Lack of Evidence?: Academic Probation Does Not Affect Student Dropout and Subsequent Performance

Revisiting the Ability, Gender, and Performance Standards: Evidence from Academic Probation (Lindo, Sanders, and Oreopoulos 2010)

Hong Shi

21 April 2021

Abstract

Many universities use academic probation as a wake-up call to ensure currently enrolled students to achieve minimum academic standards. This paper examines the causal impact on first-year students' responses in a Canadian univesity after being placed on academic probation in a sharp regression discontinuity design. The results do not suggest that placing first year students on academic probation affects the dropout decision, subsequent annual GPA and graduation. While there are heterogeneous responses across students' highschool performance, gender, first language towards academic probation. #High Level Importance

Keywords: Regression Discontinuity Design, Bandwidth, Academic Probation, Student Performance, P-value, Causality

1 Introduction

First-year students usually hold positive expectations after entering the university (McGrath and Burd 2012). Despite their high expectations, some students fail to achieve the minimum academic standards and are placed into academic probation. Academic probation serves as a wake-up call for students whose grade point average (GPA) is below a certain threshold and can lead to escalating penalties such as suspensions. Therefore, placing students on academic probation is equivalent to setting a minimum standard for their future academic performance (Lindo, Sanders, and Oreopoulos 2010), while previous studies implied that students may react quite differently to academic probation (e.g., dropout, increase academic performance, etc.)(Bénabou and Tirole 2000).

This paper examines the causal impact on first-year university students after being placed on academic probation by exploiting the discontinuous probation GPA cutoff in a sharp regression discontinuity design (RDD). The data comes form a large Canadian university with three individual campuses. It is a replication of `"Ability, Gender, and Performance Standards: Evidence from Academic Probation."` (Lindo, Sanders, and Oreopoulos 2010) with adjustment on the bandwidth (i.e., student within certain grade points of the academic probation cutoff). After applying the adjusted bandwidth, the results are NOT consistent with Lindo's findings that being placed on probation at the end of the first year increases the student dropout, improves the GPAs of the remaining students and negatively affects graduation rates. While students with different gender, highschool academic performance or first language react differently to academic probation. Furthermore, I also find heterogeneous responses to academic probation across campus that is not mentioned in the original paper (Note: May adjust the result after running the model.).

The rest of this paper is organized as follows: In Data section (Section 2), I would first explain the domain background about the academic probation. Next, I introduce the data used in the original paper. And then, I would describe the Quasi-experiment design rationale and also raise concerns about the bandwidth used in the original paper. In Model section (Section 3), I elaborate on the sharp RDD, address its assumptions and propose the model used in this study. In Results section (Section 4), I show the impact of academic probation on student dropout, subsequent performance and graduation with the adjusted bandwidth. In Discussion section (Section 5), I would first summarize the study findings, and then elaborate on the causality of the experiment, address external and internal validity and also Simpson's Paradox (also possibly Berkson's Paradox) regarding the campus variation. Finally, I conclude with some limitations and suggest future directions of this experiment.[1]

2 Data

The data comes from a large Canadian university with three individual campuses: one central campus (Campus 1) and two smaller satellite campuses (Campus 2 and Campus 3) from 1996 to 2005. I analyzed it using R (R Core Team 2020), and packages `tidyverse` (Wickham et al. 2019), `here` (Müller 2020), `rdrobust` (Calonico et al. 2021), `rdd`(Dimmery 2016), `rdpower`(Cattaneo, Titiunik, and Vazquez-Bare 2020), `haven`(Wickham and Miller 2020) . I used packages `bookdown` (Xie 2016), `kableExtra` (Zhu 2020), , `finalfit` (Harrison, Drake, and Ots 2020), `modelsummary`(Arel-Bundock 2021), `broom` (Robinson, Hayes, and Couch 2020) to format the document and referenced `Impact Evaluation in Practice` (Gertler et al. 2016) to evaluate this experiment.

2.1 About Academic Probation

The principle of academic probation in this university is simple: if a student's grade point average (GPA) is below a certain threshold, the student is placed on academic probation. At Campus 1 and Campus 2, students with a cumulative GPA below 1.5 grade points are placed on academic probation. Campus 3 has a GPA cutoff at 1.6 grade points. Students with GPA exactly at the probation cutoff are considered in good academic standing.

Since many first-year courses span the entire year, students' academic standings are evaluated at the end of first scholastic year. Academic standings are evaluated at the end of every subsequent scholastic year and summer term at Campus 1 and Campus 2, and end of every subsequent term at Campus 3[2].

Students on academic probation will be suspended in subsequent sessions if their grades do not improve. At all campuses, students on probation can avoid suspension and return to good academic standing by improving their cumulative GPA up to the cutoff. Students who fail to sufficiently improve their grades are suspended for one full academic year. If suspended students return to the university and again fail to sufficiently improve their grades, they can be suspended for three years. A third failure to meet the GPA requirement can lead to permanent suspension from all campuses. Students will be notified after being placed on academic probation. A sample letter sent to students at Campus 2 is in Appendix 1.

2.2 Original Data

The data used in the original paper includes administrative information of students in a large Canadian university as mentioned above. It covers student observations in a nine-year period from 1996 to 2005 and has cleaned the student sample in following aspects for the purpose of this experiment[3]:

1. Restrict students who entered the university before 2004 so that students can be potentially observed for two years.
2. Omit students with missing data for any variables of interest, particularly high school grades (84% of the sample).
3. Restrict students entering the university between age 17 and 21 (99% of the remaining sample).
4. Restrict students having their academic standing evaluated at the end of Year 1 (98% of the remaining sample).

[1]Code and data are available at the GitHub repo: https://github.com/honn-ishinn/RDD_academic_probation.
[2]Students also must attempt a minimum number of credits before being evaluated. The data used in the original paper omit all students who have not been evaluated by the end on first year. Potential bias issues are elaborated in Discussion section
[3]Potential bias issues about data cleaning in the original paper are elaborated in Discussion section

Table 1: Summary Statistics of Observable Characteristics and Outcomes of All Students

Description		All Students
Characteristics		
		N(%) = 44362 (100.0)
Gender	Male	16981 (38.3)
	Female	27381 (61.7)
Birth Place	North America	38633 (87.1)
	Asia	3763 (8.5)
	Other	1966 (4.4)
Study Campus	Campus 1	25915 (58.4)
	Campus 2	7695 (17.3)
	Campus 3	10752 (24.2)
First Language	English	31662 (71.7)
	Other	12469 (28.3)
		Mean (SD)
Highschool Grade Percentile		50.17 (28.86)
Credits Attempted in First Year		4.57 (0.51)
Age at Entry		18.67 (0.74)
Outcomes		
		N(%) = 44362 (100.0)
On Probation After 1st Year	Yes	7106 (16.0)
	No	37256 (84.0)
Ever On Probation	Yes	8711 (19.6)
	No	35651 (80.4)
Left University After 1st Evaluation	Yes	2175 (4.9)
	No	42187 (95.1)
Ever Suspended	Yes	3562 (8.0)
	No	40800 (92.0)
Graduated by Year 4	Yes	13407 (44.7)
	No	16610 (55.3)
Graduated by Year 5	Yes	16594 (67.5)
	No	7987 (32.5)
Graduated by Year 6	Yes	14880 (75.3)
	No	4877 (24.7)
		Mean (SD)
First Year GPA		2.44 (0.89)
Distance from Cutoff in 1st Year		0.91 (0.90)
Second Year GPA		2.55 (0.83)
Distance from Cutoff in 2nd Year		1.03 (0.84)

[1] For all characteristics and outcomes except graduation rates and distance from cutoff in 2nd year. The entire dataset consists of 44362 students. Graduation rate samples are 30017 for Year 4, 24581 for Year 5, 19757 for Year 6.38576 students have GPA observed in 2nd Year

After the data cleaning , the original data includes a total of 44362 observations and Table 1 shows the summary statistics of student observable characteristics and outcomes such as gender, age, birth place, study campus, first language, attempted course credit[4], registration status, GPA, academic standing, high school performance, and graduation status. Since the probation cutoff among campus is different (1.5 in Campus 1 and 2, 1.6 in Campus 3), the data also includes the variable of distance from the campus probation cutoff. Instead of the absolute GPA, the distance from cutoff variable will be used as the forcing variable in the regression discontinuity design which will be discussed in the following section.

2.3 Quasi-Experiment and About Bandwidth Selection

2.3.1 Regression Discontinuity Design

Instead of a randomized experiment, this study belongs to the Quasi-experiment because students' academic statues are not randomly assigned by the university administration. As the probation cutoff score is clearly defined and the GPAs affect students' academic statuses, the regression discontinuity design (RDD) is the method used to evaluate the probation impact. Therefore, the cutoff rule assigns students barely below the probation cutoff (On probation) into the treatment group and those barely above the probation cutoff (In good standing) into the comparison group to estimate impact on first-year university students after being placed on academic probation.

2.3.2 Concerns about Bandwidth Selection

In this study, since the RDD only provides the most accurate estimates around the probation cutoff where treatment and comparison are most similar (Gertler et al. 2016), it is critical to determine the bandwidth(i.e., students within certain grade points of the probation cutoff score) to maintain the balance of observed characteristics of students around the probation cutoff. In Lindo's original paper, they choose the bandwidth of 0.6 and the following table (Table 2) shows the summary statistics of student observable characteristics within 0.6 GPA.

Table 2: Summary Statistics of Students Within 0.6 GPA of Probation Cutoff

Description		In Good Standing	On Academic Probation
		N(%) = 8142 (65.1)	N(%) = 4365 (34.9)
Gender	Male	3000 (36.8)	1688 (38.7)
	Female	5142 (63.2)	2677 (61.3)
Birth Place	North America	7151 (87.8)	3734 (85.5)
	Asia	622 (7.6)	429 (9.8)
	Other	369 (4.5)	202 (4.6)
Study Campus	Campus 1	4088 (50.2)	1896 (43.4)
	Campus 2	1695 (20.8)	929 (21.3)
	Campus 3	2359 (29.0)	1540 (35.3)
First Language	English	5947 (73.4)	3041 (69.9)
	Other	2156 (26.6)	1310 (30.1)
		Mean (SD)	Mean (SD)
Highschool Grade Percentile		35.80 (23.55)	28.78 (22.03)
Credits Attempted in First Year		4.46 (0.52)	4.37 (0.55)
Age at Entry		18.71 (0.73)	18.74 (0.74)

According to Cohn et al. (2004), highschool grade is one of the key determinants of university academic success. Thus, highschool grades could somehow reveal students' study potential in the university, so it is preferable to balance student highschool performance to obtain a more accurate estimate of academic probation. However, the t-test (Table 3) conducted on highschool grade percentile suggests a large difference on highschool performance between student observations in good standing and on academic probation within

[4]This Canadian university requires at least 20 course credits to graduate.

Lindo's 0.6 GPA bandwidth. In addition, the study on the course load on academic success (Szafran 2001) implies that course credit taken by students could reflect student academic ability. Therefore, another t-test (Table 4) is conducted on total course credits attempted[5] in Year 1 between students in good standing and on probation. The result suggests a difference in the course workload between good standing and probation students within Lindo's 0.6 GPA bandwidth. As a result, the bandwidth that Lindo used in the original paper may include students with different academic potentials, reducing the local estimation accuracy of placing first-year students on academic probation.

Table 3: Two Sided T-test on Student Highschool Grade Percentile within 0.6 GPA cutoff

On Probation	In Good Standing	P-value	95% CI Low	95% CI High	Method
28.78	35.8	9.713e-61	-7.843	-6.183	Welch Two Sample t-test

Table 4: Two Sided T-test on Student Credits Attempted in Year 1 within 0.6 GPA cutoff

On Probation	In Good Standing	P-value	95% CI Low	95% CI High	Method
4.37	4.458	1.462e-18	-0.1086	-0.06907	Welch Two Sample t-test

The following table (Table 5) shows the grading theme of this Canadian university[6]. By applying Lindo's bandwidth of 0.6 GPA, observations barely above the probation cutoff roughly covers students within 60-66 average grade percentage (1.5-2.1 GPA in Campus 1 and 2, 1.6-2.2 GPA in Campus 3), and observations barely below the probation cutoff roughly covers students within 53-59 average grade percentage (0.9-1.5 GPA in Campus 1 and 2, 1.0-1.6 GPA in Campus 3). By converting the grade point value into grade percentage, it might become more obvious that there exists academic performance difference between students in good standing with 66 average grade percentage and students on probation with 53 average grade percentage. So I would argue that the bandwidth used in Lindo's is too wide to maintain similar characteristics, particularly academic potentials, between the probation and good standing students.

Table 5: Grading Scheme of the University

Grade	Grade Point Value	Grade Percentage	Definition
A+	4.0	90-100	Excellent
A	4.0	85-89	
A-	3.7	80-84	
B+	3.3	77-79	Good
B	3.0	73-76	
B-	2.7	70-72	
C+	2.3	67-69	Adequate
C	2.0	63-66	
C-	1.7	60-62	
D+	1.3	57-59	Marginal
D	1.0	53-56	
D-	0.7	50-52	
F	0.0	0-49	Inadequate; no credit obtained

As a result, I would adjust the bandwidth to:

1. Balance the observed characteristics between good standing and probation students within the bandwidth.

[5]Failing a course will not earn course credit but count as an attempted credit

[6]This is the current grading theme of this university as of 2021. The grading theme between 1996 and 2005 might be different and could not be retrieved on the university website. While the current grading theme may still serve as a reference of this study

2. Maintain a sufficient number of observations to obtain sufficient statistical power.

Table 6: Summary Statistics of Students Within 0.2 GPA of Probation Cutoff

Description		In Good Standing	On Academic Probation
		N(%) = 2361 (57.5)	N(%) = 1745 (42.5)
Gender	Male	875 (37.1)	636 (36.4)
	Female	1486 (62.9)	1109 (63.6)
Birth Place	North America	2061 (87.3)	1522 (87.2)
	Asia	191 (8.1)	141 (8.1)
	Other	109 (4.6)	82 (4.7)
Study Campus	Campus 1	1099 (46.5)	782 (44.8)
	Campus 2	524 (22.2)	361 (20.7)
	Campus 3	738 (31.3)	602 (34.5)
First Language	English	1720 (73.3)	1205 (69.4)
	Other	628 (26.7)	532 (30.6)
		Mean (SD)	Mean (SD)
Highschool Grade Percentile		32.46 (22.80)	30.82 (22.41)
Credits Attempted in First Year		4.40 (0.52)	4.40 (0.55)
Age at Entry		18.72 (0.74)	18.72 (0.75)

Since the probation cutoff GPA lies between "adequate" C- of 1.7 GPA and "marginal" D+ of 1.3 GPA, I adjust the bandwidth to be 0.2 to reduce the academic performance difference. Table 6 shows the summary statistics of 4106 student observable characteristics within 0.2 GPA. Even though the t-test (Table 7) on highschool grade percentile still suggests difference on highschool performance, the estimated highschool grade percentile difference is much less than those within 0.6 bandwidth. Besides, the t-test (Table 8) on course credit does not show a significant difference in course load within 0.2 bandwidth, so the observed characteristics become more balanced.

Table 7: Two Sided T-test on Student Highschool Grade Percentile within 0.2 GPA cutoff

On Probation	In Good Standing	P-value	95% CI Low	95% CI High	Method
30.82	32.46	0.0219	-3.032	-0.2369	Welch Two Sample t-test

Table 8: Two Sided T-test on Student Credits Attempted in Year 1 within 0.2 GPA cutoff

On Probation	In Good Standing	P-value	95% CI Low	95% CI High	Method
4.401	4.404	0.8692	-0.03591	0.03035	Welch Two Sample t-test

The power of a test is the probability that the test will reject the test null hypothesis to detect a correct alternative hypothesis (Greenland et al. 2016). In this study, as the length of the bandwidth increases, more students are included within the bandwidth. The increase in the number of student raises the power to detect the effect on placing students on probation so that the effect is less likely to happen by chance rather than by the probation cutoff rule. Therefore, I use the `rdpower`(Cattaneo, Titiunik, and Vazquez-Bare 2020) package to calculate the sufficient sample size required (i.e., effect size) to detect the probation effect on (1) Student dropout, (2) Second year GPA and (3) Graduation rate by Year 4. The result as shown below (Table 9) suggests that the bandwidth of 0.2 GPA with 4106 students (1745 on probation, 2361 in good standing) could raise sufficient statistical power. Further adjustment on the bandwidth might either unbalance student characteristics (Increase bandwidth) or decrease the power of the study (Decrease bandwidth). So I would apply the bandwidth of 0.2 GPA to examine the impact of academic probation in the following sessions.

Table 9: Sample Required to Raise Sufficient Power

Interest	Total Student	On Probation	In Goodstanding
Dropout	894	359	535
GPA in Year 2	1094	448	646
Graduation by Year 4	1375	625	750

[1] The significance level is 0.05 and desired power is 0.8.

3 Model

3.1 Test on Regression Discontinuity Design Assumptions

Before applying the regression discontinuity design in this study, it is important to test its assumptions to ensure the constructed RDD model accurately represent the impact of academic probation on first-year students. Several assumptions of RDD are:

1. The cutoff is clearly defined, unique (Gertler et al. 2016): This Canadian university clearly defined the academic probation rule (1.5 GPA in Campus 1 and 2, 1.6 GPA in Campus 3). There is also no other academic status (Other than in good standing and on probation) defined around the probation cutoff score, so this assumption is satisfied.
2. The score of the individual free of manipulation (Gertler et al. 2016): The university Code of Behavior on Academic Matters ensures that student GPA cannot be manipulated, so this assumption is satisfied.
3. The forcing variable should be continuous (Alexander 2021): as mentioned in Data Section 2.2, this study use student GPA distance from cutoff as the forcing variable. Since the GPA is a continuous variable, the GPA distance from cutoff is also continuous so this assumption is satisfied.
4. The treatment and comparison units are most similar (Gertler et al. 2016): as mentioned in Data Section 2.3.2, I would apply the adjusted bandwidth to balance the characteristics between students on probation barely below the cutoff and in good standing barely above the cutoff.

Since the academic probation cutoff rule is known, one may question that students barely below the probation cutoff would study harder to avoid being placed on academic probation. However, the survey conducted in Lindo's paper suggests that first year students are generally less familiar with university policies to know the GPA required to avoid academic probation. And Lindo also finds no evidence that students' grades are related to their knowledge about academic probation (Lindo, Sanders, and Oreopoulos 2010). So first year students are not likely to "play with" the university policy to avoid academic probation. The following figure (Figure 1) shows the distribution of the GPA distance from probation cutoff of all students. There is a smooth increase on the frequency distribution of the student GPA until reaching students' average GPA (GPA distance from probation around 1) and there is no obvious jump on either side of the probation cutoff. So students on either side of the probation cutoff is balanced as suggested in Data Section 2.3.2.

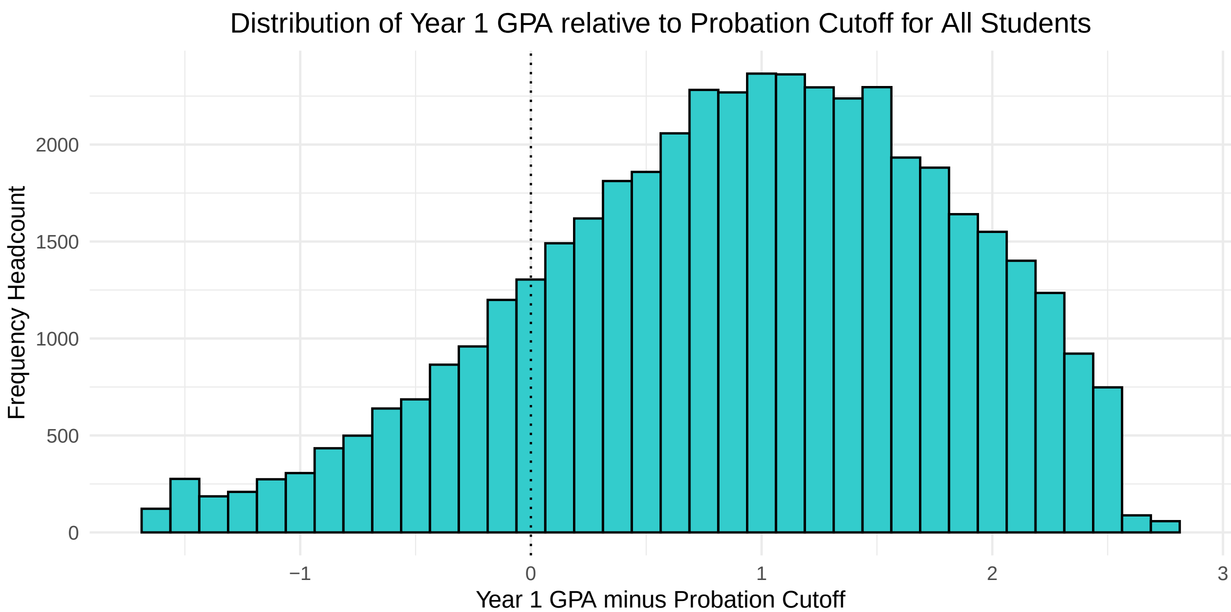

Figure 1: Distribution of Year 1 GPA relative to Probation Cutoff for All Students

3.2 Sharp RDD Model

3.2.1 Model Background

The university policy ensures that all first year students below the probation cutoff will be placed into academic probation[7]. So the discontinuity around the cutoff is "sharp".

By placing students on academic probation, the university informs these students an urge to improve their academic performance to prevent further escalating penalties such as suspension. According to *Bénabou* and Tirole's model of agents' responses to a performance standard (Bénabou and Tirole 2000), the probation policy as a negative incentive may result in the following outcomes:

1. Discourage some students to continue their study in the university, leading to an increasing chances of student dropout.
2. Encourage the rest of students to improve their GPA, leading to an improve of GPA in the second year.

Since students on probation status are allowed to take limited course credits[8], unless students actively seek to earn credits in subsequent academic year (e.g., enroll in summer course, taking more courses after going back to good standing), this policy would also:

3. Postpone the graduation, leading to an decreasing chances of graduating within certain time periods.

In addition, previous studies have suggested that:

- There is gender difference in academic performance (Dayioğlu and Türüt-Aşik 2007) and females appears to be more vulnerable to internal distress than males (Pomerantz, Altermatt, and Saxon 2002), so students with different gender may response quite differently after being placed on probation.

[7]The original data shows university administrative errors in data reporting that falsely place students either above the cutoff on probation or below the cutoff in good standing (98 out of 44362 student observations). These dirty data are removed from this study.

[8]This university has the up-to-date policy regarding to credit restrictions on probation students. However, this policy varies across campuses and the rule during the data collected period (1996-2005) could not be retrieved from the university website.

- Students with higher highschool grade are more likely to succeed in the university (Cohn et al. 2004). Even though some students with satisfactory highschool performance are unfortunately placed on academic probation (Possibly due to lack of university adaptation), they generally hold higher study potential than those with relatively poor highschool performance, triggering faster recovery from academic probation. So students with different highschool performance may response differently towards probation.
- The level of English language proficiency is important to academic success (Graham 1987). Students whose first language are not English may hold different level of English proficiency than English native speaking students. Similar to students with high highschool performance, English native speaking students may recover faster from academic probation. So students with different first language may also response differently towards probation.

3.2.2 Model Construction

Taking all of the above into consideration, I propose the following model to evaluate the impact on first year students after being placed on academic probation:

$$Y_i = \beta_0 Proation_i + f(Cutoff\ Distance_i) + \alpha_1 Gender_i + \alpha_2 HighSchool_i + \alpha_3 Language_i + \epsilon_i$$

where

- Y_i is first year student outcomes of interest for student i, which are (1) Probability of dropout after the first academic year, (2) Annual GPA in the second academic year, and (3) Probability of graduating by the fourth year after entering the university.
- $Probation_i$ is an indicator variable to show whether the student i is placed on academic probation or not in the first year. And its coefficient β_0 is the estimate of impact on first year students after being placed on academic probation that we are primarily interested in.
- $f(CutoffDistance_i)$ is the function to represent the student i's first year GPA distance from the cutoff. Since the first order polynomial is used during the power analysis to determine required student samples (Data Section 2.3.2), it will also be applied in building the sharp RDD model[9].
- $Gender_i$ is an indicator variable to show whether the student i is male or female. Its coefficient α_1 is the estimate of gender difference in response to academic probation.
- $HighSchool_i$ is an indicator variable to show whether the student i has a satisfactory highschool performance (Highschool Grade Percentile > 50) or a relatively poor highschool performance (Highschool Grade Percentile ≤ 50). Its coefficient α_2 is the estimate of highschool performance difference in response to academic probation.
- $Language_i$ is an indicator variable to show whether the student i's first language is English or other language. Its coefficient α_3 is the estimate of first language difference in response to academic probation.

Since the dropout decision of an individual student is whether leaving the university or not, it is a dichotomous variable. Therefore, I would use the logistic regression to estimate the probation impact on probability of student dropout. Similarly, the graduation decision of whether graduating by Year 4 or not is also a dichotomous variable. So I would also use the logistic regression to estimate the probation impact on probability of student graduating by Year 4.

As the GPA is an continuous variable, I would use the linear regression to estimate the probation impact on student annual GPA in the second academic year.

[9] Adjusting the order of distance from cutoff function does not affect the results. Lindo's original paper also used first order polynomial on distance from the cutoff to estimate probation outcomes

4 Results

Surprisingly, after restricting student samples within 0.2 probation cutoff GPA, the model results do not suggest significant academic probation impact on any student outcomes of interest (dropout, Year 2 GPA and graduation). This is inconsistent with Lindo's findings that there is significant academic probation impact on all outcomes of interest within 0.6 probation cutoff GPA.

4.1 Impact on Dropout

As mentioned in Model Section 3.2.2, logistic regression is used to estimate the probation impact on probability of student dropout.

$$Prob(Dropout_i) = \beta_0 Proation_i + f(Cutoff\ Distance_i) + \alpha_1 Gender_i + \alpha_2 HighSchool_i + \alpha_3 Language_i + \epsilon_i$$

In this model, the baseline characteristics of the first year students is (1) on academic probation, (2) GPA distance from the probation cutoff is 0, (3) female, (4) highschool grade below the average, (5) first language is English. The following table (Table 10) shows regression results of student samples within different GPA bandwidth (0.2 GPA, 0.3 GPA, 0.6 GPA, respectively). Within 0.2 GPA bandwidth, compared to students on academic probation, students in good standing have 0.711 times odds to leave the university, i.e., 28.9% decrease in the odds of leaving the university, while holding other baseline characteristics constant. **However**, the p-value of 0.223, which is greater than the widely defined 0.05 threshold, suggests that the first year dropout decision does not significantly differ between good standing students and probation students, holding other baseline characteristics constant. And this difference could also not be clearly detected around the probation cutoff as shown below (Figure 2). While within 0.6 GPA bandwidth used in Lindo's original paper, the results do suggest significant difference that students in good standing have 34.8% decrease in the odds of first year dropout, holding other baseline characteristics constant. Such difference is also significant within 0.3 GPA bandwidth.

Within 0.2 GPA bandwidth, there is significant difference on student dropout decision across first language: compared to English native speaking students, students whose first language are other languages have 54.5% decrease in the odds of first year dropout, holding other baseline characteristics constant. No other significant difference is found on student gender and highschool performance.

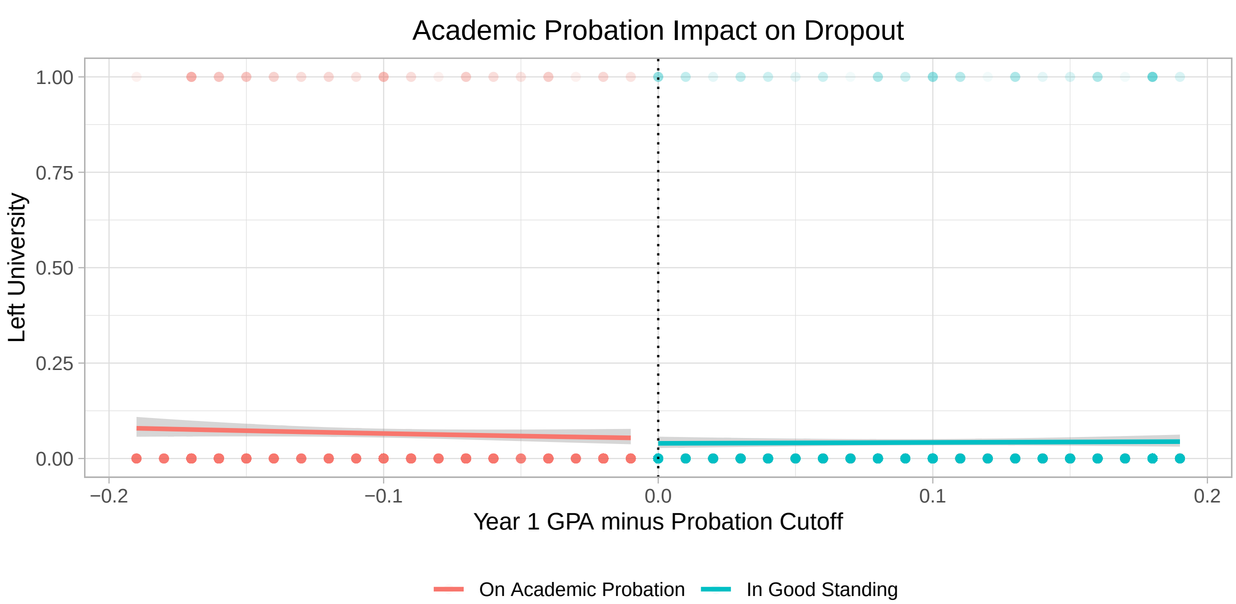

Figure 2: Academic Probation Impact on Dropout

Table 10: Logistic Regression for the Impact of Academic Probation on Student Dropout

	0.2 GPA Bandwidth	0.3 GPA Bandwidth	0.6 GPA Bandwidth
Constant	0.080***	0.079***	0.075***
	0.176 (0.000)	0.145 (0.000)	0.101 (0.000)
In Good Standing	0.711	0.627**	0.652***
	0.280 (0.223)	0.236 (0.048)	0.164 (0.009)
f(Cutoff Distance)	0.418	0.760	0.707
	1.240 (0.481)	0.702 (0.696)	0.244 (0.156)
Gender Male	1.085	1.149	1.124
	0.145 (0.573)	0.120 (0.248)	0.085 (0.167)
Highschool Above Average	0.781	0.790	0.969
	0.186 (0.183)	0.154 (0.127)	0.101 (0.759)
First Language Not English	0.455***	0.498***	0.581***
	0.191 (0.000)	0.155 (0.000)	0.105 (0.000)
Num.Obs.	4085	6101	12454
AIC	1649.9	2385.7	4817.4
Log.Lik.	-818.948	-1186.829	-2402.720

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[1] Odds Ratio in First Row. Standard Error of Log of Odds in Second Row.
P-value in Parentheses

4.2 Impact on Second Year Annual GPA

As mentioned in Model Section 3.2.2, linear regression is used to estimate the probation impact on student second year annual GPA.

$$GPAYear2_i = \beta_0 Proation_i + f(Cutoff\ Distance_i) + \alpha_1 Gender_i + \alpha_2 HighSchool_i + \alpha_3 Language_i + \epsilon_i$$

The model baseline characteristics are the same as mentioned in dropout Section 4.1. The following table (Table 11) shows regression results of student samples within different GPA bandwidth (0.2 GPA, 0.3 GPA, 0.6 GPA, respectively). Within 0.2 GPA bandwidth, compared to students on academic probation, students in good standing earn 0.055 less GPA than those on academic probation in the second year, while holding other baseline characteristics constant. **However**, the p-value of 0.245 suggests that the first year dropout decision does not significantly differ between good standing students and probation students, holding other baseline characteristics constant. And this difference could also not be clearly detected around the probation cutoff as shown below (Figure 3). While within 0.6 GPA bandwidth used in Lindo's original paper, the results do suggest significant difference that students in good standing earn 0.145 GPA less than students on probation, holding other baseline characteristics constant. Such difference is also significant within 0.3 GPA bandwidth.

Within 0.2 GPA bandwidth, there is significant difference on student second year GPA across gender and highschool performance. Compared to female students, male students earn 0.120 less annual GPA in the second year, holding other baseline characteristics constant. Compared to students with relative poor highschool academic performance, students with satisfactory academic performance earn 0.099 more annual GPA in the second year, holding other baseline characteristics constant. No other significant difference is found on student first language.

Table 11: Linear Regression for the Impact of Academic Probation on Student Year 2 GPA

	0.2 GPA Bandwidth	0.3 GPA Bandwidth	0.6 GPA Bandwidth
Constant	1.987***	2.013***	2.046***
	0.031 (0.000)	0.025 (0.000)	0.017 (0.000)
In Good Standing	-0.055	-0.120***	-0.145***
	0.047 (0.245)	0.039 (0.002)	0.027 (0.000)
f(Cutoff Distance)	0.021	0.446***	0.553***
	0.206 (0.918)	0.114 (0.000)	0.039 (0.000)
Gender Male	-0.120***	-0.104***	-0.136***
	0.025 (0.000)	0.020 (0.000)	0.014 (0.000)
Highschool Above Average	0.099***	0.112***	0.097***
	0.029 (0.001)	0.023 (0.000)	0.015 (0.000)
First Language Not English	-0.047*	-0.015	-0.038**
	0.026 (0.077)	0.022 (0.473)	0.015 (0.011)
Num.Obs.	3408	5118	10506
R2	0.013	0.014	0.049
R2 Adj.	0.011	0.013	0.048
AIC	7208.9	10799.1	21659.2
Log.Lik.	-3597.430	-5392.557	-10822.610
F	8.677	13.997	107.865

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[1] Regression Coefficient in First Row. Standard Error in Second Row. P-value in Parentheses

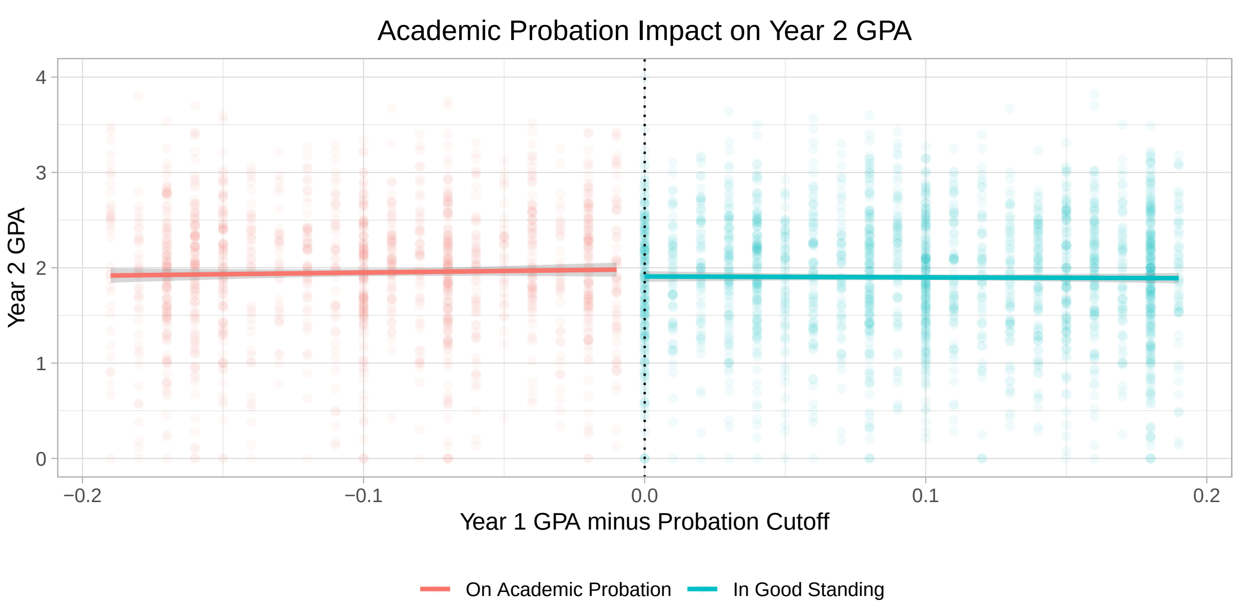

Figure 3: Academic Probation Impact on Year 2 GPA

4.3 Impact on Graduation

As mentioned in Model Section 3.2.2, logistic regression is used to estimate the probation impact on probability of student graduating by Year 4.

$$Prob(Graduation_i) = \beta_0 Proation_i + f(CutoffDistance_i) + \alpha_1 Gender_i + \alpha_2 HighSchool_i + \alpha_3 Language_i + \epsilon_i$$

The model baseline characteristics are the same as mentioned in dropout Section 4.1. As shown in the following table (Table 12), within 0.2 GPA bandwidth, compared to students on academic probation, students in good standing have 26.8% increase in the odds of graduating by Year 4, while holding other baseline characteristics constant. **However**, the p-value of 0.161 suggests that the graduation does not significantly differ between good standing students and probation students, holding other baseline characteristics constant. And this difference could also not be clearly detected around the probation cutoff as shown(Figure 4). Within 0.3 GPA and 0.6 GPA bandwidth, the results also do not suggest difference on graduation between good standing students and probation students at 0.05 significance level, holding other baseline characteristics constant.

Within 0.2 GPA bandwidth, there is significant difference on student graduation across gender: compared to female students, male students have 49.9% decrease in the odds of graduating by Year 4, holding other baseline characteristics constant. No other significant difference is found on student highschool performance and first language.

Table 12: Logistic Regression for the Impact of Academic Probation on Student Graduation by Year 4

	0.2 GPA Bandwidth	0.3 GPA Bandwidth	0.6 GPA Bandwidth
Constant	0.386***	0.363***	0.357***
	0.110 (0.000)	0.091 (0.000)	0.065 (0.000)
In Good Standing	1.268	1.267*	1.195*
	0.169 (0.161)	0.140 (0.091)	0.099 (0.073)
f(Cutoff Distance)	1.969	2.736**	3.432***
	0.737 (0.358)	0.414 (0.015)	0.144 (0.000)
Gender Male	0.501***	0.540***	0.515***
	0.092 (0.000)	0.075 (0.000)	0.052 (0.000)
Highschool Above Average	1.086	1.061	1.053
	0.104 (0.431)	0.085 (0.488)	0.057 (0.368)
First Language Not English	1.063	1.117	1.141**
	0.095 (0.518)	0.078 (0.153)	0.054 (0.015)
Num.Obs.	2899	4307	8766
AIC	3307.3	4887.6	9984.9
Log.Lik.	-1647.655	-2437.810	-4986.436

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[1] Odds Ratio in First Row. Standard Error of Log of Odds in Second Row. P-value in Parentheses

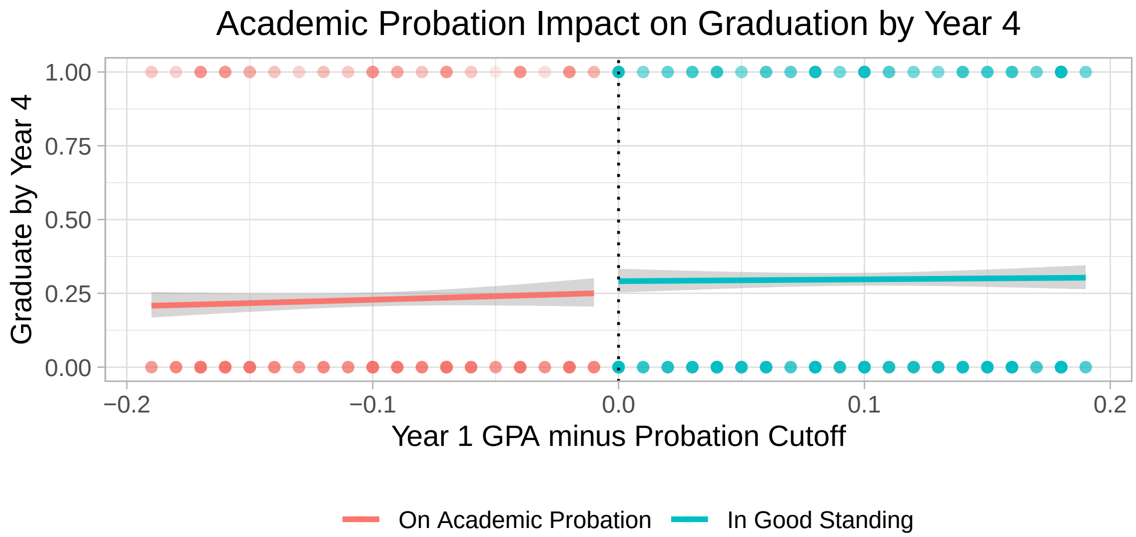

Figure 4: Academic Probation Impact on Graduation by Year 4

5 Discussion

5.1 P-value and RDD assumptions

Why does using student samples within the bandwidth of 0.6 GPA and even 0.3 GPA "has evidence" to suggest significant academic probation impact on student outcomes of interest (probability of immediate dropout, Year 2 annual GPA, probability of graduating by Year 4), while using student samples within the bandwidth of 0.2 GPA does not?

Since all of the result interpretations are based on p-values, reassessing the definition of p-value may help to explain the above inconsistent results. In simple common English, P-value means the probability that something checked wrong to happen. The lower p-value means a higher probability of something checked right to be true, and "is there an academic probation impact on first year student outcome" is the checkpoint in this study.

However, the p-value does not tell if the underlying assumptions of the checkpoint is correct or not. In the paper `Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations`, Greenland defines the p-value in a general view "as a statistical summary of the compatibility between the observed data and what we would predict or expect to see if we knew the entire statistical model (all the assumptions used to compute the P value) were correct"(Greenland et al. 2016). Therefore, reviewing RDD rationales and assumptions helps to draw correct interpretations from p-values.

In this study, the regression discontinuity design estimates academic probation impact on first year students by calculating the difference between treatment students on probation barely below the cutoff and comparison students in good standing barely above the cutoff. Its estimate is the most accurate where the treatment and comparison students are most similar and this estimate may not be generalized to students whose first year GPA are further away from the probation cutoff (Gertler et al. 2016).

According to the student summary statistics (Table 2) and t-test conducted on students highschool performance (Table 3) and credit attempted in the first year (Table 4) in Data Section 2.3.2, within the bandwidth of 0.6 GPA, students barely below the cutoff generally hold weaker academic potentials than those barely above the cutoff. And these academic differences also serves as key factors affecting student outcomes.

Even though Lindos's original paper did test whether students share similar observable characteristics around the probation cutoff GPA in Results Section A (Lindo, Sanders, and Oreopoulos 2010), it failed to access students characteristics that are further away from the cutoff. It is unlikely that a student with 1.0 GPA is as similar as one with 2.0 GPA, while they are all included within Lindo's bandwidth to estimate the impact of academic probation. As a result, using a wider bandwidth could violate the similarity assumption of RDD, and interpreting p-values from assumption violated models to seek "statistical significance" may become senseless.

5.2 Causal inference

So does placing first year students on academic probation affect student outcomes? By applying a narrower bandwidth of 0.2 GPA, student characteristics become more balanced (Table 6). Although the t-test still suggests a difference on highschool performance between treatment and comparison students (Table 3), model results in Result Section 4 has already suggestted

5.3 Bias

Within Province Data

student family background https://www.tandfonline.com/doi/abs/10.1080/15235882.1997.10162712

5.3.1 Evaluations minimum credits and NCR, late withdraw Policy (Refer I institutional Background footnote 7)

5.4 Casual Inference

5.5 External Validity and Internal Validity

5.6 Simpson's Paradox about Campus Variation

5.7 Weaknesses and next steps

Weaknesses and next steps (McGrath and Burd 2012)should also be included.

A Appendix 1 Letter Sent to Students at Campus 2

Dear < first name >:

Your academic record indicates that you are experiencing challenges with your studies at xxxxxxxxxxx. As a result, you have been placed "On Probation" at the end of the xxxxxxx session. "On Probation" is an academic status applied to a student if he or she:

1. Is having difficulty achieving a term average of at least 1.7 GPA or a yearly average of 1.5 CGPA.
2. Is having difficulty meeting performance expectations and/or deadlines as outlined by the course instructor.
3. Is having difficulty achieving the minimum grades required for graduation.

A student who at the end of any session during which they are on probation has a cumulative GPA of less than 1.5 and a sessional of less than 1.7 shall be suspended. Therefore, it is imperative that you seek assistance to improve your academic standing to avoid further sanction.

Rest assured that you can improve this status and that xxxxxxxxxxxx offers assistance at many junctions. First, you can access help by making an appointment with an academic advisor in the Office of the Registrar to develop strategies to improve your academic record. Book an appointment at xxx-xxx-xxxx or online at www.xxxxxxxxx. Second, contact xxxxxxxx for assistance with study habits, note taking, effective research, time management, study groups, and peer mentors. Finally, the xxxxxxxxx offers skills and interest testing which can help you focus on your strengths.

We know that you are capable of academic success, based on your academic record at admission. A good academic record is essential for entry to Limited Enrolment programs, graduate school, and professional schools. Let us review your goals and help you develop a plan to achieve them.

You have the opportunity and available support to be successful. Please utilize our services to insure your future success.

For further information on academic status, please refer to xxxxxxxx of the Academic Calendar or here: http://www.xxxxxxxxxxxxxx.

B Appendix 2 Regression Diag

References

Alexander, Rohan. 2021. *Causality from Observational Data*. https://www.tellingstorieswithdata.com/causality-from-observational-data.html#regression-discontinuity-design.

Arel-Bundock, Vincent. 2021. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. https://CRAN.R-project.org/package=modelsummary.

Bénabou, Roland, and Jean Tirole. 2000. "Self-Confidence and Social Interactions." National bureau of economic research.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocio Titiunik. 2021. *Rdrobust: Robust Data-Driven Statistical Inference in Regression-Discontinuity Designs*. https://CRAN.R-project.org/package=rdrobust.

Cattaneo, Matias D., Rocio Titiunik, and Gonzalo Vazquez-Bare. 2020. *Rdpower: Power Calculations for Rd Designs*. https://CRAN.R-project.org/package=rdpower.

Cohn, Elchanan, Sharon Cohn, Donald C Balch, and James Bradley Jr. 2004. "Determinants of Undergraduate Gpas: SAT Scores, High-School Gpa and High-School Rank." *Economics of Education Review* 23 (6): 577–86.

Dayioğlu, Meltem, and Serap Türüt-Aşik. 2007. "Gender Differences in Academic Performance in a Large Public University in Turkey." *Higher Education* 53 (2): 255–77.

Dimmery, Drew. 2016. *Rdd: Regression Discontinuity Estimation*. https://CRAN.R-project.org/package=rdd.

Gertler, Paul J, Sebastian Martinez, Patrick Premand, Laura B Rawlings, and Christel MJ Vermeersch. 2016. *Impact Evaluation in Practice*. The World Bank.

Graham, Janet G. 1987. "English Language Proficiency and the Prediction of Academic Success." *TESOL Quarterly* 21 (3): 505–21.

Greenland, Sander, Stephen J Senn, Kenneth J Rothman, John B Carlin, Charles Poole, Steven N Goodman, and Douglas G Altman. 2016. "Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations." *European Journal of Epidemiology* 31 (4): 337–50.

Harrison, Ewen, Tom Drake, and Riinu Ots. 2020. *Finalfit: Quickly Create Elegant Regression Results Tables and Plots When Modelling*. https://CRAN.R-project.org/package=finalfit.

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

McGrath, Shelley M, and Gail D Burd. 2012. "A Success Course for Freshmen on Academic Probation: Persistence and Graduation Outcomes." *NACADA Journal* 32 (1): 43–52.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Pomerantz, Eva M, Ellen Rydell Altermatt, and Jill L Saxon. 2002. "Making the Grade but Feeling Distressed: Gender Differences in Academic Performance and Internal Distress." *Journal of Educational Psychology* 94 (2): 396.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Robinson, David, Alex Hayes, and Simon Couch. 2020. *Broom: Convert Statistical Objects into Tidy Tibbles*. https://CRAN.R-project.org/package=broom.

Szafran, Robert F. 2001. "The Effect of Academic Load on Success for New College Students: Is Lighter Better?" *Research in Higher Education* 42 (1): 27–50.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Evan Miller. 2020. *Haven: Import and Export 'Spss', 'Stata' and 'Sas' Files.* https://CRAN.R-project.org/package=haven.

Xie, Yihui. 2016. *Bookdown: Authoring Books and Technical Documents with R Markdown.* https://github.com/rstudio/bookdown.

Zhu, Hao. 2020. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.