

Intermittent Incentives to Encourage Exercising in the Long Run¹

Ayala Arad^a, Uri Gneezy^b and Eli Mograbi^a

Abstract

We report the results of a field experiment in which students are encouraged to exercise. We compare a no-incentive control with a treatment in which participants are paid per visit, and to two intermittent incentive schemes: monetary rewards at increasing intervals and monetary rewards with unpredictable timing. The advantage of irregular incentives, as already recognized in the psychology literature, 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 live with the paradox and to cope with the frustration of sometimes not being rewarded for a proper response. This mechanism may facilitate the maintenance of the habit over the long run, after the incentives have been removed. In line with this theory, we find the two intermittent incentive schemes increased participants' frequency of exercising after the incentives were stopped and over an extended period of time more than the per-visit scheme.

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, Sally, Dotan, Yair, and participants at the Collier Conference on Behavioral Economics, Advances with field experiments conference and the ESA European conference for useful comments.

The online appendix is available at: <https://cutt.ly/ByQaLkA>

^a Collier 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 show the incentives that work best while in place are not necessarily the most effective ones 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 two major ones being present-biased preferences and self-control problems (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; Allcott and Rogers, 2014; 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., 2019², and Strohacker, Galarraga, and Williams, 2014).

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 from participants in the treatment group (Royer, Stehr, and Sydnor 2015). In the second, the post-intervention effect existed for only two weeks (Condliffe, Isgin, and Fitzgerald (2017). Using a lottery incentive scheme, Petry, Andrade, Barry, and Byrne (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) found a significant effect on the first three months after the incentivized period, but their dependent variable was not the total amount of entrances but differences from the previous quarter in the number of entrances.

The focus of the current paper is how to improve exercising in the long run. Importantly, our study is also designed to test whether the most successful incentives for the initiation of an activity in the short run are also the best in maintaining it 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 free membership to the university gym. We assigned participants into four groups: the first were not offered any incentives (*Control*); the second were offered incentives per visit (*Per-visit*); the third received frequent rewards to start with, which were aimed at overcoming the high initial cost of starting to exercise and that gradually became less frequent, with the goal that the participants would become accustomed to visiting the gym without receiving a reward (*Increasing*); and the fourth group was offered incentives on a random basis, such that not every visit was rewarded (*Unexpected*). In this last group, the reward—though random—was larger than that in the per-visit group, with the goal of adjusting participants' expectations so that they gradually become accustomed to not being rewarded for every single visit.

Although regular incentive schemes may crowd out intrinsic motivation to exercise (Gneezy et al., 2011), intermittent incentives, due to their irregular nature, may sufficiently induce the habit of exercising without creating an association between the monetary incentive and exercising. This design allows us to test whether the incentive schemes that work best in the short run are those that work best in the long run, after incentives are removed, and whether incentive schemes based on intermittent (irregular) rewards are more effective in the long run than traditional ones that are based on regularly scheduled rewards.

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 live with the paradox and 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). To our best knowledge, our study is the first to examine the success of intermittent incentives for exercising in which the irregularity is in the number of visits required for obtaining a reward, whereas the size of the reward is certain.³ Our two intermittent schemes, which have not been studied before, proved to be successful in the long run in our setting.

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

³An exception is Bachireddy et al. (2019), who paid participants for number of steps using an increasing (every two days) reward size per step or a decreasing one. Other studies such as Andrade, Barry, Litt, and Petry (2014), Patel et al. (2018), and Wing et al. (1996) used a lottery to determine whether the participant received the reward and its size. These studies did not find significant effects in the long run. Additional studies that used a lottery incentive scheme appear in Mitchell et al. (2019).

this end, all university students were invited to answer a short 3-minute lifestyle questionnaire that was distributed online (see the 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

Treatment Variable	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, Netemeyer, Spiller, and Zammit, 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, Gleicher, Boninger, and Edwards, 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. Thus, the incentives are frequent enough to induce the initiation of the exercise habit, but the habit will not become strongly associated with receiving a reward, due to its decreasing frequency.

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 when they would receive the reward. Because X is bounded, participants knew they would always get a reward by the fourth visit. 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, X was chosen by lottery.

In this scheme, participants were rewarded every two and a half visits to the gym, on average, and the number of times they were rewarded was not deterministic. They were always rewarded by the fourth visit, which reduced the likelihood of frustration over not winning a reward for too long.

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.⁴

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 Appendix). One hundred seventy-one participants attended the concluding session (roughly 85% of the remaining 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.”⁵ 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

⁴ 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.

⁵ <https://cutt.ly/MyQa0y5>

the control group. We also analyzed the 12- and 18-months follow-up questionnaires.⁶ We found that only participants in the *Unexpected* treatment exercised more than participants in *Control*. 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. These findings demonstrate that incentives that work well in the short run are not necessarily those that are more successful 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 chose to use a Tobit model for our estimations due to the large number of zero observations in our sample, which represent the individuals who did not exercise at all. We repeated the analysis using an OLS model as a robustness check, and report it in the online Appendix. We used three specifications to compare any two groups: (1) without the control variables; (2) with the control variables: gender and commute time; and (3) with the control variables as well as a treatment-gender interaction. We omitted outliers (mean ± 3 std of the dependent variable). All of the comparisons of the incentivized treatments are relative to the *Control* group. Comparisons between any pair of incentivized treatments can be found in the Appendix—Tables A1, A3, A4, and A5. 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 various membership periods. Figure 1 depicts the cumulative distribution of the number of visits to the gym in the various membership periods. Table 3 - Panel A and Figure 1 - Panel A, which relate to the incentivized period (months 1-2), show participants in the three incentivized treatments visited the gym more than the control group, with differences ranging from 6.00 to 10.36 additional visits, depending on the incentivized treatment. The incentive schemes had differential impacts during this period (Table A1 - Panel A in the Appendix).

Specifically, the *Increasing* scheme had a greater impact than the *Unexpected* scheme ($\beta = -4.33$, $SE = 2.03$, $p = 0.03$). A possible explanation for this gap is the risk element in the *Unexpected* scheme that is not present in the other treatments. This explanation is consistent

⁶ Eventually the planned 8 month's follow-up questionnaire was not carried out due to budget constraints.

with the finding that the more risk averse the individual, the less he or she exercised ($\beta=0.23$, $SE=0.09$, $p=0.01$) in the *Unexpected* treatment.⁷ When we controlled for risk preferences, the difference between *Increasing* and *Unexpected* decreased in size but remained marginally significant ($\beta=-3.81$, $SE=1.98$, $p=0.054$).

Table 3. Visits to the gym by months

	Per-visit			Increasing			Unexpected		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Months 1-2									
Treatment	8.36*** (2.11)	8.27*** (2.11)	8.71*** (2.57)	10.36*** (2.18)	10.29*** (2.17)	10.67*** (2.66)	6.00*** (2.14)	5.96*** (2.12)	6.60** (2.59)
Gender * treatment			-1.32 (4.43)			-1.14 (4.49)			-1.94 (4.48)
N	103	103	103	103	103	103	103	103	103
Panel B: Months 3-6									
Treatment	3.41 (2.63)	3.44 (2.63)	4.47 (3.34)	6.00** (2.67)	6.07** (2.68)	6.72** (3.40)	4.71* (2.59)	4.81* (2.59)	6.53** (3.25)
Gender * treatment			-2.76 (5.43)			-1.72 (5.47)			-4.85 (5.40)
N	101	101	101	104	104	104	102	102	102
Panel C: 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 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.

For the disincentivized membership period (months 3-6), Figure 2, which displays the mean amount of entrances per period, shows the *Per-visit* treatment experienced the largest

⁷ Risk preferences had no effect in the *Increasing* and *Control* treatments but did have an effect in the *Per-visit* treatment ($\beta=0.15$, $SE=0.08$).

decline in frequency of visits to the gym following the removal of incentives such that it is the closest incentivized treatment to the control group. Furthermore, 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 differences ranging from 4.71 to 6.0). This finding demonstrates the difference between the short term, where all incentive schemes were better than the control (and *Unexpected* was inferior than others), and the long term, where the two intermittent groups visited the gym more than the *Per-visit* group. We found no significant effects in a comparison between the incentivized groups for months 3-6 (Table A1 - Panel B).

Next, we examined whether the aforementioned effect was due solely to the first two months of the non-incentivized period or whether it existed throughout the four-month period. Hence, we performed a separate analysis of the behavior in months 5-6.⁸ Panel C of Figure 1 shows the cumulative distribution of the number of visits by treatment during months 5-6, revealing a similar (albeit larger) downward trend to that in Panel B relative to Panel A. 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. Panel C of Tables 3 and A1 show the participants in the *Increasing* treatment exercised more than those in the *Control* and *Per-visit* groups. These effects were significant in the presence of a gender interaction, because the difference between the groups was mainly due to women (with differences ranging from 3.96 to 4.77). 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.

⁸ An analysis of months 3-4 can be found in the appendix (Tables A2 and A3).

Figure 1: Cumulative distribution of visits to the gym by months

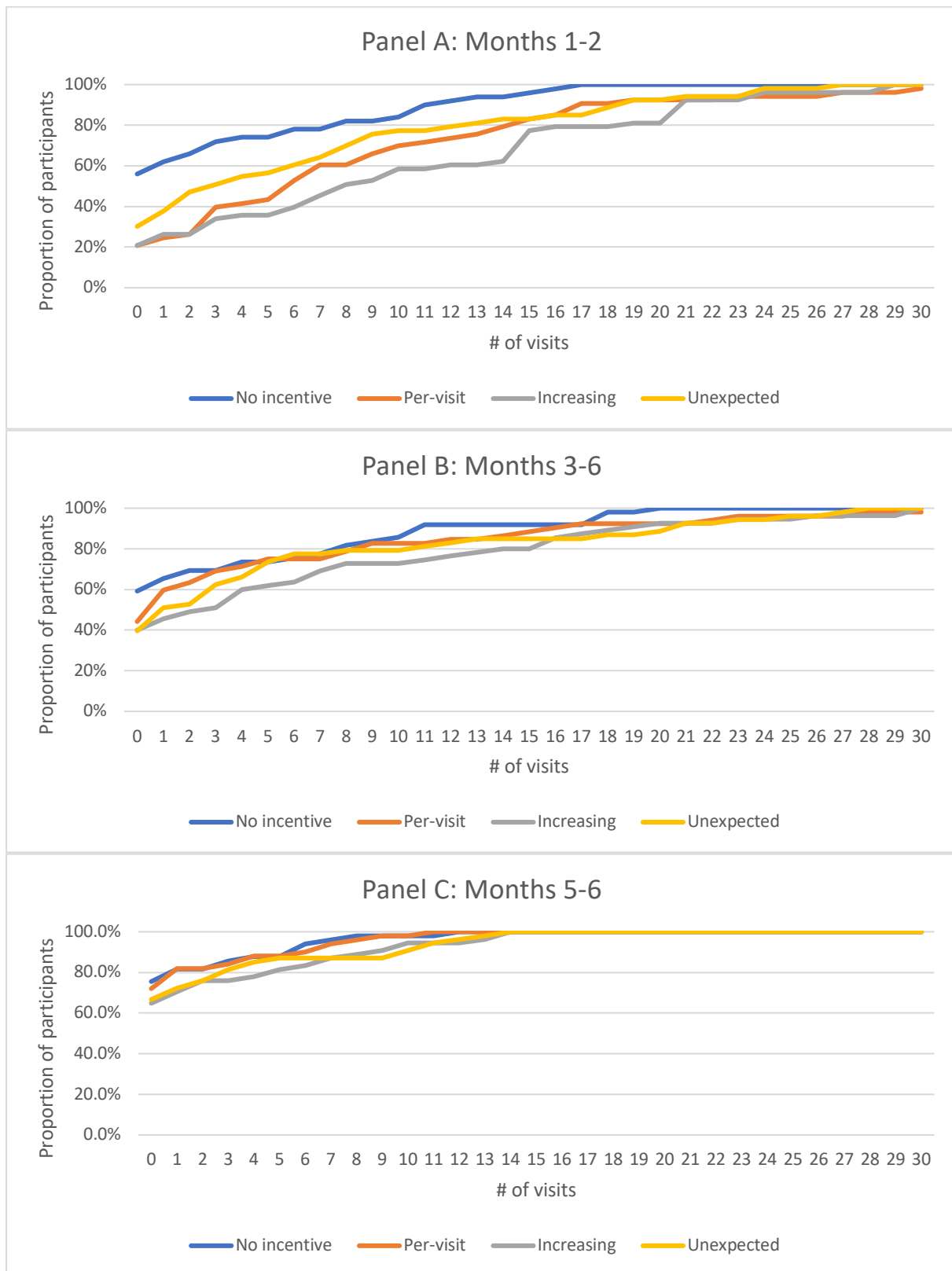


Figure 1: Cumulative distribution of the number of visits to the gym by months. Each panel represents a different time period.

We found no other significant differences between the treatments (see Panel C in Tables 3 and A1). However, in terms of absolute numbers, the difference in the number of visits between the *Unexpected* treatment and the other incentivized treatments (*Per-visit* and *Increasing*) decreased, as can be seen in Figure 2. Figure 2 also summarizes the dynamic of exercising during the membership period.

The *Per-visit* scheme is effective only during the short term; the *Increasing* scheme is best during both the short and long term; and the *Unexpected* scheme is inferior during the short term but declines the least over time, making it more effective on the long term. The results demonstrate that the incentive schemes that are effective in the short term will not necessarily be the best ones in the long term. This idea is also found in the 12- and 18-month follow-ups.

Figure 2: Number of gym visits by period

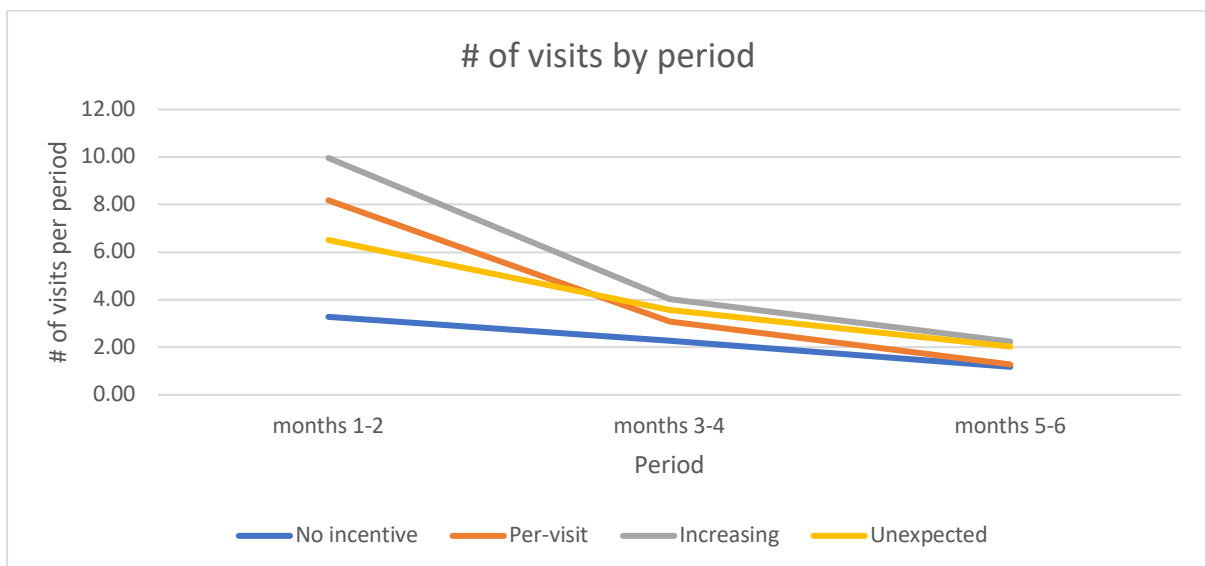


Figure 2: Number of gym visits by period.

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 4 compares each of the incentivized treatments with the *Control* group in the various follow-up periods. Figure 3 depicts the cumulative distribution of the number of visits to the gym in the various follow-up periods. Panel A of Table 4 shows that participants in the *Unexpected* treatment exercised significantly more than participants in the *Control* group. Panel A of Figure 2 shows that participants in the *Unexpected* treatment also exercised more (although not significantly) than participants in the other two incentivized treatments (see also Panel A of Table A4). Thus, whereas the effect of the incentives diminished for participants in the two incentive schemes that were the most effective during the incentivized period, namely, *Per-visit* and *Increasing*, participants under the *Unexpected* scheme, which had been less effective during the incentivized period, exhibited greater exercise perseverance. 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 3 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 4 shows a marginally significant effect 18 months after the beginning of the experiment, whereby participants in the *Unexpected* treatment exercised 0.72 more times per week than participants at the *Control* group. No other significant differences existed between the groups, although in numerical terms, participants in the *Unexpected* treatment exercised more than participants in the other two treatments (Table A4

- Panel B). The diminished effect was due to an increase in the frequency of exercise in the *Control* group, rather than a decrease in the *Unexpected* treatment.

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 and ten to sixteen months after the end of the incentivized period.

Table 4. 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 4: 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.

3.3 Persistence of incentive effects

Figure 4 presents the mean standardized Z-score for number of visits [$Z=(x-\text{mean})/\sigma$], 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 (Figure A2). 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 3: Cumulative distribution of weekly exercise frequency by period

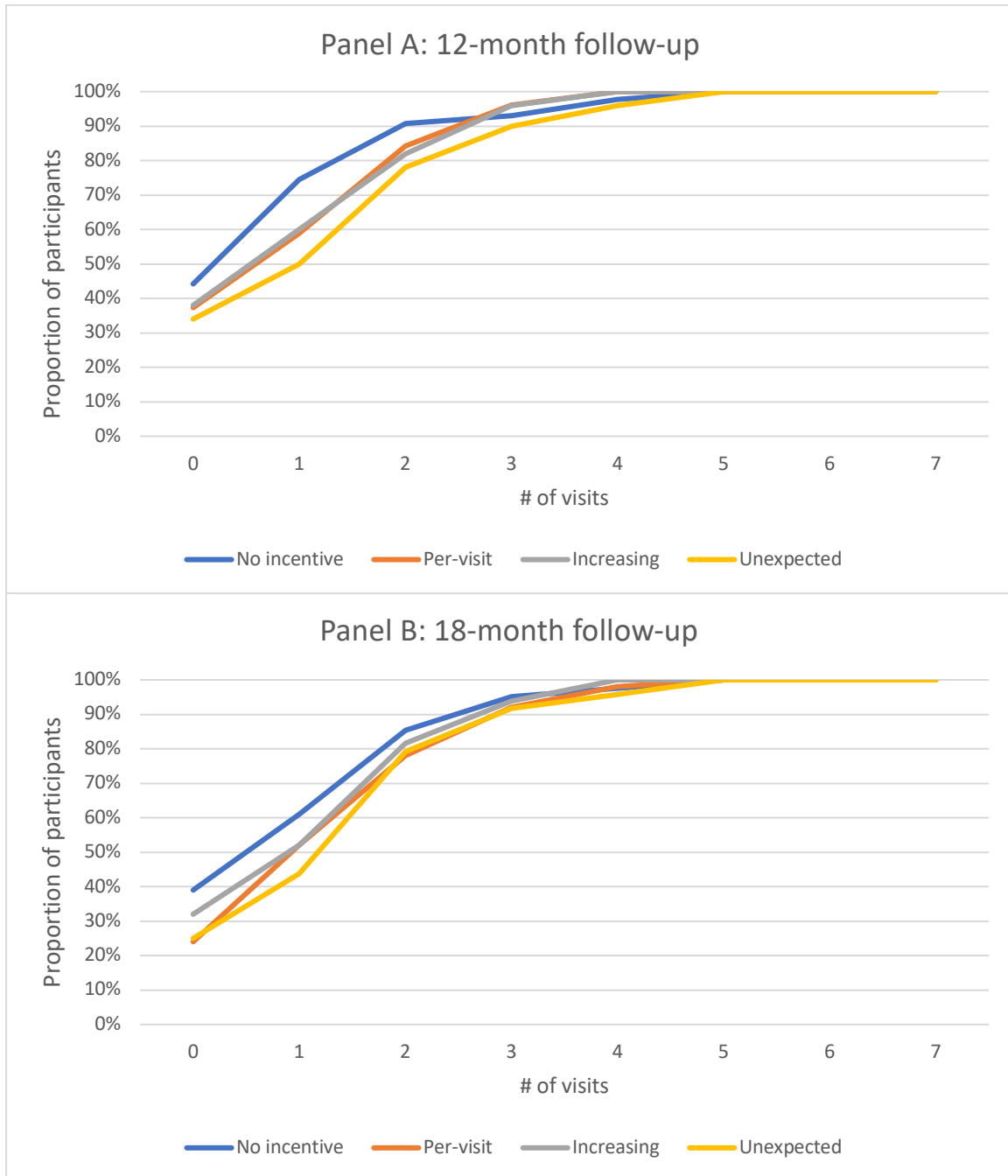


Figure 3: Cumulative distribution of the answer to the question “How many times do you exercise each week on average?” Each panel represents a different time period.

Figure 4: Standardized Z-score by period for visits to the gym

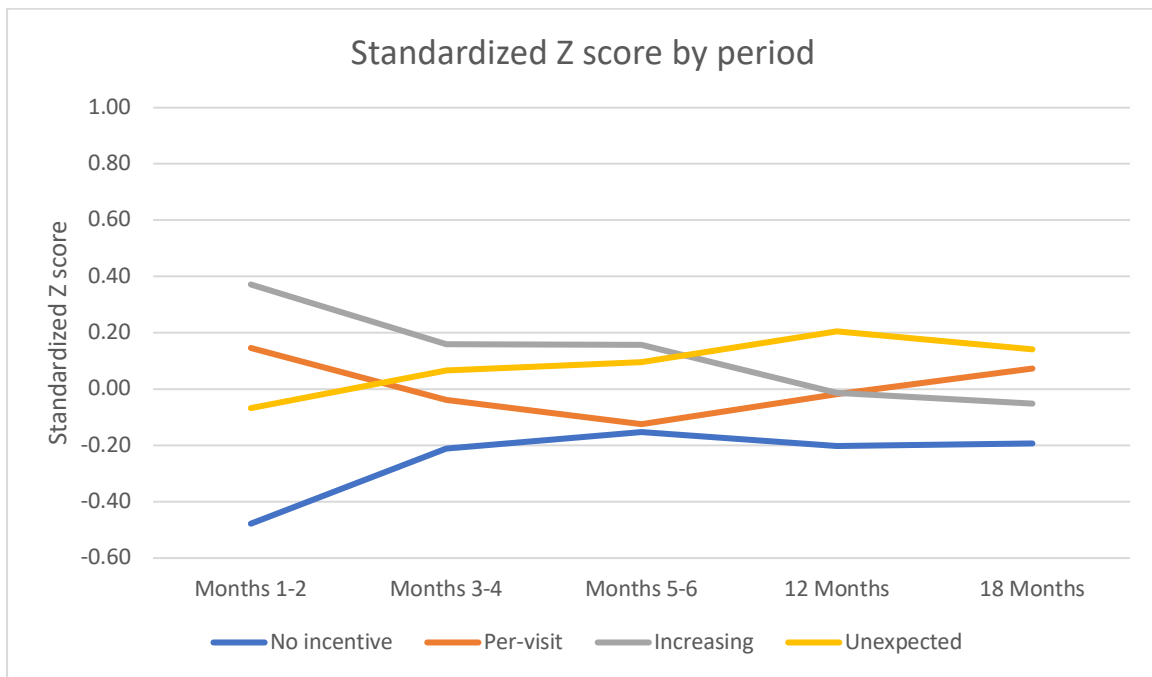


Figure 4: Standardized Z-score by period for visits to the gym and follow-ups.

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, but over time, 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 5 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 4.85 times more participants), whereas the proportions were similar among the incentivized treatments (Table A5 - Panel A).

Panel B of Table 5 presents the results of the analysis for the non-incentivized period (months 3-6). During this period, 2.29-2.32 times more participants exercised in the *Increasing* or *Unexpected* treatment than in the *Control* group. According to Panel B of Table A5, no significant differences existed between the incentivized treatments during this period. 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 5. 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 5: 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.

Risk preferences.

We also controlled for risk preferences when comparing the *Unexpected* treatment to the other treatments because it involves a certain degree of risk regarding when the reward will be received. The results can be found in the online appendix. Overall, controlling for risk preferences caused bigger differences between the *Unexpected* treatment and the *Control* group during the non-incentivized period ($\beta=5.31$, $SE=2.63$, $p=0.043$) and the 18-month follow-up ($\beta=0.84$, $SE=0.4$, $p=0.035$). Controlling for risk preferences did not affect any other comparison between treatments.

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 ($\beta=0.11$, $SE=0.04$, $p=0.005$) and improved the propensity to plan by 0.25 points ($\beta=-0.25$, $SE=0.09$, $p=0.005$).

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

Do incentives that work best in the short run continue to do so in the long run? We answer this question by comparing the effect of intermittent incentives with the effect of per-visit incentives to exercise. The goal of the intermittent incentive scheme was to get the participant to start exercising in a way that would more effectively create a long-term exercise habit. The results show differences between the short- and long-run effects. The *Unexpected* incentive scheme was less effective in the short run than both the *Per-visit* and *Increasing* incentive schemes, but outperformed them in the long run. Meanwhile, the *Increasing* incentive scheme outperformed the other two during the incentivized period and decayed more slowly than the

Per-visit scheme during the first four months after the incentives were removed. The above results together show that whereas the *Per-visit* scheme works well in the short run, when examining long-run effects after removal of incentives, the intermittent schemes are superior.

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.

Allcott, Hunt., and Rogers, Todd. (2014). “The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation.” *The American Economic Review*. Vol. 104: 3003-3037.

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.

Bachiredy, 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.

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 plan: 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. Visits to the gym by months; Comparison between treatments

	Increasing vs. Per-visit			Unexpected vs. Per-visit			Unexpected vs. Increasing		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Months 1-2									
Treatment	1.86 (2.00)	1.91 (1.99)	1.99 (2.43)	-2.42 (1.99)	-2.38 (1.98)	-2.22 (2.39)	-4.30** (2.04)	-4.33** (2.03)	-4.13* (2.47)
Gender * treatment			-0.25 (4.27)			-0.52 (4.29)			-0.59 (4.35)
N	106	106	106	106	106	106	106	106	106
Panel B: Months 3-6									
Treatment	2.52 (2.55)	2.39 (2.55)	2.19 (3.19)	1.32 (2.50)	1.35 (2.51)	1.99 (3.07)	-1.17 (2.50)	-1.16 (2.51)	-0.20 (3.09)
Gender * treatment			0.55 (5.33)			-1.93 (5.32)			-2.85 (5.35)
N	107	107	107	105	105	105	108	108	108
Panel C: Months 5-6									
Treatment	2.30 (1.98)	1.87 (1.93)	3.96* (2.35)	1.81 (2.00)	1.65 (2.00)	3.20 (2.47)	-0.52 (2.06)	-0.57 (2.03)	-1.13 (2.40)
Gender * treatment			-6.23 (4.01)			-4.57 (4.24)			1.92 (4.43)
N	104	104	104	104	104	104	108	108	108
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Table A1: Tobit estimates. Dependent variable is the number of visits to the gym in a particular period. In each group of three columns, two treatments are compared: the first is 1, and the second, which is the reference group, is 0 (e.g., in the last column, *Increasing*=0 and *Unexpected*=1). 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.

Table A2. Visits to the gym on months 3-4

	Per-visit			Increasing			Unexpected		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
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
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Table A2: Tobit Estimates. Dependent variable is the number of visits to the gym in months 3-4. 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*). 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.

Table A3. Visits to the gym months 3-4; comparison between treatments

	Increasing vs. Per-visit			Unexpected vs. Per-visit			Unexpected vs. Increasing		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	1.60 (1.64)	1.48 (1.62)	0.21 (2.01)	0.99 (1.64)	1.05 (1.64)	0.75 (2.00)	-0.61 (1.60)	-0.50 (1.59)	0.49 (1.96)
Gender * treatment			3.57 (3.39)			0.91 (3.49)			-2.85 (3.35)
N	108	108	108	106	106	106	108	108	108
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Table A3: Tobit estimates. Dependent variable is the number of visits to the gym in months 3-4. In each group of three columns, two treatments are compared: the first is 1, and the second, which is the reference group, is 0 (e.g., in the last column, *Increasing*=0 and *Unexpected*=1). 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.

Table A4. How many times do you exercise each week on average? Comparison between treatments (Tobit)

	Increasing vs. Per-visit			Unexpected vs. Per-visit		Unexpected vs. Increasing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: 12 months follow-up									
Treatment	-0.00 (0.37)	0.07 (0.37)	0.03 (0.46)	0.36 (0.39)	0.35 (0.39)	-0.02 (0.47)	0.36 (0.40)	0.28 (0.39)	0.04 (0.48)
Gender * treatment			0.12 (0.78)			1.08 (0.80)			0.73 (0.84)
N	101	101	101	101	101	101	100	100	100
Panel B: 18 months follow-up									
Treatment	-0.25 (0.34)	-0.14 (0.33)	-0.34 (0.40)	0.08 (0.34)	0.06 (0.34)	-0.04 (0.42)	0.34 (0.35)	0.25 (0.35)	0.29 (0.43)
Gender * treatment			0.64 (0.71)			0.31 (0.76)			-0.15 (0.73)
N	101	101	101	101	101	101	100	100	100
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Table A4: Tobit estimates. The dependent variable is the answer to the survey question “How many times do you exercise each week on average?” In each group of three columns, two treatments are compared: the first is 1, and the second, which is the reference group, is 0 (e.g., in the last column, *Increasing*=0 and *Unexpected*=1). 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 columns 3, 6, and 9 include controls and a treatment-gender interaction. Controls include *0.1 significance, **0.05 significance, and ***0.01 significance.

Table A5. Did the participant visit the gym at least once? Comparison between treatments

	Increasing vs. Per-visit			Unexpected vs. Per-visit			Unexpected vs. Increasing		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Months 1-2									
Treatment	1.00 [0.00] (0.48)	0.99 [-0.01] (0.48)	0.79 [-0.24] (0.59)	0.61 [-0.50] (0.45)	0.61 [-0.49] (0.46)	0.73 [-0.32] (0.57)	0.61 [-0.50] (0.45)	0.60 [-0.51] (0.45)	0.90 [-0.11] (0.56)
Gender * treatment			2.05 0.72 (1.04)			0.64 -0.44 (0.95)			0.32 -1.14 (0.98)
N	106	106	106	106	106	106	106	106	106
Panel B: Months 3-6									
Treatment	1.19 [0.17] (0.39)	1.20 [0.18] (0.40)	1.01 [0.01] (0.48)	1.21 [0.19] (0.40)	1.25 [0.22] (0.40)	1.60 [0.47] (0.49)	1.02 [0.02] (0.39)	1.00 [0.00] (0.40)	1.54 [0.43] (0.48)
Gender * treatment			1.68 0.52 (0.86)			0.47 -0.76 (0.85)			0.26 -1.33 (0.86)
N	107	107	107	105	105	105	108	108	108
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Table A5: Logistic regression. The dependent variable is whether participants visited the gym at least once during a particular period. In each group of three columns, two treatments are compared: the first is 1, and the second, which is the reference group, is 0 (e.g., in the last column, *Increasing*=0 and *Unexpected*=1). 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 the controls (gender(male), and commute time), and column 3, 6, and 9 include controls and a treatment-gender interaction. *0.1 significance, **0.05 significance and ***0.01 significance.

Figure A1: Cumulative distribution of visits to the gym – months 3-4

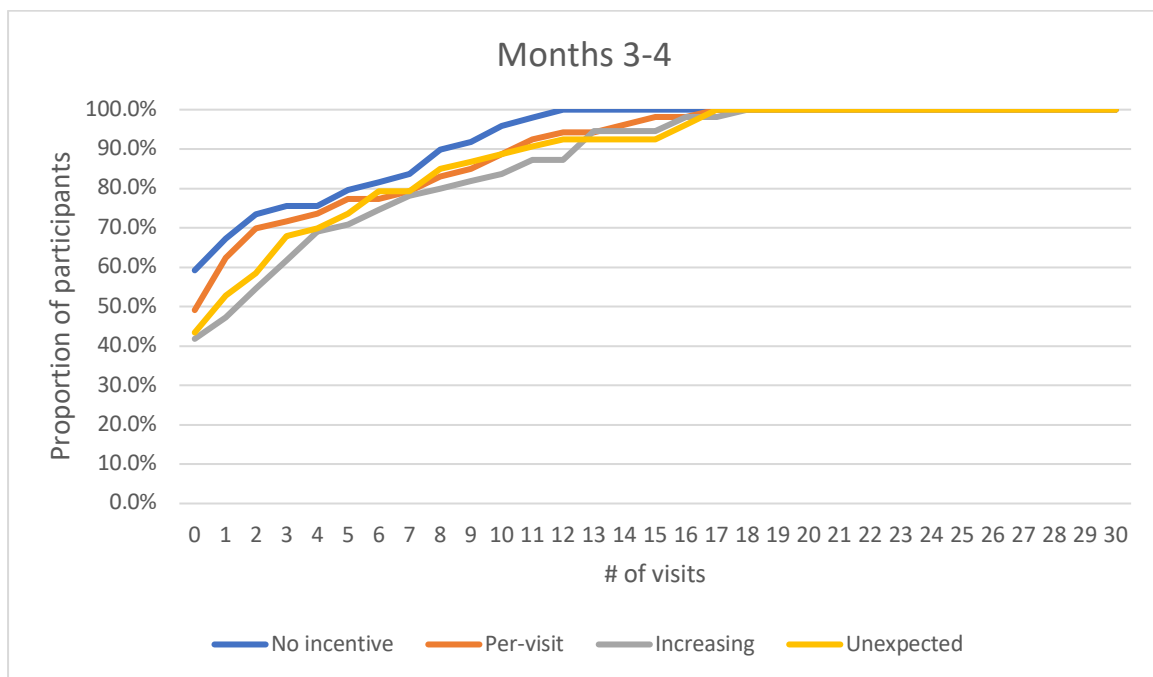


Figure A1: Cumulative distribution of the number of visits at the gym during months 3-4.

Figure A2: Standardized Z-score by period for sports center visits

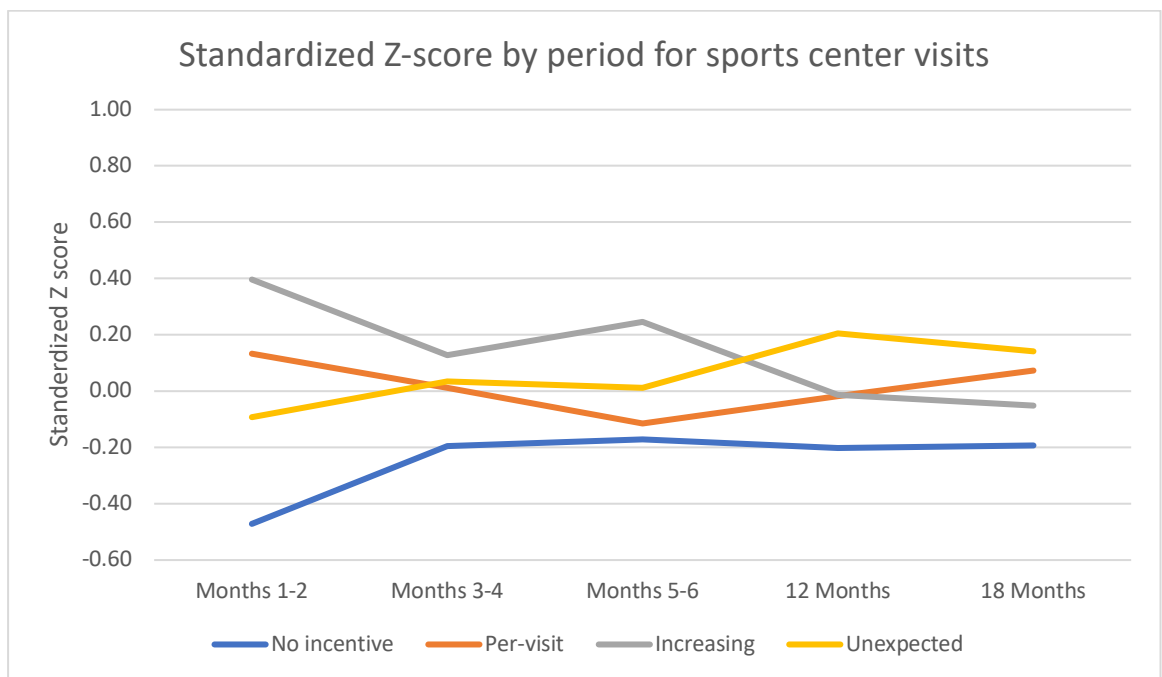


Figure A2: Standardized Z-score by period for visits to the sports center and follow-ups.