Intermittent Incentives to Encourage Exercising in the Long Run[1]

Ayala Arad[a], Uri Gneezy[b] and Eli Mograbi[a]

September 2020

Abstract

We report the results of incentivizing students to exercise. We compare a no-incentive control to a per-visit payment, and two intermittent incentive schemes: monetary rewards at increasing intervals and monetary rewards with unpredictable timing. Irregular incentives, as already recognized in the psychology literature, may facilitate the maintenance of reinforced behavior over the long run. In line with this prediction, we find that while all incentive schemes worked well during the incentivized period, only the two intermittent schemes were more effective than the control after incentives were removed and over an extended period of time.

[1] We acknowledge support from the US-Israel Binational Science Foundation (BSF) and the NSF. We thank Neta Klein, Aviv Halak, May Kenet, Guy Tikochinsky, and Itamar Amir for their research assistance, and the Tel Aviv University sports center, especially Pavel Kolosovsky, Shiri Eitan, and Yael Keilin, for their cooperation. We also thank Adi Shany, Sally Sadoff, Dotan Persitz, Yair Antler, and participants at the Coller Conference on Behavioral Economics, Advances with field experiments conference and the ESA European conference for useful comments. This study is registered in the AEA RCT Registry and the unique identifying number is: AEARCTR-0005873. The online appendix is available at: http://www.tau.ac.il/~aradayal/Online appendix Intermittent incentives.pdf
[a] Coller School of Management, Tel Aviv University, Sderot Henry Gildred, Tel Aviv-Yafo, Israel.
[b] Rady School of Management, UC San Diego, 9500 Gilman drive, La Jolla, CA 92093, USA

1 Introduction

Can incentives be used to encourage the initiation and maintenance of good habits? In this paper, we study the effect of incentives on exercising in the short and long run, with the goal of increasing long-term impact. We develop novel intermittent incentive schemes that, unlike regular incentive schemes, are effective in the long run, after incentives are removed. Increasing exercising has far-reaching consequences in view of the fact that physical inactivity is ranked as the fourth most important health risk factor worldwide. The estimated global annual cost of physical inactivity is $54 billion in direct healthcare expenditure and another $14 billion in productivity loss (WHO, 2018).

A possible reason for physical inactivity is that individuals simply do not want to exercise. Alternatively, people may want to exercise but fail to do so. This may have multiple causes, with a major one being a lack of self-control which is often described as present-biased preferences (DellaVigna and Malmendier 2006; Laibson, 1997; and O'Donoghue and Rabin, 1999). In these cases, incentives that increase utility in the present may be able to help overcome the failure to exercise (Charness and Gneezy, 2009). Moreover, incentivizing behavior in the short run might create exercising habits, as in Becker and Murphy's (1988) model in which the stock of exercise (i.e., how much one exercised in the past) may enhance the utility from exercising in the future. Thus, if individuals are incentivized to exercise in the short run, they may find exercising in the long run is easier or more enjoyable and may continue to exercise even after the incentives are removed.

Findings from the research on incentives to exercise show they are successful in motivating individuals to start exercising, but are less effective in overcoming the challenge of sustaining the effect in the long run, after incentives have been removed (Acland and Levy, 2015; Babcock and Hartman, 2010; Charness and Gneezy, 2009; Roemmich et al., 2012; Royer, Stehr, and Sydnor, 2015; see two recent reviews by Mitchell et al., 2019[2], and

[2] This review reports on only four studies that found post-intervention effects. One study might have suffered from a selection problem - the treatment group with long-term effects was willing to sign commitment contracts, which requires a larger motivation to exercise (Royer, Stehr, and Sydnor 2015). In the second, participants were incentivized for three weeks and the post-intervention effect existed for only two weeks (Condliffe, Isgin, and Fitzgerald (2017). Using a lottery incentive scheme, Petry et al. (2013) found an effect three months after the intervention ended among 20 adults per group over the age of 55. However, they only tested one particular week, three months after the intervention, and reminded the participants before the beginning of that week, possibly causing a commitment effect for the incentivized group. Finally, Rohde and Verbeke (2017) examined gym attendance among participants who had paid for a membership and thus were likely to have a strong intrinsic motivation to exercise. Following a six-month intervention, they did not find that intervention had a significant effect on the total number of visits during the first three months following the incentivized period. However, they did find less of a decrease in the number of visits (relative to the previous quarter) in the treatment group. This effect disappeared in the subsequent quarter.

Strohacker, Galarraga, and Williams, 2014). The focus of the current paper is on how to induce exercising in the long run.

Our research involved a field experiment in which students were incentivized to exercise in a university gym. All participants received a six-month free membership to the university gym. We examine their exercise frequency throughout the membership period as well as 12 and 18 months after the beginning of the experiment. We assigned participants into four groups, where three of them received monetary incentives for the first two months of the membership. Participants in the first group were not offered any incentives (*Control*) while those in the second group were offered per-visit incentives (*Per-visit*). In past studies, per-visit incentives encouraged exercising while in place, but after their removal behavior returned to baseline levels (Hardman, Horne and Lowe, 2011; Royer, Stehr and Sydonr, 2015). We used this incentive scheme in order to ensure that behavior in our setting is consistent with that observed in other studies and thus we did not expect it to have long-run effects. The third group received frequent rewards to start with, which were aimed at overcoming the high initial cost of starting to exercise, and they gradually became less frequent but larger (*Increasing*). The fourth group was offered incentives on a random basis, such that not every visit was rewarded. In order to limit the amount of uncertainty, the randomness was introduced by way of the number of visits required to receive the reward (1, 2, 3 or 4), rather than the standard lottery scheme used in previous studies (*Unexpected*). Further details on the incentive schemes and comparison to past studies can be found in Section 2.3.

Although regular incentive schemes may crowd out the intrinsic motivation to exercise (Gneezy, Meier and Rey-Biel, 2011), intermittent incentives such as those provided to the *Increasing* and *Unexpected* groups may succeed—due to their irregular nature—in establishing the habit of exercising without creating an enduring association between it and the monetary incentive. This is achieved by not getting the participants accustomed to receiving a reward after every single visit, with the goal of encouraging the maintenance of the exercise habit in the long run, after removal of incentives.

The idea of intermittent rewards builds on a strand in the psychology literature. Ferster and Skinner (1957) introduced the concept of a schedule of reinforcement and demonstrated experimentally, using lab animals, that changing the reinforcement schedule affects the initiation and maintenance of the desired behavior. In light of the empirical evidence suggesting intermittent schemes lead to better maintenance of the reinforced behavior after the reinforcement is removed, Amsel (1958) developed what is known as frustration theory, whereby the use of intermittent incentives leads individuals to hold two contradictory

expectations simultaneously: that of being rewarded and that of not being rewarded. Individuals continue to pursue the reinforcement because the anticipation of a reward is greater than the anticipation of no reward. With repeated experience, they learn to cope with the frustration of sometimes not being rewarded for a proper response. Thus, the reinforced response also becomes conditioned in the situation of no reward, increasing the likelihood that the behavior that was reinforced will continue after the incentives are removed (Amsel, 1992; Pittenger, 2002; Papini, 2003; Domjan, 2010).

The phenomenon whereby intermittent rather than regularly scheduled incentives maintain a behavior for a more extended period is known as the partial reinforcement extinction effect or PREE (Hochman and Erev, 2013; Humphreys, 1939; Pittenger, 2002). In a number of laboratory experiments, participants who received continuous reinforcement for some initial period (by being rewarded for every choice they made) were less likely to keep playing after "failures" in which they did not receive a reward, compared with participants who had received intermittent reinforcements in the initial period (Golz, 1992; Hogarth and Villeval, 2014; see Pittenger, 2002, for a review).

Our two intermittent schemes, which have not been studied before, proved to be successful in the long run in our setting. In contrast, the regular incentive scheme was more effective than the control only during the incentivized period. These findings demonstrate the importance of examining the effectiveness of incentives over the relevant horizon, since not all incentives that work while in place are also successful in the long run.

2 Experimental Design

The participants consisted of 213 students attending Tel Aviv University who were selected after a pre-screening process. They were offered the chance to participate in an experiment aimed at increasing their physical activity. Participants received a six-month membership (January-June 2018) to Tel Aviv University sports center. Financial incentives for exercising were provided only for the first two months (January-February 2018), allowing us to observe whether the effects of the incentives persisted during the subsequent four months. We also observed whether participants continued to exercise at two six-month intervals, after the free membership had ended (12 and 18 months after the start of the experiment).

2.1 Recruitment

Past research (e.g., Charness and Gneezy, 2009) usually found the most significant effects among participants who did not previously exercise, which guided us in the recruitment. To this end, all university students were invited to answer a short 3-minute lifestyle questionnaire that was distributed online (see the online appendix). To incentivize the students to answer the questionnaire, they were told that 5% of them would be randomly selected to receive 50 NIS (at the time of the experiment, 3.6NIS=$1). The questionnaire included several questions about lifestyle, including filter questions used to determine who was eligible to participate in the experiment.

The students filling out the questionnaire did not know the details of the experiment and its incentives and thus did not have an incentive to lie. We designed the selection criteria to identify individuals who (1) did not exercise at all or exercised only once a week but not in a gym and not swimming, (2) had a commute time of up to 120 minutes from their residence to the university (participants who live closer to the gym can more easily exercise there), and (3) wanted to exercise more. Initially, our criteria were supposed to be stricter, with criterion (1) being not exercising at all, and criterion (2) being a commute time of up to 30 minutes. However, we did not have enough students who answered our stricter criteria. Of the 1,115 participants who completed the pre-study questionnaire, we invited 420 to participate in the experiment based on the lenient criteria.

We then sent the selected students invitations to participate. We told them that based on their answers to the lifestyle questionnaire, they were eligible to participate in a research project aimed at encouraging them to exercise in the university gym and that they would be given a free six-month membership (a value of 1,500 NIS). Of the 420 students invited to participate, 229 showed up to the introductory session. Of the 229 that showed up, 10 were removed and 6 dropped out, leaving us a total of 213 participants (see explanations below).

2.2 Introductory session

Participants attended a 90-minute introductory session. The 29 sessions were held between December 17 and December 28, 2017, in the Interactive Decision-Making Lab at Tel Aviv University. Each session was attended by participants from the same treatment in order to minimize the chance of exposure to participants from other treatments. We balanced the treatment assignment according to age, gender, and commute time (see Table 1 below, which

includes all 213 participants), resulting in groups of between 50 and 55 participants in each treatment.

Before being informed about the incentives, participants were asked to sign a consent form to participate in the experiment; 223 agreed to sign the consent form (six participants dropped out at this stage). By signing the form, participants gave consent not only to participate in the study, but also for us to access their future grade transcripts, their Israeli SAT scores, and their matriculation exam scores. After signing the consent form, participants received the experiment's instructions for the treatment to which they were assigned. The instructions were read aloud to them and fully explained. The instructions included the incentive they would be offered to exercise, how to use the mobile app for the experiment, how rewards would be distributed during the experiment, and so on. Participants also filled out a medical questionnaire, as required by Israeli law, in order to ensure exercising was safe for them.

Table 1: Balanced assignment to treatments

Variable \ Treatment	Control	Per-visit	Increasing	Unexpected
Participants #	50	54	55	54
Age	23.77 (2.32)	24.32 (2.77)	23.68 (2.91)	24.36 (4.17)
Commute time (minutes)	36.46 (28.97)	38.13 (26.27)	38.4 (27.58)	35.81 (26.21)
Gender (Male%)	0.36 (0.48)	0.33 (0.48)	0.35 (0.48)	0.33 (0.48)

Table 1: Group means of the variables used to balance the groups. Standard deviations appear in parentheses. Asterisks indicate a difference of means (compared to the *Control* group) at *0.1 significance, **0.05 significance, and ***0.01 significance.

Participants then answered the following psychological questionnaires: (1) the Propensity to Plan scale (Lynch et al.,2010), which measures an individual's tendency to plan, a trait that is typically necessary in order to exercise on a regular basis; (2) CFC—Consideration of Future Consequences (Strathman, et al., 1994), which estimates the extent to which participants take into account future consequences, and in this case, those of the decision to exercise; (3) DOSPERT—Domain-Specific Risk-Taking (Blais, and Weber 2006; Weber, Blais, and Betz, 2002), which measures risk tolerance and may be relevant in this context because one of the intermittent incentive schemes in the study involved a certain degree of risk; and (4) a happiness questionnaire, based on a subset of questions from the Oxford Happiness Questionnaire (Hills and Argyle 2002), that measures differences in levels of happiness

between people who exercise and people who do not. Participants also answered additional questions about general lifestyle. While participants were answering the questionnaire, they were taken out one by one to a separate room where a nursing student measured their medical indicators (pulse, weight, and body fat percentage) using non-invasive devices—an Omron Body Composition Monitor BF511 for weight and fat percentage, and an Omron M3 device for pulse and blood pressure.

Table 2 shows no major or significant differences in these indicators between the different treatments, suggesting the aforementioned assignment to treatments induced balanced groups. The table includes the 213 participants who participated in our study. The measurement of the psychological and physiological indicators during the introductory session were repeated at the end of the six-month sports center membership.

Table 2: Balanced assignment to treatments – other variables

Variable \ Treatment	Control	Per-visit	Increasing	Unexpected
CFC	3.69 (0.57)	3.65 (0.57)	3.54 (0.52)	3.44 (0.58)
Propensity to plan	56.4 (13.69)	55.25 (15.55)	56.2 (14.41)	56.81 (12.15)
Risk	99.44 (21.48)	91.39 (17.36)	96.89 (19.29)	91.94 (17.24)
Happy	45.06 (6.69)	43.87 (7.86)	44.18 (7.49)	45.77 (6.35)
Pulse	74.94 (9.31)	75.88 (12.33)	75.21 (10.59)	76.02 (11.97)
Weight	65.21 (14.99)	65.47 (13.86)	67.77 (20.17)	64.55 (14.74)
Body fat	27.21 (8.78)	29.34 (10)	30.45 (8.96)	29.29 (9.7)

Table 2: Group means of variables measured during the introductory session. Standard deviations appear in parentheses. Asterisks indicate the significance of the differences between the means (compared to the *Control* group): *0.1 significance, **0.05 significance, and ***0.01 significance.

Participant ID numbers were reported to the university sports center for registration purposes. In line with our inclusion criteria, from the 223 participants, 10 participants who were members of the sports center during the previous year (2017) were paid 100 NIS and removed from the study, leaving a total of 213 students.

2.3 Treatments

To facilitate comparison, we set the mean reward per-visit to the gym at 20 NIS for the three incentivized treatments.

We implemented the following incentive schemes:

1. *Control*: No monetary incentives were given to visit the gym.
2. *Per-visit*: Participants received 20 NIS for each visit to the gym.
3. *Increasing*: The number of visits required to receive the next reward and the size of the next reward increased after the payment of each reward. A reward of 20 NIS was paid after the first visit, a reward of 40 NIS after two additional visits (i.e., after the third visit), a reward of 60 NIS after three additional visits (i.e., after the sixth visit), and so on. Participants were aware of this setup, and information on the timing of the next reward and its size was available to them from the mobile app. Under this incentive structure, participants can more easily receive a reward early on, when the need to compensate for the high initial costs of exercising is bigger. Over time, the reward is paid less frequently, and participants gradually get used to visiting the gym without a reward. In other words, the incentives are frequent enough to establish the exercise habit, but not to the extent that it becomes strongly associated with receiving a reward. This is accomplished by means of the reward's decreasing frequency, thus mitigating crowding-out effects.

Our study is the first to use an incentive scheme in which participants must exercise more each time in order to receive the next reward. A recent study by Bachireddy et al. (2019) had an incentive scheme which they also referred to as *Increasing*. However, in their case the size of the reward increased every two days regardless of the participant's behavior. In addition, their reward was paid out on a daily basis, based on an average of $0.0002 per step measured by a pedometer. Thus, the amount of exercising required to obtain a reward does not increase, as it does in our study. Their setup for the increasing-reward scheme is not likely to mitigate the crowding-out effect because exercising is rewarded every day, and indeed participants in this group did not exercise more than those in other groups during and after the incentivized period.

4. *Unexpected*: Participants received 50 NIS after X visits to the gym, where X is a number between 1 and 4 and determined randomly by lottery. After a participant in this group received a reward, a new random number was drawn, and a new count of visits began. Participants knew only that a number between 1 and 4 had been drawn, but did not know which and therefore could not know exactly when they would receive the reward. Previous studies

such as Andrade et al. (2014), Patel et al. (2018), and Wing et al. (1996) incentivized individuals by rewarding them with a lottery ticket each time they exercised. However, the lottery left participants with uncertainty as to whether and when they would receive a reward. In contrast, because X was bounded in our scheme, participants knew they would always get a reward by the fourth visit, thereby eliminating part of the uncertainty that existed in previous studies. The purpose of this feature was to help participants begin forming the habit of exercising by providing them with certainty that they would eventually receive the monetary reward and by avoiding the frustration over not receiving a reward for an overly long period. On their first visit, these participants received 20 NIS in order to overcome any feelings of uncertainty that they would indeed receive a reward, and to help establish the researchers' credibility among the students. After that visit, the first X was determined by lottery.

At the end of the experiment, we discovered that due to a technical error, the lotteries had drawn a random X of between 1 and 3 rather than 1 and 4. Nonetheless, participants likely believed they were receiving a reward once every four visits when they decided whether to visit the gym during the incentivized period. Either way, despite the (unintentional) offer of a higher expected value, the results reported below show this incentive scheme had the weakest effect among the three treatments during the incentivized period.

In all treatments, no more than one entrance to the gym per day was counted. Thus, the maximal number of visits eligible for a reward is 59, the number of possible days during the incentivized period (January–February 2018).

2.4 Sports center membership

As part of the membership, each participant received one personal training session, during which they received an exercise program suited to their needs. The goal of providing an individualized program was to enhance the exercise's effectiveness (Jeffery, 2012).

The experiment had an associated mobile app (a customized website for smartphones). A direct link to the app was uploaded to all of the participants' smartphones during the introductory session. The app contained information regarding the participant's exercise program and membership, including the number of visits, number of visits in the previous week, the rewards that had already been received, the size and timing of their next reward, and so on. Each participant also received a text message on the last day of every week telling them how many times they had exercised in the previous week and whether they had reached their weekly goal (which was set at the recommended three visits). For example, a text might read

as follows: *"You exercised at the gym twice this week. Good job! You can do even better next week. The recommendation is to exercise at least 3 times a week."*

Visits to the gym were recorded by chip swipes at the entrance to the sports center and to the gym. An employee of the sports center at the entrance verified that the chip used belonged to the individual that swiped it (a picture of the member pops up when the chip is swiped). This system applies to all members of the sports center.

When participants were eligible for a reward for a particular visit, they received a text message approximately 15 minutes after they swiped the chip. We set this delay so participants would still be at the gym if they were actually exercising, in order to prevent participants from swiping the chip to enter, receiving the reward, and leaving immediately. An example of such a message follows: *"You are entitled to a reward of 20 shekels for exercising at the gym today. You can pick up the money at the gym reception desk. In order to receive the payment, please enter the app or the link in order to confirm receiving the money."*

After receiving this message, participants had until the end of the day to pick up their reward from the gym reception desk. If a participant forgot to collect the payment on the same day, he or she could pick it up from the experiment's administrator.

During the four months following the incentivized period (March–June 2018), participants could still access the gym and use the app, and they continued to receive weekly text messages, but they did not receive any rewards. Of the 213 participants who started the experiment, eight canceled their participation at some point during the six months: three from *Control*, one from *Per-visit*, two from *Increasing*, and two from *Unexpected*. Thus, 205 participants remained in the study until its completion.[3]

2.5 Post-membership period

During the first two weeks of June 2018, all the participants were invited to a concluding lab session similar to the introductory one held in December, and received 100 NIS for attending. During the session, participants answered the same questions as in the original questionnaire, along with some additional ones (see the online appendix). One hundred seventy-one participants attended the concluding session (roughly 85% of the remaining participants).

[3] Why they decided to drop out is unclear. We counted them as participants who did not exercise or attend the gym during the membership period. They were not included in the two follow-ups, because they did not answer the questionnaires. The qualitative results hold regardless of whether we include these participants.

Participants were sent two short follow-up questionnaires in which they were asked whether they had continued to exercise and where. The first online questionnaire was sent in January 2019, 12 months after the beginning of the incentivized period, and the second was sent in June 2019, 18 months after the beginning of the incentivized period and prior to the university's exam period. The first questionnaire had a response rate of roughly 95% (194 responses), whereas the second had a response rate of 92% (189 responses). Participants received 50 NIS for answering each questionnaire.

3 Results

The analysis follows the plan outlined in our proposal of the Binational Science Foundation (BSF) grant supporting this research, called "Intermittent Incentives for Encouraging Physical Activity."[4] As planned, the results are discussed according to time period.

During the incentivized period (months 1-2), participants in the three incentivized treatments visited the gym more often than participants in *Control*. During the non-incentivized period (months 3-6), participants in the two intermittent-schemes treatments exercised more than the *Control* group, whereas participants in the *Per-visit* group did not exercise more than the *Control* group. We also analyzed the 12- and 18-months follow-up questionnaires and found that only participants in the *Unexpected* treatment exercised more than participants in *Control*.[5] Thus, over time, after incentives were removed, the *Per-visit* incentive scheme became inferior. The *Unexpected*, although not performing as well as the other schemes during the incentivized period, became superior in the long run.

For the dependent variable of the number of visits during the membership period, no more than one entrance to the gym per day was counted, as in the payment of rewards. We omitted outliers (mean ±3 std of the dependent variable). We estimate the effect of treatment relative to the *Control* group using a Tobit model. That model was chosen in view of the large number of zero observations in our sample, which represent individuals who did not exercise at all.[6] Our main specification includes the control variables of gender and commute time.[7] It is augmented with two additional specifications, one with a treatment-gender interaction and

[4] www.267c5708-a14e-40fe-aa99-aa8d5d15b476.filesusr.com/ugd/08b008_4f13793501f04c67b60f497e54283dc3.pdf
[5] Eventually the planned 8 month's follow-up questionnaire was not carried out due to budget constraints.
[6] We repeated the analysis using an OLS model as a robustness check, and report it in the online appendix.
[7] As previously mentioned, we were planning to screen participants based on a commute time of less than 30 minutes. Since this would have severely limited sample size, this was not done and commute time was included as a control variable.

the other without any control variables at all. The focus of the paper is to determine the long-term effects on the groups that received intermittent incentives relative to the non-incentivized *Control* group. Comparisons between any pair of incentivized treatments can be found in the online appendix. In what follows, we analyze the membership period, the follow-up, and the dynamic of exercising over time, ending with a number of robustness checks.

3.1 Membership period

Table 3 compares each of the incentivized treatments with the *Control* group in the incentivized and non-incentivized membership periods. Figure 1 depicts the cumulative distribution of the number of visits to the gym in these membership periods. Table 3 - Panel A and Figure 1 - Panel A, which relate to the incentivized period (months 1-2), show that participants in the three incentivized treatments visited the gym more than the *Control* group, with Tobit estimators ranging from 5.96 to 10.67 additional visits, depending on the incentivized treatment. The mean number of visits was 9.96 (SE=1.18) in *Increasing*, 8.19 (SE=1.12) in *Per-visit*, 6.5 (SE=1.04) in *Unexpected*, and 3.28 (SE=0.71) in *Control*.

Table 3. Visits to the gym during the membership period

	Per-visit			Increasing			Unexpected		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Months 1-2									
Treatment	8.36***	8.27***	8.71***	10.36***	10.29***	10.67***	6.00***	5.96***	6.60**
	(2.11)	(2.11)	(2.57)	(2.18)	(2.17)	(2.66)	(2.14)	(2.12)	(2.59)
Gender * treatment			-1.32			-1.14			-1.94
			(4.43)			(4.49)			(4.48)
N	103	103	103	103	103	103	103	103	103
Panel B: Months 3-6									
Treatment	3.41	3.44	4.47	6.00**	6.07**	6.72**	4.71*	4.81*	6.53**
	(2.63)	(2.63)	(3.34)	(2.67)	(2.68)	(3.40)	(2.59)	(2.59)	(3.25)
Gender * treatment			-2.76			-1.72			-4.85
			(5.43)			(5.47)			(5.40)
N	101	101	101	104	104	104	102	102	102
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Table 3: Tobit estimates. Dependent variable is the number of visits to the gym in a particular period. Each group of three columns compares one of the treatments to the *Control* group (1-3 *Per-visit*, 4-6 *Increasing*, and 7-9 *Unexpected*). Each panel represents a different time period. Standard errors appear in parentheses. Columns 1, 4, and 7 do not include controls, columns 2, 5, and 8 include controls (gender(male), and commute time), and columns 3, 6, and 9 include controls and a gender-treatment interaction. * 0.1 significance, ** 0.05 significance, ***0.01 significance.

Figure 1: Cumulative distribution of visits to the gym during membership period

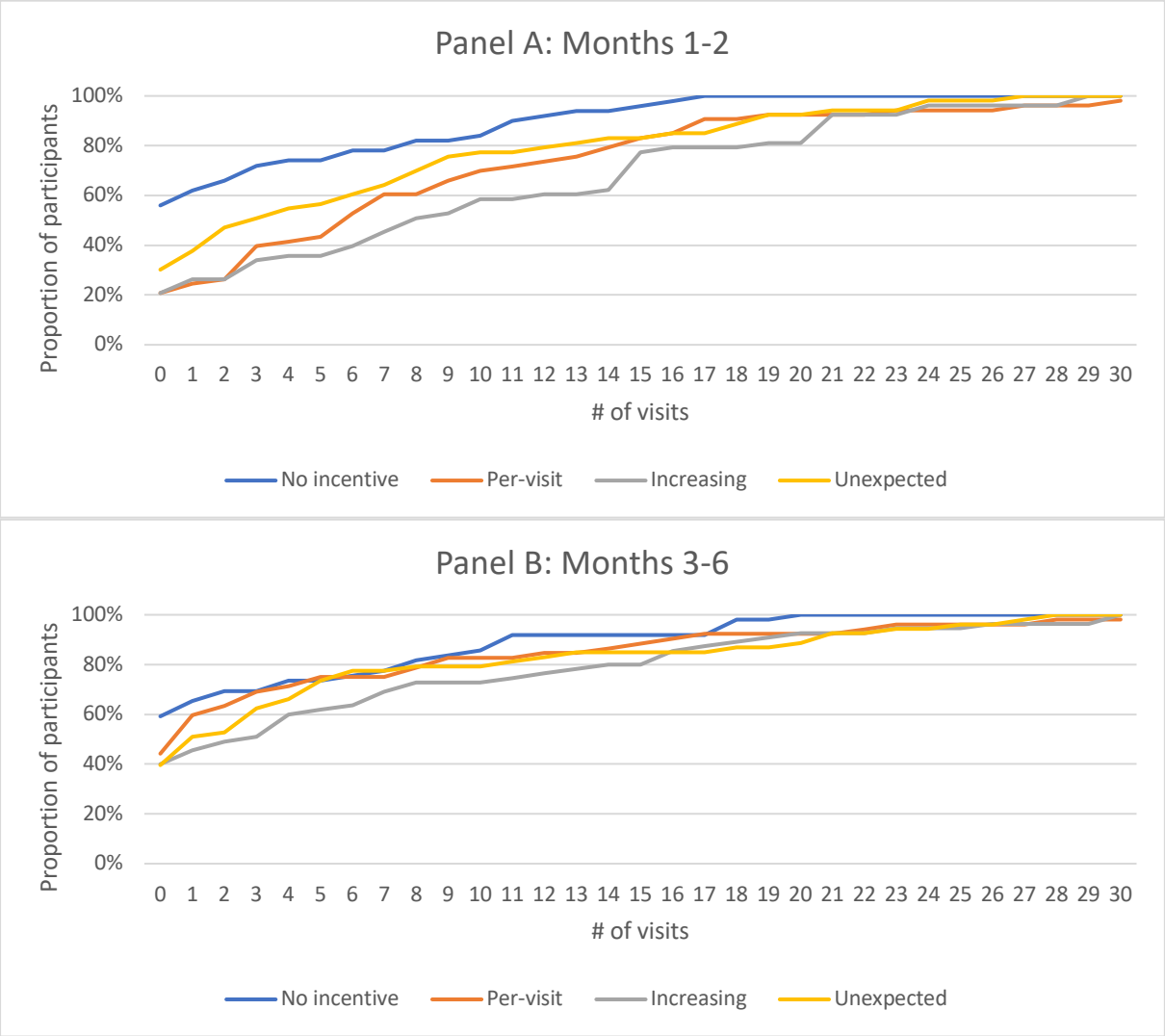

Figure 1: Cumulative distribution of the number of visits to the gym during membership period. Each panel represents a different time period.

A possible explanation for the inferiority of the *Unexpected* scheme in this period is its risk element, which is not present in the other treatments. This explanation is consistent with the finding that the more risk averse the individual, the less he or she exercised (β=0.23, SE =0.09, p=0.01) in the *Unexpected* treatment. Moreover, the difference between *Unexpected* and *Control* increases after adding a control for risk attitude (β=6.60, SE =2.13, p=0.002). Given the importance of risk attitude for behavior in the *Unexpected* group, we performed an additional analysis of the *Unexpected* scheme while controlling for risk attitude in the non-

incentivized periods as well (see Table A1). All the effects of the *Unexpected* scheme become stronger relative to the *Control* group after controlling for risk attitude.[8]

Our main interest is in behavior during the non-incentivized membership period (months 3-6). Panel B of Table 3 shows that during this period, the results for the *Per-visit* treatment were not significantly different from the *Control* group, whereas the number of visits to the gym in both the *Increasing* and *Unexpected* treatments were significantly larger than in the *Control* group (with Tobit estimators of the differences ranging from 4.71 to 6.72). Moreover, the difference between *Unexpected* and *Control* increases after adding a control for risk attitude (β=5.31, SE =2.63, p=0.04). These differences are also evident in Figure 1 - Panel B. In terms of the mean number of visits, *Increasing* exercised 3 more times than *Control* (whose members exercised 3.47 times during this period) and *Unexpected* exercised 2 more times while *Per-visit* exercised only 1.3 more times than *Control*. Qualitatively speaking, the difference between the intermittent groups and *Control* is approximately twice as large as the difference between *Per-visit* and *Control*. The findings demonstrate the divergence between the short term, in which all the incentive schemes did better than *Control*, and the long term, in which the two intermittent groups visited the gym significantly more than *Control* while the *Per-visit* group did not.[9]

Next, in order to better understand the source of the aforementioned effect, we separately examine the first and the second half on the non-incentivized membership period. In months 3-4, participants in the intermittent incentives groups exercised significantly more than those in the *Control* group: β=3.47 (SE=1.66, p=0.037) in *Increasing* and β=2.91 (SE=1.66, p=0.08) in *Unexpected*. The magnitude of the effect and the coefficient of the *Unexpected* group increased when controlling for risk preferences (β=3.34, SE=1.67, p=0.045). In contrast, the *Per-visit* group did not exercise significantly more than the *Control* group (see Table 4 - Panel A). The above differences are evident in Figure 2 - Panel A, which shows the cumulative distribution of the number of visits by treatment during months 3-4. Comparing this figure to Figure 1- Panel A reveals a downward trend relative to the behavior in months 1-2.

[8] Controlling for risk preferences did not affect the comparisons of the other incentivized treatments (*Increasing* and *Per-visit*) to the *Control*.

[9] Note that the sample size provides us with 80% power in identifying an effect of size that is f^2=0.1 or larger (with 0.05 significance).

Table 4. Visits to the gym during the non-incentivized membership period

	Per-visit			Increasing			Unexpected		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Months 3-4									
Treatment	1.84 (1.70)	1.84 (1.71)	2.73 (2.16)	3.43** (1.68)	3.47** (1.66)	2.84 (2.11)	2.80* (1.68)	2.91* (1.66)	3.47* (2.09)
Gender * treatment			-2.44 (3.56)			1.62 (3.38)			-1.55 (3.46)
N	102	102	102	104	104	104	102	102	102
Panel B: Months 5-6									
Treatment	0.60 (1.89)	0.63 (1.88)	0.50 (2.40)	3.03 (2.10)	2.83 (2.07)	4.77* (2.55)	2.54 (2.12)	2.46 (2.13)	3.95 (2.65)
Gender * treatment			0.33 (3.85)			-5.72 (4.28)			-4.37 (4.51)
N	99	99	99	103	103	103	103	103	103
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Table 4: Tobit estimates. Dependent variable is the number of visits to the gym in a particular period. Each group of three columns compares one of the treatments to the *Control* group (1-3 *Per-visit*, 4-6 *Increasing*, and 7-9 *Unexpected*). Each panel represents a different time period. Standard errors appear in parentheses. Columns 1, 4, and 7 do not include controls, columns 2, 5, and 8 include controls (gender(male), and commute time), and columns 3, 6, and 9 include controls and a gender-treatment interaction. * 0.1 significance, ** 0.05 significance, ***0.01 significance.

We continue with an analysis of the behavior in months 5-6. Panel B of Table 4 shows that participants in the *Increasing* group exercised more than those in the *Control* group, although the effect is significant only in the presence of a gender interaction (β=4.77, SE=2.55, p=0.06) due to the fact that the difference between the groups is mainly due to the female participants. Although we suspected the incentives might affect men and women differently, we did not have a specific hypothesis regarding the gender differences in the long-term influence of incentives. This larger effect could be due to men being affected more by monetary incentives (see Croson and Gneezy, 2009, for a review). Thus, their intrinsic motivation is crowded out more than for women, even when provided with intermittent incentives. The *Unexpected* and the *Per-visit* did not show significant difference from the *Control*.[10] Panel B of Figure 2 shows the cumulative distribution of the number of visits during months 5-6, revealing a similar (albeit larger) downward trend to that in Panel A of Figure 2 (months 3-4) relative to Panel A of Figure 1 (months 1-2).

[10] After controlling for risk attitude, the difference between *Unexpected* and *Control* is still insignificant.

Figure 2: Cumulative distribution of visits to the gym during the non-incentivized membership period

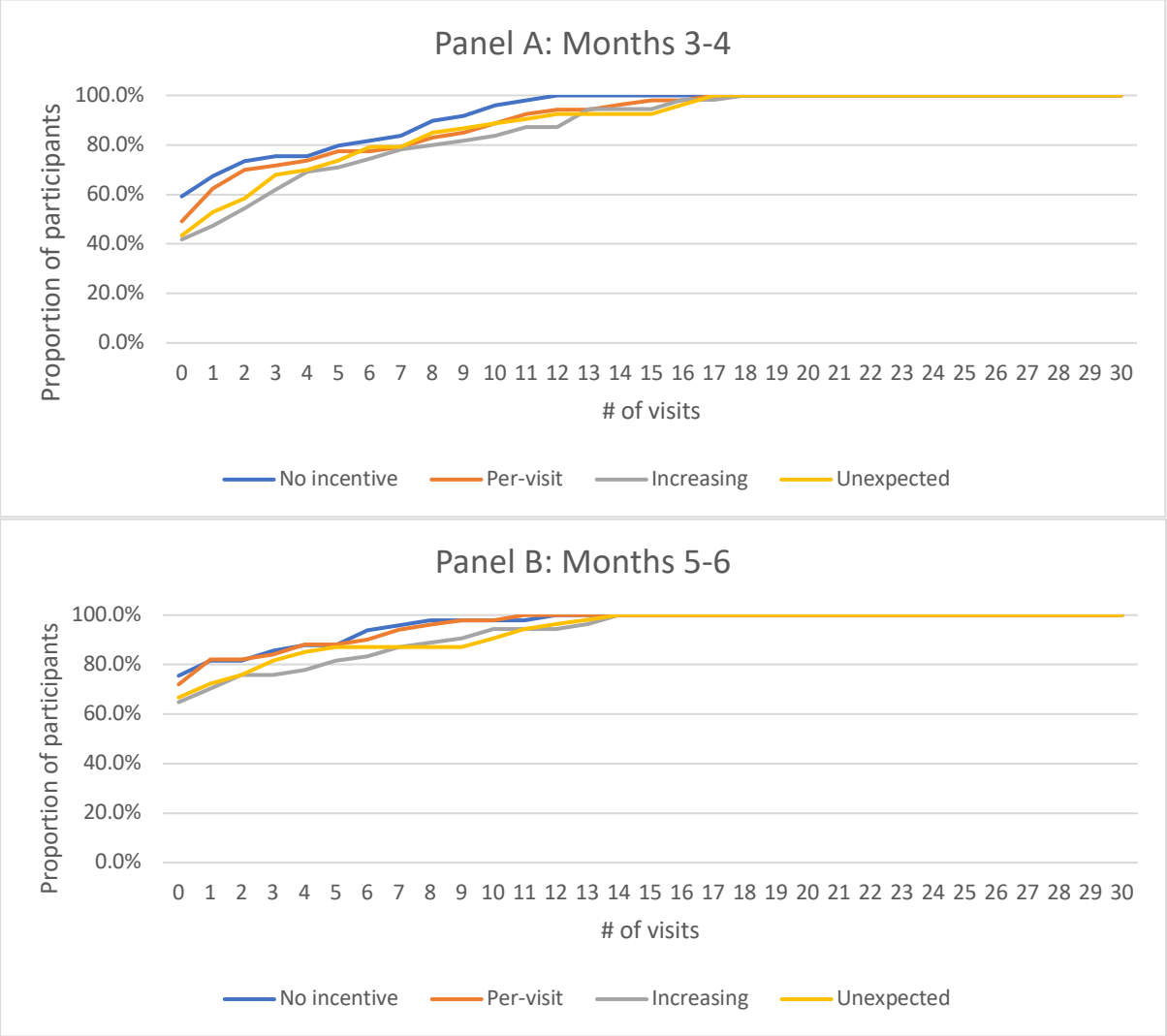

Figure 2: Cumulative distribution of the number of visits to the gym during the non-incentivized membership period. Each panel represents a different time period.

Figure 3: Number of gym visits by period

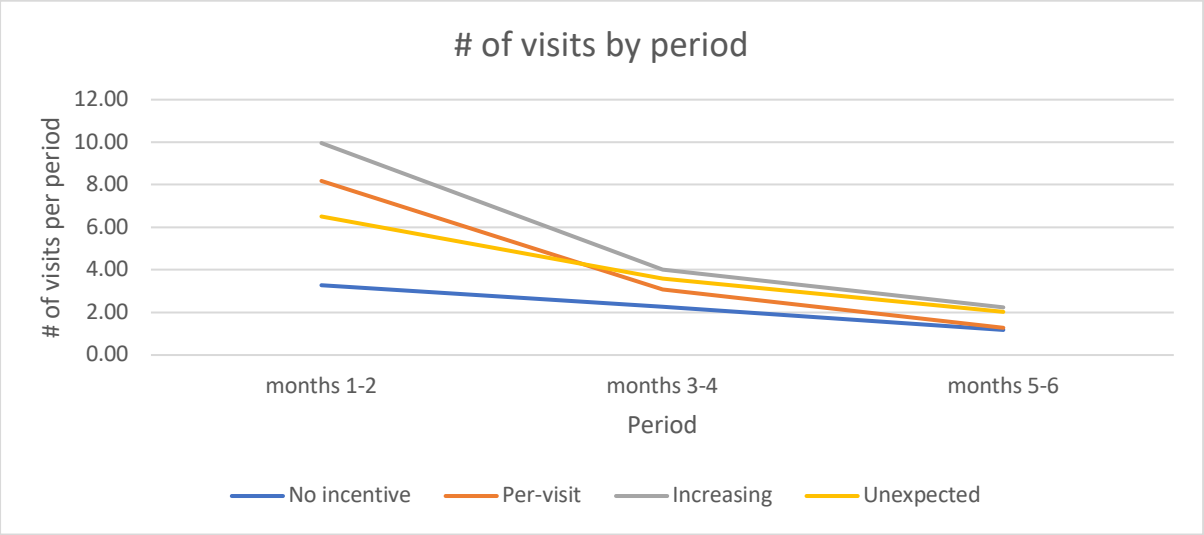

Figure 3: Number of gym visits by period.

Finally, Figure 3 summarizes the dynamic of exercising during the membership period. It shows that the number of visits declined in all four groups, with the *Per-visit* treatment showing the largest decline—down to the level of the *Control* group (only 0.1 visits more in months 5-6), and the *Unexpected* showing the smallest decline—almost at the level of the *Increasing* group (0.2 visits less in months 5-6). Overall, our analysis indicates that the success of the intermittent incentives in the non-incentivized membership period (months 3-6) is driven by both their effect in months 3-4 and their effect in months 5-6, although the former was larger.

In conclusion, the *Per-visit* scheme is effective only in the short term; the *Increasing* scheme is effective in both the short term and the long term; and the *Unexpected* scheme is the least effective during the short term although its effect declines the least over time, a pattern that becomes important in the 12- and 18-month follow-ups.

3.2 Follow-up

We used the 12- and 18-month follow-up questionnaires to determine whether participants continued to exercise after the free six-month sports center membership had expired. Participants received 50 NIS to answer each questionnaire. Both questionnaires had a high response rate that was similar across groups (the 12-month questionnaire had a 95% response rate that ranged between 91.3% and 98.1% across groups; the 18-month questionnaire had a 92% response rate that ranged between 87.2% and 96% across groups). The main question was

"How many times a week do you exercise on average?" and we used the responses as the dependent variable in the analysis. We chose not to limit the analysis to exercising in the sports center, because only about 15% of the students continued their membership after the six-month free membership period, which also meant we no longer had any reason to use commute time as a control variable. Furthermore, some of the participants graduated by the time they received the follow-up questionnaires and therefore may have had more or less time to exercise. Therefore, they were asked to report the number of hours they work each week and the number of hours they study each week, and we added these variables as controls in the analysis in addition to the previously used gender control. Of the questions in the follow-up questionnaire, only these two questions were viable as controls. (In the other four questions of the follow-up questionnaires, one of the possible answers was "I do not exercise." Therefore, the answers to these questions are correlated with the answers to the whether-you-exercise question, leading to an endogeneity problem if these questions are used.)

Table 5 compares each of the incentivized treatments with the *Control* group in the various follow-up periods. Panel A of Table 5 shows that participants in the *Unexpected* treatment exercised significantly more than participants in the *Control* group (β=1.01, SE =0.43, p=0.018). The size of the effect increases after controlling for risk attitude (β=1.06, SE =0.43, p=0.010). In contrast, the *Increasing* and *Per-visit* did not exercise significantly more than the *Control*. Figure 4 depicts the cumulative distribution of the number of visits to the gym in the various follow-up periods. The mean number of exercise sessions per week was 1.52 (SE=0.20) in *Unexpected*, 1.24 (SE=0.17) in *Increasing*, 1.23 (SE=.16) in *Per-visit* and 1.00 (SE=0.18) in *Control*.

Thus, whereas the effect of the incentives for participants in the *Per-visit* and *Increasing* groups diminished considerably over time, for participants in the *Unexpected* scheme there was less of an effect during the incentivized period, but it diminished less subsequently. This difference between the incentivized groups also mitigates the concern regarding misreporting, which should not be different between groups.

For the 18-month follow-up questionnaire, comparing Panel A to Panel B in Figure 4 shows that participants in all four groups exercised more after 18 months than after 12 months. This improvement can be attributed to the different seasons, because individuals in Israel tend to exercise less in the winter than in the summer (the two follow-ups were conducted in January and June, respectively). Panel B of Table 5 shows a marginally significant effect 18 months after the beginning of the experiment, whereby participants in the *Unexpected* treatment

exercised more than participants at the *Control* group per week (β=0.72, SE =0.40, p=0.07). The difference between the groups increased after controlling for risk attitude (β=0.84, SE =0.40, p=0.035). The diminished effect was due to an increase in the frequency of exercise in all groups, rather than a decrease in the *Unexpected* treatment. There was no significant difference between the other incentivized treatments and *Control*. The mean number of exercise sessions per week was 1.65 (SE=0.19) in *Unexpected*, 1.56 (SE=0.18) in *Per-visit*, 1.4 (SE=0.17) in *Increasing*, and 1.21 (SE=0.19) in *Control*.

Table 5. How many times do you exercise each week on average?

	Per-visit			Increasing			Unexpected		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: 12 months follow-up									
Treatment	0.37 (0.41)	0.45 (0.43)	0.41 (0.55)	0.37 (0.42)	0.39 (0.44)	0.27 (0.54)	0.74* (0.44)	1.01** (0.43)	0.66 (0.55)
Gender * treatment			0.10 (0.85)			0.31 (0.87)			0.88 (0.88)
N	94	94	94	93	93	93	93	93	93
Panel B: 18 months follow-up									
Treatment	0.54 (0.38)	0.57 (0.38)	0.56 (0.48)	0.29 (0.39)	0.32 (0.39)	0.17 (0.49)	0.63 (0.39)	0.72* (0.40)	0.78 (0.52)
Gender * treatment			0.02 (0.78)			0.41 (0.82)			-0.15 (0.84)
N	91	91	91	91	91	91	89	89	89
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Table 5: Tobit estimates. The dependent variable is the answer to the survey question "How many times do you exercise each week?" Each group of three columns compares one of the treatments to the *Control* group (1-3 *Per-visit*, 4-6 *Increasing*, and 7-9 *Unexpected*). Each panel represents a different time period. Standard errors appear in parentheses. Columns 1, 4, and 7 have no controls, columns 2, 5, and 8 include controls (gender(male), time spent studying, and time spent working), and column 3, 6, and 9 include controls and a gender-treatment interaction. *0.1 significance; **0.05 significance; ***0.01 significance.

The findings of the long-term follow-ups are encouraging in view of the fact that the exercise effect persisted more than one year after the beginning of the experiment (ten months after the end of the incentivized period).

Figure 4: Cumulative distribution of weekly exercise frequency by period

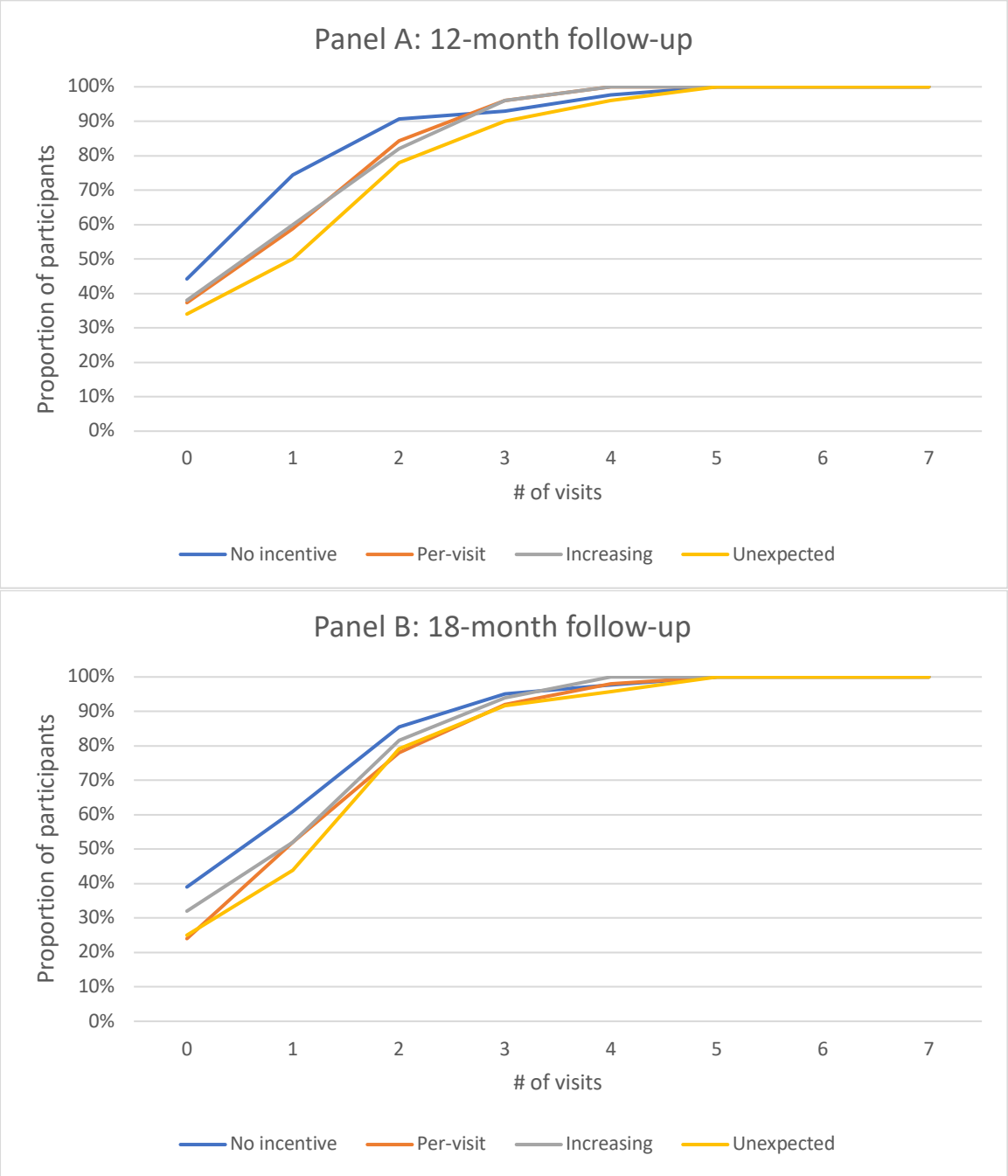

Figure 4: Cumulative distribution of the answer to the question "How many times do you exercise each week on average?" Each panel represents a different time period.

3.3 Persistence of incentive effects

Figure 5 presents the mean standardized Z-score for number of visits [Z=(x-mean)/σ], for each group in each period (months 1-2, months 3-4, months 5-6, 12-month follow-up, and 18-month follow-up), and shows the difference in trends between the treatments. We used Z-scores

because behavior was measured using different scales during the membership period and the follow-up period: for the membership period, we used the number of visits to the gym, although the graph looks similar when we use the number of entrances to the sports center (see the online appendix). Further explanation regarding the number of entrances to the sports center can be found in section 3.4 below. For the follow-up period, we used the answer to the question "How many times do you exercise each week on average?"

Figure 5: Standardized Z-score by period for visits to the gym

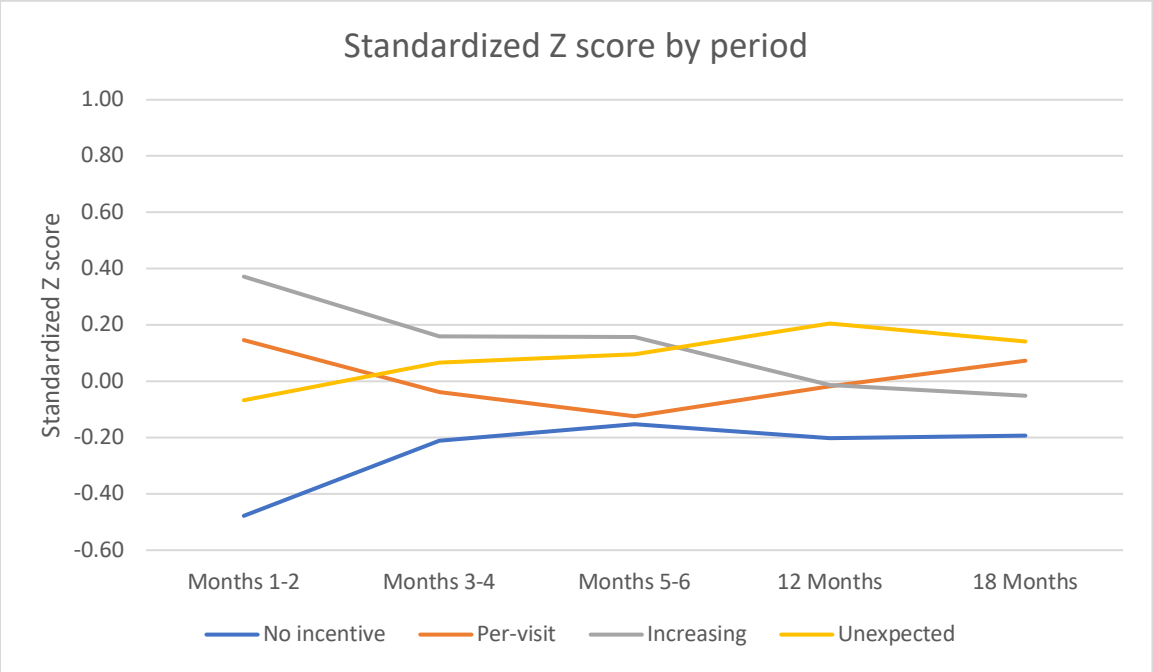

Figure 5: Standardized Z-score by period for visits to the gym and follow-ups.

Figure 5 indicates that the *Control* group had the lowest Z-score during the incentivized period. After the incentives were removed, the *Control* group's Z-score increased and then remained stable. Participants in the *Per-visit* treatment exercised to a large extent when incentives were present but suffered from a sharp decrease in gym visits after the incentives were removed. Participants in the *Increasing* treatment visited the gym more than the other groups during the incentivized period and the non-incentivized membership period. However, over time, after the membership period, participants in the *Unexpected* group visited the gym more. The effect of the *Unexpected* treatment, which was the least effective among the incentivized groups in getting participants to start exercising, steadily improved relative to the other treatments. Thus, although it was less effective in getting the participants to start exercising in the short term, it led to greater exercise perseverance in the long run.

3.4 Additional analysis

This section presents the results of several robustness checks and an analysis of additional data collected in the experiment.

Exercised or not

A different way to determine whether the incentives were able to encourage exercise is to use the metric of whether participants exercised at least once per period. Thus, we use a logistic regression to examine the effect of the treatment on whether an individual exercised at least once during each period. Although in the pre-study questionnaire the participants indicated they wanted to exercise more and although they had attended a one-hour instruction session, a proportion ranging from 20% to 60% of each group did not exercise at all during the incentivized period (Figure 1 - Panel A). The highest proportion (60%) occurred in the *Control* group, demonstrating that free membership was insufficient to get most of this group exercising. By contrast, the proportions of participants who did not exercise at all in the other three incentivized treatments is between 20% and 30%. Panel A of Table 6 shows that during the incentivized period, a larger proportion of participants in all the treatments visited the gym at least once relative to the *Control* group (odds ratio ranging from 2.94 to 5.00 times more participants).

Panel B of Table 6 presents the results of the analysis for the non-incentivized period (months 3-6). During this period, 2.18-3.35 times more participants exercised in the *Increasing* or *Unexpected* treatment than in the *Control* group, and there were no significant differences between the *Per-visit* group and the *Control* group. These results, like the previous ones, show a rapid decline in exercising under the *Per-visit* incentive scheme, which demonstrates its inferiority in the long run, after incentives are removed.

Sports center entrances

We next consider spillover effects. Besides use of the gym, the sports center membership also includes swimming, yoga, Pilates, and other activities. Although participants were only incentivized to attend the gym, they may also have taken advantage of the other activities at the sports center. Thus, we performed a similar analysis to the one described above, with number of entrances to the sports center instead of number of visits to the gym as the dependent variable. Every visit to the gym is also a visit to the sports center (the gym is inside the sports center), but not every visit to the sports center was necessarily a visit to the gym. During the

incentivized period, the differences in number of visits within each group were between 0.2 and 0.4, a small and similar amount demonstrating that participants in the non-incentivized group did not replace the gym with other activities due to the lack of incentives. Overall, the results remained quite similar and are available in the online appendix.

Table 6. Did the participant visit the gym at least once?

	Per-visit			Increasing			Unexpected		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Months 1-2									
Treatment	4.85*** [1.58] (0.44)	4.85*** [1.58] (0.45)	5.00*** [1.61] (0.56)	4.85*** [1.58] (0.44)	4.81*** [1.57] (0.44)	3.82** [1.34] (0.54)	2.94*** [1.08] (0.41)	2.94*** [1.08] (0.42)	3.49** [1.25] (0.53)
Gender * treatment			0.91 [-0.09] (0.94)			2.03 [0.71] (0.96)			0.62 [-0.47] (0.87)
N	103	103	103	103	103	103	103	103	103
Panel B: Months 3-6									
Treatment	1.82 [0.60] (0.40)	1.84 [0.61] (0.41)	2.12 [0.75] (0.52)	2.18* [0.78] (0.40)	2.32** [0.84] (0.41)	2.08 [0.73] (0.51)	2.20** [0.79] (0.40)	2.25** [0.81] (0.41)	3.35** [1.21] (0.52)
Gender * treatment			0.68 [-0.39] (0.86)			1.38 [0.32] (0.87)			0.33 [-1.12] (0.86)
N	101	101	101	104	104	104	102	102	102
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Table 6: Logistic regression. The dependent variable is whether participants visited the gym at least once during a particular period. Each group of three columns compares one of the treatments to the *Control* group (1-3 *Per-visit*, 4-6 *Increasing*, and 7-9 *Unexpected*). Each panel represents a different time period. The first line presents the odds ratio, the B value appears in square brackets, and standard errors appear in parentheses. Columns 1, 4, and 7 have no controls, columns 2, 5, and 8 include controls (gender(male), and commute time), and columns 3 ,6, and 9 include controls and a treatment-gender interaction. *0.1 significance, **0.05 significance, and ***0.01 significance.

Physiological and psychological indicators

Finally, we calculated the changes in the physiological and psychological indicators between the introductory session and the concluding session (body fat, weight, pulse, consideration of future consequences, propensity to plan, risk preferences, and happiness), and compared these changes across treatments (controlling for these variables did not change the results). Eighty-two percent of the participants attended the concluding session, with substantial differences across groups (74.5% in *Control*, 84.9% in *Per-visit*, 92.5% in *Increasing*, and 78.8% in

Unexpected). Although for a small proportion of the comparisons, the differences between the treatments were mildly significant, their interpretation is problematic considering the large number of tests and the different attendance rates across treatments. The full results appear in the online appendix.

Pooling the results of all the treatments together to examine the effect of exercising on the seven measures mitigates the above concerns. We found that an increase of one visit to the gym during the non-incentivized period (the last four months before the follow-up session) reduced the participant's body fat percentage by 0.11 percentage points (β=0.11, SE =0.04, p=0.005) and improved the propensity to plan by 0.25 points (β=-0.25, SE =0.09, p=0.005).

These results indicate participants who exercised more lowered their body fat percentage as expected (we found no differences in weight, which was expected due to the increase in muscle mass), and also improved their ability and propensity to plan. We expected the last improvement, because exercise requires planning (if one plans to go to the gym tomorrow, one needs to arrange the day accordingly), thereby assisting in forming a habit of planning.

4 Conclusion

Can intermittent incentives mitigate the crowding-out effect associated with standard incentive schemes and induce individuals to both establish a habit and maintain it after the incentives are removed? We attempt to answer this question by examining the short- and long-run effects of two novel intermittent incentive schemes and a per-visit incentive scheme on exercising. Although all three incentives schemes were effective while in place, only the *Increasing* and *Unexpected* incentive schemes induced more exercising than in the *Control* group after incentives were removed. The *Unexpected* treatment's effect lasted for 10 months (or more) after the end of the incentivized treatment and the results are even more pronounced after controlling for risk attitude. Together the results indicate that whereas the *Per-visit* scheme works well in the short run, the intermittent schemes are superior in the long run.

The results point to an innovative way to incentivize participants, which can potentially improve the long-run success of interventions. The results are particularly important in designing incentives-based policies to encourage the creation of new habits. More generally, the findings indicate we should be cautious when judging the effectiveness of incentives. If the

horizon we are examining is different from the horizon we care about, testing the intervention over only a short period of time may produce misleading conclusions.

5 Bibliography

Acland, Dan., and Levy, Matthew. (2015). "Naiveté, projection bias, and habit formation in gym attendance". *Management Science.* Vol. 61: 146-160.

Amsel, Abram. (1958). "The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin.* Vol. 55: 102-119.

Amsel, Abram. (1992). "Frustration theory. An analysis of dispositional learning and memory". New York, NY: Cambridge University Press.

Andrade, Leonardo, F., Barry, Danielle., Litt, Mark, D., and Petry, Nancy, M. (2014). "Maintaining high activity levels in sedentary adults with a reinforcement-thinning schedule." *Journal of Applied behavioral analysis.* Vol. 47:523-536.

Babcock, Philip S., Hartman, John L. (2010). Exercising in herds: treatment size and status specific peer effects in a randomized exercise intervention, Mimeo.

Bachireddy, Chethan., Joung, Andrew., John, Leslie, K., Gino, Francesca., Tuckfield, Bradford., Foschini, Luca., Milkman, Katherine, L. (2019). "Effect of different financial incentive structures on promoting physical activity among adults: A randomized clinical trial." *Jama Netw Open.* 2(8):e199863.

Becker, Gary S., and Murphy, Kevin M. (1988). "A Theory of Rational Addiction," *Journal of Political Economy.* Vol. 96: 675–700.

Blais, Ann-Renee., and Weber, Elke U. (2006). "A Domain-Specific Risk-Taking (DOSPERT) Scale for Adult populations." *Judgment and Decision Making*. Vol. 1:33-47.

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

Condliffe, Simon., Isgin, Ebru., and Fitzgerald, Brynne. (2017). "Get thee to the gym! A field experiment on improving exercise habits." *Journal of Behavioral and Experimental Economics.* 70:23-32.

Croson, Rachel., and Gneezy, Uri. (2009). "Gender differences in preferences." *Journal of Economic Literature.* Vol. 47(2): 448-474.

DellaVigna, Stefano., and Malmendier Ulrike. (2006). "Paying not to go to the gym." *The American Economic Review.* 96(3): 694-719.

Domjan Michael. (2010). “The Principles of Learning and Behavior”. (6th ed.). Belmont CA; Wadsworth/Cengage Learning

Ferster, Charles S., and Skinner, Burrhus .F. (1957). “Schedules of Reinforcement”. Appleton-Century-Crofts, New York, NY.

Golz, Sonia M. (1992). “A sequential learning analysis of decisions in organization to escalate investments despite continuing costs or losses”. *Journal of Applied Behavior Analysis*. Vol. 25: 561–574.

Hardman, Charlotte, A., Horne, Pauline, J., and Lowe, C, Fergus. (2011). “Effects of rewards, peer-modelling and pedometer targets on children’s physical activity: A school based intervention study.” *Psychology and Health.* 26(1):3-21.

Gneezy, Uri., Meier, Stephan, and Rey-Biel, Pedro. (2011). “When and why incentives (don’t) work to modify behavior.” *Journal of Economic Perspectives.* 25(4): 191-210.

Hills, Peter. and Argyle, Michael. (2002). “The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being”. *Personality and Individual Differences.* Vol. 33: 1073-1082.

Hochman, Guy. and Erev, Ido. (2013). “The partial-reinforcement extinction effect and the contingent-sampling hypothesis.” *Psychonomic Bulletin & Review.* Vol. 20: 1336-1342.

Hogarth, Robin M., and Villeval, Marie C. (2014). “Ambiguous incentives and the persistence of effort: experimental evidence”. *Journal of Economic Behavior & Organization.* Vol. 100: 1-19.

Humphreys, Lloyd G. (1939). “The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *Journal of Experimental Psychology.* Vol. 25: 141-158.

Jeffery, Robert W. (2012). “Financial incentives and weight control”. *Preventive Medicine*. Vol. 55: S61–S67.

Laibson, David. (1997) “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics.* Vol. 112: 443–77.

Lynch, John G., Netemeyer, Richard G. Spiller, Stephen A. and Zammit, Alessandra. (2010). “A generalizable scale of propensity to plane: the long and the short of planning for time and for money”. *Journal of Consumer Research.* Vol. 37(1): 108-128.

Mitchell, Marc, S., Orstad, Stephanie, L., Biswas, Aviroop., Oh, Paul, I., Jay, Melanie., Pakosh, Maureen, T., and Faulkner, Guy. (2019). “Financial incentives for physical activity in adults: systematic review and meta-analysis.” *British Journal of Sports Medicine.* 1-12.

O'Donoghue, Ted., and Matthew, Rabin. (1999) "Doing It Now or Later." *American Economic Review.* Vol. 89: 103–24.

Rohde, Kirsten, I.M., and Verbeke, Willem. (2017). "We like to see you in the gym – A field experiment on financial incentives for short and long term gym attendance." *Journal of Economic behavioral & Organization.* Vol. 134: 388-407.

Papini, Mauricio R. (2003). "Comparative psychology of surprising nonreward". *Brain, Behavior and Evolution.* Vol. 62: 83-95.

Patel, Mitesh, S., Volpp, Kevin, G., Rosin, Roy., Bellamy, Scarlett, L., Small, Dylan, S., Heuer, Jack., Sproat, Susan., Hyson, Chris., Haff, Nancy., Lee, Samantha, M., Wesby, Lisa., Hoffer, Karen., Shuttleworth, David., Taylor, Devon, H., Hilbert, Victoria., Zhu, Jingsan., Yang, Lin., Wang, Xingmei., Asch, David, A. (2018). "A randomized, controlled trial of lottery-based financial incentives to increase physical activity among overweight and obese adults." *American Journal of Health Promotion.* 32(7):1568-1575.

Petry, Nancy, M., Andrade, Leonardo, F., Barry, Danielle., and Byrne, Shannon. (2013). "A randomized study of reinforcing ambulatory exercise in older adults." *Psychology and Aging.* 28(4): 1164-1173.

Pittenger, David J., (2002). "The Two Paradigms of Persistence". *Genetic, Social, and General Psychology Monographs.* Vol. 128: 237-268.

Roemmich, James N., Lobarinas, Christina L., Barkley, Jacob E., White, Tressa M., Paluch, Rocco., and Epstein, Leonard H. (2012). "Use of an open-loop system to increase physical activity". *Pediatric Exercise Science.* Vol. 24: 384-398.

Royer, Heather., Stehr, Mark., and Sydnor, Justin. (2015). "Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune 500 company." *American Economic Journal: Applied Economics.* Vol. 7: 51-84.

Strohacker, Kelley., Galarraga, Omar., and Williams, David M. (2014). "The impact of incentives on exercise behavior: a systematic review of randomized controlled trials". *ann. behav. med.* Vol. 48: 92-99

Strathman, Alan. Gleicher, Faith. Boninger, David S. and Edwards, Scott. (1994). "The consideration of Future consequences: Weighting Immediate and Distant Outcomes of Behavior." *Journal of Personality and Social Psychology.* Vol. 66(4): 742-752.

Weber, Elke U. Blais, Ann-Renee. and Betz, Nancy E. (2002). "A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors." *Journal of Behavioral Decision Making.* Vol. 15:263-290.

Wing Rina, R., Jeffery, Robert, W., Pronk, Nicolaas., and Wendy, Hellerstedt, L. (1996). “Effects of a personal trainer and financial incentives on exercise adherence in overweight women in a behavioral weight loss program.” *Obesity Research.* Vol. 4: 457-462.

World Health Organization (WHO). (2018). “Global action plan on Physical Activity: More Active People for a Healthier World.” Retrieved online at October 13 2019 from: https://www.who.int/publications-detail/global-action-plan-on-physical-activity-2018%E2%80%932030

Appendix

Table A1. Gym visits by months and long run follow-ups, Controlling for risk attitude

	Unexpected column 1			Unexpected column 2		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Months 1-2				Panel B: Months 3-6		
Treatment	6.48***	6.60***	7.28***	5.24**	5.31**	6.95**
	(2.15)	(2.13)	(2.58)	(2.64)	(2.63)	(3.26)
Gender * treatment			-2.09			-4.66
			(4.41)			(5.37)
N	103	103	103	102	102	102
Panel C: Months 3-4				Panel D: Months 5-6		
Treatment	3.28*	3.34**	3.83*	2.80	2.75	4.26
	(1.69)	(1.67)	(2.08)	(2.17)	(2.17)	(2.70)
Gender * treatment			-1.37			-4.42
			(3.41)			(4.52)
N	102	102	102	103	103	103
Panel E: 12-month follow-up				Panel F: 18-month follow-up		
Treatment	0.81*	1.06**	0.71	0.74*	0.84**	0.91*
	(0.44)	(0.43)	(0.55)	(0.39)	(0.40)	(0.51)
Gender * treatment			0.86			-0.19
			(0.88)			(0.82)
N	93	93	93	89	89	89

Table A1: Tobit estimates. Dependent variable is the number of visits to the gym in a particular period (for Panels A-D); Dependent variable is the answer to the survey question "How many times do you exercise each week on average?"(for Panels E-F). The *Unexpected* treatment is compared to the *Control* group in the two *Unexpected* columns. Each panel represents a different time period. Standard errors appear in parentheses. Columns 1, and 4 include only a risk attitude control, columns 2 and 5 include gender(male), commute time, and risk attitude as control variables, and columns 3 and 6 include these three control variables and a gender-treatment interaction. * 0.1 significance, ** 0.05 significance, ***0.01 significance.