Rebates as Incentives:

The Effects of a Gym Membership Reimbursement Program

By TATIANA HOMONOFF, BARTON WILLAGE AND ALEXANDER WILLÉN*

Can financial incentives induce healthy behaviors? We examine this question by evaluating a large-scale wellness program at a major American university. The program offers gym membership reimbursements for students who attend the gym at least 50 times in a six-month period. Our analysis exploits individual-level administrative data on daily gym attendance for the universe of students over a five-year period: the three years that the policy was in place, one year before implementation, and one year after termination. This provides us with 100,000 student-year observations and 1.5 million gym visits. Using a combination of bunching methods and difference-in-difference strategies, we provide four empirical results. First, we document significant bunching at the 50-visit threshold in years when the policy is in place. Second, we show that this effect translates into a statistically significant and economically meaningful increase in gym attendance: the program increased average gym visits by almost five visits per semester, a 20% increase from the mean. Third, we show that the policy not only motivated students who were previously near the threshold, but that it increased attendance across the entire visit distribution. Finally, we show that approximately 50% of the effect persists after program termination, providing strong evidence of habit formation. Taken together, these results suggest that rebate-framed incentives with a high attendance threshold can successfully induce healthy behaviors in the short-term, and also create new habits in the long-run.

1. Introduction

In the United States, less than 5 percent of adults engage in the recommended level of daily physical activity (Troiano et al. 2008). The high level of inactivity has raised medical costs, lowered labor productivity and reduced well-being (Jones et al. 2018). These costs are often born by third parties such as employers, health insurers and state governments. In response to the rising costs of physical inactivity, these organizations have initiated a number of different wellness incentive programs to encourage physical activity, with financial subsidies and rebates being the most common ones (NBGH 2011, Reis 2012).

In this paper, we examine how financial incentives affect individual gym attendance, exploiting the introduction and subsequent discontinuation of a large-scale wellness program at a major American university. Conditional on obtaining health insurance through the Student Health Plan (SHP), the program offers full reimbursements of university fitness membership fees ($75) for students who attend the gym at least 50 times during the semester. The SHP requirement means that the program disproportionately benefits graduate students at the university: while almost all of the university's graduate students have SHP, far fewer undergraduate students rely on this form of health insurance. Both the rebate-framing nature of this incentive program and the high attendance threshold closely resemble many recent fitness programs implemented by US health insurers, state governments, and higher education institutions, making this a particularly interesting program to evaluate.[1]

To perform our analysis, we exploit individual-level administrative data on gym memberships and daily gym attendance for the universe of students over a five-year period: the three years that the policy was in place, one year before implementation, and one year after termination. Our data includes 100,000 student-year observations and more than 1.5 million gym visits. Not only do we have access to a much larger sample and longer time frame than the traditional field experiments on this topic, but the fact that we are not actively recruiting participants also means that we are examining a group of individuals more relevant to policymakers.

We begin by documenting significant bunching at the 50-visit threshold among graduate students in years when the policy is in place, consistent with a rational response to a non-linear incentive scheme with a large change in incentives at 50 visits. Specifically, using the nonparametric bunching method developed in Chetty et al. (2011), we find a large and statistically significant excess mass right above the 50-visit threshold. We find no evidence of bunching among graduate students in the year

---

[1] Part of this may be in response to the Affordable Care Act, which raised the cap on employer-provided, health-contingent incentives from 20 to 30 percent of the cost of insurance coverage. Health insurance companies offering this type of fitness incentive program include Blue Cross Blue Shield, Aetna and UnitedHealthcare; state governments include Maine, New York and New Hampshire; higher education institutions include University of Minnesota, Binghamton University and University of North Dakota. See Appendix Table 1 for links to the relevant webpages that contain information on each of these programs.

before the policy took effect nor in the year after the policy was discontinued. In addition, we do not observe bunching for undergraduate students, a group that is largely ineligible for the reimbursement.

The bunching estimator relies on strong assumptions regarding the exclusion window, functional form, and counterfactual distribution (e.g. Dekker et al. 2016; Blomquist and Newey 2017; Aronsson et al. 2017; Marx 2018; Bertanha 2018). To ensure that our results are not driven by the assumptions underlying this estimator, we take advantage of two unique features of our data to estimate the effect of the program using a difference-in-difference design – an estimation strategy that relies on an entirely different set of assumptions. This analysis exploits the fact that we have both pre- and post-policy data as well as a population with almost universal exposure to the policy (graduate students) and a population with very limited exposure (undergraduate students). Results from this analysis reveal that the policy led to a 4 percentage point increase in the likelihood of just crossing the 50-visit threshold. We also find that the introduction of the policy increased overall gym attendance by almost 5 visits per semester, a 20% increase from the mean. We find no effects of the policy on the extensive margin (i.e., becoming a gym member).

After identifying a large effect of the policy on gym attendance across all eligible students, we exploit the panel structure of our data to investigate whether these effects are driven by students who were low- or high-frequency gym users prior to the policy's introduction. We find the largest effects on just crossing the reimbursement threshold among students who were just below the threshold in the pre-period (a 6 percentage point increase). However, we also find significant bunching among students who attended the gym less than 10 times in the year prior to the policy (a 3 percentage point increase). This suggests that even though the threshold for reimbursement was quite high, it still motivated students at the bottom of the attendance distribution. In addition, we see no evidence of bunching among students who attended the gym more than 50 times in the pre-period, but large positive effects on their overall attendance. Thus, the threshold did not discourage gym attendance at the right-tail of the gym visit distribution by providing a reference point that may have been lower than what these students would have chosen absent of the policy.

In our final analysis, we take advantage of the unexpected discontinuation of the program in 2017 to study habit formation. Using a difference-in-difference framework similar to our main specification, we compare the post-policy/pre-policy difference in gym attendance of graduate students with that of undergraduate students. We find that the average gym attendance was 2 visits higher in the post-policy period relative to the pre-policy period among graduate students compared to undergraduate students. This finding suggests that roughly half of the program effect persisted in the year after the policy was terminated, suggesting a higher degree of habit formation than observed in prior studies (Acland and Levy 2012; Royer et al. 2015). Given the high attendance requirement under our policy, our findings are consistent with the theoretical model in Carrera et al. (2017), which suggests that crossing a habit "threshold" contributes to generating sustained effects on gym attendance.

This paper makes several important contributions to the literature. First, while a number of papers have evaluated the effectiveness of incentives for gym attendance (Gneezy and Charness 2009; Acland and Levy 2012; Royer et al. 2015; Carrera et al. 2017; Cappelen et al. 2017; Carrera et al. 2018), this literature largely relies on results from field experiments. While these experiments have several benefits, namely that they allow the authors to cleanly test the relative effectiveness of different incentive designs, they also suffer from a few shortcomings. In particular, these studies are often small in scale with sample sizes as low as a few hundred participants, the incentives provided are often short-lived with incentive periods typically ranging from 4 to 16 weeks, and the sample population must be actively recruited to participate in the study, raising questions about external validity. In contrast, our analysis is based on a large-scale natural experiment involving almost 100,000 students in which incentives were available for several years.

Second, to our knowledge, this is the first paper to explore the effects of a gym incentive program framed as a rebate. This type of program closely models what many institutions and employers have implemented in recent years, both in design (membership reimbursement conditional on attendance) and the attendance threshold level (50 visits in six months), making our results particularly informative about the effectiveness of existing programs. Additionally, while other studies have examined the effect of providing free memberships on attendance (Cappelen et al. 2017; Carrera et al. 2017), the loss-framing nature of the reimbursement program may be more effective at encouraging attendance (Kahneman and Tversky 1979; Levitt et al. 2016; Field 2009; Hossain and List 2010; Rees-Jones 2018).[2] This incentive scheme may also be more cost-effective since payments are only made to individuals who meet the attendance requirements. We also show that incentive programs with a very high threshold for reimbursement can increase attendance even among low-frequency gym goers.

Third, our paper contributes to a large literature evaluating the benefits of workplace wellness programs. While a meta-analysis of the effectiveness of these wellness programs shows substantial cost savings in the form of reduced medical costs and worker absenteeism (Baicker et al. 2010), a recent study by Jones et al. (2018) finds no effects of a large university's wellness program on health expenditures and health behaviors (including gym attendance). One key difference between the wellness program studied in Jones et al. (2018) and the program studied in our setting, is that their program was much more comprehensive, providing financial incentives for a wide variety of wellness activities. Our results suggest that programs that target a specific activity, such as gym attendance, may be more successful at changing behaviors. Additionally, recent results show that increased gym

---

[2] Fricke et al. (2017) presents results of a field experiment at a Swedish university in which gym incentives were designed explicitly to evoke loss aversion by paying students upfront, then deducting payments if they failed to go to the gym twice per week. This stands in contrast to earlier studies that frame the incentives as gains by only paying participants once they have attended the gym (Gneezy and Charness 2009; Acland and Levy 2012; Royer et al. 2016).

attendance leads to improved academic performance (Cappelen et al. 2017), suggesting that gym attendance may be a particularly important activity to target, especially for a student population.

Lastly, our results contribute to a debate on the existence of habit formation in gym attendance. While some studies in the gym attendance literature find evidence of habit formation (Gneezy and Charness 2009), others show that effects fade shortly after the incentives are removed (Acland and Levy 2012; Royer et al. 2015). We find that roughly half of the treatment effect persists in the year after the program was discontinued, providing evidence of strong habit formation. This is particularly encouraging since prior studies that showed persistent effects of gym incentive programs only considered a follow-up period of a few months (Gneezy and Charness 2009; Acland and Levy 2012).[3] The results from our analysis suggest that periodic incentives may provide a cost-efficient alternative to permanent financial subsidies, consistent with findings from Carrera et al. (2017).

This rest of this paper is organized as follows: Section 2 provides institutional background and economic intuition. Section 3 introduces our data and empirical strategy. Section 4 presents the main results on bunching and gym attendance. Section 5 investigates which types of students responded to the policy. Section 6 presents results on habit formation. Section 7 concludes.

2. Institutional Background and Economic Intuition
i. SHP Membership Reimbursement Program

In 2014, a major American university launched a gym reimbursement program to incentivize physical activity among students. This initiative emerged from a collaboration between the university's fitness facilities and the SHP provider. The program's stated objective was to promote healthy behaviors and help enhance student well-being. The university decided to discontinue the program in 2017, three years after its inception.

The gym reimbursement initiative was a bi-annual rebate program, operating both in the Spring/Summer (March through August) and in the Fall/Winter (September through February). Program participants were eligible for a reimbursement of 50% of the annual gym membership ($75) conditional on attending the gym 50 times during one of the two reimbursement periods, approximately 2 gym visits per week. Participants who attended the gym 50 times in both reimbursement periods received a reimbursement of 100% of their annual gym membership ($150). All students were able to track their daily gym attendance in the current membership period on the university's fitness website by using their student login (Appendix Figure 1).

---

[3] One exception is the study by Royer et al. (2015) which follows participants for three years after the removal of the incentives. The authors find that participants who were offered financial incentives to attend the gym in addition to the ability to contribute to a voluntary commitment device showed persistent effects on gym attendance, though the effects were roughly one quarter of the increase in attendance while receiving incentives. In contrast, an incentive-only group showed no evidence of habit formation after the first few post-intervention months.

Students with SHP were automatically enrolled in the program, while non-SHP students were ineligible to participate. This eligibility requirement was imposed because the SHP provider – and not the university – financed the initiative and reimbursements. In practice, this meant that the program disproportionately benefited graduate students at the university: while more than 95% of the university's graduate students have SHP, less than 30% of undergraduate students rely on this form of health insurance.

To receive the reimbursement, a student had to submit a simple form to the Office of Student Health Benefits (SHB) that included her name, student ID, and a statement certifying that she had attended the gym at least 50 times in the six-month period. To verify that the student had met the required visit threshold, SHB used the student's ID to retrieve data on gym attendance for the semester from the Fitness Center Database, which tracks daily gym attendance for all students. Following the verification process, a check was sent to the student for the reimbursement amount.

ii. Economic Intuition and Predictions

While there exists a rich literature examining various wellness and fitness incentive programs, the design of the incentive considered in this paper differs from those previously evaluated. Earlier studies commonly provide participants free gym memberships (Royer et al. 2015; Cappelen et al. 2017) or pay study participants for each gym visit (Gneezy and Charness 2009; Acland and Levy 2012; Royer et al. 2015). In contrast, individuals in our study must pay for their membership upfront and only receive the incentive after they have met the 50-visit attendance requirement. Fricke et al. (2017) study a gym incentive program that most closely resembles the loss-framed incentives considered in our study; however, in their study, participants are still rewarded for every gym visit rather than conditioning payment on meeting an attendance threshold as in our setting. As a result, the behavioral response to our policy might be quite different from responses to gym incentives previously documented in the literature.

We first investigate whether the rebate program led to bunching at the 50-visit threshold. Since eligible students must attend the gym 50 times in a six-month period to obtain the reimbursement, the rebate program incentivizes graduate students to attend the gym 50 times, but provides no additional incentive for attending the gym more than 50 times. Thus, one likely implication of the program is that the fraction of students who attend the gym exactly 50 times increases.

The second question we examine is whether the reimbursement program had a positive impact on average gym attendance. This effect may be small or large even in the presence of significant bunching at the 50-visit threshold. For example, if the policy only induces a behavioral response from students who would have been very close to the policy threshold in the absence of the rebate, we would observe bunching at the threshold, but only small increases in overall attendance. In contrast, if low-

attendance students also are incentivized by the policy, the effects on average gym attendance may be quite large, with the potential for extensive margin effects (i.e., increases in gym memberships) as well.

A third, related question, asks who responds to the incentive. We analyze heterogeneous treatment effects across the pre-reform gym attendance distribution. While individuals close to the 50-visit threshold only need to increase attendance by a few visits, those far below the threshold will have to greatly increase their gym attendance to reap the benefits of the incentive program. We predict an increase in gym attendance for students who would have gone to the gym less than 50 times in the absence of the program. The predictions for students who would have attended the gym more than 50 times are less clear. Specifically, individuals who typically attend the gym more than 50 times may interpret the 50-count threshold as a reference point or a sign of what constitutes a healthy amount of exercise, and adjust their gym behavior downwards. Because of this possibility, we also investigate if high-frequency gym users decrease their gym attendance down to the 50-visit threshold.

The final question we investigate is whether the rebate program led to a change in gym behavior that persists even after the termination of the program. The rationale underlying this question is that the potential increase in gym attendance among graduate students may cause these individuals to form lasting habits. If graduate students develop exercise habits, their post-policy gym count would be higher than if they had never been exposed to the rebate program. The theoretical model in Carrera et al. (2017) suggests that habit formation might be particularly likely in our setting, since the high attendance requirement to receive the reimbursement requires that students develop a sufficient "habit stock" which they argue is necessary for sustained increases in physical activity. However, even if the incentive program leads to habit formation that persists after program elimination, we would not necessarily expect to observe bunching at the 50-visit since the financial incentive for reaching that particular threshold is removed.

3. Data and Empirical Methodology
i. Data

To evaluate the effect of the reimbursement program on gym attendance, we rely on individual-level administrative data. These data contain information on gym membership and daily gym attendance over a five-year period. Our data spans academic years 2013-14 to 2017-18, covering one year before the rebate policy was implemented, the three years in which the reimbursement was available, and one year after the policy was discontinued. The data include the universe of undergraduates and graduate students with gym memberships. During our analysis period, university student enrollments averaged just under 20,000 student per year, with undergraduates comprising roughly three quarters of the student body.

The membership data set includes a unique identifier for each student, student type (undergraduate or graduate), membership start date, and whether the membership was for the Fall, Spring, or full academic year. While we do not have data on students who did not purchase gym

memberships, we obtained annual Fall enrollments for each student type, allowing us to calculate annual membership rates for both graduate and undergraduate students.

We link the membership data to visit-level data on gym attendance. These data come from the Fitness Center Database, which records each time a student swipes his student ID card at a university facility. These data include a unique student identifier, visit date, and location. Since the data include the student identifier, we are able calculate individual-level attendance measures within each semester and across years. In total, our data set includes approximately 100,000 student-year observations and more than 1.5 million gym visits. Our data set is substantially larger than most other studies that have examined responses to gym incentives, both in terms of sample size and study period.[4]

While this data set includes very detailed data on gym attendance, it does not include information on whether a student is enrolled in SHP, so we unable to determine who is eligible for the gym reimbursement. However, while a smaller fraction of undergraduate students enroll in SHP, nearly all graduate students are SHP members. Therefore, in our difference-in-difference analysis, we proxy for rebate eligibility using the graduate student population, and we use undergraduates as a control.[5]

ii. Empirical Methodology

To examine potential bunching in response to the reimbursement program's 50-visit threshold, we first rely on the nonparametric bunching method developed in Chetty et al. (2011).[6] This method assumes that any observed bunching around the 50-visit threshold is due to the reimbursement incentive, and that the density distribution of gym visits would have been smooth around this cut-off in the absence of the financial reward. Under this assumption, one can compare the actual mass of observations in the density distribution around the 50-visit threshold with the mass that would have been in this interval in a counterfactual distribution without the incentive. This comparison provides a measure of "excess mass" around the threshold (as a percent of the average height of the counterfactual distribution), and estimates the extent of bunching caused by the policy.

While the actual mass in the bunching interval can be observed, what the mass would have been in the absence of the policy cannot. To remove the influence of the incentive threshold from the distribution and obtain a counterfactual distribution, we follow Chetty et al. (2011) and use nonparametric methods. We begin by choosing an analysis window which specifies the sample that we use to estimate bunching and the counterfactual distribution. We then fit a flexible polynomial to this

---

[4] We do not have information on student demographic characteristics apart from whether the student is a graduate student or an undergraduate student.

[5] Less than one third of undergraduates are eligible for the rebate, and there are several reasons why these students may be less likely to respond to the policy. First, undergraduates may be more likely to bill their gym membership to their parents than graduate students; at the same time, undergraduates who do not rely on their parents' finances may find it difficult to pay the upfront fees. Additionally, student athletes receive gym memberships for free. For these reason, the reimbursement policy may be a weaker incentive for undergraduate students even among those who are eligible for the program.

[6] Based on the bunching method originally proposed by Saez (2010).

distribution, excluding observations in a small visit band around the threshold which encompasses all individuals that bunch. Finally, estimate the excess mass around the threshold by comparing the observed density in the excluded window with the estimated counterfactual density in the excluded window.[7]

The advantage of the bunching method is that it can nonparametrically identify behavioral responses at the threshold using a single cross-section of data. However, the assumptions required to construct the counterfactual distribution are a major limitation of this method. Specifically, the counterfactual distribution is obtained using nonparametric polynomial smoothing based on a visual identification of the region of excess mass. The counterfactual density in the specified bunching range is thus an out-of-sample prediction that may not be well fitted, especially if the policy leads to behavioral changes in the density distribution far from the threshold. Many of these concerns have been discussed extensively in the literature (e.g. Dekker et al. 2016; Blomquist and Newey 2017; Aronsson et al. 2017; Marx 2018; Bertanha 2018).

One key benefit of our data and institutional setting is that we do not need to rely on a single cross-section of treated individuals to construct a counterfactual distribution. Specifically, we have data from before and after the rebate program, and we have information for a treatment group (graduate students) as well as a control group (undergraduate students). These two features permit us to examine the effect of the policy using a difference-in-difference framework, comparing gym attendance during periods in which the policy was available to periods in which it was not for graduate students relative to undergraduate students. An additional advantage of this method is that it allows us to estimate an effect that is less local to the threshold compared to the formal bunching method, and to estimate the effect of the policy on overall gym attendance.

To estimate the causal effect of the rebate program on individual gym behavior using our difference-in-difference design, we rely on the following equation:

(1) $Y_{it} = \beta_0 + \beta_1[Grad_i * PolicyOn_t] + \beta_2 Grad_i + \beta_3 PolicyOn_t + \varepsilon_{it}$,

where Y_{it} is one of the outcomes listed above for individual i at time t. The dichotomous variable $Grad_i$ takes the value of 1 if person i is a graduate student, and $PolicyOn_t$ is an indicator variable that equals

---

[7] We define the range used to calculate the counterfactual distribution to be between 20 and 70 gym visits, and the region of excess mass to be between 50 and 60 gym visits. We use a seventh-degree polynomial and impose the integration constraint which ensures that the area under the counterfactual distribution is equal to the area under the observed distribution. The intervals we have chosen are based on a visual inspection of the distribution of gym visits, though Appendix Table 2 considers alternative intervals for the region of excess mass and the interval used to calculate the counterfactual distribution. Standard errors are obtained using a bootstrapping method. For additional details regarding the bunching estimator, see Kleven (2015)

1 if the rebate program was active at time t. β_1 is the parameter of interest, and measures the intent-to-treat effect of the rebate program on gym behavior. Our estimates are clustered on the individual level, and our results are robust to using bootstrapped standard errors.[8]

4. Effect of Rebate on Gym Attendance
A. Bunching at the 50-visit Threshold
i. Graphical Depiction

In Section 2, we note that the rebate program may incentivize individuals to bunch at the 50-visit threshold. The rationale underlying this hypothesis is that the program rewards individuals who attend the gym 50 times, but provides no additional incentive to students who visit the gym more than 50 times. If students minimize the effort required to earn the rebate, we would expect a high level of bunching just above the 50-visit threshold.

To investigate evidence of bunching, we first evaluate the graphical evidence at the reimbursement threshold among our graduate student population, nearly all of whom were enrolled in the SHP. Panel A of Figure 1 plots the density of gym visits among graduate students for the years in which the program was in effect, and overlays the same density when rebates were not available. While the distribution of gym visits is smooth across the threshold in years without rebates, the figure shows clear evidence of bunching just above the 50-visit threshold in the years when reimbursements were available.

Panel B in Figure 1 replicates this analysis for undergraduate students, most of whom were ineligible for the rebate. The distributions of gym attendance for undergraduates versus graduates are similar in the pre-period, both in terms of level and shape. However, the distribution for undergraduates is nearly identical for the policy-on and policy-off periods, and shows no evidence of bunching at the 50-visit count.

ii. Bunching Estimation

Panel A of Figure 2 provides evidence of bunching among graduate students at the 50-visit threshold using the bunching estimator developed by Chetty et al. (2011). In this figure, the solid line depicts the observed density as shown in the data. The dotted line shows the constructed counterfactual density function, obtained by fitting a flexible polynomial through the observed density excluding the region of excess mass. The results in Panel A of Figure 2 suggest that there is a large and statistically significant excess mass just above the 50-visit threshold. Specifically, the observed mass as a percent of the average height of the simulated counterfactual distribution is 954%.[9] This provides strong

---

[8] These results are available from the authors upon request.
[9] We find large and statistically significant bunching using alternative intervals for the region of excess mass and the range of data used to calculate the counterfactual distribution as well (Appendix Table 2).

evidence that the program induced graduate students to attend the gym just often enough to obtain the rebate.[10]

      A strength of our data and setting is that we have information on gym attendance from before and after the rebate program was implemented. To ensure that our results in Panel A of Figure 2 are not driven by potentially unobserved confounders, we take advantage of this feature and perform the bunching estimation on graduate students when the rebate is not available. As can be seen in Panel B of Figure 2, there is no evidence of bunching in years when the reimbursement was not available, suggesting that the bunching behavior identified in Panel A is not driven by unobserved confounders.

      Another unique feature of our data and setting is that we also have the same data for a group of students that largely was unexposed to the policy. We take advantage of this feature of our setting and perform the bunching estimation on undergraduate students, both during the policy years (Panel C) and in years without the policy (Panel D). In both cases, we would not expect to see any bunching behavior around the 50-visit threshold. Looking at the two last panels of Figure 2, we find small and not statistically significant estimates for bunching at the 50-visit threshold.

iii. Difference-in-Difference

      Our estimates in the previous section suggest that students who were eligible for the rebate responded to the reimbursement policy by bunching just above the 50-visit threshold. However, as discussed in Section 3, the bunching estimator relies on a relatively strong set of assumptions. In this section, we take advantage of both data from periods without reimbursement and a population of undergraduate students who were (largely) unaffected by the policy. This type of counterfactual data is often unavailable in studies using bunching estimation, and allows us to verify the results obtained from the bunching estimator.

      Difference-in-differences results based equation (1) are presented in Table 1, where the outcome is a binary variable for attending the gym between 50 and 60 times in the six-month period. We show results from comparing the rebate period to both the pre- and post-period (column i), comparing only the pre-period and the rebate period (column ii), and comparing only the post-period with the rebate period (column iii).

      The results in Table 1 suggest that the rebate program led to a 4 percentage point increase in the probability of attending the gym between 50 and 60 times. This result is robust across all three comparisons we make. Taken together, this analysis confirms the evidence shown in the previous section that the policy induced a significant increase in attendance just above the rebate threshold.

---

[10] The McCrary Density Test – another common test for examining discontinuities in densities around certain thresholds – provide very similar results. See Appendix Figure 2

B. Number of Gym Visits

In this section, we examine if the program led to an increase in the number of gym visits conditional on having a gym membership. As mentioned in Section 2, the effect on the average number of visits could be large or small, even in the presence of significant bunching, depending on the counterfactual attendance of the bunchers.

Figure 3 plots the average visit count over time for graduate students (Panel A) and undergraduate students (Panel B). This figure provides evidence that graduate students visit the gym more often when the policy is in effect (AY 2014 - AY 2016), while no such behavior can be observed among undergraduate students. Specifically, Panel A of Figure 3 shows that graduate students attended the gym an average of 25 times per semester prior to the introduction of the rebate program. This number increased slightly in the first year of the program, and increased significantly more in second and third years of the program (around 30 visits). The gradual increase in gym attendance over time could be due to imperfect information about the existence of the program in the first year. Panel B of Figure 3 shows that undergraduate attendance remained constant throughout the entire period.

Table 2 shows the difference-in-differences results obtained from estimating equation (1) using number of gym visits as the dependent variable. In column (i), we compare the rebate period to both the pre- and the post-period. The results demonstrate that graduate students increase gym visits when rebates are available. Specifically, the gym reimbursement program led to a significant increase in the number of gym visits of approximately 3.5 visits per six-month period.

In column (ii) of Table 2, we restrict the comparison to the pre-period and the rebate period, and in column (iii) we only compare the post-period with the rebate period. We find that the magnitude of the point estimate is larger when we restrict the comparison to the pre-period and the rebate period (4.6 visits) compared to when we compare the post-period with the rebate period (2.4 visits). These findings are consistent with a model in which the reimbursement program led to habit formation in the year after the policy was discontinued, an outcome we address in Section 6. Therefore, results in column (i) that use both the pre- and post-policy period are likely to be downward biased. Thus, we consider the specification underlying the results in column (ii) to be our preferred specification.

C. Gym Membership

The analyses above estimate the intensive margin effect of the rebate program, i.e., the effect of the rebate program among students with gym memberships. However, the reimbursement program may also serve as an incentive for eligible students to purchase a membership. Specifically, eligible students who may have been deterred from joining the gym due to the cost of the membership, but who expect to meet the 50-visit threshold if they had a membership, may be incentivized to enroll. If the rebate policy has a separate effect on purchasing a gym membership, our intensive margin estimates will suffer from selection bias. For example, students who purchase memberships in response to the

11

rebate policy may be more likely to bunch than those who purchased memberships in the absence of the policy. While there exists a large literature examining the effectiveness of various fitness incentive programs through field experiments, the majority of these studies either provide free gym memberships or incentives to join for all experiment participants (Royer et al. 2015), or target populations that already have gym memberships (Charness and Gneezy 2009; Carrera et al. 2018). Therefore, our ability to look at potential extensive margin effects represents an important contribution to the literature.

Suggestive evidence of the program's effect on membership take-up is shown in Figure 4, which plots the gym membership rate for graduate students (Panel A) and undergraduate students (Panel B) over time. While the figure shows that membership rates differed across the two groups – 26% of graduate students had a membership compared to 42% of undergraduates – the within-group rates remained stable throughout the study period.

To formally estimate if the program led to an increase in membership take-up, Table 3 shows difference-in-differences results obtained from estimating equation (1) with membership enrollment as the dependent variable. In column (i) we use all years of data comparing both the pre- and post-period to the rebate period, in column (ii) we restrict the comparison to the pre-period and the rebate period, and in column (iii) we only compare the post-period with the rebate period. The results in Table 3 are consistent with Figure 4; the policy's effect on membership take-up is very small and not statistically significant. These results suggest that any potential effects on the intensive margin are not driven by changes in sample composition.

5. Heterogeneity by Prior Gym Attendance

A unique feature of our panel data is that we have information on pre-policy gym attendance for both graduate and undergraduate students. This allows us to separately estimate the behavioral response to the policy conditional on various levels of pre-policy gym attendance. In other words, we can determine whether we observe changes in gym attendance only among students who were previously close to the 50-visit threshold, or if the policy induced students from across the pre-policy distribution to increase their gym attendance.

To explore this question, we group students into bins based on how often they visited the gym in the Fall prior to policy implementation: less than 10 visits, 10-19, 20-29, 30-49, or 50+ visits. Since this analysis relies on observing gym attendance in the pre-period, these regressions are restricted to students who were gym members in the Fall of 2013 as well as during the period in which the reimbursement was available. We compare the likelihood of attending the gym between 50 and 60 times and average gym attendance for graduates and undergraduates during years in which the rebate was available separately by pre-policy gym attendance. The use of undergrads as a counterfactual allows us to control for factors such as mean reversion and other changes in attendance over time.

Panel A of Table 4 presents results on the likelihood of attending the gym between 50 and 60 times. We find that for students who attended the gym less than 50 times in the pre-period, the point estimates are economically meaningful, statistically significant and monotonically increasing in prior gym attendance. This monotonic increase is consistent with a model in which the policy has a larger effect on students who were already close to the reimbursement threshold. However, somewhat surprisingly, we still observe large and significant bunching among graduate students who attended the gym very infrequently in the pre-period. Specifically, we observe a statistically significant 3 percentage point increase in the likelihood of attending the gym 50 to 60 times among students who previously attended the gym less than 10 times in the semester. This is approximately half the size of the effect of the policy on bunching among students who previously attended the gym between 30 and 49 times. In contrast, we observe a small and not statistically significant effect on bunching for students who previously attended the gym 50 or more times.

Panel B of Table 4 presents results for average gym attendance. Here we find statistically significant and economically meaningful results for all groups. We observe increases in average attendance between 3 and 6 visits per semester for those who went to the gym less than 20 times in the pre-period, and increases between 8 and 10 visits for those who attended the gym 20 or more times, including for those who previously met the 50-visit requirement.[11] The increase in overall attendance coupled with the absence of bunching among the 50+ group suggests that the 50-visit threshold did not discourage attendance by providing a reference point that was lower than what the student would have chosen in the absence of the policy.

The results in Table 4 suggest that although the policy presents a high bar for participants to meet in order to claim the reimbursement, it still generates behavioral responses among previously low-attendance students.[12]

6. Habit Formation

In Section 2, we note that the gym reimbursement program may have a habit formation effect if the policy-induced change in gym behavior is sufficiently strong. Data on gym attendance both before program implementation and after its discontinuation allow us to examine this question directly.

A subset of the prior literature on fitness incentive programs has examined habit formation, with mixed results. For example, Charness and Gneezy (2009) present evidence of habit formation in gym attendance among university students in the two months following a one-time, 4-week gym incentive. Acland and Levy (2015) replicate these findings using a similar experiment, but find that the

---

[11] The large effect for students with high pre-policy visit counts could be because graduate students are incentivized to maintain high gym attendance, while undergraduate students potentially decrease their gym attendance over time due to mean reversion.

[12] An interesting implication of this result is that one of the assumption required for the formal bunching estimator discussed in Section 2 – that the policy did not induce a behavioral response from far outside the bunching interval – is violated regardless of the interval used, potentially biasing the results.

effects fade when using a slightly longer follow-up period.[13] Royer et al. (2015) analyze a gym incentive program at a large company over a substantially longer post-intervention period (three years) and find that the habit formation effects of financial incentives alone persist for only two months.[14] To our knowledge, we are the first paper to complement these experimental studies by looking for evidence of habit formation in a non-experimental context.

To test for habit formation, we rely on a difference-in-difference approach in which we compare the difference in gym attendance before the policy was introduced and after the policy was discontinued among graduate students to that same difference among undergraduate students (excluding the policy-on period). We examine both the change in visit count and the probability of attending the gym between 50 and 60 times. If the policy led to habit formation, we would expect a persistent increase in visits after the end of the financial incentive, but not necessarily an increase in the likelihood of bunching since the financial incentive to bunch above the 50-visit threshold was removed.

Column (i) in Table 5 presents clear evidence of an economically meaningful and statistically significant habit formation effect. Specifically, we estimate an increase in gym attendance of 2.3 visits per semester. This effect suggests that approximately 50% of the program effect persists after the policy has been discontinued.

With respect to post-policy bunching behavior, the results in column (ii) show a precisely estimated zero for the likelihood of attending the gym between 50 and 60 times in the year after the policy was discontinued. Figure 5 further illustrates this point by plotting the density of gym visits among graduate students (Panel A) and undergraduate students (Panel B) prior to the rebate program, and overlays the same densities for the post-program period. While graduate students attend the gym more frequently in the post-program period compared to the pre-program period, there is no sign of bunching at the 50-visit threshold. The distribution for undergraduates is similar across both periods.

7. Discussion and Conclusion

Physical inactivity represents a major problem for policymakers, contributing to rising medical costs, lower labor productivity and reduced well-being. In response to the rising costs of physical inactivity, employers, health insurers and state governments, have introduced a number of different wellness incentive programs to encourage physical activity. This paper evaluates the effect of the introduction and subsequent termination of one such wellness program at a major American university. This program provides gym membership reimbursements to members who attend the gym 50 times in a six-month period.

---

[13] Carrera et al (2018a) test the effect of different gym incentive structures on habit formation and finds that sporadic payments are more effective at encouraging habit formation in the eight weeks post-intervention than front-loaded incentives or constant incentives, while Carrera et al (2018b) find no effects of incentives for new gym members on habit formation.

[14] However, Royer et al. (2015) find the combination of financial incentives plus a voluntary commitment device increased gym attendance for over a year after the incentive was removed.

We document significant bunching at the 50-visit threshold in years when the policy is in place, and that this bunching effect translates into a statistically significant and economically meaningful effect on overall gym attendance. Specifically, we find that exposure to the reimbursement program increased students' gym visits by 4.6 visits per semester, representing a 20% increase from the mean. This effect is driven exclusively by the intensive margin; we find no effect on membership take-up. Examining heterogeneous treatment effects by pre-policy gym attendance behavior, we find increases in gym visits among both low- and high-frequency gym attendees. This suggests that our bunching results are not solely driven by students who were previously close to the attendance threshold. Additionally, the fact that we see increases in attendance among those who had previously met the reimbursement requirement before the policy was implemented suggests that the threshold does not discourage these students from attending the gym more than 50 times.

Our results demonstrate that rebates as incentives can be successful in not only inducing healthy behaviors in the short-term, but also in creating new habits in the long-run. Specifically, we show that approximately half of the program effect persists after program termination, providing evidence of habit formation. These effects are considerably larger than many of those identified in the existing literature on fitness incentive programs. These large estimates of habit formation are consistent with models of accumulation of habit stock (Becker and Murphy 1988), especially those which suggest that this habit stock must cross a certain threshold in order to lead to sustained activity (Carrera et al. 2017).

In addition to the health benefits associated with increased physical activity, previous studies have documented a causal relationship between gym attendance and academic performance at the university level. For example, Cappelen et al. (2017) perform a field experiment in which they randomly provide students without gym memberships access to fitness facilities, and find that this program (with an average cost of $110 per student) leads to a 5.7 visit increase in gym attendance per semester and a 0.3 standard deviation increase in total grade points. The per student cost of the program we examine is significantly lower than the cost of the Cappelen et al. (2017) experiment ($17.5), and the effect on gym attendance is quite similar (4.6 versus 5.7 visit increase).[15] If we assume that the effect of exercise on student achievement is the same in our setting, the rebate is a more cost-efficient method of improving student educational outcomes. Specifically, the per dollar effect on student total grade point is approximately 5 times larger in our setting.[16] These results are encouraging in light of the recent efforts to induce healthy behaviors through fitness rebates undertaken by US health insurers, state governments, and higher education institutions.

---

[15] The low per student cost in our setting is due to the fact that only graduate students who attend the gym more than 50 times during the policy years (23.3%) receive $75, while every treated student in Cappelen et al. (2017) receives $110.
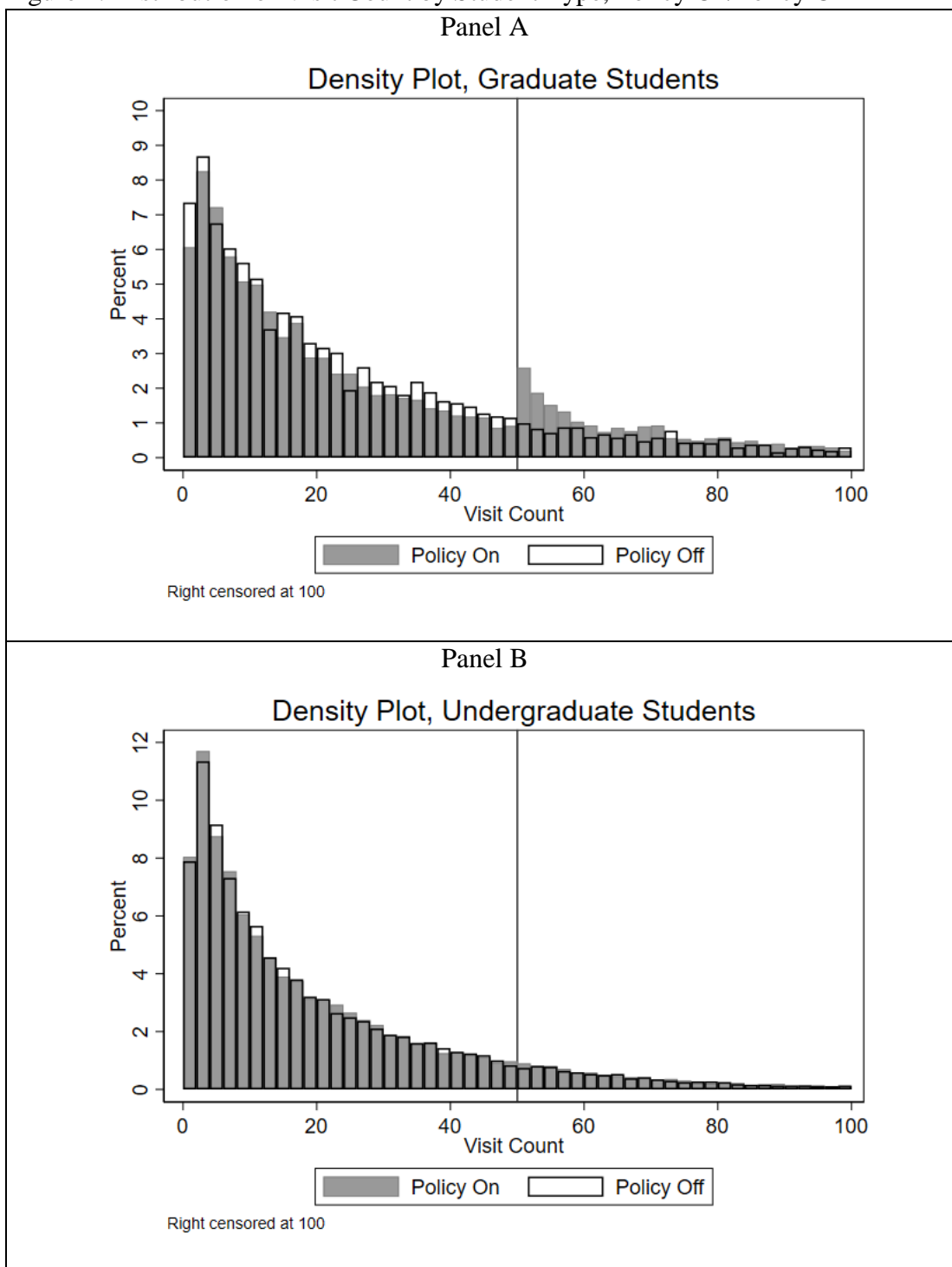
[16] This is likely an upper bound since the effects in our study are exclusively coming from individuals that already are exercising, while the participants in Cappelen et al. (2017) do not have memberships prior to the intervention. Thus, to the extent that the effect on academic achievement in Cappelen et al. (2017) is driven by students having more structure and routine in their lives, this will likely not transfer to our setting.

References

Acland, Dan, and Matthew Levy (2015). "Naiveté, Projection Bias, and Habit Formation in Gym Attendance." *Management Science* 61(1): pp. 146-160

Aronsson, Thomas, Katharina Jenderny and Gauthier Lanot (2018). "Alternative parametric bunching estimators of the ETI" *Mimeo*

Baicker, Kathrine, David Cutler and Zirui Song (2010). "Workplace Wellness Programs Can Generate Savings." *Health Affairs* 29(2): pp. 1-8

Bertanha, Marinho, Andrew McCallum and Nathan Seegert (2018). "Nicer Notching, Better Bunching" *Mimeo*

Blomquist, Soren, and Whitney K. Newey (2017). "The Bunching Estimator Cannot Identify the Taxable Income Elasticity" *CESifo Working Paper Series 6736*

Cappelen, Alexander, Gary Charness, Mathias Ekström, Uri Gneezy and Bertil Tungodden (2017). "Exercise improves academic performance." *IFN Working Paper No 1180*

Carrera, Mariana, Heather Royer, Mark Stehr, and Justin Sydnor (2017). "The structure of health incentives: Evidence from a field experiment" *NBER Working Paper 23188*

Carrera, Mariana, Heather Royer, Mark Stehr, and Justin Sydnor (2018). "Can financial incentives help people trying to establish new habits? Experimental evidence with new gym members" *Journal of Health Economics* 58: pp. 202-214

Charness, Gary and Uri Gneezy (2009). "Incentives to exercise" *Econometrica* 77(3): pp. 909-931

Chetty, Raj, John Friedman, Tore Olsen, and Luigi Pistaferri (2011). "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: evidence from Danish tax records" *Quarterly Journal of Economics* 126(2): pp. 749-804

Dekker, Vincent, Kristina Strohmaier, and Nicole Bosch (2016). "A data-driven procedure to determine the bunching window: an application to the Netherlands" *Hohenheim Discussion Papers 05-2016*

Field, E. (2009). "Educational debt burden and career choice: Evidence from a financial aid experiment at NYU Law School" *American Economic Journal: Applied Economics 1*(1): pp. 1-21

Fricke, Hans, Michael Lechner and Andreas Steinmayr (2018). "The effect of physical activity on student performance in college: an experimental evaluation" University of St. Gallen *Discussion Paper No. 2017-07*

de Haan, Thomas, and Jona Linde (2018). "Good nudge lullaby: choice architecture and default bias reinforcement." *The Economic Journal 128:* pp. 1180-1206

Hossain, T., & List, J. A. (2012). "The behavioralist visits the factory: Increasing productivity using simple framing manipulations" *Management Science 58*(12): pp. 2151-2167

Jones, D., Molitor, D., & Reif, J. (2018). "What do workplace wellness programs do? Evidence from the Illinois workplace wellness study" *NBER Working Paper No. 24229*
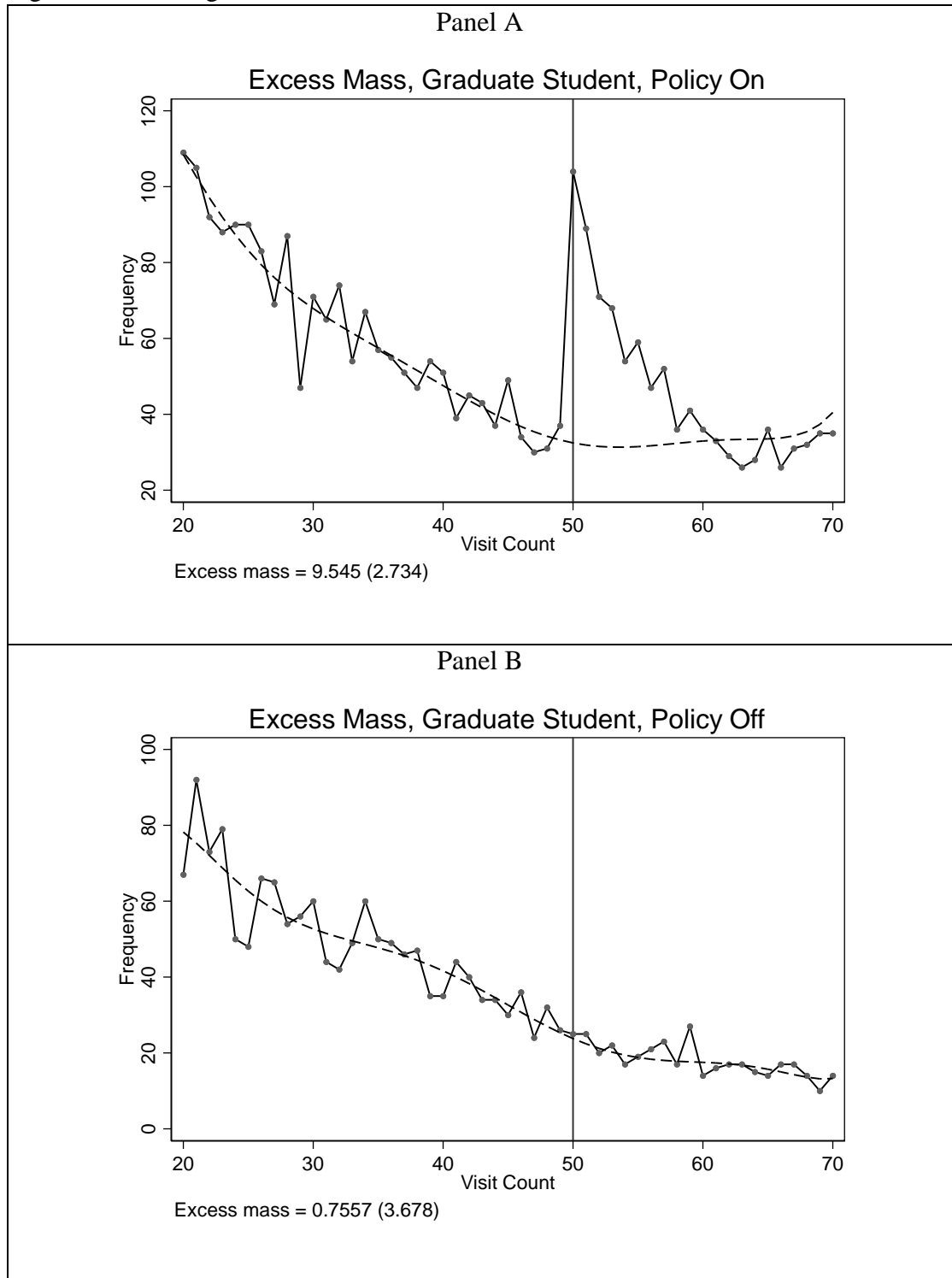
Kahneman, D., & Tversky, A. (1979). "Prospect Theory: An Analysis of Decision Under Risk" *Econometrica* 47: pp. 263-291

Kleven, Henrik Jacobsen (2016). "Bunching." *Annual Review of Economics* (8): pp. 435-464

Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). "The behavioralist goes to school: Leveraging behavioral economics to improve educational performance" *American Economic Journal: Economic Policy* 8(4): pp. 183-219

Marx, Benjamin M. (2018). "Dynamic Bunching Estimation with Panel Data," *MPRA Paper 88647*

NBGH (2011). *Employer investments in improving employee health*. Available at https://jointhehealthjourney.com/images/uploads/channel-files/StudyEmployeeHealth Programs.pdf

Rees-Jones, Alex (2018). "Quantifying loss-averse tax manipulation" *Review of Economic Studies* 85: pp. 1251-1278

Reis, Nola (2012). "Financial incentives for weight loss and healthy behaviors" *Healthcare policy* 7(3): pp. 23-28

Saez, Emmanuel (2010). "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3): pp. 180-212

Troiano, R, D. Berrigan, K. Dodd, L. Masse, T. Tilert and M. McDowell (2008). "Physical activity in the United States measured by accelerometer" *Medical Science Sport Exercise* 40(1): pp. 181-8

Figure 1: Distribution of Visit Count by Student Type, Policy On/Policy Off

Panel A

Density Plot, Graduate Students

Right censored at 100

Panel B

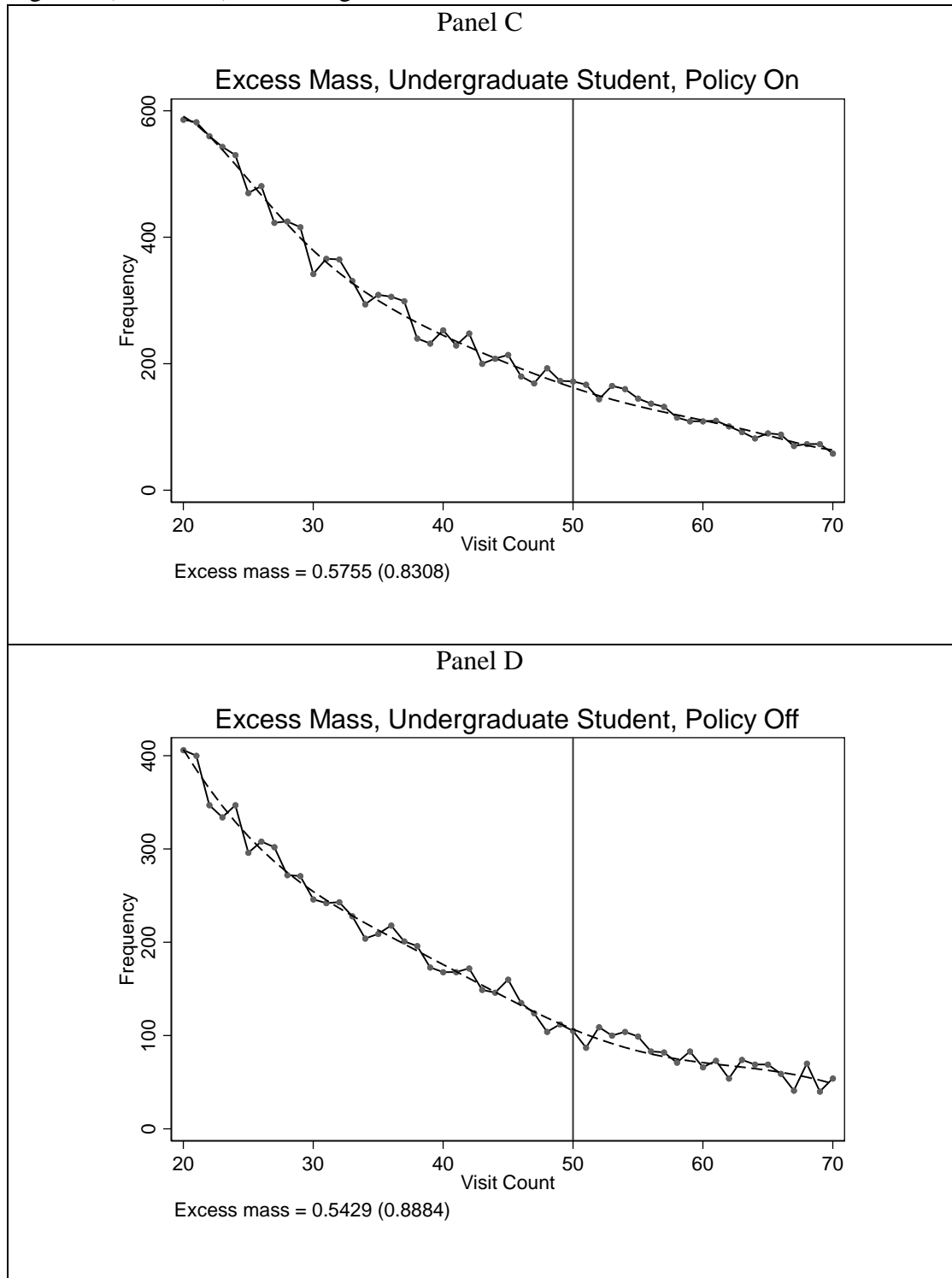Density Plot, Undergraduate Students

Right censored at 100

Graph represents percent of students with different levels of visit counts. Policy on is Fall 2014-Spring 2017, policy off is Fall 2013-Spring 2014 (pre-period) and Fall 2017-Spring 2018 (post-period). Vertical line indicates the 50-visit threshold that students need to reach in order to obtain the reimbursement.

Figure 2: Bunching Estimation

Panel A

Excess Mass, Graduate Student, Policy On

Excess mass = 9.545 (2.734)

Panel B

Excess Mass, Graduate Student, Policy Off

Excess mass = 0.7557 (3.678)

Graph represents frequency of students with different levels of visit counts. Solid line-dots is actual distribution; dash line is smoothed counterfactual distribution. Vertical line indicates the 50-visit threshold that students need to reach in order to obtain the reimbursement. Bootstrap standard errors in parentheses.

Figure 2 (continued): Bunching Estimation

Panel C

Excess Mass, Undergraduate Student, Policy On

Excess mass = 0.5755 (0.8308)

Panel D

Excess Mass, Undergraduate Student, Policy Off
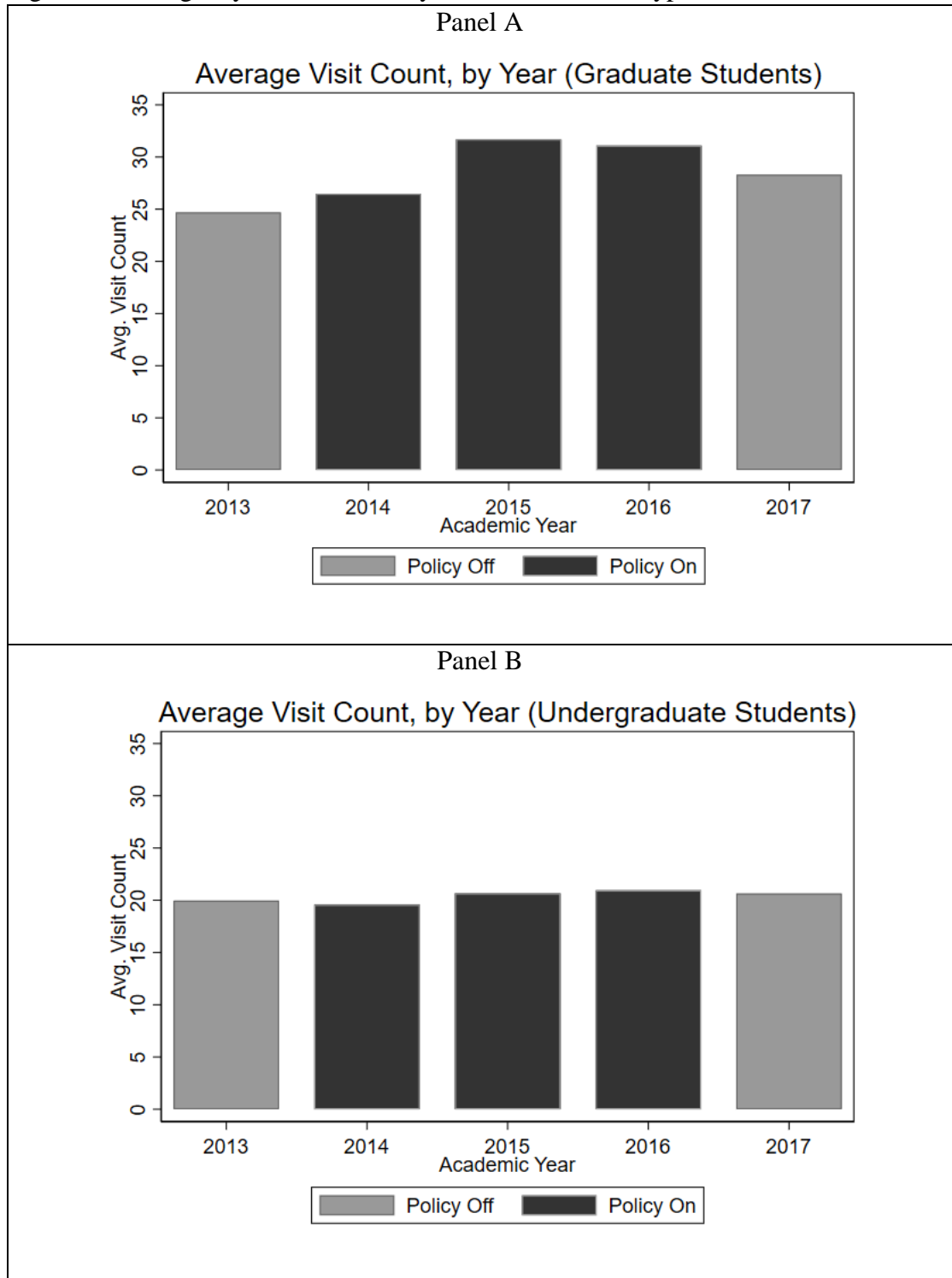

Excess mass = 0.5429 (0.8884)

Graph represents frequency of students with different levels of visit counts. Solid line-dots is actual distribution; dash line is smoothed counterfactual distribution. Vertical line indicates the 50-visit threshold that students need to reach in order to obtain the reimbursement. Bootstrap standard errors in parentheses.

Figure 3: Average Gym Attendance by Year and Student Type

Panel A

Average Visit Count, by Year (Graduate Students)

Panel B

Average Visit Count, by Year (Undergraduate Students)

Each bar represents the average number of visits for of each academic year.

Figure 4: Gym Membership Rate by Year and Student Type

Panel A

Membership Rate, by Year (Graduate Students)

Panel B

Membership Rate, by Year (Undergraduate Students)
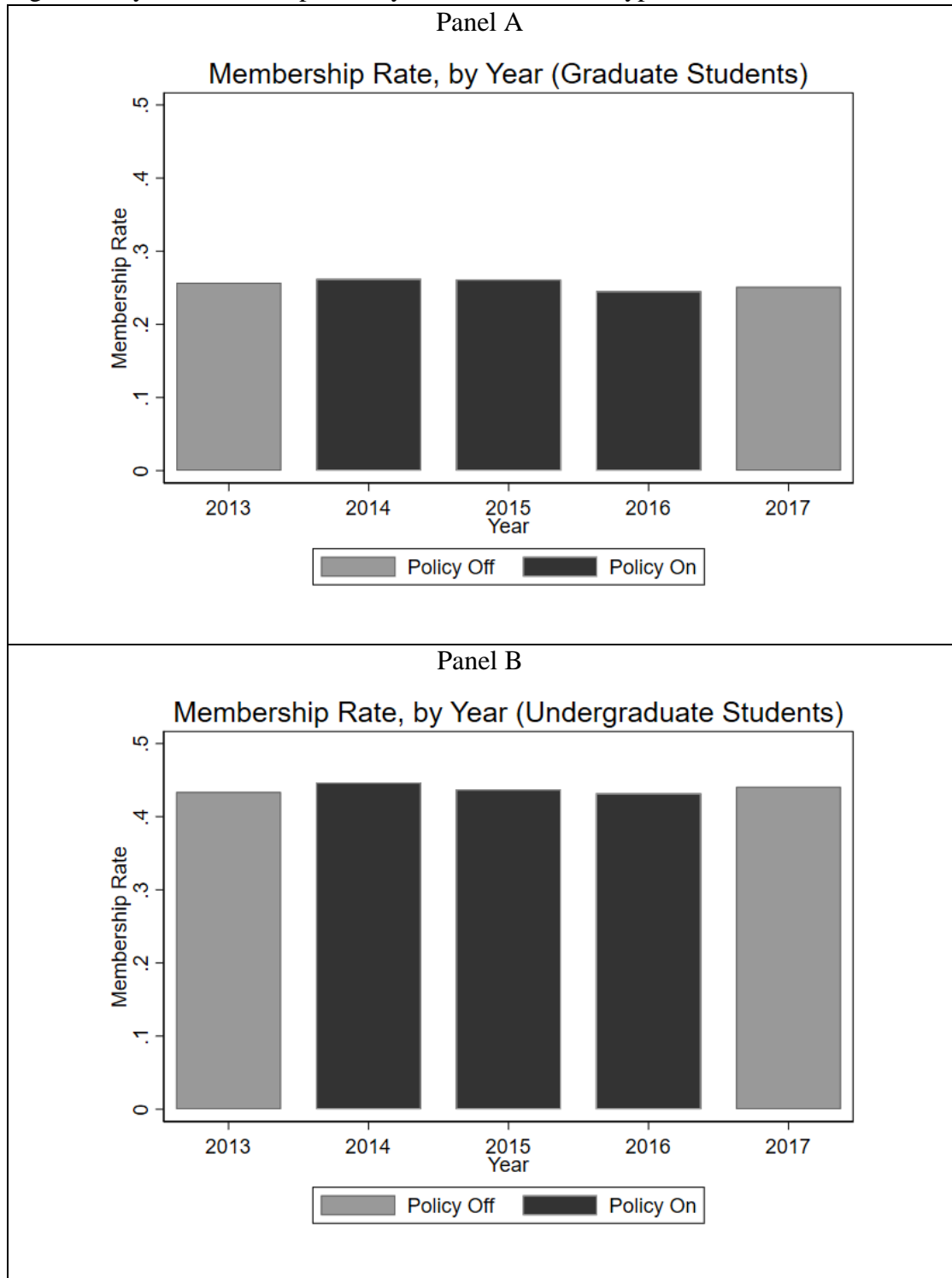
Each bar represents the membership rate for the Fall of each academic year.

Figure 5: Habit formation - Distribution of Visit Count by Student Type, Pre-Policy/Post-Policy
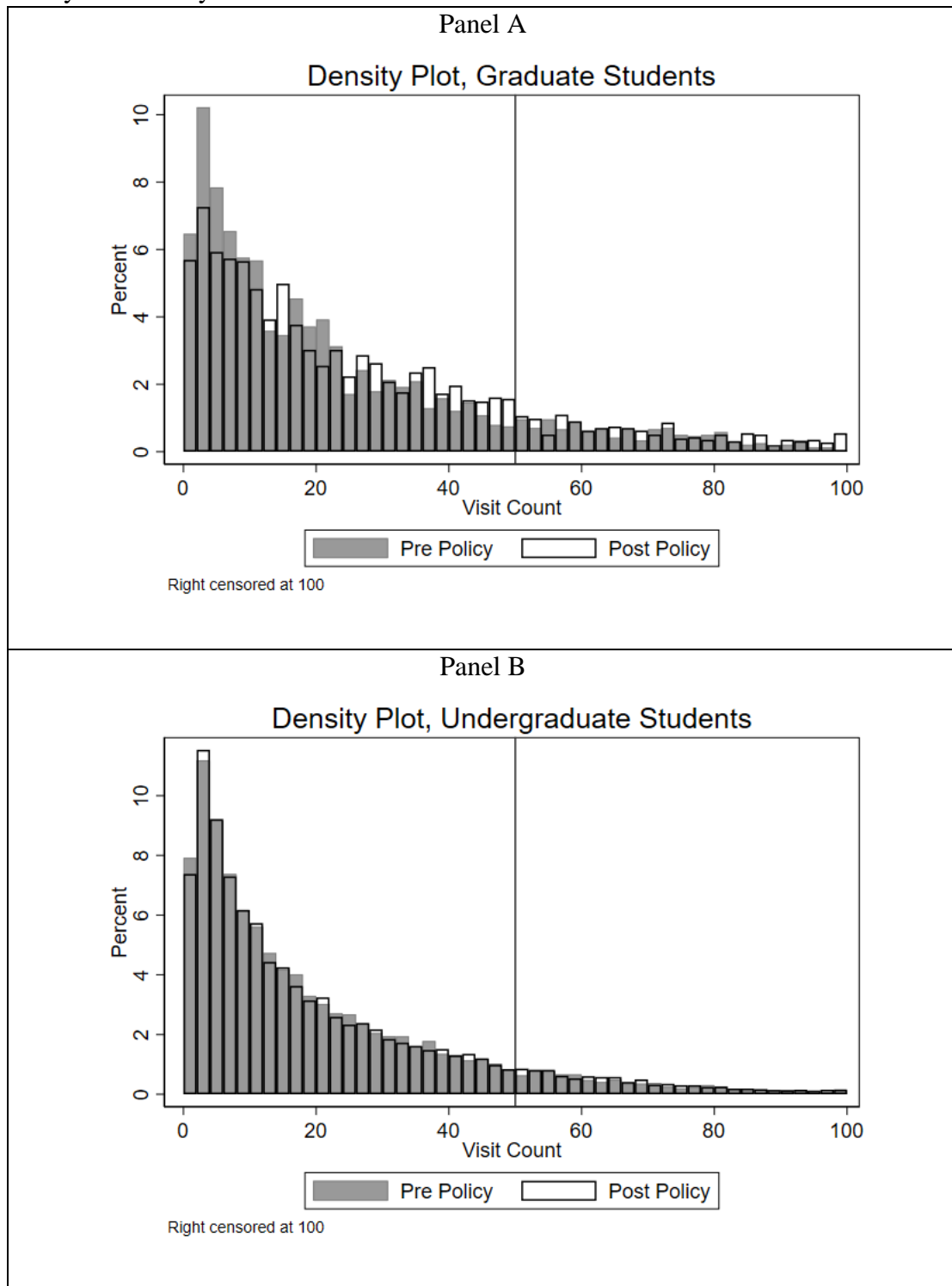
Graph represents percent of students with different levels of visit counts. Pre Policy is Fall 2013-Spring 2014, and Post Policy is Fall 2017-Spring 2018. Vertical line indicates the 50-visit threshold that students need to reach in order to obtain the reimbursement.

Table 1: Effect of Reimbursement Policy on Attending the Gym 50 to 60 Times

	(i) Full Period	(ii) Pre/On	(iii) On/Post
Grad x Policy On	0.040***	0.039***	0.040***
	(0.005)	(0.006)	(0.006)
Policy On	0.003*	0.004*	0.002
	(0.002)	(0.002)	(0.002)
Grad	0.007**	0.008	0.007
	(0.003)	(0.005)	(0.005)
Outcome Mean	0.045	0.046	0.047
Observations	75887	60129	60873

* p<0.10, ** p<0.05, *** p<0.01

Standard errors in parentheses, clustered at individual level

Each column presents results from a separate difference-in-differences regression:

$Y_{it} = \beta_0 + \beta_1[Grad_i * PolicyOn_t] + \beta_2 Grad_i + \beta_3 PolicyOn_t + \varepsilon_{it}$.

Pre-period is Fall 2013-Spring 2014, policy-on period is Fall 2014-Spring 2017, post-period is Fall 2017-Spring 2018. Column (i) compares policy-on period to both pre-period and post-period. Column (ii) excludes post-period and compares policy-on period to pre-period. Column (iii) excludes pre-period and compares policy-on period to post-period. Outcome: indicator for attending the gym 50 to 60 times in a six-month period.

Table 2: Effect of Reimbursement Policy on Overall Gym Attendance

	(i) Full Period	(ii) Pre/On	(iii) On/Post
Grad x Policy On	3.450***	4.640***	2.376***
	(0.606)	(0.806)	(0.786)
Policy On	0.127	0.418	-0.151
	(0.193)	(0.256)	(0.259)
Grad	5.914***	4.724***	6.988***
	(0.546)	(0.761)	(0.739)
Outcome Mean	21.644	21.629	21.871
Observations	75887	60129	60873

* p<0.10, ** p<0.05, *** p<0.01

Standard errors in parentheses, clustered at individual level

Each column presents results from a separate difference-in-differences regression:

$Y_{it} = \beta_0 + \beta_1[Grad_i * PolicyOn_t] + \beta_2 Grad_i + \beta_3 PolicyOn_t + \varepsilon_{it}$.

Pre-period is Fall 2013-Spring 2014, policy-on period is Fall 2014-Spring 2017, post-period is Fall 2017-Spring 2018. Column (i) compares policy-on period to both pre-period and post-period. Column (ii) excludes post-period and compares policy-on period to pre-period. Column (iii) excludes pre-period and compares policy-on period to post-period. Outcome: number of gym visits in a six-month period.

Table 3: Effect of Reimbursement Policy on Gym Membership

	(i) Full Period	(ii) Pre/On	(iii) On/Post
Grad x Policy On	0.001	-0.006	0.007
	(0.007)	(0.009)	(0.008)
Policy On	0.001	0.005	-0.002
	(0.004)	(0.005)	(0.005)
Grad	-0.183***	-0.177***	-0.190***
	(0.005)	(0.007)	(0.007)
Outcome Mean	0.389	0.390	0.390
Observations	98972	78460	79556

* p<0.10, ** p<0.05, *** p<0.01
Robust standard errors in parentheses
Fall semester only
Each column presents results from a separate difference-in-differences regression:
$Y_{it} = \beta_0 + \beta_1[Grad_i * PolicyOn_t] + \beta_2 Grad_i + \beta_3 PolicyOn_t + \varepsilon_{it}.$
Pre-period is Fall 2013-Spring 2014, policy-on period is Fall 2014-Spring 2017, post-period is Fall 2017-Spring 2018. Column (i) compares policy-on period to both pre-period and post-period. Column (ii) excludes post-period and compares policy-on period to pre-period. Column (iii) excludes pre-period and compares policy-on period to post-period. Outcome: indicator for having a gym membership.

Table 4: Heterogeneity in Policy Effects by Pre-Policy Visit Count

Panel A Attended Gym 50-60 Times					
	(i) <10	(ii) 10-19	(iii) 20-29	(iv) 30-49	(v) 50+
Grad	0.030**	0.036**	0.044**	0.062***	0.017
	(0.014)	(0.015)	(0.021)	(0.023)	(0.018)
Observations	2493	1779	1196	1546	1862
Panel B Visit Count					
	(i) <10	(ii) 10-19	(iii) 20-29	(iv) 30-49	(v) 50+
Grad	6.057***	2.891*	9.764***	8.448***	9.190***
	(1.705)	(1.653)	(2.856)	(2.460)	(3.104)
Observations	2493	1779	1196	1546	1862

* p<0.10, ** p<0.05, *** p<0.01
Standard errors in parentheses, clustered at individual level
Fall semester only
Each cell presents the coefficient on $Grad_i$ from a separate regression:
$Y_{it} = \beta_0 + \beta_1 Grad_i + \varepsilon_{it}.$
Each column is based on pre-policy visit count. Outcome: indicator for attending the gym 50-60 times (Panel A) and number of gym visits (Panel B) in a six-month period.

Table 5: Effect of Reimbursement Policy on Habit Formation

	(i) Visit Count	(ii) 50-60 Visits
Grad x Post Policy	2.264**	-0.000
	(1.030)	(0.007)
Post Policy	0.569*	0.002
	(0.340)	(0.002)
Grad	4.724***	0.008
	(0.761)	(0.005)
Outcome Mean	21.224	0.040
Observations	30772	30772

* p<0.10, ** p<0.05, *** p<0.01

Standard errors in parentheses, clustered at individual level

Each column presents results from a separate difference-in-differences regression:

$Y_{it} = \beta_0 + \beta_1[Grad_i * PostPolicy_t] + \beta_2 Grad_i + \beta_3 PostPolicy_t + \varepsilon_{it}$.

Pre-period is Fall 2013-Spring 2014, post-period is Fall 2017-Spring 2018. Outcome: indicator for attending the gym 50-60 times (column i) and number of gym visits (column ii) in a six-month period.

Appendix Figure 1: Gym Visit Count Website

The gym visit count website, which displays the current number of visits in the six-month time period.
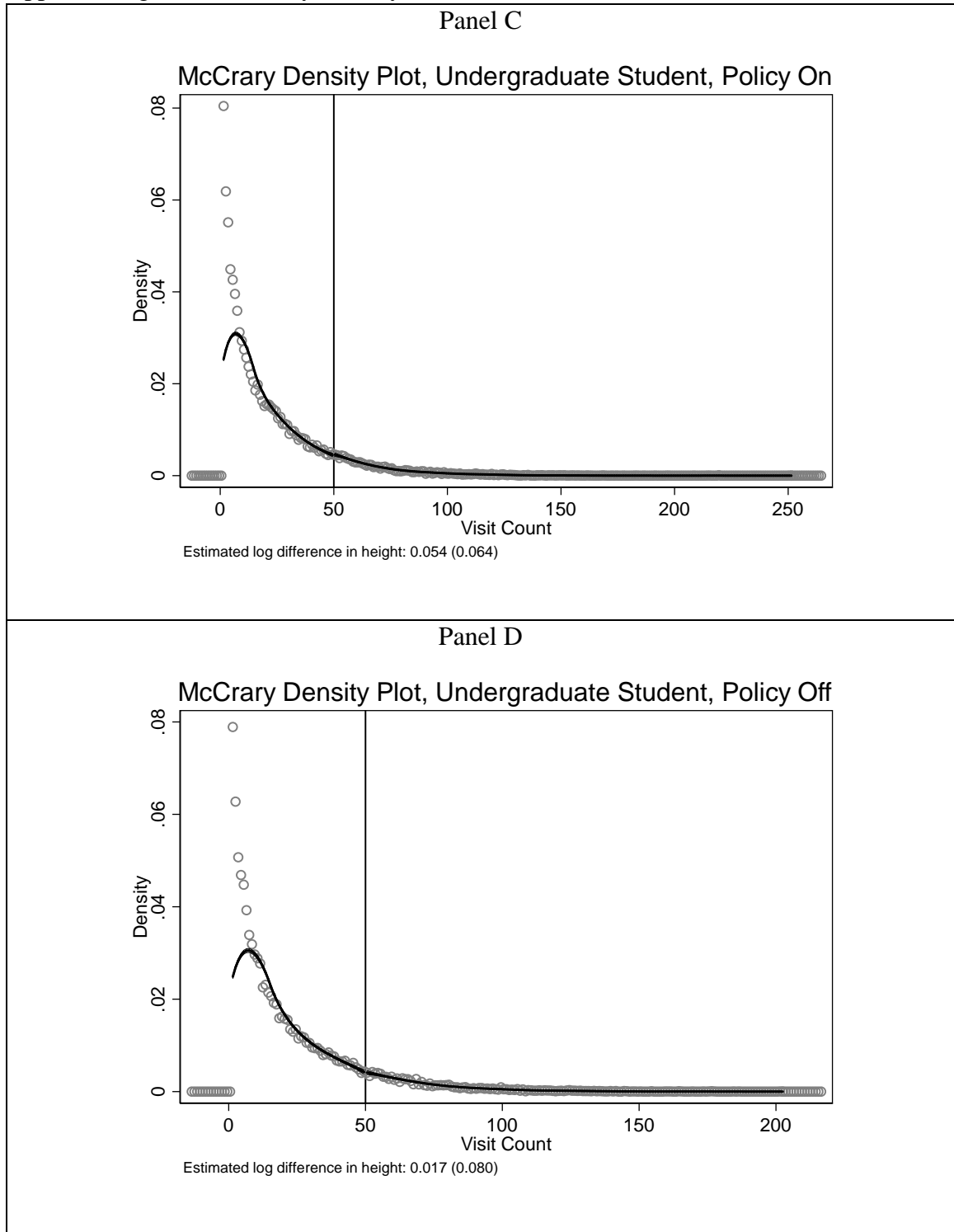
Appendix Figure 2: McCrary Density Tests
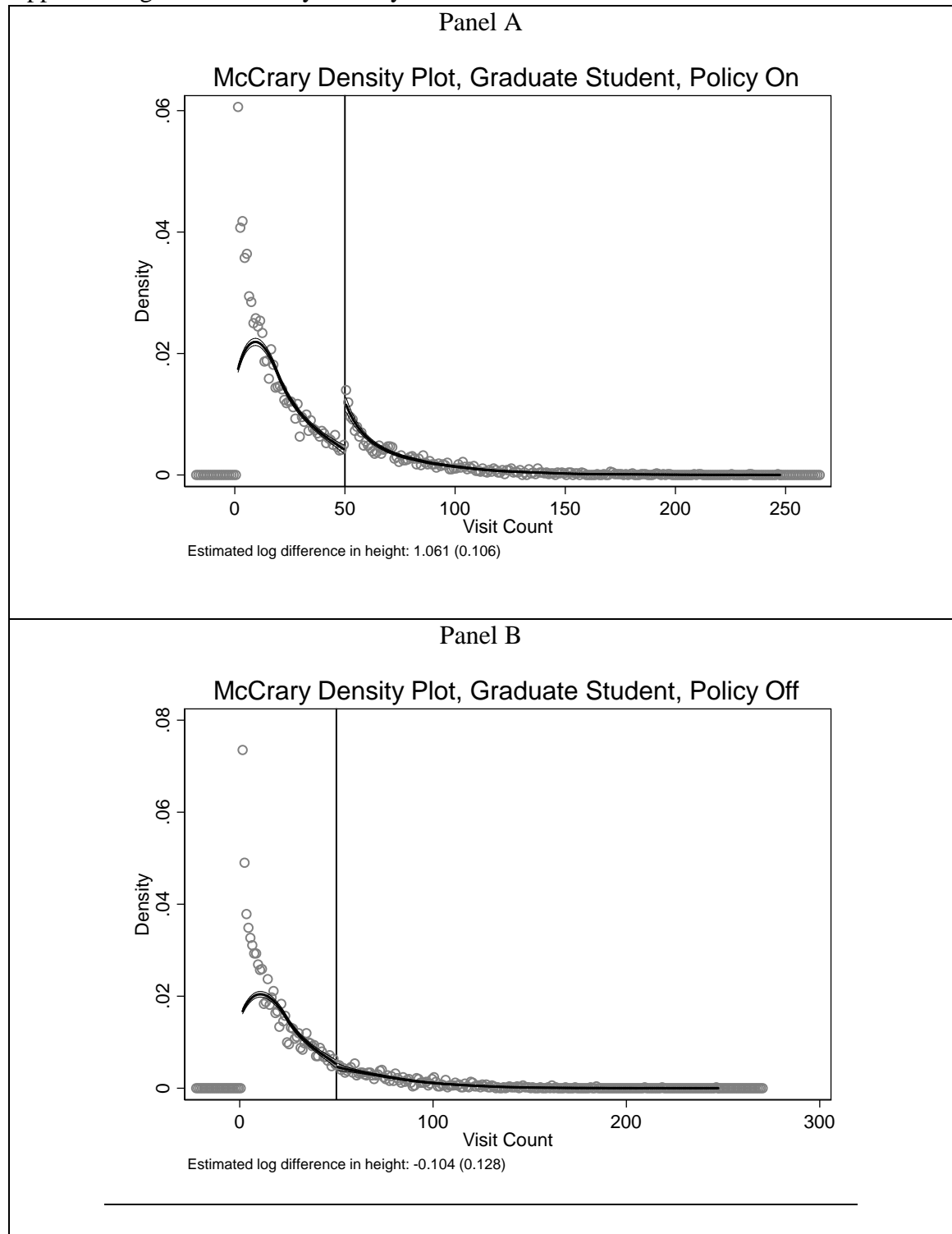
Panel A

McCrary Density Plot, Graduate Student, Policy On

Estimated log difference in height: 1.061 (0.106)

Panel B

McCrary Density Plot, Graduate Student, Policy Off

Estimated log difference in height: -0.104 (0.128)

Graph represents percent of students with different levels of visit counts (dots) and the fitted distribution. Any discontinuity at the cut-off (50 visits) indicates bunching. Vertical line indicates the 50-visit threshold that students need to reach in order to obtain the reimbursement. Bootstrap standard errors in parentheses.

Appendix Figure 2: McCrary Density Tests (continued)

Panel C

McCrary Density Plot, Undergraduate Student, Policy On

Estimated log difference in height: 0.054 (0.064)

Panel D

McCrary Density Plot, Undergraduate Student, Policy Off

Estimated log difference in height: 0.017 (0.080)

Graph represents percent of students with different levels of visit counts (dots) and the fitted distribution. Any discontinuity at the cut-off (50 visits) indicates bunching. Vertical line indicates the 50-visit threshold that students need to reach in order to obtain the reimbursement. Bootstrap standard errors in parentheses.

Appendix Table 1: Organizations and Sources of Fitness Incentive Programs

Name	Type	Website
Blue Cross Blue Shield	Health Insurer	https://mss.empireblue.com/ny/nyny_ep_gymreimbursementbrochure_eng.pdf
Aetna	Health Insurer	http://www.aetna.com/employer/commMaterials/documents/Roadmap_to_Wellness/fitness-reimbursement-member-postenroll-flyer-hcr-ny.PDF
UnitedHealthcare	Health Insurer	https://broker.uhc.com/assets/Fitness%20Reimbursement%20Flyer.pdf
Maine	State Government	https://www.maine.gov/bhr/oeh/wellness/gym-membership-reimbursement
New York	State Government	https://www.cs.ny.gov/employee-benefits
New Hampshire	State Government	https://das.nh.gov/wellness/Docs/Anthem_Approved_Fitness_Facilities_for_State_of_NH_Employees.pdf
University of Minnesota	Higher Education Institute	http://recwell.umn.edu/member-services/gym-reimbursement
Binghamton University	Higher Education Institute	https://www.binghamton.edu/campus-recreation/memberships/pilot-program.html
University of North Dakota	Higher Education Institute	https://www1.und.edu/health-wellness/workwell/programs/hcc.cfm

Appendix Table 2: Robustness Test of Bunching Estimator

	Analysis Region (Visits)		
	20 - 70	15 - 75	10 - 80
Region of Excess Mass			
50 - 60 Visits	9.545***	11.620***	9.088***
	(2.734)	(3.107)	(2.240)
50 - 55 Visits	5.047***	5.646***	5.418***
	(0.892)	(1.136)	(0.993)
50 - 57 Visits	7.142***	8.433***	7.512***
	(1.380)	(1.852)	(1.477)

Bootstrap standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Each column presents results from a separate bunching analysis.

Analysis region refers to range of visits for fitting the counterfactual distribution.

Our preferred specification uses a region of excess mass of 50-60 visits and an interval considered of 20-70 visits.