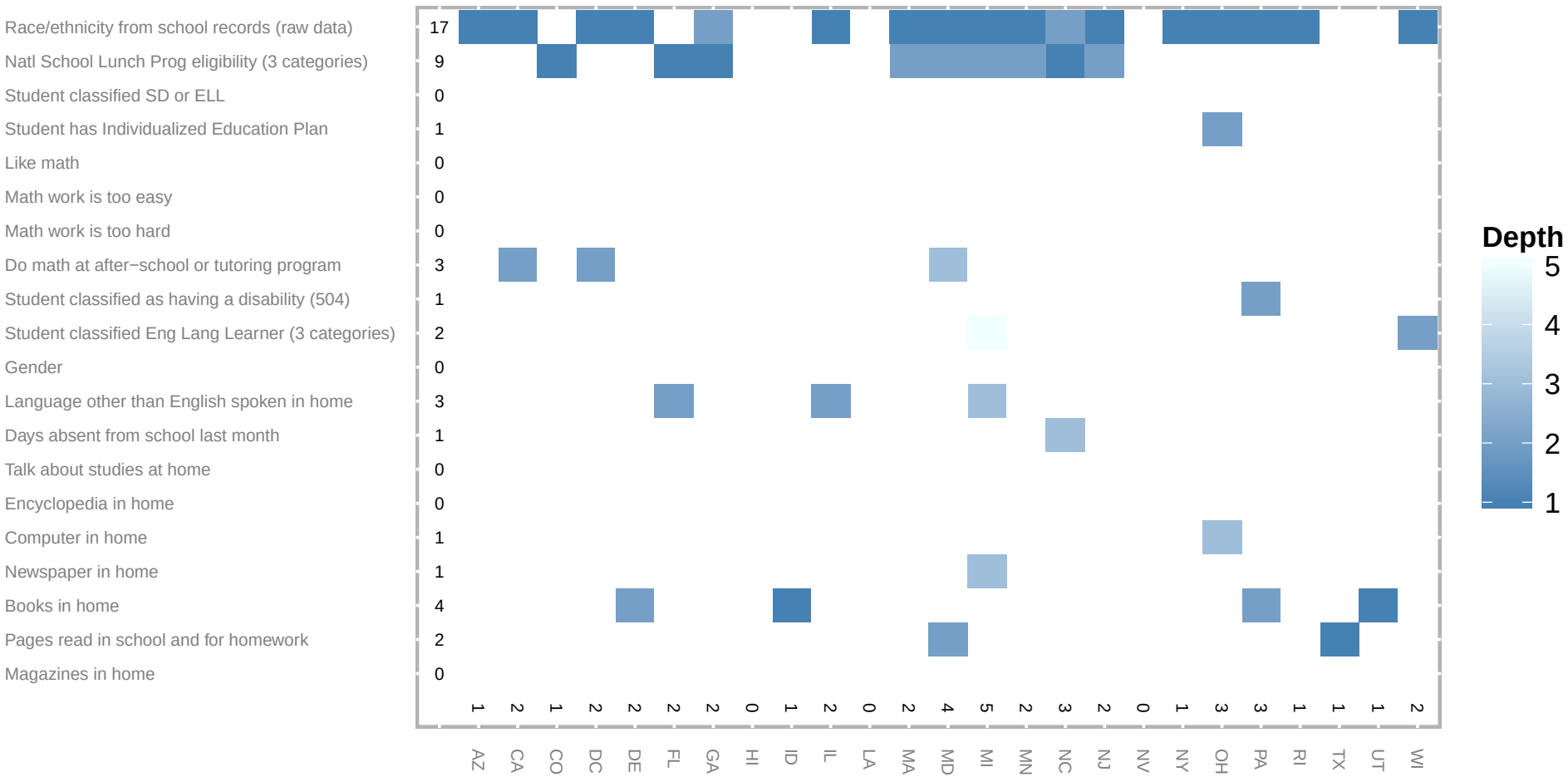
Matrix Plot of NAPCS Quality of Charter Law Scores and NAEP Effect Sizes



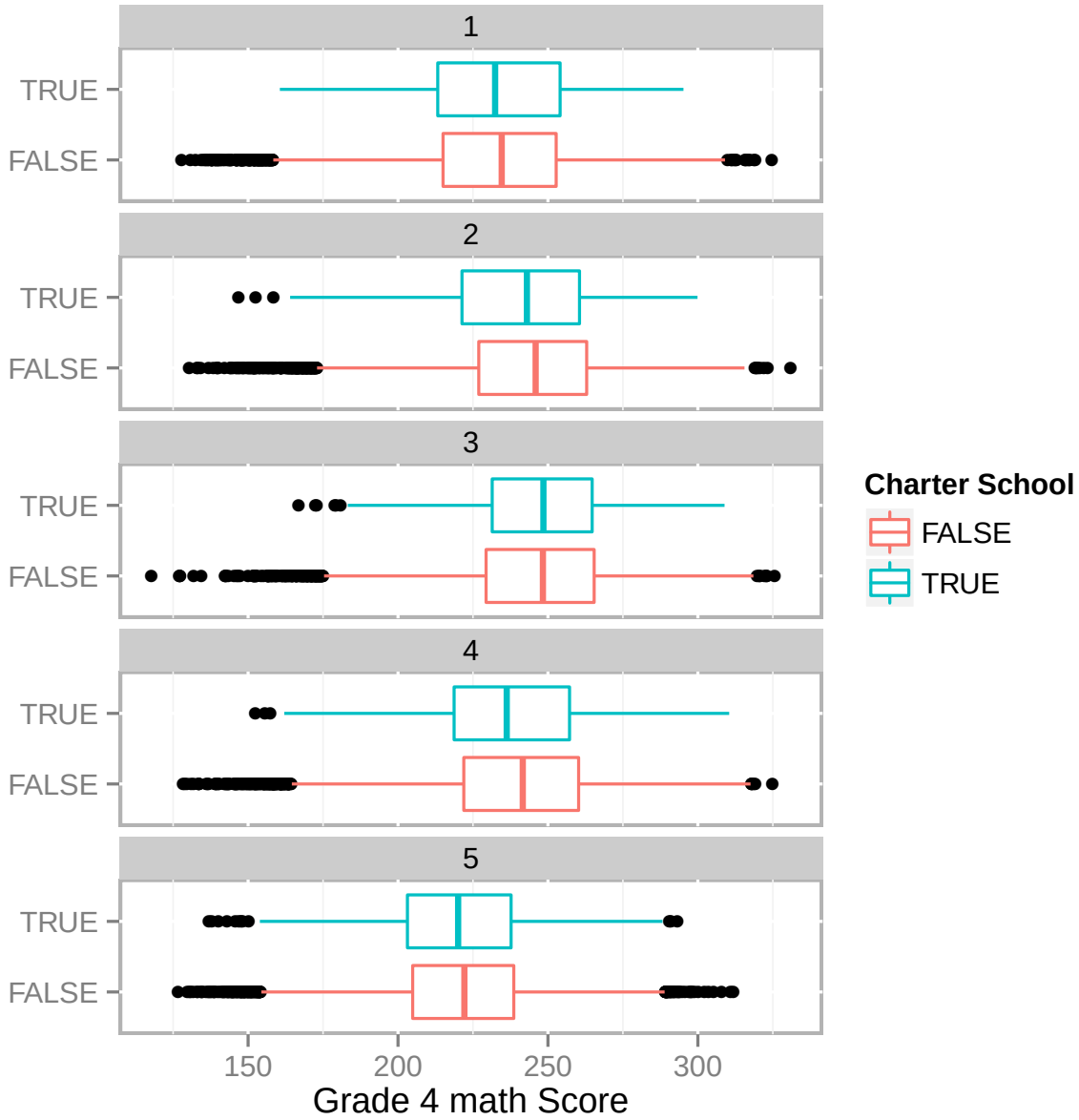
Missing Data

For the logistic regression models the `MICE` (van Buuren & Groothuis-Oudshoorn, 2011) package was used to impute missing values.

Covariate Heat Map for Stratification using Conditional Inference Trees



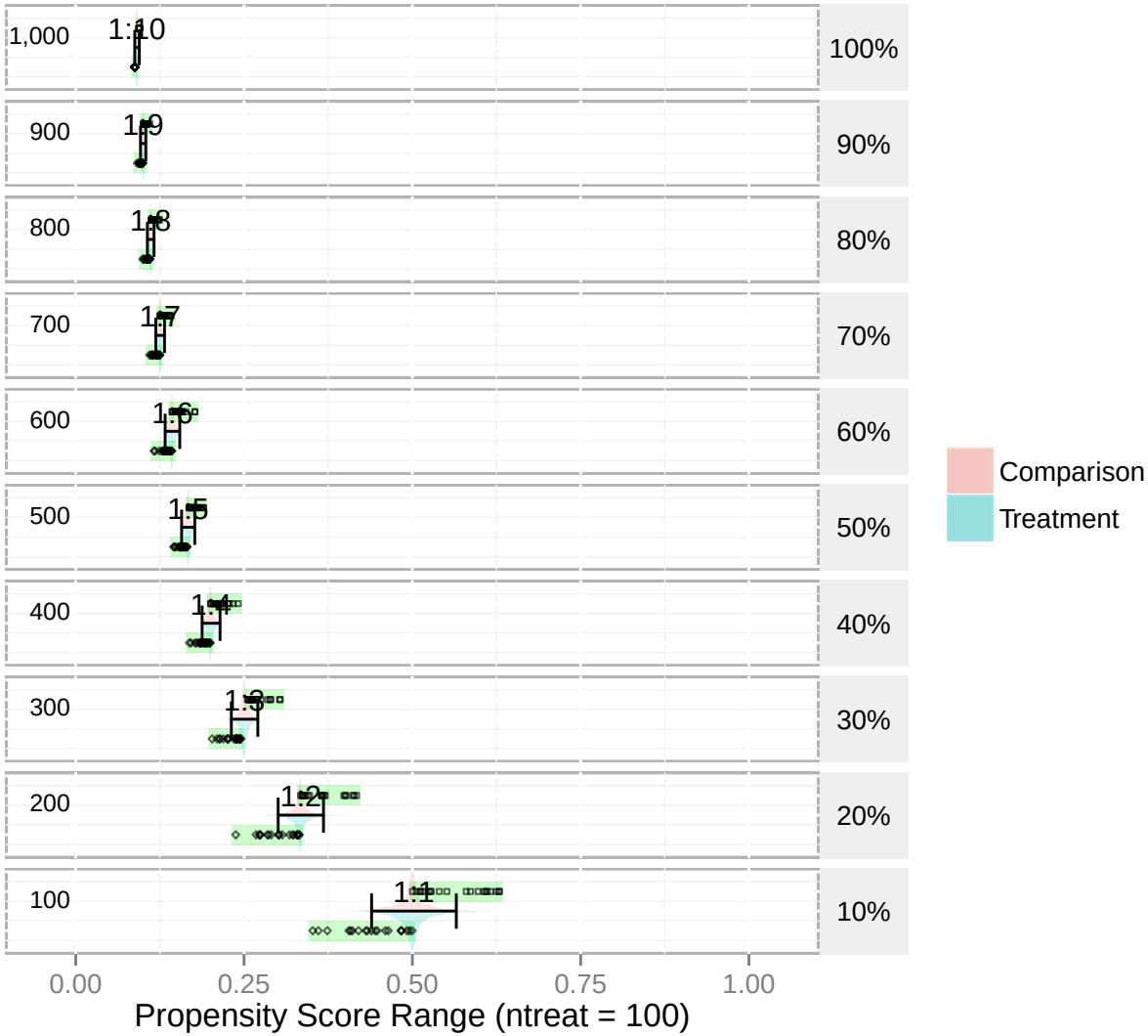
Boxplot of Means by Strata: Grade 4 Math



Propensity Score Ranges for Various Treatment-to-Control Ratios: Perfect Overlap

Parameters used for simulating dependent variables:

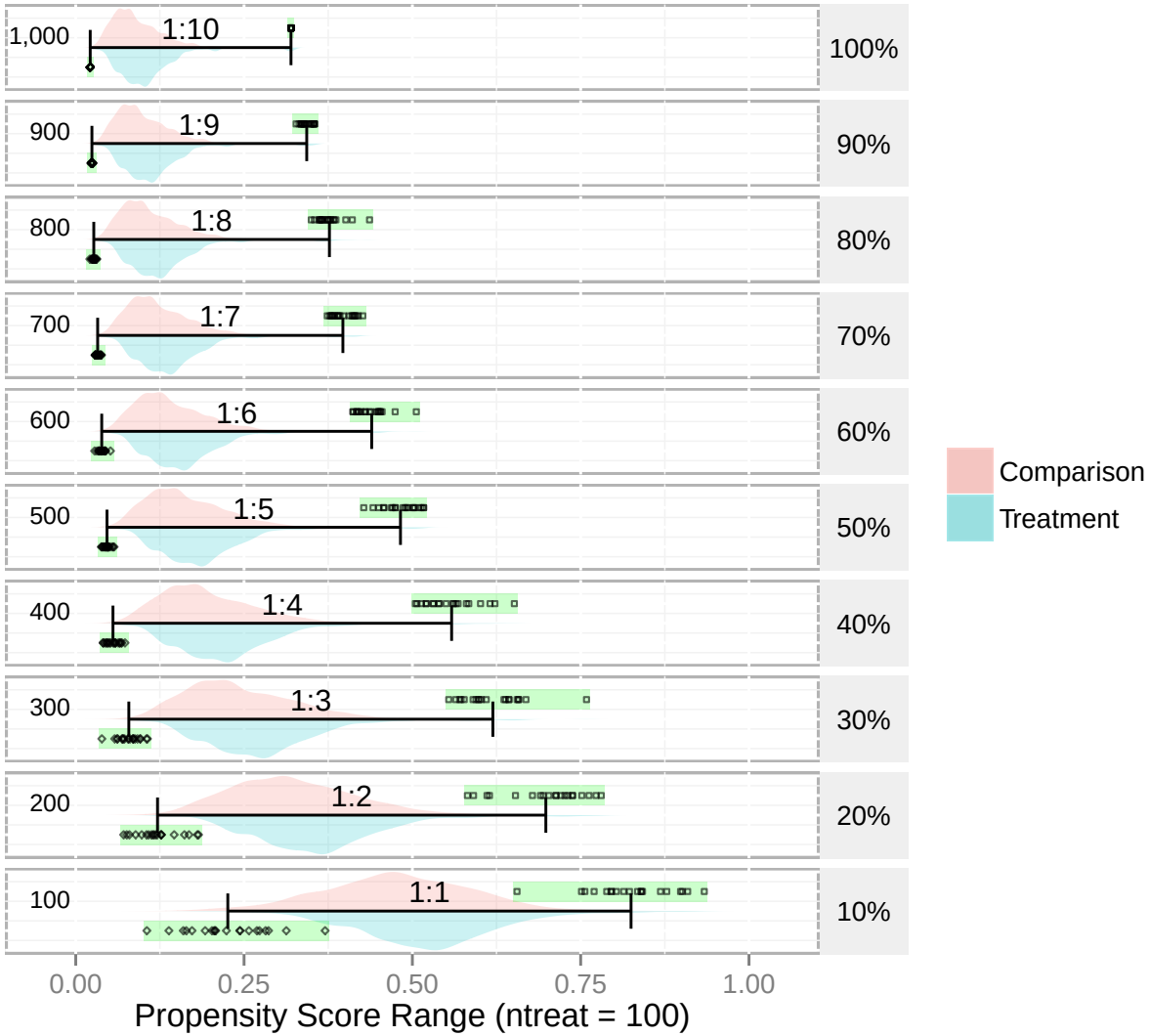
$$\mu_T = 0.5$$
$$sd_T = 0.4$$
$$\mu_C = 0.5$$
$$sd_C = 0.4$$



Propensity Score Ranges for Various Treatment-to-Control Ratios: Moderate Overlap

Parameters used for simulating dependent variables:

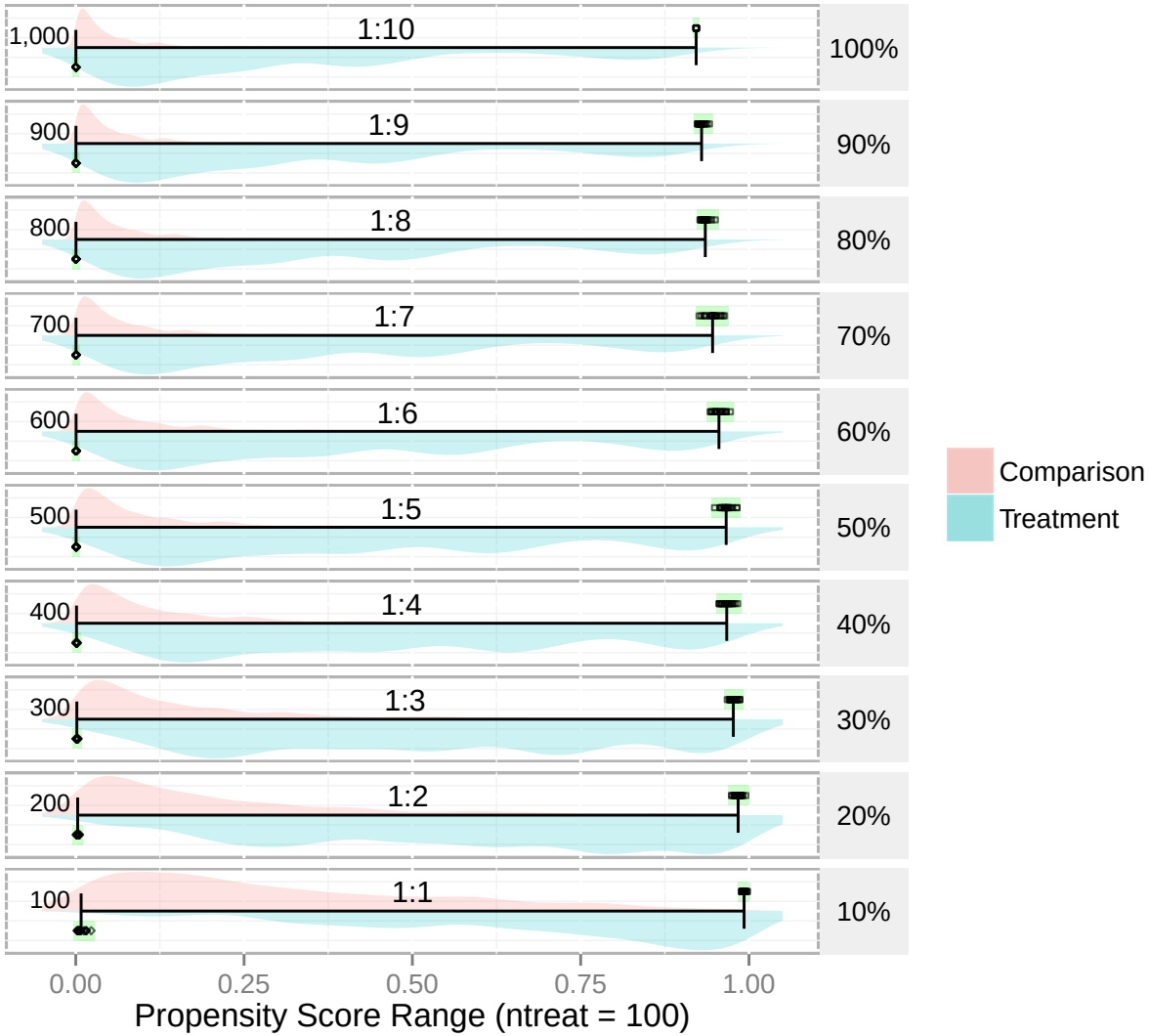
$$\mu_T = 0.6$$
$$sd_T = 0.4$$
$$\mu_C = 0.4$$
$$sd_C = 0.4$$



Propensity Score Ranges for Various Treatment-to-Control Ratios: No Overlap

Parameters used for simulating dependent variables:

$$\mu_T = 0.2$$
$$sd_T = 0.4$$
$$\mu_C = 0.8$$
$$sd_C = 0.4$$



Propensity Score Ranges for Canada (source: PISA)