## Abstract Title Page

*Not included in page count.*

**Title:**
Improving Balance in Regression Discontinuity Design by Matching: Estimating the Effect of Academic Probation After the First Year of College

**Authors and Affiliations:**
Olivia L. Chi (Harvard University)
Aaron W. Dow (Harvard University)

## Abstract Body

*Limit 4 pages single-spaced.*

### Background / Context:

In the study *Ability, Gender, and Performance Standards: Evidence from Academic Probation*, Lindo, Sanders, and Oreopoulos (2010) use regression discontinuity design to examine students' responses to being placed on academic probation after their first year of college. Specifically, they examine the impact of academic probation on the decision to return to the university in a following term, subsequent GPAs, and graduation rates. They find that being placed on academic probation has a discouragement effect for some students, leading them to leave the university, while academic probation placement has an encouragement effect for others, resulting in higher subsequent GPAs. In addition, they find negative impacts on graduation rates, particularly for students with highest high school grades.

However, Lindo, Sanders, and Oreopoulous do not account for imbalance in their regression discontinuity design, in which dissimilarity exists between the pre-treatment characteristics of the treatment and control groups. Such imbalance has attenuated their estimates of the impact of academic probation on students' outcomes.

### Purpose / Objective / Research Question / Focus of Study:

Our study focuses on how matching, a method of preprocessing data prior to estimation and analysis, can be used to reduce imbalance between treatment and control group in regression discontinuity design. To examine the effects of academic probation on student outcomes, we replicate and expand upon research conducted by Lindo, Sanders, and Oreopolous in their 2010 study *Ability, Gender, and Performance Standards: Evidence from Academic Probation*. In replicating the results of Lindo et al. (2010), we find imbalance in observable pre-treatment characteristics between the treatment and control groups in the study. Such imbalance may indicate that randomization may not have been properly approximated through the regression discontinuity design. To improve balance and better approximate a randomized experiment, we preprocess Lindo et al.'s data set by performing exact matching on pre-treatment covariates, such as high school grade percentile and native language. We then re-estimate the impact of academic probation on student outcomes.

### Setting:

A large Canadian university made up of one central campus and two smaller satellite campuses.

### Population / Participants / Subjects:

The original sample in the dataset created by Lindo, Sanders, and Oreopolous consists of 12,530 first year students. All of these first year students had their academic standing evaluation at the end of their first year, and they had all obtained GPAs which are within ±0.6 points of their campus' GPA cutoff for academic probation. These students entered the university between ages 17 and 21 in the school years 1996-97 through 2003-04. After preprocessing the data through exact matching, our sample consists of a subset of the original sample, with 6,506 first year students.

**Intervention / Program / Practice:**

University students with enough credits to qualify for an academic standing evaluation at the end of their first year are placed under academic probation if their GPA falls below their campus' GPA cutoff. (The GPA cutoff is 1.5 at two of the university's campuses and 1.6 at the other university campus.)

**Significance / Novelty of study:**

In replicating the results of Lindo, Sanders, and Oreopolous (2010), we find covariate imbalance between the treatment and control groups. Such imbalance may indicate that randomization may not have been properly approximated through the regression discontinuity design. Figure 1 illustrates an example of imbalance in a pre-treatment characteristic (please insert Figure 1 here). The high school grade percentile distributions of the treatment group (represented by the dotted red line) and the control group (represented by the solid blue line) suggest that the group of individuals receiving the treatment of academic probation had lower prior academic achievement than those just above the cutoff for academic probation. The left side of Table 1 describes the imbalance measures between the treatment and control group in the original sample used by Lindo et al. We calculated a multivariate L1 statistic value of 0.399.[*] Such evidence suggests appreciable nonrandom sorting at the GPA cutoff (please insert Table 1 here).

Our study proposes the use of a matching method to preserve and improve the ability to exploit quasi-experimental designs to obtain valid estimates. Matching allows us to preprocess data and ensure that the actual relationship between pre-treatment student characteristics and treatment is eliminated, without introducing bias in our findings (Ho, Imai, King, and Stuart, 2011). By matching individuals in the treatment and control groups based on pre-treatment characteristics, we are able to reduce multivariate imbalance in pretreatment characteristics, better approximate a random experiment, and obtain improved estimates of the treatment effect.

To this date, we do not know of any education related studies that have capitalized on matching as a preprocessing method prior to estimation in a regression discontinuity design. Given the popularity of regression discontinuity design in education research and program evaluation, the preprocessing method of matching can be exploited to reduce imbalance and model dependence, thereby generating more accurate estimates of treatments.

**Statistical, Measurement, or Econometric Model:**

We propose the use of matching, a nonparametric method of controlling for the confounding influence of pre-treatment control variables in observational data (Ho, Imai, King, and Stewart, 2007; Icarus, King, and Porro, 2012). Matching is a preprocessing method performed on a dataset prior to estimation. As explained by Ho et al. (2007), the goal of matching is to eliminate or reduce the relationship between the treatment indicator $\boldsymbol{T_i}$ and pre-

[*] The multivariate L1 statistic, a comprehensive measure of global imbalance, is the difference between multivariate histograms of pretreatment covariates in the treatment and control groups. A multivariate L1 value of 0 would suggest complete overlap between multivariate histograms of treatment and control groups, while a multivariate L1 value of 1 would suggest no overlap between the multivariate histograms (Icarus, King, and Porro, 2012).

treatment covariates $\boldsymbol{X_i}$ while inducing little bias and inefficiency. The matching process prunes observations from the dataset such that the remaining observations have improved balance between treatment and control groups (i.e., the distributions of pre-treatment covariates in the groups are more similar). In essence, the preprocessed dataset will only include a selected subset of the full dataset for which this relationship holds:

$$p(X \mid T = 1) = p(X \mid T = 0),$$

where $p(\bullet)$ refers to the observed empirical density of the data (Ho et al., 2007). By performing matching prior to estimation in a regression discontinuity design, we are able to remove imbalance, reduce model dependence, and reduce attenuation in the estimates of the treatment in the original study.

**Usefulness / Applicability of Method:**

Given the large sample size of over 12,530 first year students in the original dataset, we elected to use exact matching, which matches all control units with exactly the same covariate values as each of the treatment units (Ho et al., 2007). Exact matching ensures that, within the new sample, students from the treatment group are paired with students from the control group who match exactly on the pretreatment covariates of interest: high school grade percentile, location of campus, gender, age at entry, total credits attempted, and binary indicators for native English language speaker and born in North America.

After performing exact matching, there is no imbalance present in these pre-treatment covariates and we have reduced our sample size to 6,506 students. Figure 2 plots the densities of the high school grade percentiles of students in the treatment and control groups in our pruned dataset (please insert Figure 2 here). The density plots are overlaying, indicating no imbalance in this covariate. This contrasts greatly with Figure 1 described above. The right-side panel of Table 1 also illustrates that the multivariate imbalance measure L1 has transformed into 0.00 in our exact-matched sample. Through exact matching, we are able to ignore baseline student characteristics and conduct local linear regressions around the threshold for academic probation placement to obtain the effect of academic probation on subsequent student outcomes.

**Research Design:**

Regression discontinuity design with matched data.

**Data Collection and Analysis:**

We utilize the dataset collected and created by Lindo, Sanders, and Oreopoulous (2010). The dataset was downloaded from http://www.aeaweb.org/articles.php?doi=10.1257/app.2.2.95 on March 11, 2013. After reducing imbalance by preprocessing the data with an exact matching method on pre-treatment covariates (described above), we estimate the impact of being placed on academic probation by using the regression discontinuity design employed by Lindo et al. (2010). We estimate the discontinuity using local linear regressions with rectangular kernel weights using a bandwidth of 0.6 grade points. As in Lindo et al. (2010), we cluster the standard errors on GPA, since the GPA data are discrete in hundredths of a grade point (please insert Figure 3 and Table 3 here).

**Findings / Results:**

Our results indicate that the effect sizes of academic probation on the decision to leave the university, as well as the effect on subsequent grades, were previously underestimated by Lindo et al. (2010). As displayed in Table 4, we find that the overall effect size of being placed on academic probation following the first year of college is almost double what Lindo et al. (2010) had found: we estimate a 3.5% increase in the rate of leaving for students near the cutoff that are placed under academic probation (please insert Figure 4 and Table 4 here). We also find larger effect sizes on the decision to leave for both males and native English speakers, while finding insignificant effects for females and nonnative English speakers. Contrary to the findings of Lindo et al. (2010), we find a significant impact on academic probation placement for students with high school grades below the median. In this subgroup, students under probation appear to be more likely to leave the university after their first year. Our results can be visualized in Figure 5, where we used simulations to plot the densities of the expected values of the rate of leaving for the treatment and control groups (please insert Figure 5 here).

As displayed in Table 5, across all groups, except for the subgroup of students with high school grades above the median, we again find larger estimated effects of being placed on academic probation on students' GPA in the next term (please insert Figure 6 and Table 5 here). This continues to suggest that academic probation and the threat of suspension may serve as an incentive to improve grades for students who choose to return for another term. Our results can be visualized by simulation in Figure 7 (please insert Figure 7 here).

As displayed in Table 6, performing the analysis on our matched sample reveals no significant impact of being placed under academic probation on the probability of graduating in 4, 5, or 6 years (please insert Figure 8 and Table 6 here). Lindo et al. (2010) had concluded that academic probation has negative effects on graduation rates, particularly for students with the highest high school grades. We, however, were unable to affirm their conclusion. Our results can be visualized by simulation in Figure 9 (please insert Figure 9 here).

**Conclusions:**

By utilizing the matching method of preprocessing data prior to estimation and analysis, we reduced imbalance between treatment and control groups in a regression discontinuity design. By improving balance, we were able to reduce bias in the estimates of the effects of being placed on academic probation on student outcomes and uncover larger discouragement and encouragement effects. We recommend the adoption of preprocessing data through matching when encountering imbalance in regression discontinuity designs, as well as other quasi-experimental methods.

## Appendices

*Not included in page count.*

### Appendix A. References

*References are to be in APA version 6 format.*


Angrist, Joshua David, and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press.

Caughey, Devin and Jasjeet S. Sekhon. 2011. "Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942-2008." *Political Analysis*, 19:385-408.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis*, 15(3): 199-236, http://gking.harvard.edu/_les/abs/matchp-abs.shtml.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. 2007b. "Matchit: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software*, http://gking.harvard.edu/matchit/.

Iacus, Stefano M., Gary King and Giuseppe Porro, 2012. "Causal Inference Without Balance Checking: Coarsened Exact Matching." *Political Analysis*, 20(1):1-24.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science*, 44(2):341-355.

Lee, David S. and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48: 281-355.

Lindo, Jason M., Nicholas J. Sanders, and Philip Oreopoulous. 2010. "Ability, Gender, and Performance Standards: Evidence from Academic Probation." *American Economic Journal: Applied Economics*, 2(2):95-117.

## Appendix B. Tables and Figures

*Not included in page count.*

Figure 1: Pre-matching: Imbalance in High School Grade Percentile Distributions

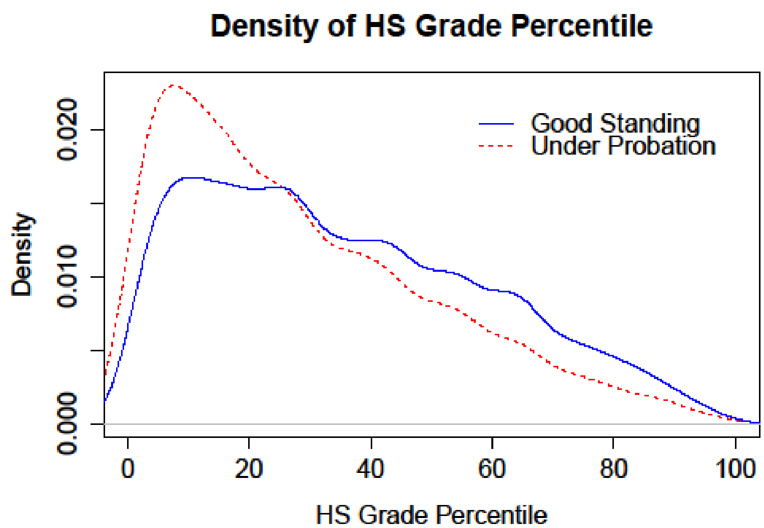



Figure 2: Post-matching: Balance in High School Grade Percentile Distributions

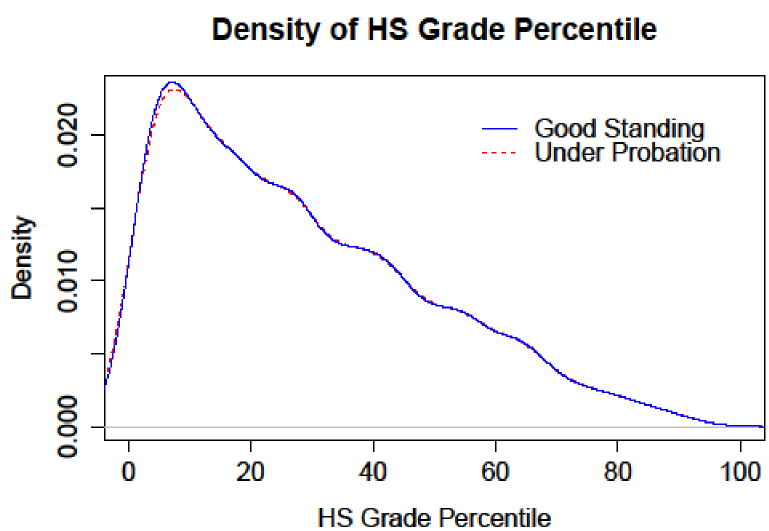
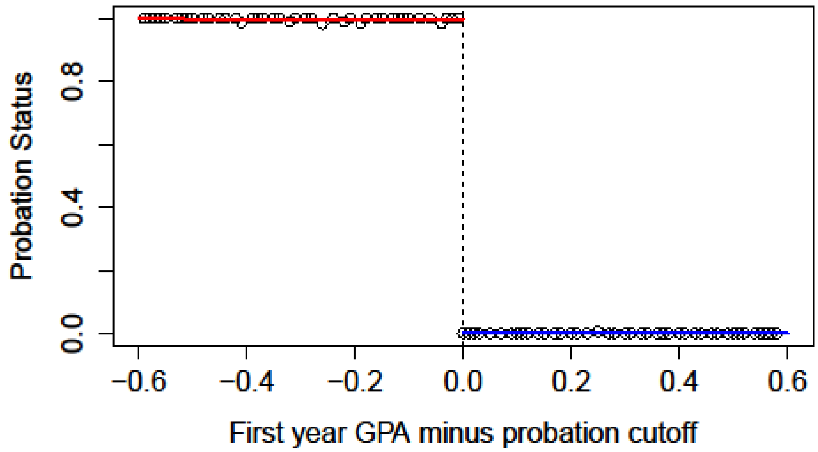
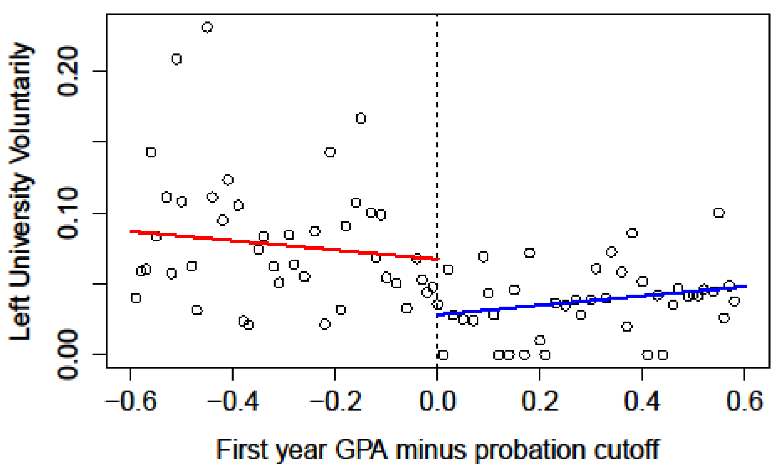


Figure 3: Post-matching: Probation Status at the End of the First Year



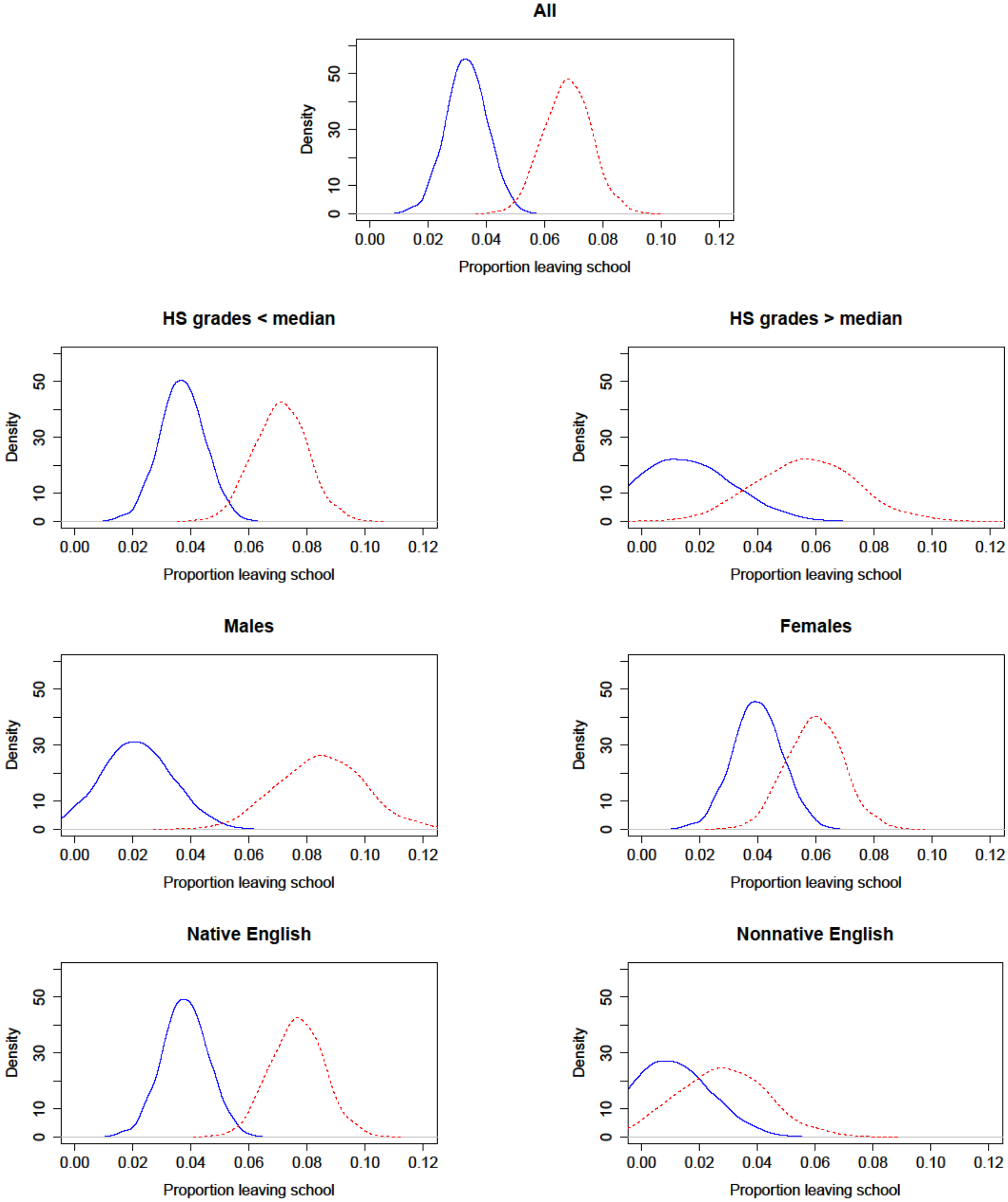

Notes: Each hollow circle is the mean of the outcome in an interval of 0.01 around the point (including the lower, but not the upper endpoint). After exact matching, the curve is predicted from local linear regressions with a bandwidth of 0.6 using rectangular kernel weights.

Figure 4: Post-matching: Voluntarily Leaving at the End of the First Year



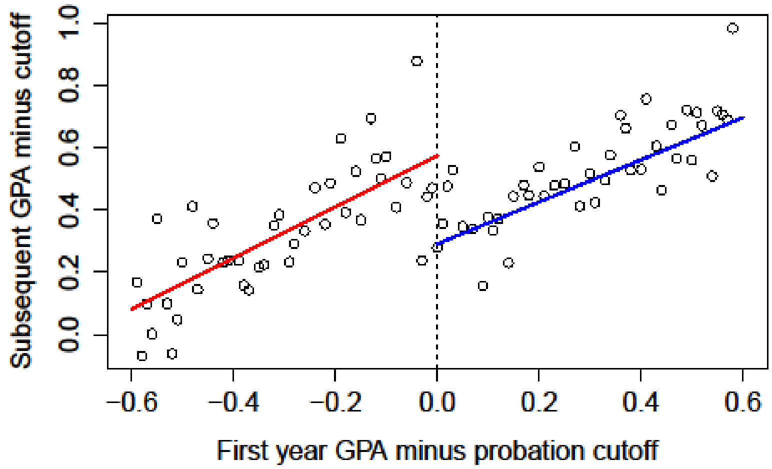

Notes: Each hollow circle is the mean of the outcome in an interval of 0.01 around the point (including the lower, but not the upper endpoint). After exact matching, the curve is predicted from local linear regressions with a bandwidth of 0.6 using rectangular kernel weights.

Figure 5: Simulated Levels of Leaving University Voluntarily by Group



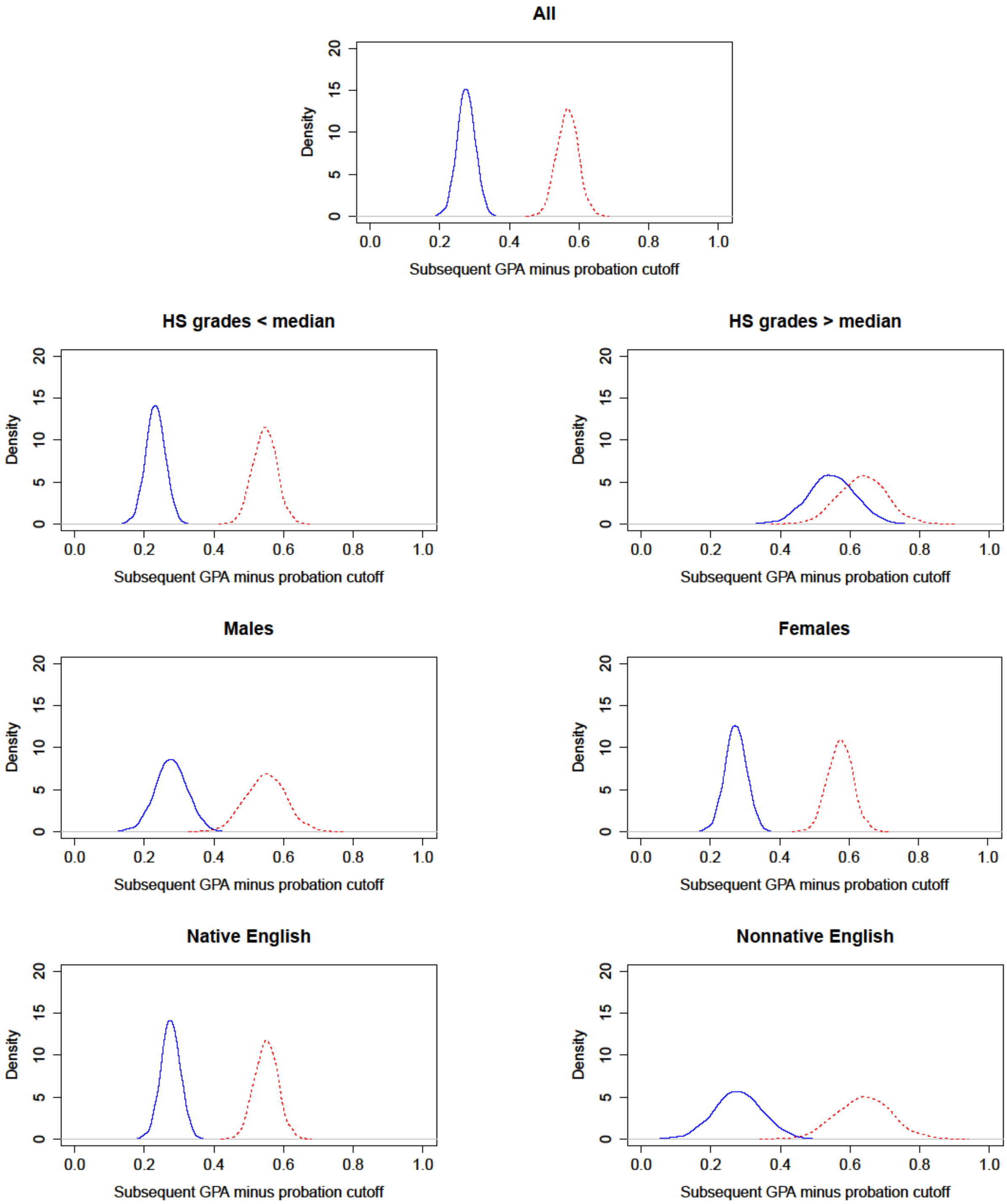

Note: These panels contain density estimates of expected levels of voluntarily leaving university at the end of the first year for: students who are just above the cutoff for academic probation and are not under academic probation (blue solid curve); and for students who are just below the cutoff for academic probation and receive academic probation (red dotted curve).

Figure 6: Post-matching: GPA in Next Enrolled Term



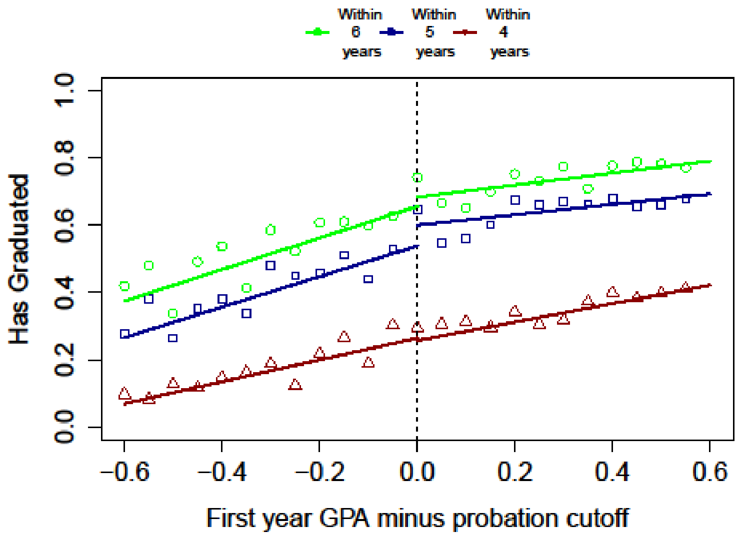

Notes: Each hollow circle is the mean of the outcome in an interval of 0.01 around the point (including the lower, but not the upper endpoint). After exact matching, the curve is predicted from local linear regressions with a bandwidth of 0.6 using rectangular kernel weights.

Figure 7: Simulated Subsequent GPA by Group



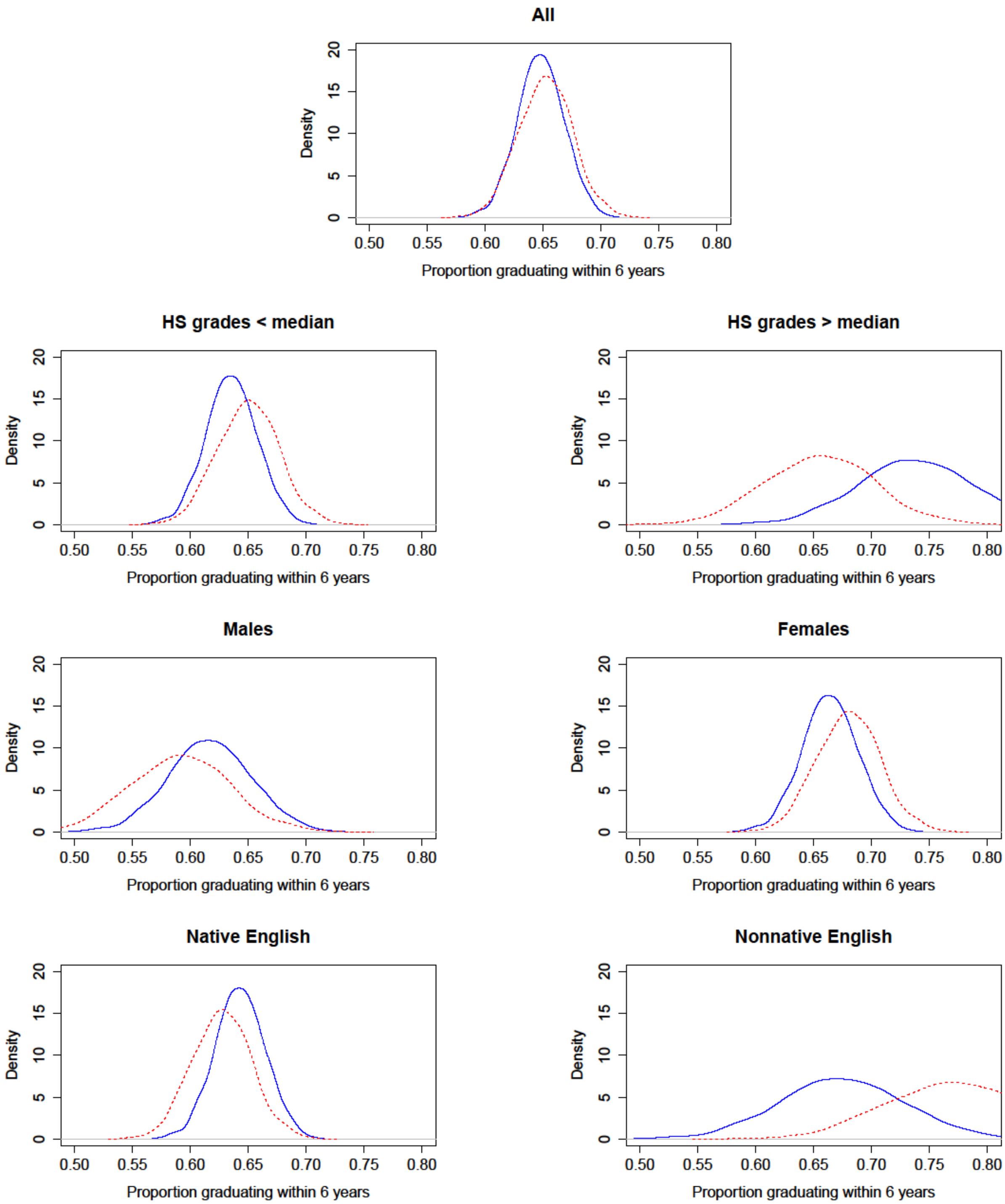

Note: These panels contain density estimates of expected values of subsequent GPA (minus probation cutoff) for: students who are just above the cutoff for academic probation and are not under academic probation (blue solid curve); and for students who are just below the cutoff for academic probation and receive academic probation (red dotted curve).

Figure 8: Post-matching: Graduation Rates




Notes: Each hollow circle/square/triangle is the mean of the outcome in an interval of 0.01 around the point (including the lower, but not the upper endpoint). After exact matching, the curve is predicted from local linear regressions with a bandwidth of 0.6 using rectangular kernel weights.

Figure 9: Simulated Levels of Graduating Within 6 Years by Group




Note: These panels contain density estimates of expected levels of graduation within 6 years for: students who are just above the cutoff for academic probation and are not under academic probation (blue solid curve); and for students who are just below the cutoff for academic probation and receive academic probation (red dotted curve).

TABLE 1. Imbalance Measures in Original and Matched Samples

| Student Characteristic | Original Sample | | Exact Matched Sample | |
|---|---|---|---|---|
| Statistic | Difference in Means | L1-Stat | Difference in Means | L1-Stat |
| High School Grade Percentile | 7.035 | 0.127 | 0.000 | 0.000 |
| Total Credits, Year One | 0.090 | 0.24 | 0.000 | 0.000 |
| Age at Entry | -0.032 | 0.006 | 0.000 | 0.000 |
| Male | -0.020 | 0.020 | 0.000 | 0.000 |
| Native-English Speaker | 0.034 | 0.034 | 0.000 | 0.000 |
| Born in North America | 0.023 | 0.023 | 0.000 | 0.000 |
| Attended Campus 1 | 0.068 | 0.068 | 0.000 | 0.000 |
| Attended Campus 2 | -0.004 | -0.004 | 0.000 | 0.000 |
| Attended Campus 3 | -0.064 | -0.064 | 0.000 | 0.000 |
| | Multivariate L1-Stat | | | |
| Overall Imbalance | 0.399 | | 0.000 | |

Notes: Imbalance measured in pre and post-matched sample at a bandwidth of 0.6. Difference in means are calculated as the difference in mean covariate values between treatment and control groups. L1 statistics are created by overlaying histograms of covariate values between treatment and control groups.

TABLE 2. Summary Statistics

| | Mean | St. Dev. |
|---|---|---|
| *Characteristics* | | |
| High school grade percentile | 29.98 | 21.13 |
| Credits attempted in first year | 4.46 | 0.49 |
| Age at entry | 18.73 | 0.53 |
| Male | 0.33 | 0.47 |
| English is first language | 0.86 | 0.35 |
| Born in North America | 0.97 | 0.17 |
| At Campus 1 | 0.54 | 0.50 |
| At Campus 2 | 0.22 | 0.42 |
| At Campus 3 | 0.24 | 0.43 |
| *Outcomes* | | |
| Distance from cutoff in 1st year | 0.08 | 0.33 |
| On probation after 1st year | 0.39 | 0.49 |
| Ever on academic probation | 0.49 | 0.50 |
| Left university after 1st evaluation | 0.05 | 0.23 |
| Distance from cutoff at next evaluation | 0.47 | 0.79 |
| Ever suspended | 0.17 | 0.38 |
| Graduated by year 4 | 0.28 | 0.45 |
| Graduated by year 5 | 0.55 | 0.50 |
| Graduated by year 6 | 0.66 | 0.47 |

Notes: For all variables except graduation rates and next evaluation distance from the cutoff, the sample consists of 6,506 students within 0.6 grade points of the cutoff in their first year after using exact matching. 5,826 students are observed with a GPA following their first evaluation. Graduation rate samples are 4,769 for four years, 3,955 for five years, and 3,303 for six years.

Table 3. Estimated Discontinuities in Probation Status

| Relevant group | All (1) | HS grades <median (2) | HS grades >median (3) | Male (4) | Female (5) | Native English (6) | Nonnative English (7) |
|---|---|---|---|---|---|---|---|
| *Dependent variable: On academic probation after first evaluation* | | | | | | | |
| First year GPA <cutoff | 0.996*** | 0.997*** | 0.993*** | 0.996*** | 0.996*** | 0.995*** | 1.000*** |
| | (0.002) | (0.002) | (0.006) | (0.002) | (0.002) | (0.002) | (0.000) |
| Constant (control mean) | 0.001 | 0.001 | 0.000*** | 0.000 | 0.001 | 0.001 | 0.000*** |
| | (0.001) | (0.001) | (0.000) | | (0.001) | (0.001) | (0.000) |
| Observations | 6,506 | 5,270 | 1,236 | 2,137 | 4,369 | 5,593 | 913 |
| *Dependent variable: Ever on academic probation* | | | | | | | |
| First year GPA <cutoff | 0.654*** | 0.642*** | 0.730*** | 0.653*** | 0.654*** | 0.656*** | 0.640*** |
| | (0.020) | (0.021) | (0.033) | (0.025) | (0.027) | (0.021) | (0.040) |
| Constant (control mean) | 0.343*** | 0.356*** | 0.263*** | 0.344*** | 0.343*** | 0.340*** | 0.360*** |
| | (0.020) | (0.021) | (0.033) | (0.025) | (0.027) | (0.021) | (0.040) |
| Observations | 6,506 | 5,270 | 1,236 | 2,137 | 4,369 | 5,593 | 913 |

Notes: Estimated standard errors, clustered on GPA, are displayed in parentheses. Estimates are calculated after exact matching and based on linear regression with rectangular kernel weights and a bandwidth of 0.6. *p<0.1; **p<0.05; ***p<0.01

Table 4. Estimated Effect on the Decision to Leave After the First Evaluation

| Relevant group | All (1) | HS grades <median (2) | HS grades >median (3) | Male (4) | Female (5) | Native English (6) | Nonnative English (7) |
|---|---|---|---|---|---|---|---|
| First year GPA <cutoff | 0.035*** | 0.034** | 0.044* | 0.064*** | 0.020 | 0.039*** | 0.018 |
| | (0.012) | (0.014) | (0.026) | (0.022) | (0.014) | (0.013) | (0.019) |
| Constant (control mean) | 0.033*** | 0.037*** | 0.012 | 0.020* | 0.039*** | 0.038*** | 0.008 |
| | (0.007) | (0.009) | (0.016) | (0.010) | (0.009) | (0.008) | (0.011) |
| Observations | 6,506 | 5,270 | 1,236 | 2,137 | 4,369 | 5,593 | 913 |

Notes: Estimated standard errors, clustered on GPA, are displayed in parentheses. Estimates are calculated after exact matching and based on linear regression with rectangular kernel weights and a bandwidth of 0.6. *p<0.1; **p<0.05; ***p<0.01

TABLE 5. Estimated Discontinuities in Subsequent GPA

| | All | HS grades <median | HS grades >median | Male | Female | Native English | Nonnative English |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A: Dependent variable: Next term GPA* | | | | | | | |
| First year GPA <cutoff | 0.307*** | 0.330*** | 0.099 | 0.289*** | 0.316*** | 0.290*** | 0.378*** |
| | (0.053) | (0.053) | (0.124) | (0.080) | (0.074) | (0.060) | (0.091) |
| Constant (control mean) | 0.268*** | 0.224*** | 0.541*** | 0.271*** | 0.264*** | 0.267*** | 0.269*** |
| | (0.036) | (0.038) | (0.088) | (0.048) | (0.056) | (0.041) | (0.074) |
| Observations | 5, 826 | 4, 688 | 1, 138 | 1, 879 | 3, 947 | 4, 976 | 850 |
| *Panel B: Dependent variable: Probability of improving GPA in next term* | | | | | | | |
| First year GPA ¡ cutoff | 0.137*** | 0.152*** | 0.034 | 0.087** | 0.165*** | 0.135*** | 0.147** |
| | (0.031) | (0.034) | (0.062) | (0.042) | (0.040) | (0.032) | (0.061) |
| Constant (control mean) | 0.651*** | 0.632*** | 0.772*** | 0.697*** | 0.625*** | 0.653*** | 0.642*** |
| | (0.024) | (0.027) | (0.050) | (0.033) | (0.032) | (0.026) | (0.050) |
| Observations | 5, 826 | 4, 688 | 1, 138 | 1, 879 | 3, 947 | 4, 976 | 850 |

Notes: Estimated standard errors, clustered on GPA, are displayed in parentheses. Estimates are calculated after exact matching and based on linear regression with rectangular kernel weights and a bandwidth of 0.6. *p<0.1; **p<0.05; ***p<0.01

TABLE 6. Estimated Effects on Graduation

| Relevant group | All (1) | HS grades <median (2) | HS grades >median (3) | Male (4) | Female (5) | Native English (6) | Nonnative English (7) |
|---|---|---|---|---|---|---|---|
| *Panel A: Dependent variable: graduated after four years* | | | | | | | |
| First year GPA <cutoff | 0.020 | 0.017 | −0.007 | −0.048 | 0.055 | −0.0004 | 0.118 |
| | (0.031) | (0.031) | (0.072) | (0.049) | (0.040) | (0.033) | (0.079) |
| Constant (control mean) | 0.244*** | 0.229*** | 0.345*** | 0.213*** | 0.259*** | 0.242*** | 0.257*** |
| | (0.021) | (0.023) | (0.052) | (0.039) | (0.022) | (0.023) | (0.061) |
| Observations | 4,769 | 3,892 | 877 | 1,573 | 3,196 | 4,129 | 640 |
| *Panel B: Dependent variable: graduated after five years* | | | | | | | |
| First year GPA <cutoff | −0.015 | −0.013 | −0.070 | −0.010 | −0.019 | −0.029 | 0.062 |
| | (0.053) | (0.058) | (0.071) | (0.061) | (0.063) | (0.050) | (0.099) |
| Constant (control mean) | 0.553*** | 0.539*** | 0.655*** | 0.502*** | 0.581*** | 0.535*** | 0.641*** |
| | (0.044) | (0.050) | (0.051) | (0.054) | (0.049) | (0.041) | (0.084) |
| Observations | 3,955 | 3,213 | 742 | 1,307 | 2,648 | 3,432 | 523 |
| *Panel C: Dependent variable: graduated after six years* | | | | | | | |
| First year GPA <cutoff | 0.011 | 0.023 | −0.077 | −0.017 | 0.023 | −0.008 | 0.106 |
| | (0.048) | (0.053) | (0.086) | (0.064) | (0.060) | (0.047) | (0.096) |
| Constant (control mean) | 0.644*** | 0.632*** | 0.735*** | 0.614*** | 0.660*** | 0.639*** | 0.669*** |
| | (0.042) | (0.048) | (0.064) | (0.050) | (0.047) | (0.038) | (0.085) |
| Observations | 3,303 | 2,691 | 612 | 1,095 | 2,208 | 2,870 | 433 |

Notes: Estimated standard errors, clustered on GPA, are displayed in parentheses. Estimates are calculated after exact matching and based on linear regression with rectangular kernel weights and a bandwidth of 0.6.
*p<0.1; **p<0.05; ***p<0.01