NBER WORKING PAPER SERIES

INCENTIVES, COMMITMENTS AND HABIT FORMATION IN EXERCISE: EVIDENCE FROM A FIELD EXPERIMENT WITH WORKERS AT A FORTUNE-500 COMPANY

Heather Royer
Mark F. Stehr
Justin R. Sydnor

Working Paper 18580
http://www.nber.org/papers/w18580

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2012

We are thankful for funding from the National Science Foundation, the Upjohn Institute, and the Case Western Reserve University ACES fund. Royer also thanks the RAND corporation for support through the NIA. We are appreciative for the outstanding research assistant work of Stephen Cabrera, Andrew Chang, Vishal Chauhan, Tina Chen, Jon Evans, Natalie Greene, Brian Jameson, Victor Marta, Rachel Smith, and Bert Wagner. We appreciate the comments and suggestions of Nava Ashraf, John Beshears, Eric Bettinger, Tanguy Brachet, John Cawley, David Clingingsmith, Stefano DellaVigna, Uri Gneezy, Dean Karlan, Nicola Lacetera, and Jason Lindo along with those of various seminar and conference participants. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Incentives, Commitments and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company
Heather Royer, Mark F. Stehr, and Justin R. Sydnor
NBER Working Paper No. 18580
November 2012
JEL No. D03,D9,I1

ABSTRACT

Rapidly growing health-care costs have fueled interest in using financial incentives to improve health behaviors. Most of the research on financial incentives outside of clinical studies has been observational, limiting our ability to make causal inferences on their effectiveness. The few carefully-designed studies have generally found little lasting effect on behavior after the incentive program ended. We report on a large field experiment with employees of a Fortune 500 company which offered incentives for using the company gym. In addition to understanding the effects of incentives alone, we investigate a novel approach to generate lasting behavior change using self-funded commitment contracts. At the end of incentive period, half of the incentive group were offered the opportunity to create a self-funded commitment contract to motivate their own behavior. Workers responded strongly during the incentive period, doubling their rate of use of the company gym. After the incentive period ended, we find that those offered incentives only continued to attend at higher rates, but the effect was quite modest in magnitude. The availability of a commitment contract, however, substantially improved the long-run effects of the incentive program both during the commitment period and well beyond, offering a promising new approach to increasing the long-run effect of incentive programs.

Heather Royer
Department of Economics
University of California, Santa Barbara
2127 North Hall
Santa Barbara, CA 93106
and NBER
royer@econ.ucsb.edu

Mark F. Stehr
Drexel University
LeBow College of Business
Matheson Hall 504E
3141 Chestnut Street
Philadelphia, PA 19104-2875
stehr@drexel.edu

Justin R. Sydnor
University of Wisconsin - Madison
975 University Avenue
Madison, WI 53706
jsydnor@bus.wisc.edu

An online appendix is available at:
http://www.nber.org/data-appendix/w18580

American lifestyles are characterized by poor diet and a lack of physical activity.[1] Since many Americans receive medical insurance through government or group-rated policies, these unhealthy behaviors likely generate large externalities and substantial budget burdens for government programs such as Medicare and Medicaid (Finkelstein et al., 2009). Additionally, work in behavioral economics argues that unhealthy behaviors often stem from time inconsistency, where impatience leads to adverse behaviors that impose "internalities" on one's long-run self.

Concerned with both the externalities and "internalities" of poor health behaviors, firms and policy makers have shown increasing interest in using financial incentives to motivate healthier behaviors (Baicker, Cutler and Song, 2010). For example, many companies and insurers are experimenting with "wellness incentives" that encourage employees to complete health assessments, attend gyms or consume healthy food. As part of the Patient Protection and Affordable Care Act, employers have greater scope to use rewards and penalties to target specific behaviors or health outcomes. Despite the widespread interest in incentive programs, however, there has been limited research that carefully assesses how health behaviors respond to these workplace incentive programs.

We fill this gap by implementing one of the first large-scale randomized field experiments of a workplace health-incentive program.[2] Our experiment was conducted with 1,000 employees at the headquarters of a Fortune 500 company. In the treatment arm, randomly-selected employees were offered a one-month financial incentive to attend the company's onsite exercise facility ($10 per visit for up to 3 visits each week). After the completion of this incentive period, this incentive group was randomized into two groups – one offered the opportunity to create a self-funded "commitment contract" and the other not. This commitment contract allowed individuals to pledge that they would

[1] As of 2009, only 14% ate the recommended amounts of fruits and vegetables (USA Today, 2009). According the Centers for Disease Control, in 2010, only 20.4% of adults met the CDC's muscle-strengthening and aerobic exercise recommendations. See http://www.cdc.gov/nchs/fastats/exercise.htm

[2] Baicker, Cutler and Song (2010) survey studies of workplace wellness programs and estimate a large return to these investments, but note that few studies have used randomized evaluations and many use un-incentivized health assessments. Cawley and Price (2011) study the effect of incentives for weight loss in a workplace wellness program without random assignment but with a large sample size. The program had a high rate of attrition, but even taking this attrition into account, they find little evidence that incentives help to generate lasting weight-loss effects. In one of the few studies both to offer financial incentives and employ random assignment among employees, Finkelstein et al. (2007) found that incentives for weight loss increased weight loss among obese employees. However their study also faced some selection issues, as their recruitment criteria required that potential subjects be obese, and they observed significant sample attrition.

continue to use the gym over the 2 months following the original incentive period. Individuals could put as much of their money at stake as they wished. If the employee kept to the commitment, her money was refunded and otherwise it was donated to the United Way.

The design of this study -- a temporary financial incentive potentially followed by the opportunity to create a self-funded commitment contract -- is motivated by insights from the economics literature on time inconsistency.[3] High "startup costs" are often an obstacle to behavioral change. In the context of gym attendance, these costs may include needing to join and learn about the gym, rearranging schedules, and an initially high level of physical discomfort. Those high startup costs may lead to sub-optimal behavior from individuals with either present-biased preferences or projection bias. Individuals with present bias may continually procrastinate beginning a beneficial behavior such as exercise (O'Donoghue and Rabin, 1999, 2001). A person with projection bias may fail to realize that some of the high costs (e.g., physical discomfort) of initial exercise are likely to fall as one continues to engage in the activity, and hence might give up on the activity too soon (Loewenstein, O'Donoghue and Rabin, 2003). In both of these situations, a temporary incentive could in theory provide the kick start a person with time inconsistency needs to establish lasting behavior change. However, time inconsistency also suggests that in many cases a kick start alone may not be enough to generate such change. Activities like exercise, with present costs and future benefits, can generate recurring struggles for individuals with present bias. For instance, DellaVigna and Malmendier (2006) find that most gym members, even those who had been paying a membership for a long time, used the gym very little. Our combined treatment is motivated by a desire to help individuals overcome these two types of obstacles. It offers a temporary (one month) financial incentive to motivate individuals through the initial cost of attending the gym and the option to create a self-funded commitment contract to address the persistent self-control problems that may ensue after the temporary incentive has ended.

Our study has four primary findings. First, we find that the response to financial incentives is strong. The incentives doubled the rate at which employees used the company gym while they were in effect. The interpretation of this increase depends on whether the incentives generated new exercise or

[3] See Strotz (1955-56), Phelps and Pollak (1968), and Laibson (1997) for foundational work on time inconsistency in economic models of discounting. See also Frederick, Lowenstein and O'Donoghue (2002) for a review.

simply changed the location of exercise. Drawing on complementary survey data, our analysis suggests that at least 70% of the treatment effect is new exercise. Our second primary finding is that the temporary one-month incentive had a long-run, albeit modest, effect. Beyond the incentive period, the long-run increase in gym attendance was 16% of its incentive period size. Third, the availability of commitment contracts substantially improved the lasting impact of the incentive program. In the two months following the end of the incentive period, the group offered a combination of the incentives and the commitment contract had gym attendance rates that were 47% of the size of their in-treatment effect. Furthermore, this effect was persistent – lasting sizable effects are observed 1 year following the experiment. This is an especially important finding given the high cost and often-disappointing long-run effects of health-behavior incentive programs (Gneezy, Meier, Rey-Biel 2011). Fourth, we find that survey measures designed to capture awareness of self-control problems fail to predict take-up of commitment contracts. This finding is intriguing since it suggests that commitment contracts may appeal to individuals for reasons other than a desire to address a recognized self-control problem.

This study contributes to a growing literature exploring the ability of financial incentives to generate lasting health-behavior change. In a significant recent study, Charness and Gneezy (2009) showed that among college students, a short-term financial incentive to use the campus gym resulted in lasting increases in gym use even after the incentive was removed. That finding was largely confirmed by Acland and Levy (2011), who also found that many students showed projection bias and failed to predict how readily the exercise habit could be established.[4] Our study complements these papers in at least two ways. First, the measurement of long-run effects in university settings is interrupted by the end of the term. Since there are no such breaks for employees, we are able to identify effects of our program a full year out. Second, our study is conducted with a working population that on many health dimensions (e.g., rates of obesity) looks similar to the broader population and suggests that employees, like undergraduates exhibit lasting impacts of a temporary incentive. However, in our study, the effect of the incentives alone is much more modest than in these studies. This result is shared to some extent by the

[4] Other exercise studies with undergraduate populations include Babcock and Hartman (2010) who look at the role of spillover effects in the context of incentives and Babcock et al. (2011) who examine the effect of team versus individual incentives for exercise and studying.

literature on financial incentives for health behaviors in clinical settings, which frequently finds that while incentives motivate behavior change, most patients revert back to old behaviors after the incentives end (e.g., Volpp et al., 2008; John et al., 2011).

Our work also contributes to the literature on commitment contracts as tools for addressing time inconsistency problems. A number of studies have shown that various commitment technologies can help increase savings (Ashraf et al., 2006; Benartzi and Thaler 2004; Beshears et al., 2011). Work in the health context has shown that commitments can be effective tools for increasing exercise (Milkman, Minson and Volpp, 2012), smoking cessation (Jeffrey et al., 1990; Gine, Karlan and Zinman, 2010) and weight loss (Jeffery et al., 1990; Volpp et al., 2008; John et al., 2011). The Milkman et al. study finds that individuals benefit from and are willing to pay for a commitment program that gives them access to enjoyable audio books only when they are exercising at the gym. The other existing health-related studies of commitments have focused on "deposit contracts" broadly similar to the contracts in our study. Our study is the first to test the effect of these types of commitment contracts in a non-clinical setting in the U.S. We offer commitment contracts to a broad population, rather than targeting a sub-population with clear behavioral issues (e.g., smokers or the obese).[5] This leads us to some interesting findings about the demand for commitment; specifically, we show evidence that the expected determinants of commitment demand hold little predictive power. This study is also the first to explore the use of commitment contracts as a tool for extending the impacts of an incentive program rather than as a stand-alone program, and we find that this coupling can improve the long-run impact of incentive programs.

Beyond our primary findings, the large sample size and rich setting of our field experiment allow us to explore the heterogeneity of the response to incentives in this population on a number of dimensions. In particular, the company requires employees who want to use the fitness center to complete a basic health assessment and to pay a modest membership fee. Prior to our experiment 38% of the employees in the sample were members of the gym and 62% were non-members. We find a lasting effect of the incentive only treatment (when not paired with a commitment option) exists only

[5] Goldhaber-Feibert, Blumenkranz, and Garber (2010) focus on the effect of nudges and anchors on commitment contract take-up for exercise.

for the non-members. That pattern is consistent with the idea that a temporary incentive might generate lasting changes primarily by helping individuals to overcome the high startup costs. The availability of a commitment contract after the incentive period, however, increases attendance significantly for both members and non-members. Among members, there is a particularly large effect of the commitment contract offer for those who were moderate users of the gym pre-experiment. These patterns are consistent with the idea that commitment technologies could be most helpful to those who face consistent struggles to maintain their exercise routine.

The take-up rate of commitment contracts among those offered the option in our study was 12% overall and 23% among employees who had used the company gym at least once during the incentive period. We find a number of interesting patterns in the correlates of demand for commitment. First, we find that men and young employees are substantially less likely to make commitments than female and older employees. Second, we find that measures of self-awareness of self-control, which theory leads us to believe would be important predicators of commitment contract demand, are not correlated with take-up. In contrast to survey measures of self-control, we find that indicators based on observed attendance behavior are predictive. In particular, those who exercised regularly during the treatment, but fell short of earning the maximum possible incentives were the most likely to make commitments. We also find that those who (over-optimistically) said they planned to increase attendance in the post-treatment period relative to the incentive period were more likely to commit. Finally, surprisingly, we find substantial take-up of commitment contracts even for those who exercise very regularly prior to our study. This finding suggests that commitment contracts might be attractive alternatives to other forms of self-control regulation that people already use. As such, offering commitment options might improve welfare even without any observable changes in behavior.

The remainder of the paper proceeds as follows: In Section 2 we describe the experimental design. Section 3 provides summary statistics from our initial survey and an overview of the data. In Section 4 we present the main empirical results of the experiment along with a series of heterogeneity cuts and an analysis of substitution effects. Finally, in Section 5 we conclude with a discussion of some of the potential implications for this research and highlight a number of open questions that this study could not address and that might motivate future research.

2. Experimental Design

2.1 Subject recruitment

The experiment took place at the headquarters of a Fortune 500 company located in the Midwest. At this location, there are approximately 1,900 employees holding a variety of jobs. The headquarters has a fitness center located on site that has the usual amenities of a modern gym. In order to use the gym, employees must become members of the wellness center and pay a membership fee of $12.96 every 2 weeks that is automatically withdrawn from their paychecks.[6,7] Upon entry to the gym, employees log in at a computer terminal and these computerized log-ins serve as our primary data.

We began the experiment in February 2009 and enrolled our last participants in March 2011. We ran the experiment in 15 waves, with modest-size cohorts, to ensure that the gym staff could accommodate new gym member signups and that our results were not specific to a particular time of the year. Appendix Figure 1 describes the timeline of the experiment. We detail the number of participants along each step of the experiment in Appendix Figure 2. Survey Appendix A provides copies of our communication with subjects and Survey Appendix B contains our survey instruments.

To recruit subjects for each cohort, we first randomly drew a sample of employees from the company's full list of employees at the headquarters site, excluding high-level executives, members of the human resources team, and gym staff privy to the details of the research. Then we sent the employees an invitation via e-mail to participate in two online wellness surveys (initial and follow-up) spaced 5 weeks apart (see Survey Appendix for a copy of these surveys). We described the experiment as a university study supported by the corporation. The employees were compensated with a $25 payment conditional on completion of both surveys. The initial survey collected a range of information on demographics, self-assessed fitness levels, exercise patterns, and subjective wellbeing. Response rates for this survey averaged 62% (see Appendix Figure 2).[8] We view this response rate as relatively

[6] There are no start-up fees or contracts and employees can cancel their membership at any time with no penalty.
[7] The gym is open Monday through Friday from 6:00 a.m. to 7:00 p.m.
[8] Response rates do vary some across cohorts although in a regression of whether or not an individual responded on cohort fixed effects, we are unable to reject that the cohort fixed effects are jointly equal to zero. Moreover, the fraction of responders who are gym members is not changing systematically over time. If word spread rapidly through the company about the details of our experiment, we would expect that response rates and their fraction who are gym members would vary across cohorts.

high; for the Card et al (2012) study of peer pay of UC employees, the survey response rate was just above 20%. Subjects were informed that none of their individual responses to any surveys or other information would be shared with anyone at the corporation. Since employees were aware they were participating in a study, this experiment is a "framed field experiment" in the lexicon of List (2009).

Our pool of experimental subjects consists of those 1000 employees who responded to our initial survey. Upon completion of this survey, we randomized individuals into treatment and control groups. The treatment group was eligible to receive financial incentives for gym attendance for a 4-week period whereas the control group was not; we elaborate on these treatments in more detail below. Because we anticipated that the response to incentives was likely to be heterogeneous, we stratified the randomization into 4 groups: a cross of a) gym membership with b) current exercise above or below their target level. After the completion of the incentive period, the incentive-eligible subjects were divided into two treatment groups, detailed below. During the final week of the incentive program, all subjects who responded to the initial survey (including the control group) were asked to complete our follow up survey. This survey largely asked the same questions as the initial survey (omitting demographics). The response rate to this survey was 91.4% (see Appendix Figure 2).

Since the subject pool consisted exclusively of individuals who responded to the initial survey and are thus unlikely to be a random sample of the employees at the firm, caution is warranted when extrapolating our results to the broader population of employees.[9] In principle, we could use characteristics on non-responders to predict what the treatment effects would have been for the overall population, but we are hesitant to do so for several reasons. First, the crucial assumption underlying this approach is that these characteristics adequately characterize selection. Second, we believe that in light of the response emails we received, a significant fraction of the non-response comes from those who were away from work (on vacation or work travel). Third, at the end of our experiment, we contacted non-responders to our initial survey and assigned them to different treatment arms without having to

[9] From the limited data on employees from the company we have (essentially departmental unit, position, and gym membership status), we can characterize response rates. Gym members responded to the initial survey at a slightly higher rate than gym non-members – 74% versus 57%. While departmental unit is predictive of the response rate, there is no department characteristic that is clearly associated with higher response rates. Specifically, departments that have low response rates appear to be neither highly low-skill nor highly high-skill departments.

fill out the initial survey. The treatment effects in this case were rather small. Thus, we believe that the conservative approach for extrapolating our estimates to the entire company's employees would be to assume that survey non-responders have no response to financial incentives. Since our survey response rates are rather high, while this adjustment would lower our estimated treatment effects, it would not change our conclusions that the incentive program was effective in changing exercise behavior. Adjusting for non-response in this way would also not change any of our conclusions related to the value of incentive programs coupled with commitment contracts relative to the program without a commitment option.

2.2 First-level treatment: Financial incentives

Incentive-eligible participants could earn \$10 for each visit (up to 3 visits per week and only 1 visit per day) to the corporate wellness center over a specified 4-week period. The treatment group also received a free gym membership during the incentive period (a value of \$25.92). Additionally, since joining the gym involves a 1-hour new membership assessment, we offered \$20 to new members to join. Since all treatment groups included both per-use incentives and the membership reimbursements/bonus, while the control group received neither, the incentive program is a package of incentives.[10] To ensure that the incentives were salient to participants, we informed treatment subjects via both email and via a physical letter to their company mailbox. Based on evidence from follow-up surveys, lack of information about the incentive program was not an impediment to participation.

We measure gym use via the login records described above. As is common at most gyms (including those studied in previous research on exercise incentives), the gym only uses a log-in process and does not require individuals to log out when leaving. As such, it is not possible to know how long the employee exercised or the nature of that exercise. In theory, there is some scope for employees to cheat on the program by logging in and not exercising, but our research assistants, who we asked to discretely monitor the gym, reported no such behavior. In addition, the gym staff -- who were aware of the program but did not know who was offered incentives -- reported no increases in suspicious logins

[10] In pilot experiments at the company prior to this experiment, there was essentially zero response to a treatment offering only a free membership.

and did not observe increases in employees showing up at the gym without exercising. Additionally, while such behavior could in theory be a concern during the incentive period when visits earn money for the participants, our primary interest is behavior after the incentive program ends, when there is no incentive for this cheating.

Much of the interest on health-incentive programs to date has focused on incentivizing weight loss. For this study, we decided to incentivize gym-attendance rather than weight loss for several reasons. Most importantly, our interest in this study is in understanding how incentives interact with *behaviors* in situations where time inconsistency may matter. Exercising less often than one desires is a standard example of a behavior that may result from time inconsistency. Weight-loss, in contrast, is a desired *outcome* that could be achieved through a range of behaviors (some of which, e.g., use of diuretics, are unhealthy). Another reason for our focus on gym attendance is that while reducing rates of obesity is an important goal of health-promotion, there are clear and direct benefits to physical activity itself, including improved cardiac health, mental health, productivity, etc. Furthermore, the benefits of exercise are important to the broad population, both the obese and non-obese, which fits well with company-wide health promotion efforts. Incentive programs that focus on a single outcome such as obesity may not realize the larger set of benefits that come from incentivizing exercise. Fryer (2010) has made the point – in the context of educational incentives – that in general incentivizing positive behaviors may be more effective than incentivizing outcomes in situations where either the production function mapping inputs to outcomes is not clear. Finally, it is possible in an experimental setting to observe gym attendance in a non-obtrusive way, whereas studies focusing on weight-loss generally require repeated weigh-ins and often suffer from high levels of attrition (e.g., Cawley and Price 2011).

2.2. Second-level treatment: Self-funded commitment contract

At the end of the 4-week incentive period, members of the treatment group were randomized into a second-level treatment, in which roughly half of the incentive eligible subjects were offered the chance to create a commitment contract. We refer to these two groups as the incentive-only and incentive+commit groups.[11] Up until the commitment contract offer, we treated these groups the same. Throughout incentive+commit denotes the group *offered* the commitment option and is an "intention-to-treat" grouping that does not depend on whether the subject actually made a commitment. The commitment contract for this study was a pledge not to go more than 14 calendar days without attending the company gym over an 8-week period. Participation was voluntary. Participants who decided to create a commitment contract could decide to put as much money as they wanted towards the commitment. All commitments were self-funded, meaning that participants placed their own money at stake and reaped no external financial rewards. Subjects who successfully completed their commitment were returned their committed money. In the case of a failed commitment, the committed money was forfeited to the United Way. To ensure an active response showing either interest or no interest to the commitment offer, the offer of a commitment contract was made when subjects were asked for their mailing address for their gym incentives and survey payment. In principle, individuals could commit more than they earned in the incentive program by writing a check made out to the United Way that was held until the end of the commitment period and returned if they successfully completed the commitment. Importantly, all payments for the gym-attendance incentive were mailed after this 8-week commitment period, so a subject who decided to create a commitment contract would not see a delay in receiving his or her incentive payment.

In principle, it would have been ideal to allow each subject the opportunity to determine the level of attendance targeted by the commitment contract. However, we used the fixed commitment of not missing more than 14 days in a row at the gym to keep the program simple so that it could be described briefly via email and an online survey. The commitment here involves a low attendance target, essentially a visit every other week. This low level was chosen to ensure that it would not be too ambitious for employees with work-related travel or vacation. In addition, we anticipated that this low-

[11] In order to ensure balance between the incentive-only and the incentive+commit groups, we re-randomized during this step until a p-value on the test of the equality of the in-treatment effects between the two incentive groups exceeded 0.10. For the first few cohorts, we made these random sub-treatment assignments prior to observing exercise behavior from the incentive period. Given the relatively small sample size of cohorts, we observed some imbalance between the incentive-only and incentive+commitment groups in terms of gym visits during the treatment period. For that reason we decided to change the protocols and conduct the randomization after the incentive period.

level attendance target for the commitment would be attractive to those most on the margin of establishing an effective exercise routine. Naturally, having a fixed contract with a modest goal, likely made the contract less desirable to some participants. While we think understanding optimal commitment contract design is important, we leave this area for future research.

Subjects in the incentive-only group were sent a nearly identical email that encouraged them to commit themselves to not missing more than 14 days in a row at the gym over the following 8 weeks. This email did not, however, mention putting money at stake for that goal. Thus, the difference during the commitment period between the incentive-only and incentive+commit groups measures the effect of the offer of commitment rather than the combined effect of the encouragement and offer of commitment.

Section 3. Data

Table 1 provides the means for key variables from our initial survey. The table is split in two panels by gym membership status prior to treatment. Columns (1) and (4) show means for the control group with standard deviations of continuous variables for the control group in parentheses. To explore whether randomization provided balance in these characteristics across the different groups, we also display estimated mean differences between the control and incentive-only group (columns (2) & (5)) and between the control and incentive+commit group (columns (3) & (6)).[12] The last two columns in each panel are the p-values from two tests: first, the equivalence of the means across the 3 randomized groups and second, the equivalence of the means across the incentive-only and the incentive+commit groups. Overall, the groups are fairly well balanced across the different treatments; none of the pre-treatment differences examined in Table 1 are statistically different from zero at the 5% level.

Our subject pool is on average 40 years old, roughly equally divided across genders, and is well-educated (more than 65% have a college degree or more). In comparison, overall in the United States in

[12] These estimated mean differences come from simple regressions that include strata fixed effects (a combination of gym membership, exercise relative to target and cohort), which are included in all regressions throughout. Including strata fixed effects ensures that results are not biased by fluctuations across cohorts in the shares of employees randomly sorted into control and treatment groups.

2009, just under 30% of adults aged 25 and older had at least a college degree.[13] Possible time constraints are measured by marital status, presence of children at home, and commute times. Although marital status and presence of children at home are comparable to overall US patterns, commute times are significantly longer.[14] Company employees are on average somewhat less unhappy than in the US as a whole (14.3% report being unhappy in the 2010 General Social Survey).[15] Based on self reports of height and weight, 69% of our subjects are either obese or overweight, statistics that resemble those at the national level.[16] Both existing members of the gym and non-members report on average being around 20 lbs. heavier than their personal target weight.

We asked subjects in the initial survey to report their current exercise activities and their targets for how often they would like to exercise. The average difference between targeted and self-reported exercise is 1.5 days/week for gym members and 2 days/week for non-gym members, implying that individuals want to increase their exercise and that incentives for exercise may move them closer to their target level. Given diminishing health returns to exercise, those who are inactive are likely to reap the largest returns. In our subject pool, rates of inactivity are high; this is true even among the gym members, as evidenced by the large fractions of individuals reporting no exercise in a typical week. Thus, our subjects likely have much to gain from increased exercise.

Section 4. Results

4.1 Overall graphical analysis

Figure 1 depicts the weekly time-series of the fraction of subjects with at least one visit to the company gym in a given week by treatment status. We combine the data for each cohort such that incentive week 1 is the first week of the incentive program for each cohort (i.e., week 1 on Figure 1) and all weeks are numbered relative to that week. We mark the incentive period with solid vertical bars at weeks 1 and 4. The other bar at week -2 denotes when subjects took the initial survey. Treated subjects

[13] This is based on authors' calculations using the 2010 Census.
[14] This is based on authors' calculations using the 2010 Census.
[15] Source of statistic is http://sda.berkeley.edu/cgi-bin/hsda?harcsda+gss10.
[16] http://www.cdc.gov/obesity/data/adult.html.

learned about the incentives at the end of week -1, and week 0 was set aside for non-members to become members of the gym.

This figure allows us to examine gym attendance over three distinct time periods. First, prior to the intervention, all 3 groups show similar rates of attendance at the company gym; approximately 20% attend the gym in any given week. The overlap of the attendance of all three groups gives us faith that randomization worked well. Second, during the incentive period, both incentive groups roughly double their use of the gym. As the announcement of the commitment contract offer occurred following the incentive period, we should see similar patterns for the two incentive groups during the incentive period, which we do. Third, after the end of the incentive period, both incentive groups reduce their frequency of exercise relative to their incentivized levels. However, the two incentive groups diverge during this post-treatment period. The attendance of the incentive+commit group remains clearly elevated relative to the control, signifying some lasting effects on behavior. The incentive-only group exhibits much more muted differences relative to the control. Without the commitment contract opportunity, it appears the overall long-run impact of the incentives is weak.

4.2 Regression framework

To supplement our graphical analysis and quantify our results, we run regressions using data from the pre-incentive, incentive, and post-incentive periods. Our regression models are of the following form:

$$\begin{aligned} y_{itw} = \alpha_0 + \alpha_1(IO) + \alpha_2(IC) + \delta_0 intreatment + \delta_1(IO) \times intreatment + \delta_2(IC) \times intreatment \\ + \beta_0 posttreatment + \beta_1(IO) \times posttreatment + \beta_2(IC) \times posttreatment + \mu_s \\ + \pi_w + \varepsilon_{itw} \end{aligned}$$

where y_{itw} is an outcome measure, such as an indicator for attendance, for subject *i* in incentive week *t,* and calendar (not experiment) week *w. IO* is a dummy variable for the incentive-only group, *IC* is a dummy variable for the incentive+commit group*, in-treatment* is a dummy variable for the in-treatment period, *post-treatment* is a dummy variable for the initial post-treatment period (weeks 5-13)*,* and μ_s

are strata fixed effects (i.e., "exercise vs target" x "cohort fixed effects").[17] Since there are weekly observations on the same individuals, we adjust the standard errors for clustering at the individual level.

The regression above combines the effects of the 3 time periods of interest – pre-intervention, intervention, and post-intervention into 1 regression, allowing for concurrent comparisons of effects. α_0 measures the mean outcome for the control group in the pre-intervention period. Thus, α_1 and α_2 measure differences for the incentive-only and incentive+commit groups relative to the control group, respectively in the pre-intervention period and should be near 0 due to randomization. δ_0 measures the mean level of the outcome for the control group during the incentive period relative to its pre-incentive period mean. Our "in-treatment effects" are given by δ_1 and δ_2, which are difference-in-difference parameters measuring the extent to which differences in the mean outcome for the incentive-only and incentive+commit groups, respectively, between the intervention and the pre-intervention periods differ from the analogous difference for the control group. We refer to $\hat{\delta}_1$ and $\hat{\delta}_2$ as our estimates of the effect of the incentives for the incentive-only and incentive+commit groups. We expect their values to be very similar since these treatments only differ in the post-intervention period. β_1 and β_2 are difference-in-difference parameters analogous to δ_1 and δ_2, except that they measure the extent to which differences in the mean outcome for the incentive-only and incentive+commit groups between the post-intervention and the pre-intervention periods differ from the analogous difference for the control group. Since in the post-treatment period, the incentive-only and the incentive+commit groups differ in treatment (i.e., this is the commitment period for the incentive+commit group), we interpret β_1 as the effect of the incentives on behavior in the post-treatment period and β_2 as the effect of the incentives and the commitment contract in this period. Thus, β_2 - β_1 is the effect of the commitment contract offer.

4.3 Results by membership status

We stratified the randomization based on gym-membership status prior to the experiment and present our results separately for members and non-members.

[17] We estimate the equation above for gym members and non-members, so we do not consider membership as part of the strata for this regression.

Members: We start our analysis by membership status with Figure 2a, the time-series analogue to Figure 1 for ex-ante gym members. To quantify these patterns, the first two columns of regression estimates in Table 2 display estimates of the regression above separately for two outcomes: any visit in a particular week and average number of weekly visits.[18] At the bottom of Table 2, we display the p-values comparing the incentive-only and the incentive+commit group coefficients in the pre-incentive, incentive, and post-incentive periods.

Among existing members, we see a substantial share not attending the gym in a given week before our intervention (around 40%), a finding that is not without precedent. DellaVigna and Malmendier (2006) also find that many gym members are fairly inactive with 20% attending less than once per month. The shares not attending are roughly balanced across our three groups giving us assurance that randomization worked well; indeed none of the estimated coefficients for the pre-incentive period are statistically distinguishable from zero.

The incentives increased the weekly attendance rate for members by approximately 20 percentage points (60% to 80%), and significantly raised their average number of weekly visits by 0.8-0.9 (roughly a 50% increase relative to the control group). As expected, the effect sizes for the incentive-only and incentive+commit groups are roughly equal during the incentive period.

The effects for the post-treatment period covering weeks 5 through 13 allows us to explore the lasting response to the incentives and the relative importance of the availability of the commitment contract. Existing members in the incentive-only group fall back rapidly to match the control group averages after the incentives are removed. That is, among existing members there is no lasting effect of the incentive by itself. On the other hand, rates of attendance are substantially higher after the incentive period for those in the incentive+commit group relative to those in the control and incentive-only groups; for the any visit outcome, the incentive+commit group attends the gym at 0.10 higher rate than the control and has 0.30 more visits per week than the control group does. These effects are

[18] For ease of interpretation, we present OLS estimates of these regressions. We also estimated probit models to take into account the binary nature of the dependent variable, “any visit,” and these models produced similar results.

roughly one-third to one-half of the size of the in-treatment effects for positive visits and average visits, respectively. Both of these effects are highly statistically significant and suggest that commitment contracts can increase the long-run effects of an incentive program. The incentive-only and incentive+commit difference in the post-incentive period, our estimate of the commitment contract offer on positive weekly visits, is 0.07. This difference has a p-value of 0.03 as shown at the bottom of Table 2. The p-value for average weekly visits is 0.16. These intent-to-treat effects suggest the commitment contract had a large effect for the 23% of members who created a contract (see bottom of Table 2). From Figure 2a, it is apparent that the positive effect of the commitment contract offer is also relatively stable – the gap between the incentive-only and the incentive+commit groups is long-lasting and quite constant.

Non-members: Looking at non-members, we note that prior to the intervention, the rates of attendance for these employees were, of course, zero. The incentive program motivated a fifth of the employees who were not already users of the gym to attend, with weekly attendance rates around 20% during the treatment period. Overall, combining the incentive-only and incentive+commit groups, we find that 22% of non-gym members joined the gym and attended at least once during our treatment period. Among the control group, only 1.5% joined the gym on their own over the same period. The treatment effects during the incentive period are highly significant. While the incentive effect on positive visits is similar to that of members, not surprisingly the effect on average visits is much less – 0.4 to 0.5 visits per week.

One may note that the incentive+commit group responded more to the incentive than the incentive-only group during the in-treatment period. This is an unexpected random difference in response rates. As a sensitivity check in a later table (discussed below), we address this concern. While the effects of commitment contracts diminish slightly in this analysis, they are still substantial and statistically significant.

Unlike the results for members, the post-treatment period effects for non-members reveal modest positive and statistically-significant effects of the incentive alone. During this period, non-members offered incentives were 4 percentage points more likely to attend than the control group. Measured relative to the 15 percentage point effect while incentives were in place, this would suggest that for non-members 25% of the in-treatment differences persisted after incentives were removed. The lasting effects are stronger for the incentive+commit group. We find that non-members in the incentive+commit group were 9 percentage points (a 15% increase relative to the control) more likely to attend the gym during this period than control. Measured relative to the in-treatment effect, the post-treatment effect for this group is 45% of the in-treatment effect size. Although the intent-to-treat impacts are not too different from those for members, the effect of the incentive+commit treatment for non-members is all the more impressive given that the commitment contract takeup rate is 0.06, a much smaller figure than that for members. The commitment contract effect on positive visits, that is, the difference between the incentive-only and incentive+commit groups in the post-treatment period, is 0.05 with a p-value of 0.02. Indeed for both members and non-members, we see that the commitment contract is particularly powerful at raising gym attendance beyond the incentive period.

4.4 Impact of the commitment-contract option

To hone in more closely on the effect of the commitment contract and address the possible concern about the differences observed between the incentive-only and the incentive+commit during the incentive period among the non-members, we offer an alternative approach to the regressions of Table 2. In Table 3, we present estimates of the commitment period difference in gym attendance of the incentive+commit group and the incentive-only group after controlling for whether or not the subject attended the gym in each week of the incentive period. The first and third columns of estimates are our estimates of the effect of the commitment contract offer on gym attendance for members and non-members, respectively. For both groups, these estimates are similar to the analogous estimates in Table 2 and they are highly statistically significant. The commitment contract offer effect for members in Table 3 is 0.08 whereas in Table 2, the incentive+commit and incentive difference is 0.07. For non-members, the Table 3 estimate is 0.04 and the Table 2 estimate is 0.05.

To further assure that these differences are due to the commitment contract offer, we limit the sample in columns 2 and 4 to those subjects attending the gym at least once during the incentive period. This eliminates any differences in attendance due to the differences along the extensive margin during the incentive period between the incentive-only and incentive+commit group. Not surprisingly, the effect of the commitment contract offer is larger for this sample and again highly significant. Overall the results of this Table imply that the commitment contract offer raised gym attendance.

The results in this table are "intention to treat" effects - effects for the entire group offered the chance to commit and not conditional on whether or not they actually committed. We can use the take-up rates to check the plausibility of the size of intention-to-treat estimates. If we assume that the entire difference in behavior between these groups is driven by the behavior of those who create commitments, we can divide the estimates in Table 3 by the commitment take-up rate to get an implied IV estimate. These estimates at the bottom of the table imply that the commitment contracts increased weekly gym attendance rates by 0.43 for members and 0.83 for non-members, quite sizable impacts on attendance. Given that the commitment required at least one visit every other week (0.5 in the outcome measure), these "treatment-on-the-treated" estimates are generally sensible.[19]

4.5 Longer-run effects

In Figures 3a and 3b, we investigate the very-long run impacts of the incentive program, following differences out a full year after the end of the incentive program (and hence 10 months after the end of the commitment program). In these figures the units of the x-axis are months rather than weeks where a month equals exactly four weeks. We present these effects for months 2 to 13 following the start of the incentive period. Months 2 and 3 encompass the commitment contract period. These figures display the difference in the probability of attending the gym in a given month between each of the incentive groups and the control. Figure 3a shows results for existing members of the gym. For members, gym attendance among the incentive-only group is only slightly elevated relative to the control group in the first month after the incentive period. By month 3, the attendance of the incentive-only group is below that of the control group and from then on, the difference remains near zero,

[19] We are, however, reluctant to interpret these estimates in the context of treatment on the treated. Although it is tempting to assume that any effect of the commitment-contract treatment would have to come through those making commitments, in this context the exclusion restriction need not apply. Subjects in the incentive+commit group were given the option to create a specific financial commitment contract that was monitored by us. It is possible that some subjects who chose not to create this commitment nonetheless were influenced by the idea of creating such a commitment to form their own commitment contract (e.g., with a friend) outside of the experiment. More generally, by suggesting the possibility of the financial commitment, the treatment may have changed how subjects thought about the importance of committing.

indicating little persistence in the effect of the incentives. For the incentive+commit group, the patterns are much different. The effect of the incentive+commit intervention is quite long-lasting. Those offered commitments see strongly-elevated visit frequency during months 2 and 3 when the commitment is in place, with frequencies more than 10 percentage points higher than control. While these differences also fade after the end of the commitment program, they remain elevated relative to control until about month 11.

Figure 3b is the comparable figure for non-members. For this group, the incentive effects are more persistent and stable. For those offered incentives but not commitment contracts, there is a stable difference of around 3 percentage points in the fraction using the company gym relative to control. The difference between the incentive+commit and control group is more pronounced. In particular, those offered commitments attend at frequencies 8-10 percentage points higher than control during months 2 and 3 when the commitment contract is in place. The attendance differential falls after the end of the commitment-contract period and stabilizes at a frequency 5-6 percentage points higher than the control group a year after the incentive program began.

4.6 Accounting for substitution

Our results above show that there were meaningful and lasting effects (especially for the incentive+commit group) of incentives on attendance at the company gym. An obvious concern with these estimates, however, is that some of the changes in behavior might reflect substitution. For example, in our case, subjects might simply start exercising at the company gym as a substitute for their exercise elsewhere. Substitution is an important issue with most incentive programs, since most target a particular measurable behavior, but the degree to which substitution occurs for these types of programs is largely unknown.[20] To test for substitution, we use data from the follow-up survey, which includes questions about overall exercise and exercise at the corporate gym during the incentive period.

[20] At least one other study in this area has attempted to measure substitution, but the conclusions are unclear. Charness and Gneezy (2009) ask participants to fill out an exercise log. The log includes questions about overall exercise, exercise at a gym, and exercise outside of a gym. As they state, the self-reported data in their case do not seem to be reliable. For example, the effect on gym use for the main incentive group is 0.04 gym visits/week whereas that measured via administrative data is 1.22 university gym visits/week. The difference in these

Table 4 presents the relevant estimates dividing the sample along the lines implied by our stratification variables for randomization (i.e., membership status and level of exercise relative to target level of exercise). We provide estimates for members and non-members separately at the top of the table (Panel A). The remaining panels delve further into possible heterogeneity, dividing the sample by whether an individual's pre-intervention overall exercise was below their target (Panel B) or at or above their target (Panel C). We postulated that for individuals at or above their target level of exercise, the incentives would not lead to increases in overall exercise but substitution of the location of exercise, especially in the case of non-members since for many of them earning incentives would only require that they move their existing exercise to the company gym.[21]

Before discussing the substitution results in terms of overall exercise, it is useful to examine whether the impact of the incentives using the self-reported levels of exercise at the corporate gym match well with those derived from computerized gym records reported in Table 2. Since the gym data are very likely to be accurate, this comparison allows us to check the validity of the survey data. The first column in each set of estimates is derived from the computerized gym records. We combine the incentive-only and incentive+commit groups for this analysis because we are examining effects during the four-week incentive period when these two groups were treated identically. The second column is the equivalent measure using the self-reported data. The sample sizes differ across the two sets of regressions because of non-response to the follow up survey; regressions estimated using the computerized gym data on the sample of follow up survey responders give similar estimates.

There are two main conclusions from the contrast of weekly visits as measured by gym data versus survey data: a) existing gym members tend to overstate their frequency of visits in the survey data relative to that observed in the actual login data and b) despite that overstatement, the estimated effects of the treatment on gym attendance are generally quite similar across the two data sources. Essentially, it looks as though self-reports are inflated, but that inflation seems to hold in broadly similar ways for both the control group and the treatment groups, leading to generally sensible estimated

---

estimates could reflect considerable measurement error in the exercise logs or significant substitution (i.e., substitution of other gyms for the university gym).

[21] The cost of moving their location of exercise could be large.

treatment effects from the survey data. For example, looking at Table 4a Panel B, for members below their target level of exercise, the control group had average weekly visits of 0.97 in the login data. This group had a self-reported visit average of 1.46 visits per week, suggesting that their self-reports are inflated by roughly 50% on average. However, the effect of the incentives is similar using either measure – a 0.88/week increase in visits and a 0.76 increase in visits using the actual login data. This similarity between treatment effects using actual data and self-reports holds for all of the subgroups in Table 4, except for members at/above their target level of exercise.[22] Excluding this group, the average margin of error between the login data and the self-reported data is fairly low – 16%. For the members at/above target, however, the discrepancy between the gym and survey data suggests that the self-reported data may not be very accurate and thus, we may want to be cautious about interpreting the substitution effects for that group.

A natural method of assessing the degree of substitution is to compare the treatment effects for overall exercise to those for company gym exercise. That is, we can simply divide column 3 by column 2 in each of the panels of Table 4 to get an estimate of the fraction of the overall exercise effect that is attributable to company gym visits. If the two estimated treatment coefficients are the same (a ratio of 1), it would suggest that effects on company gym exercise represent one-for-one increases in overall exercise and there was no substitution. Focusing on the overall effects in Panel A, we see that this ratio is 82% for existing members and 63% for non-members. The weighted average of those two effects suggests that 70% of the overall increase in exercise at the company gym from the treatment represented new exercise. Looking at Panel B, for those reporting low levels of exercise prior to the experiment, most of the increases in company-gym attendance appear to have been new exercise. Among existing members, 100% of the company-gym effect was new exercise and for non-members 70% was new exercise. For the company gym members, this finding is not surprising since their main place of exercise is the company gym. For the non-gym members, the result is also not too surprising since many had little pre-existing exercise away from which to substitute. For non-members at or above their target level of exercise, there appears to be considerable substitution. Taken literally, 75% of the

[22] For this group, how often they go (3 versus 4 times) may be less salient than for a less frequent user.

effect is substitution. The effects for members at/above target are difficult to interpret because of the measurement issues discussed above. Since roughly 70% of our subjects were below their target level of exercise, even if we assume minimal effects for those at or above target, the program generated a real change in exercise behavior for the majority of our subjects. Furthermore, since the health benefits from increased exercise are likely greater among those who exercise less, the increase in exercise was concentrated amongst those who stood to reap the largest health benefits.

This substitution analysis is based on information from our follow-up survey. Although the response rates to that survey are high (91.4%), we did observe some statistically-significant differential attrition between the treatment group and control group among non-members as seen in Appendix Table 1. In this table, we display estimates from regressions of whether or not an individual responded to the survey as a function of treatment status. Non-members offered the incentive program were somewhat less likely to respond to the second survey than non-members in the control group (87% vs. 96%). To address the possible non-response bias for non-members in such analyses, we estimate the degree to which non-response might affect our substitution estimates in Appendix A. We show that given reasonable assumptions about the behaviors of non-responders, the implied substitution effects in Table 4 could understate the degree of substitution by roughly 5-10%. Specifically, the Table 4 estimates imply that 63% of the treatment effect for non-members is new exercise whereas estimates accounting for the possibility of non-response bias imply a percentage closer to 61%. Thus, it is reasonable to conclude that most of the treatment effect on company gym visits for non-members below target is not substitution whereas for those at/above target, the treatment effect is mostly substitution.

4.7 Heterogeneity of response to incentives

In this subsection, we discuss the heterogeneity of the response to the incentive-only and incentive+commit treatments in order to gain more insight into how financial incentives interact with time inconsistency. The theoretical literature on time inconsistency suggests two potential sources of heterogeneity that could affect whether incentives lead to lasting behavior change in our setting: a) variation in the size of the "start-up" costs a person faces in using the gym and b) variation across individuals in the degree of present bias, which could cause variation in the ability to establish a lasting habit even once the initial cost of using the gym are overcome. If start-up costs are the key impediment to exercise, then we would expect to see greater post treatment effects for the non-members since members have already overcome this hurdle by joining the gym. In support of this point, the post-treatment effects for the incentive-only group are stronger for non-members when the post treatment effects are measured as percentages of the in-treatment effects. The commitment contract may be an effective tool primarily for individuals with a strong present bias, because this group may struggle more with the day-to-day challenge of exercising. Consistent with that idea, Table 2 showed a strong lasting effect for the group offered the commitments relative to the control group for both members and non-members.

To further examine the possible heterogeneous effects to our treatments, in Table 5 we consider subsamples based on pre-intervention exercise behavior. For members, we divide subjects based on terciles of the distribution of the fraction of weeks in the pre-period with positive visits to the gym. The "low" group visited the gym 2 or fewer weeks in the 6-week period before the intervention, the "middle" group 3-5 weeks, and the "high" group all 6 weeks. For non-members, measures of actual attendance are zero, so we conduct a similar tercile cut based on their self-reported overall exercise in the pre-treatment period. We divide the sample as follows: the "low" group reports exercising on average 0 to 0.5 times per week, the "mid" group 1-2.5 times per week, and the "high" group 3-7 times per week. It is, of course, important to recognize that attributing the differences in behavior to pre-intervention differences in exercise is problematic, as other factors, besides their pre-intervention behavior, could explain the patterns.[23] Nonetheless, we feel this analysis can provide useful suggestive evidence about the mechanisms behind the main effects.

Among members, the in-treatment effects are statistically significant for each of these sub-groups, but naturally are larger for those who were not exercising much prior to the study. Interesting patterns emerge when looking at the post-treatment effects. Specifically, the long-run effects of the incentive-only group are statistically insignificant for all groups but show larger effects for the lowest

[23] We did not pre-specify the groupings for pre-treatment exercise. As such, this investigation is an exploratory analysis that should be interpreted with some caution.

level exercisers. The "low" group here is analogous to the infrequent gym users identified in DellaVigna and Malmendier's (2006) study of gym use. The fact that this group has a somewhat larger longer-run response to the incentive could indicate that their low levels of exercise have left them with "startup costs" associated with re-initiating their use of the gym. Turning to the effects of the commitment contract paired with incentives, we see that the incentive+commit group had significant post-incentive effects for all groups except the high-exercise group. For the mid-exercise group, the size of the post-incentive effects is only slightly smaller than the incentive effects (0.18 versus 0.22), suggesting that commitments are particularly effective for this group. That finding is interesting because it is consistent with the possibility that commitment technologies might be most useful for those whose level of present bias allows them to exercise periodically but not consistently.

The parallel analysis for non-members reveals that the incentive response was fairly similar across all of these groups. We also find little meaningful differences among these groups looking at the post-treatment period, though the low-exercisers do not appear to have had any lasting effect from the incentive alone. Confirming our prior analysis, the availability of the commitment contract increased gym-attendance over the incentive alone across all groups. In fact without the commitment contract, in the "low" and "high" groups we observe no significant effects on long-run attendance.

Another insightful heterogeneity cut is that by gender. Later, when looking at commitment contract demand, we find that women demand commitment more than men. Looking at Table 5, in general, men increased their exercise during the treatment period in response to incentives more than women did. That difference is most pronounced among those who were already members of the gym. Looking at the post-treatment effects, there are modest differences between men and women showing that men had slightly higher long-run effects to both the incentive-only and incentive+commit treatments. Big post-treatment differences between men and women emerge for the existing members. Male members of the gym show a significant lasting response to receiving the incentive alone (i.e., with no commitment option). In contrast, women offered only incentives have a statistically insignificant and negative post-treatment effect. The availability of the commitment option raises the post-treatment effects only slightly for male members. For female members, however, there is a substantial increase in the post-treatment effects when commitments are offered relative to only incentives.

We should also note that we have also explored heterogeneity in these effects by education level. For the sake of brevity, we do not report them here. We postulated that the effects could vary by education because of differences in the value of the incentive. For the less educated worker, the incentive likely is a higher percentage of their earnings. However, the treatment effects do not differ significantly across education level.

4.8 Cross-contamination effects

One of the challenges in conducting a randomized workplace intervention is that since workplaces are usually closed environments, subjects in the experiment will often interact with each other.[24] Two different types of interactions that might be important when interpreting the results of an experiment, like ours, that is conducted in a closed setting. The first issue is that individuals may talk to one another, what we might term the "cross-talk issue." Individuals who are not in the experiment or in the control group may learn about the treatment. The second issue, what we call the "spillovers issue", is that even in the absence of "cross-talk," the social nature of exercise may cause there to be an interdependence of behaviors among individuals. In this section we consider these possibilities and how they might affect the interpretation of our study.

The potential for cross talk at the workplace presents two possible challenges to the interpretation of our results. First, we expect that over time many people at the company learned about the possibility of being selected for the incentive treatment in the study. That leak of information could in theory generate selection bias, because one could imagine that those most interested in being paid to use the company gym would respond more to our initial survey. However, we doubt that this type of selection occurred because it should lead survey response rates to increase over time, which they did not. Also, this type of selection might cause the share of existing gym members who respond to the survey to grow over time, which it does not. Finally, the treatment effect does not increase across cohorts, a finding we would expect if interest in the program was higher among later cohorts.

[24] The standard treatment effects literature assumes the existence of the stable-unit-treatment-value (SUTVA) assumption (Cox 1958) and such cross-contamination effects would be a violation of this assumption.

The second concern with cross talk among employees is awareness of the incentives in the control group (especially in later cohorts). It is possible that disappointment with the random selection or jealousy of the treatment group could have discouraged the control group from attending the gym. That type of decrease in control-group use would bias up our estimated treatment effects. However, the control group attendance does not change from the pre-incentive to the incentive period.

If exercise has positive social spillovers, incentivizing one person may also increase exercise of those they interact with. With positive spillovers, increased exercise by the treatment group could induce exercise of the control group. When we compare the treatment group to the control group, we would then understate the increase in exercise induced by the treatment program.

A recent related study, Babcock and Hartman (2011), explored the nature of spillovers of financial treatments. They find that incentive effects are stronger when treated individuals have treated best friends, suggesting some positive spillovers among treated individuals. However, the treatment status of the control group's best friends does not affect behavior and treated individuals do not appear to respond differently when they have casual acquaintances who are also treated, suggesting the extent of spillovers may be fairly limited.

Our experiment was conducted with small cohorts, so that at any one time only around 3% of the company was eligible for incentives. As such, there was limited scope in our setting for spillover effects to emerge. Nonetheless, we have analyzed our data to look for these types of spillover effects and did not detect any significant patterns. Specifically, we looked at whether the treatment effects varied by the fraction of one's department unit incentivized, suspecting that the higher the fraction incentivized the larger the treatment effect. Department unit is the most natural social grouping in our data.[25] We assessed the size of possible spillover effects by estimating our basic regression models but including an interaction term of treatment status with the fraction of the individual's unit who was treated in that cohort. The mean fraction incentivized for a particular cohort was 11.9% (median 6.8%). For the in-treatment period, the interaction term was not statistically different from zero for both positive visits and average number of visit outcomes. Thus, it does not appear that individuals respond

[25] The 236 department units vary in size from 1 person to 121 persons.

more to the incentives when more of those they work with are incentivized. We also do not find any effect of changes in visits for the control group based on the fraction of their department that is incentivized at the same time.[26]

4.9 Understanding take-up of commitment contracts

In this subsection we explore the correlates of the demand for a commitment contract. This subsection is somewhat speculative as the analysis is based on cross-sectional regressions that rely on non-experimental variation. Nevertheless, given that our commitment treatment led to a lasting effect of the incentive program, and given the limited number of studies on the demand for commitment, we think that this analysis can be informative.

Overall among the 346 subjects in the incentive+commit group, 12.4% chose to make a commitment and on average these committers placed $58 at stake. Among ex-ante gym members, the take-up rate of commitments was 23%. For those who were not members of the gym prior to the study, the overall take-up rate was 6%, but take-up was 21% for those making at least one visit during the incentive period. Sixty-three percent of those who created commitments successfully maintained the commitment of not missing more than 14 days in a row at the gym. Although these take up rates are somewhat modest, they are in line with existing studies of commitment contracts. For example, Gine, Karlan and Zinman (2010) saw 11% take-up of their smoking-cessation commitment device in the Philippines. Ashraf, Karlan and Yin (2006) saw 28% take-up of their commitment savings product.

Table 6 shows the results from regressions examining the correlates of commitment contract demand. For this analysis we restrict the sample to those subjects who were offered the commitment option (incentive+commit group) and stated on the follow up survey (conducted during the last week of the incentive program)that they had interest in using the company gym over the following weeks (67%

[26] Also, there is a possibility of crowding the gym, a type of negative spillover. Since our intervention increased attendance, the gym as a result may have been more crowded, possibly making to less attractive to others. Part of the reason for conducting our experiment in small cohorts was to avoid overcrowding at the gym at any one point. Given the stability of the attendance trends among the control group over time, we are not too concerned about this overcrowding concern.

of the sample or 231 subjects).[27] In this way we focus on those who had some possibility of committing, because (unsurprisingly) none of those without interest in using the gym decided to make a commitment. Limiting the analysis to follow-up survey responders leaves us a sample of 224 subjects. The overall take up rate of commitment in this group was 19%.

In each column of Table 6 we include an indicator for whether the subject was a member of the gym prior to the study and control for the average number of visits per week the subject made to the gym during the incentive period. Being a member ex-ante is correlated with a slightly higher rate of commitment, but that difference is modest and not statistically significant once we have controlled for visit patterns during the treatment period. Visit patterns matter. In particular, the highest take-up rates are for those who used the gym 2-3 times a week during the treatment period (i.e., those who fell just short of reaching the maximal incentives).[28]

All columns also include demographic controls for gender, age, children at home, and a college degree. We find that men are significantly less likely (17 percentage points) to make commitment contracts than women. This gender effect aligns with our findings in Table 5, where we saw a larger difference for women than for men in post-treatment effects when commitment contracts were offered. We also find a large age effect. Employees in the bottom quartile of age (age < 33) are 15-20 percentage points less likely to make a commitment contract than older employees.[29] The presence of children at home and one's education level are not significant correlates of commitment contract demand.

The expected predictors of the desire for commitment, motivated by models of time inconsistency, do not predict commitment demand. Specifically we investigate whether three different measures are related to commitment demand: exercise relative to one's own target as measured by our initial survey (column 1), self-reported self-control for exercise (column 2), and self-reported intrinsic

[27] The follow up survey with these measures of future desired gym use was conducted before subjects learned about the commitment option.
[28] The p-value on the difference in commitment rates between those with weekly visits between 2 and 3 versus those with over 3 visits hovers around 0.05 for each specification.
[29] These results hold if we include dummies for different age quartiles. We present the simple comparison of "young" employees versus older employees to simplify the exposition. There are little differences in take up rate by age among the higher quartiles of age.

motivation (column 3). We anticipated that each of these measures might be positively correlated with the demand for commitment as they all seem to capture some level of meta-awareness of a self-control problem for exercise. Instead we find that each is slightly negatively correlated with take up of commitment (i.e., strangely perverse) but is not statistically significant.

In addition to these questions from the initial survey, as part of the follow up survey, we asked subjects about their expectations for how often they would attend the company gym over the subsequent 8 weeks. We also asked them about how often they would ideally attend if they had no problems with motivation or self-control. We use those answers to construct two measures that might predict demand. In column (4), we include an indicator for whether the subject expects to go less than they ideally would like to (expect < ideal) in the post-period. This is our most direct measure of what appears to be an awareness of a self-control problem. Like our other self-reported measures related to self-control awareness, we again see a slightly negative and statistically insignificant effect. In column (5), we present a rather different measure. Here we include an indicator for whether the subject expected to attend on average more in the post-treatment period than he reported actually attending during the incentive period (expect > actual). Since these subjects are planning to go even more than they report they did while incentivized, we label them "extreme optimists." Thirty-nine percent of subjects of the sample are classified as "extreme optimists." As a comparison, among the control group subjects who also stated a desire to use the company gym on the follow up survey, we find that 47% are extreme optimists. This extreme optimism measure has a sizable (0.12) and marginally statistically significant (p-value = 0.08) partial correlation with take-up of commitment. These results suggest that simple measures of self-control awareness do little to predict the demand for commitment, but there is some predictive power from a measure that captures an (potentially unrealistic) expectation of improvement in behavior.[30]

Models of time inconsistency (e.g., O'Donoghue and Rabin 2001) generally predict that those who are aware of a self-control problem (i.e., "sophisticates") will demand commitment, while those

[30] One reasonable interpretation of this result is that the "extreme optimists" are revealing directly that they were disappointed in their behavior during the incentive period. Furthermore, by stating that they expect to increase exercise, they may be revealing that they believe their low attendance was truly avoidable.

who are overly optimistic about their future behavior may not see the need for commitment - two predictions that are not met in our data. That could simply be because it is very difficult empirically to identify those who theoretically have a demand for commitment. On the other hand, it might suggest that the demand for commitment has more subtle motivations than the literature has previously considered. Another challenge to the traditional "sophisticated present-bias" view of the demand for commitment is that we see a take-up rate of commitment of 22% even among those who were high-use members of the gym prior to the study (i.e., those attending on average at least 3 times per week prior to our study). These individuals show no obvious need for this commitment device. The commitment is unlikely to bind for them because the contract only requires an average 0.5 visits per week. In the final section we discuss some ways in which the findings here might suggest new avenues for research into the demand for commitment.

Section 5. Discussion and Conclusion

5.1 Relationship to prior literature

In our field experiment, working adults responded strongly to financial incentives for gym use and this incentive had a long lasting, although modest, impact beyond the incentive period. This long-run effect of the incentive program was concentrated among employees who were not using the gym prior to the study. This suggests that the high start-up costs of exercise are a barrier to people with time inconsistency and that incentive programs could be a useful tool for overcoming that start-up barrier. However, despite the stronger effects for non-members, the overall lasting impact of the incentive-only program was minimal. Seen in that light, these results add to a large list of settings where health interventions have shown little ability to generate lasting changes in behavior.

To gauge our long-run effect size, we can compare both our incentive and post-incentive results in Table 2 to those of Charness and Gneezy (2009) and Acland and Levy (2011), both of whom report average visit measures. The sample of Acland and Levy (2011) is university students and staff at UC Berkeley who self-report that they had never attended a gym regularly whereas for Charness and Gneezy (2009), it is university students, regardless of their exercise habits. In both cases, subjects were incentivized to attend the university gym for $25 once during the first week of the experiment and then were incentivized to go to the gym 8 more times over the succeeding 4 weeks. For Charness and Gneezy (2009), gym membership was free to students, and in the case of Acland and Levy (2011), the researchers paid the membership ($10 cost) for each subject and filled out the required paperwork for non-members. Thus, the length and the incentive structure of these two studies are similar to our own albeit with important differences: a) gym membership is not free (or nearly free) for our subjects as it may not be for many workers and b) our incentive structure is a per-visit incentive whereas in the other studies, it has a threshold structure.

The in-treatment incentive effects for Charness and Gneezy (2009) and Acland and Levy (2011) imply that the incentives increase attendance by 1.2 visits per week. Our estimates are much more modest – 0.56 visits per week, suggesting that employees are less responsive to incentives than university students. The post-treatment effect for Charness and Gneezy (2009) is 0.59 whereas for Acland and Levy (2011), it is 0.26. Our estimate of post-treatment effects for employees offered only incentives is again substantially smaller at 0.11.

5.2 Insights for corporate wellness programs

Our study was designed to test the response of a health behavior, gym attendance, to incentives and not as a program evaluation of an implemented corporate wellness program. An appropriate cost-benefit calculation would take into account the effect of the intervention on health in addition to absenteeism, productivity, and the myriad of other outcomes that exercise may influence. We do not have access to data on absenteeism, productivity, worker well-being or health-care spending for the employees who took part in our study. Nonetheless, it is possible to use the results of our study to generate back-of-the-envelope calculations that may provide some benchmarks for those interested in employing wellness incentives.

First, while it is clear that employees respond to financial incentives to exercise, it is important to note that relatively little of the money spent on incentives goes to new exercise. Taking into account pre-treatment exercise levels and our estimates of substitution effects, we conclude that approximately

35% of the cost of the incentive program was spent on new behavior, while 65% paid employees for exercise they would have done without the incentive.

Our temporary incentive program coupled with a commitment contract option at the end is likely to have a better cost-benefit ratio than a consistent-incentive program, because it takes advantage of lasting changes in behavior that do not require continued payments from the company. On average, the program cost $57 per individual for those who were offered incentives. Using annual health care costs per employee of $5,049 (Kaiser Family Foundation), the cost of the program equals 1.1% of health care costs. During the 13 weeks it was in effect and the 39 weeks following, the program produced an increase in average weekly visits to the company gym of 37%. During the treatment period, across all subjects, 70% of gym visits represented new exercise. Outside the treatment period, it is likely that a larger fraction of gym visits represent new exercise since subjects did not gain monetarily by switching exercise to the gym. Nevertheless, if we conservatively assume that 70% of gym visits represented new exercise over the year, then the program produced an increase in exercise of 26%. Thus, the program pays for itself in terms of reduced health care costs if a 26% increase in exercise frequency results in a 1.1% decrease in health care costs. Put another way, the program would be cost effective in terms of health-care-cost reductions if the absolute value of the elasticity of health care costs with respect to exercise is at least 0.04.

Baicker, Cutler and Song (2010) estimate that reductions in absences from work are a key channel for the benefits of workplace wellness programs and use a figure of $20 per hour (or ~ $160 per day) to value an absence. Based on that rate, the $57 per-employee cost of an incentive program like ours could be made up through reduced absences if it reduced the yearly per-employee absences by 0.36. That is, if roughly 1 in 3 employees experience one day less of absence due to the program, the program would be paid for through that channel. Taken together, these back-of-the-envelope calculations suggest to us that a temporary incentive program coupled with a commitment contract to improve lasting effects would likely be quite cost effective for an employer.

5.3 Role of commitment contracts

We find that coupling the incentive program with an option for subjects to create their own self-funded commitment contracts substantially improves the long-run effects. In particular, when offered the chance to commit, employees who exercised some, but irregularly, prior to the intervention saw substantial improvements in exercise behavior after the incentive was removed. This pattern is consistent with the existence of important time-inconsistency problems that limit people's ability to maintain an exercise routine even after they overcome the initial start-up costs of exercise. More generally, this is the first study to show that commitment contracts can be an effective way of improving the long-run effect of health behavior interventions.

Our commitment contract program is unique in that it is preceded by an incentive program. Just as commitments increase the lasting effect of incentives, it is also possible that incentives increase the effectiveness of commitment contracts.[31] The incentive program may condition individuals to exercise and in the process, help them learn that incentives may be an effective method of encouragement, leading later to higher take-up rates of commitment contracts than in absence of such incentives. As future research, it would be interesting to test whether demand for commitment contracts can be augmented by offering commitment contracts following an intervention. An increase in commitment demand may be desirable given the low cost of commitment contracts.

Although this study finds promising results on the effectiveness of commitment contracts, it also leaves many open questions about the nature of the demand for these contracts. For example, we find large gender differences in the take-up of commitment, but more research will be needed to better understand those differences. Our results also suggest that the demand for commitment may be more subtle than the theoretical literature on time inconsistency would have predicted. One possibility that seems worth exploration in future research is that financial commitment contracts could serve as substitutes for other forms self-control. For example, research on the concept of ego depletion (Baumeister et al., 1998, 2000) suggests that self-control is like a muscle and fatigues when used. It may be that having a financial commitment in place helps people to exert less personal effort to maintain

[31] In a small pilot study, the commitment contract takeup rate in absence of incentives was quite low.

self-control. In such situations it is even possible that commitment contracts could improve welfare even without measurably changing observed behavior. Since our study was the first to offer commitments broadly to even those with no apparent need for commitments, little is known about this issue. More research is needed to understand the boundaries of where the use of commitment contracts can be valuable.

References:

Acland, Dan and Matthew Levy. 2011. "Habit Formation and Naiveté in Gym Attendance: Evidence from a Field Experiment." Working Paper.

Ashraf, Nava, Dean Karlan and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *Quarterly Journal of Economics,* 121(2): 673-697.

Babcock, Philip and John Hartman. 2010. "Networks and Workouts: Treatment Size and Status Specific Peer Effects in a Randomized Field Experiment" NBER Working Paper No. 16581.

Babcock, Philip, Kelly Bedard, Gary Charness, John Hartman, and Heather Royer. 2011. "Letting Down the Team? Evidence of Social Effects of Team Incentives" NBER Working Paper No. 16687.

Baicker, K., D. Cutler, and Z. Song (2010). "Workplace Wellness Programs Can Generate Savings." *Health Affairs,* 29(2): 304-311.

Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice. 1998. "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology*, 74(5): 1252-1265.

Baumeister, R. F., M. Muraven, M., and D.M. Tice 2000. "Ego Depletion: A Resource Model of Volition, Self-Regulation, and Controlled Processing" *Social Cognition* 18(2): 130-150.

Benartzi, Shlomo and Richard Thaler. 2004. "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy*, 112(1): S164-S187.

Beshears, John, James J. Choi, David Laibson, Brigitte Madrian, and Jung Sakong. 2011. "Self Control and Liquidity: How to Design a Commitment Contract."

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. 2012. "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction." *American Economic Review* 102(6): 2981-3003.

Cawley, John and Joshua Price. 2011. "Outcomes in a Program that Offers Financial Rewards for Weight Loss." In *Economic Aspects of Obesity*, Michael Grossman and Naci H. Mocan, editors, National Bureau of Economic Research.

CDC. 2011. http://www.cdc.gov/nchs/fastats/exercise.htm.

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

Cox DR. 1958. *Planning of Experiments*. Wiley; New York.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review*, 96: 694-719.

Finkelstein, Eric, Laura Linnan, Deborah Tate, and Ben Birken. 2007. "A Pilot Study Testing the Effect of Different Levels of Financial Incentives on Weight Loss among Overweight Employees." *Journal of Occupational and Environmental Medicine* 49(9): 981-989.

Finkelstein, Eric A., Justin G. Trogdon, Joel W. Cohen and William Dietz (2009). "Annual medical spending attributable to obesity: payer-and service-specific estimates." *Health Affairs*, 28(5):822-831.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, 40(2): 351-401.

Fryer, Roland. 2010. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." National Bureau of Economic Research Working Paper No. 15898.

Giné, Xavier, Dean Karlan, and Jonathan Zinman. 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." *American Economic Journal: Applied Economics*, 2(4): 213–35.

Gneezy, Uri, Stephen Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *The Journal of Economic Perspectives* 25(4): 191—209.

Goldhaber-Feibert, Jeremy, Erik Blumenkranz, and Alan M. Garber. 2010. "Committing to Exercise: Contract Design for Virtuous Habit Formation." NBER Working Paper w16624.

Jeffery, Robert, Hellerstedt, Wendy L., and Schmid, Thomas L. 1990. "Correspondence programs for smoking cessation and weight control: A comparison of two strategies in the Minnesota Heart Health Program." *Health Psychology*, 9(5): 585-598.

John, Leslie K, Loewenstein, George, Troxel, Andrea B, Norton, Laurie, Fassbender, Jennifer E., and Kevin G. Volpp. 2011. "Financial Incentives for Extended Weight Loss: A Randomized, Controlled Trial." *Journal of General Internal Medicine*, 26(6): 621-626.

Laibson, D. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics,* 112(2); 443–477.

List, John. 2009. "An Introduction to Field Experiments in Economics." *Journal of Economic Behavior and Organization* 70(3): 439-442.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *The Quarterly Journal of Economics* 118(4): 1209-1248.

Milkman, Katherine L, Julia A Minson, and Kevin G.M. Volpp. 2012. "Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling." Working paper.

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review*, 89(1): 103-124.

O'Donoghue, Ted and Matthew Rabin. 2001. "Choice and Procrastination," *Quarterly Journal of Economics*, 116(1): 121-160.

Phelps, E. and R. Pollak. 1968. "A second-best national saving and game-equilibrium growth." *Review of Economic Studies,* 35; 185–199.

Strotz, R.H. 1955-56. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies,* 23 (3): 165–180.

Volpp, Kevin, John, Leslie, Troxel, Andrea, Norton, Laurie, Fassbender, Jennifer, and George Loewenstein. 2008. "Financial Incentive-Based Approaches for Weight Loss: A Randomized Trial." *JAMA,* 300(22): 2631-2637.

Figure 1. Fraction with positive gym visits by treatment status

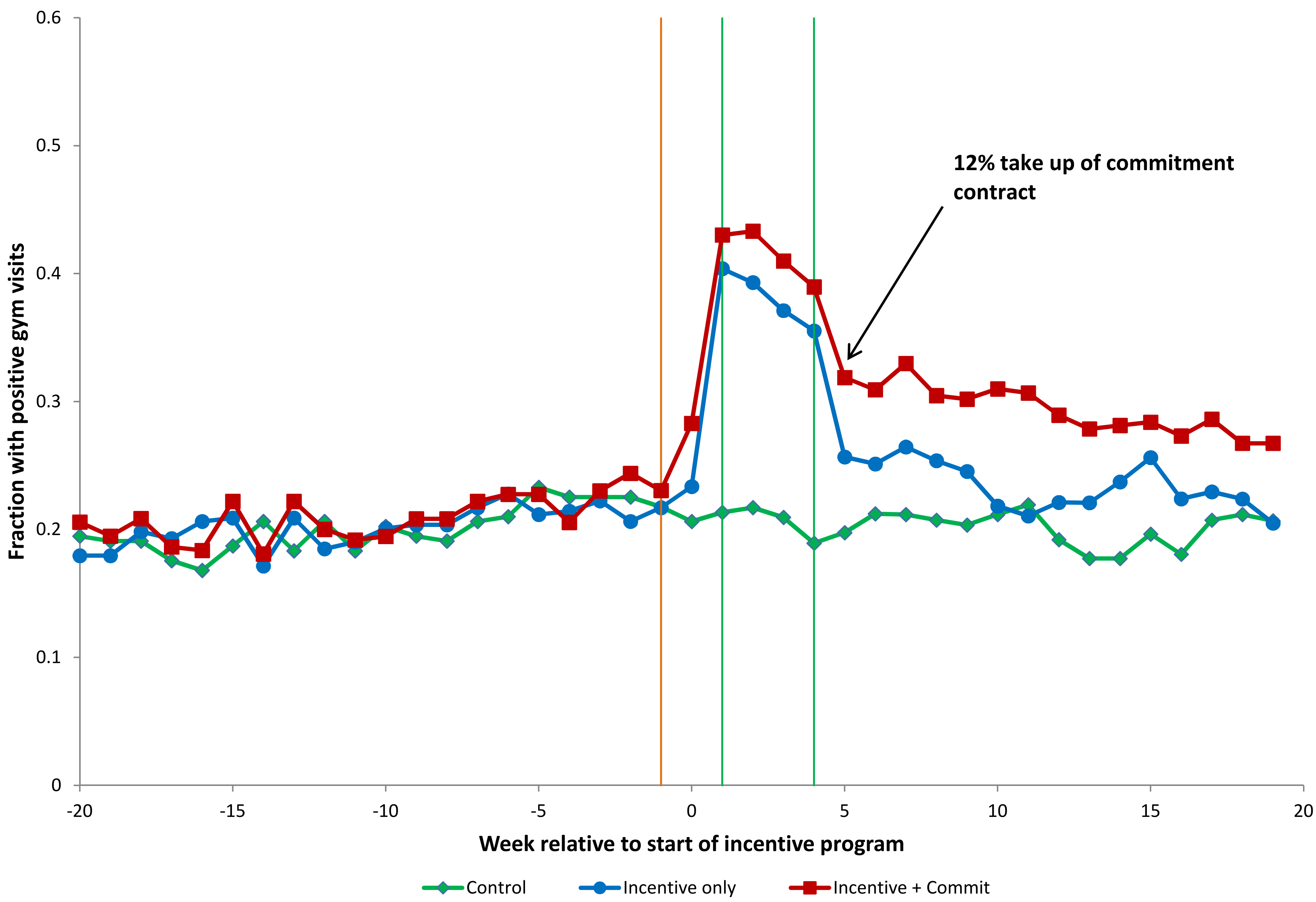


Figure 2a. Fraction with positive gym visits by treatment status (members only)

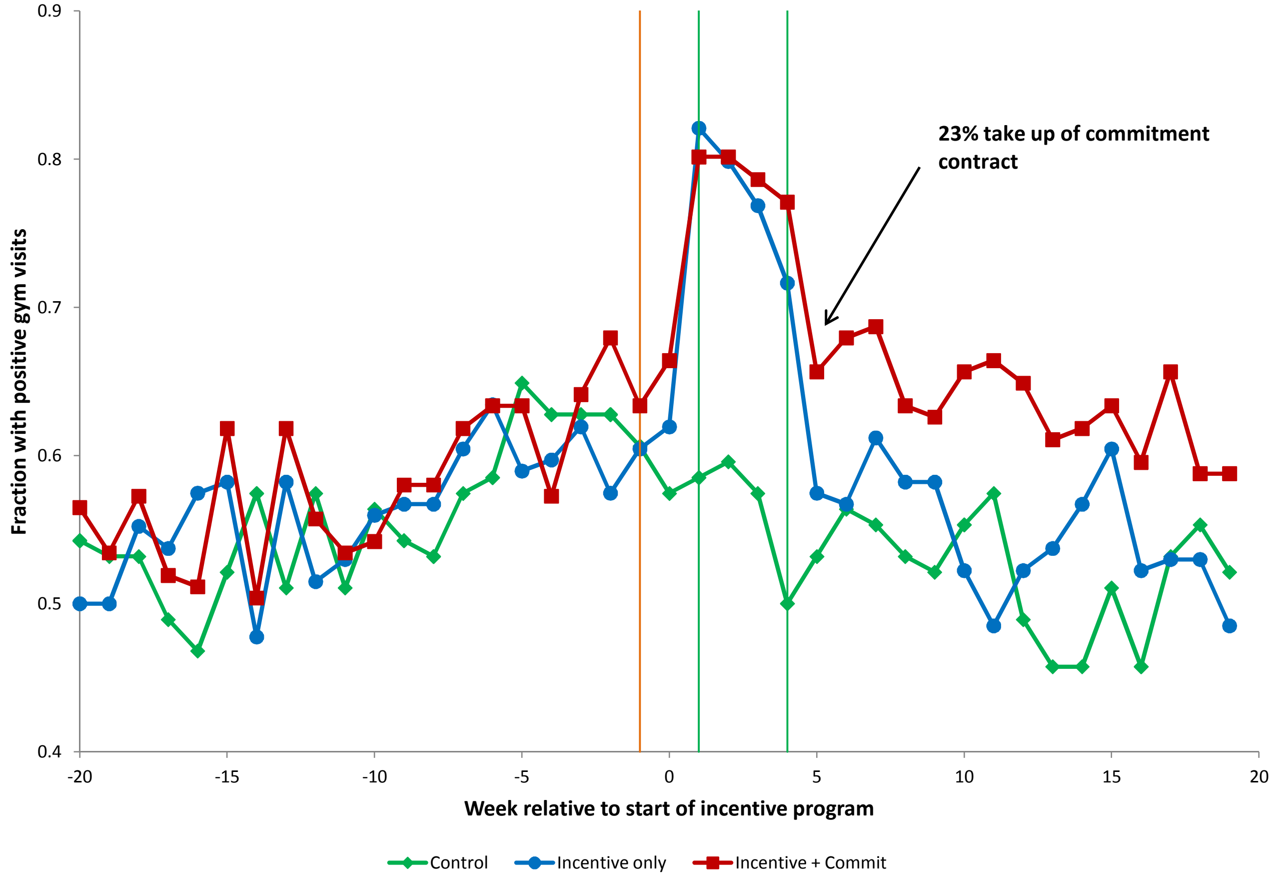


Figure 2b. Fraction with positive gym visits by treatment status (non-members only)

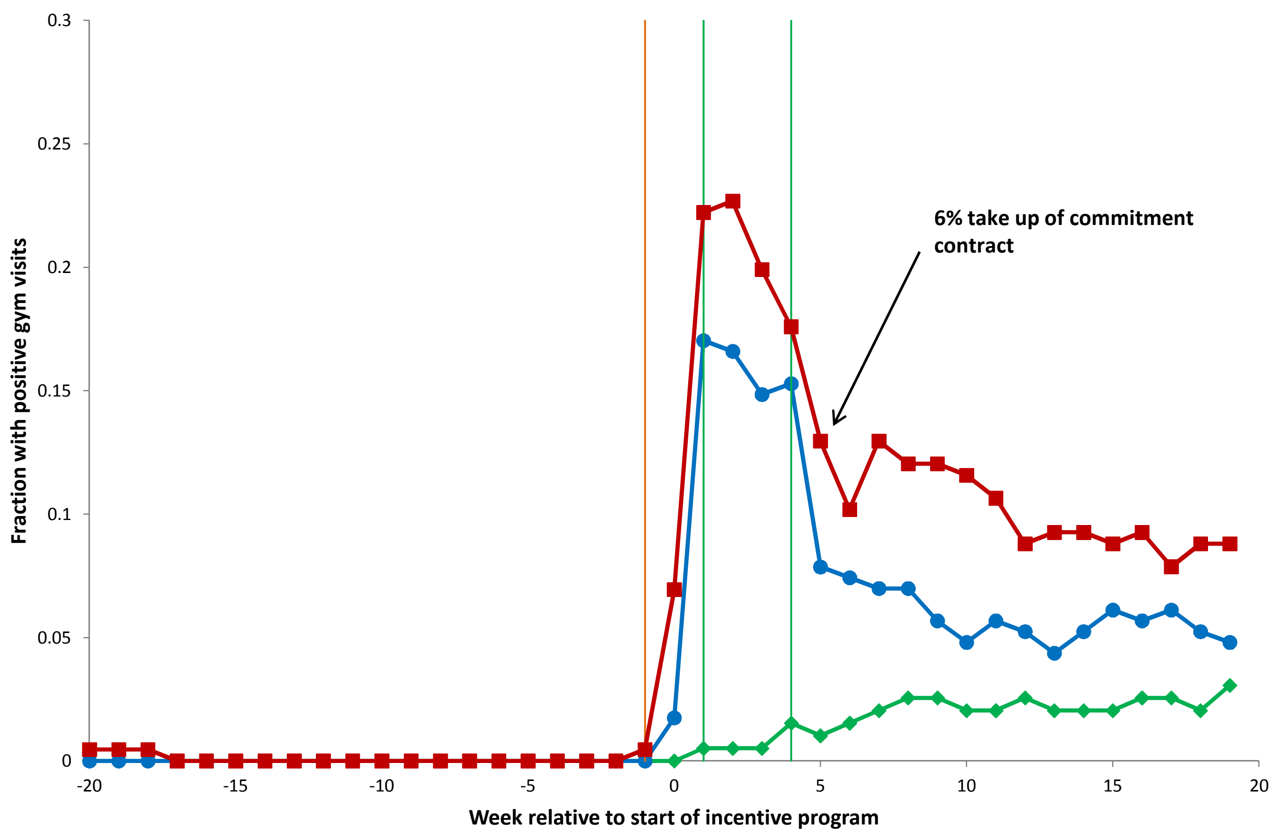


Figure 3a. Long run treatment effects (members only)

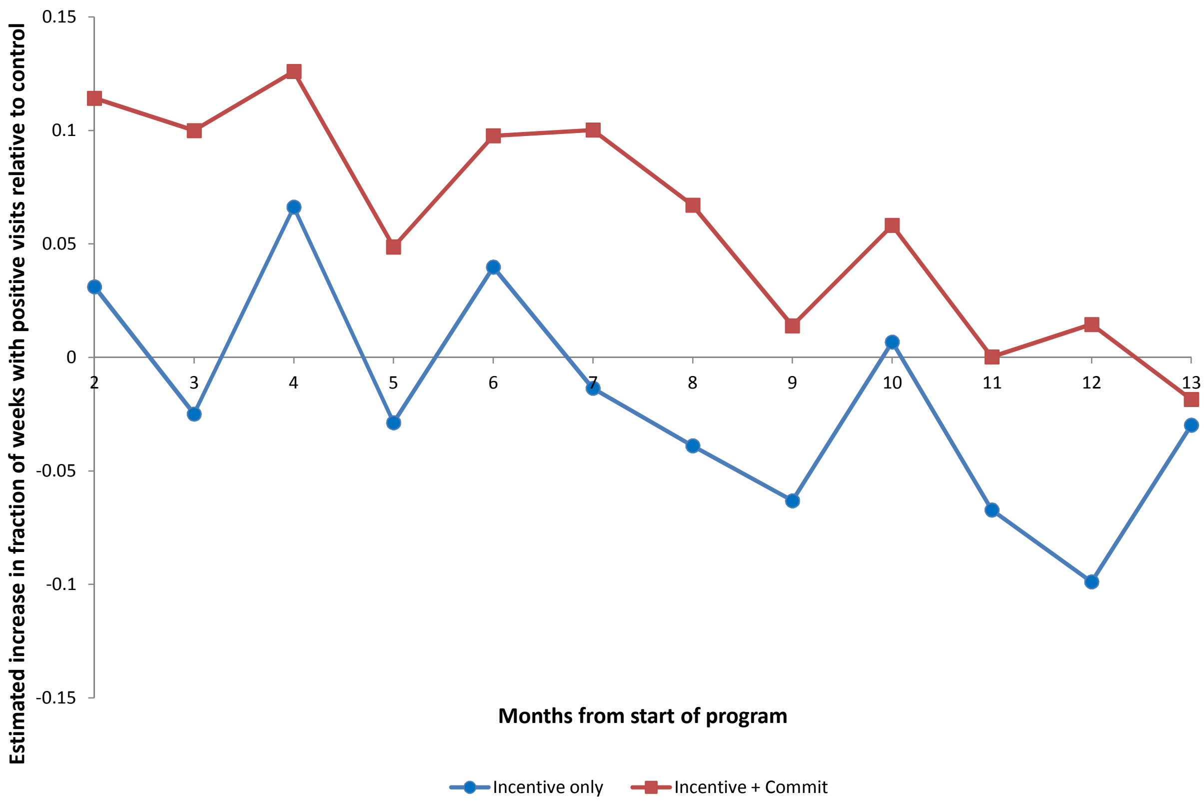


Figure 3b. Long run treatment effects (non-members)

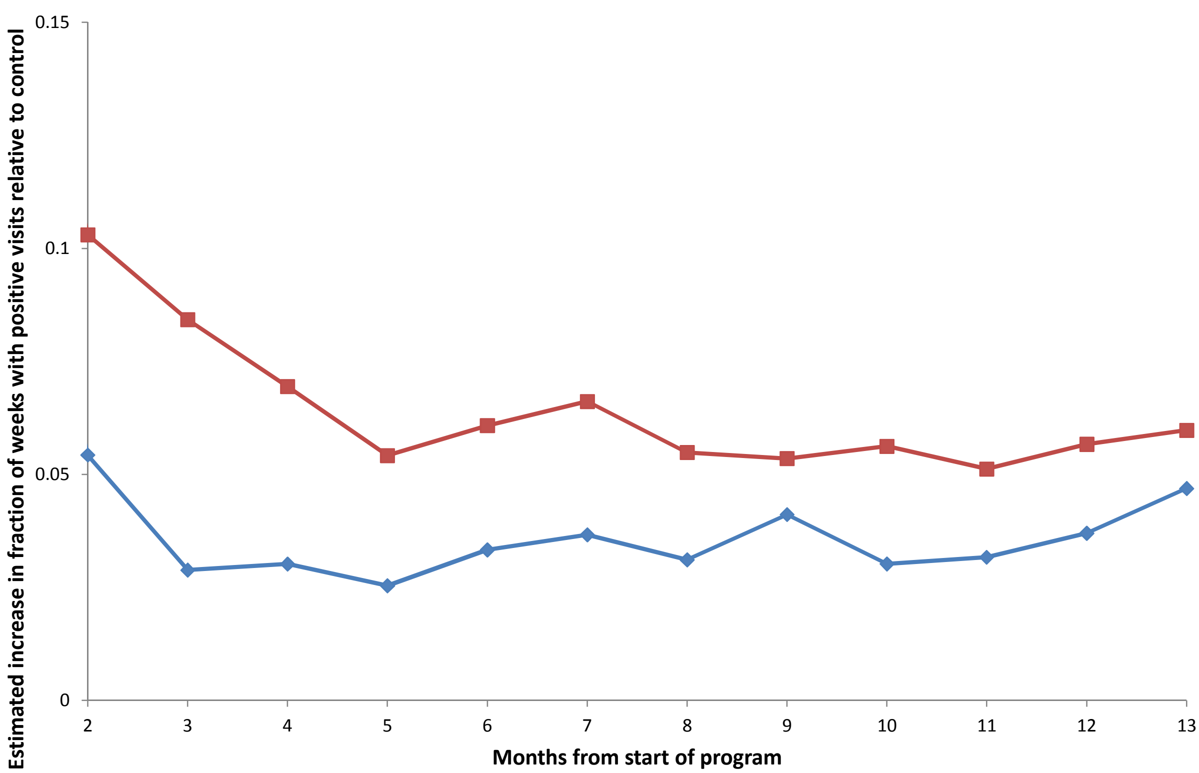


Table 1. Pre-Treatment Descriptive Statistics

	Members					Non-Members				
	(1) Control Mean	(2) Incentive-Only Diff	(3) Incentive+ Commit Diff	p-value (2)=(3)=0	p-value (2)=(3)	(4) Control Mean	(5) Incentive-Only Diff	(6) Incentive+ Commit Diff	p-value (2)=(3)=0	p-value (2)=(3)
Basic Demographics										
Age	40.12	-1.17	-0.01	0.61	0.37	39.62	-0.75	0.16	0.65	0.36
	(10.63)					(10.82)				
Male	0.46	0.04	0.08	0.53	0.63	0.52	0.01	-0.01	0.90	0.71
College Degree or More	0.61	0.09	0.05	0.37	0.52	0.64	0.03	0.10	0.06	0.09
Living Situation										
Married	0.68	0.02	0.04	0.78	0.71	0.67	-0.03	0.01	0.65	0.31
Has at Least 1 Kid at Home	0.45	0.08	0.07	0.46	0.89	0.48	-0.05	0.04	0.19	0.06
One-Way Commute (Minutes)	37.82	-2.35	-0.21	0.48	0.28	38.03	0.76	-0.85	0.69	0.41
	(21.23)					(20.54)				
Subjective Wellbeing										
Unhappy with Life	0.07	0.00	0.02	0.89	0.70	0.11	0.01	0.01	0.91	0.86
Unhappy with Fitness	0.34	0.05	0.01	0.64	0.53	0.54	-0.10	-0.08	0.08	0.72
Unhappy with Weight	0.58	-0.01	-0.08	0.33	0.23	0.55	-0.04	-0.05	0.61	0.90
Health and Fitness										
Pounds over Target Weight	20.28	-1.63	1.42	0.69	0.37	22.72	-0.91	2.00	0.58	0.29
	(18.68)					(28.59)				
BMI	28.31	-0.60	-0.29	0.68	0.67	28.22	-0.49	0.40	0.36	0.13
	(5.52)					(6.51)				
Overweight	0.43	-0.04	0.03	0.50	0.29	0.42	-0.06	-0.04	0.44	0.62
Obese	0.30	0.00	-0.06	0.47	0.30	0.30	-0.02	0.03	0.43	0.18
Takes Blood Pressure Meds	0.12	0.03	0.00	0.78	0.57	0.13	0.00	-0.02	0.75	0.49
Exercise										
Average Days of Overall Exercise	3.36	-0.10	0.12	0.34	0.16	1.98	-0.13	-0.09	0.54	0.73
	(1.65)					(1.73)				
Target Days of Exercise	4.79	0.03	0.19	0.26	0.17	4.05	-0.18	0.00	0.30	0.20
	(1.08)					(1.33)				
0 Days of Overall Exercise	0.05	0.03	-0.03	0.16	0.07	0.24	0.01	0.02	0.86	0.76
Number of Observations	94	134	131			195	228	215		

Notes: Columns (1) and (4) are the control group means. Columns (2), (3), (5), and (6) are the mean differences between that group and the control group; these are estimated via regressions that include strata fixed effects. Standard deviations for continuous variables are presented with parentheses.

Table 2. OLS Regression results

Dependent variables: Any visit = 0/1 indicator whether individual attended gym in a given week; Weekly visits=number of visits an individual had in a given week	Members		Non-Members	
	Any Visit	**Weekly Visits**	**Any Visit**	**Weekly Visits**
Control mean of dep var in pre-period	0.62	1.80	0	0
Incentive only	-0.02	-0.19	-	-
	(0.05)	(0.16)		
Incentive + Commit	0.01	-0.03	-	-
	(0.05)	(0.16)		
In-treatment period (weeks 1-4)	0.03	0.14	-	-
	(0.04)	(0.13)		
(Incentive only) x (In-treatment)	0.23***	0.86***	0.15***	0.38***
	(0.04)	(0.13)	(0.02)	(0.07)
(Incentive + Commit) x (In-treatment)	0.21***	0.94***	0.20***	0.53***
	(0.04)	(0.13)	(0.03)	(0.08)
Posttreatment (weeks 5-13)	0.03	0.20	0.03**	0.06*
	(0.05)	(0.14)	(0.01)	(0.03)
(Incentive only) x (Post-treatment)	0.03	0.15	0.04**	0.09*
	(0.03)	(0.11)	(0.02)	(0.05)
(Incentive + Commit) x (Post-treatment)	0.10***	0.30***	0.09***	0.20***
	(0.03)	(0.10)	(0.02)	(0.06)
Strata and week fixed effects	X	X	X	X
Subject-week observations	6,821	6,821	8,333	8,333
Number of subjects	359	359	641	641
P-values test of equal effects -- incentive-only vs. incentive + commit:				
Pre-treatment	0.46	0.25	NA	NA
In-treatment (weeks 1-4)	0.72	0.53	0.15	0.13
Post-treatment (weeks 5-13)	0.03	0.16	0.02	0.08
Commitment contract takeup rate:	0.23	0.23	0.06	0.06

Robust standard errors clustered by individual in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3. Estimated post-treatment differences between incentive groups

Dependent variable: Indicator for whether the subject attended the company gym that week during post-treatment
Sample is restricted to subjects offered incentives during treatment

	Members		**Non-Members**	
	All	> 0 visit during Incentive Period	All	> 0 visit during Incentive Period
Mean for incentive-only in weeks 5-13	0.55	0.61	0.06	0.31
Incentive + Commit	0.08***	0.10***	0.04**	0.15*
	(0.03)	(0.04)	(0.02)	(0.08)
Subject-week observations	2,385	2,133	4,005	864
Number of subjects	265	237	445	96
Commitment contract takeup rate	0.23	0.24	0.06	0.21
Implied IV estimate	0.43	0.50	0.83	0.71

Robust standard errors clustered by individual in parentheses. All regression estimates control for strata and calendar week fixed effects. These regressions control for the whether or not the subject attended the gym in each week of the incentive period. The implied IV estimate is simply a scaling that divides the treatment effect by the commitment contract takeup rate.
*** p<0.01, ** p<0.05, * p<0.1

Table 4a. Substitution Analysis for Members - In-Treatment Effects

Panel A: All subjects

Dependent Variable:	Weekly visits Gym data	Weekly visits Survey data	Overall exercise Survey data
Incentive-only or inc+commit	0.79***	0.49***	0.40**
	(0.15)	(0.17)	(0.18)
Observations	359	335	337
Mean control	1.59	2.26	3.25

Panel B: Subjects reporting exercise below their target in pre-survey

Dependent Variable:	Weekly visits Gym data	Weekly visits Survey data	Overall exercise Survey data
Incentive-only or inc+commit	0.88***	0.76***	0.76***
	(0.18)	(0.21)	(0.22)
Observations	209	190	192
Mean control	0.97	1.46	2.32

Panel C: Subjects reporting exercise at/above their target in pre-survey

Dependent Variable:	Weekly visits Gym data	Weekly visits Survey data	Overall exercise Survey data
Incentive-only or inc+commit	0.60**	0.03	-0.16
	(0.26)	(0.29)	(0.29)
Observations	150	145	145
Mean control	2.58	3.51	4.71

Table 4b. Substitution Analysis for Non-Members - In-Treatment Effects

Panel A: All subjects

Dependent Variable:	Weekly visits Gym data	Weekly visits Survey data	Overall exercise Survey data
Incentive-only or inc+commit	0.45***	0.60***	0.38***
	(0.05)	(0.07)	(0.13)
Observations	641	571	572
Mean control	0.03	0.07	2.09

Panel B: Subjects reporting exercise below their target in pre-survey

Dependent Variable:	Weekly visits Gym data	Weekly visits Survey data	Overall exercise Survey data
Incentive-only or inc+commit	0.47***	0.65***	0.45***
	(0.06)	(0.08)	(0.15)
Observations	499	446	447
Mean control	0.003	0.03	1.58

Panel C: Subjects reporting exercise at/above their target in pre-survey

Dependent Variable:	Weekly visits Gym data	Weekly visits Survey data	Overall exercise Survey data
Incentive-only or inc+commit	0.37***	0.43**	0.11
	(0.14)	(0.19)	(0.31)
Observations	142	125	125
Mean control	0.11	0.20	3.96

Notes: Dependent variable is average of weekly visits or exercise over the treatment period. Regressions include study cohort fixed effects. Standard errors are clustered by individual.

Table 5. Heterogeneity Cuts on Treatment Effects

Dependent variable: Any visit = 0/1 indicator whether individual attended gym in a given week

		Members			Non-Members	
Heterogeneity Cut		Incentive only	Incentive + Commit		Incentive only	Incentive + Commit
Low preperiod exercise	**In-treatment period**	0.32***	0.28***	**In-treatment period**	0.14***	0.18***
	Control Mean = 0.07	(0.09)	(0.09)	Control Mean = 0.004	(0.04)	(0.04)
	Post-treatment period	0.07	0.16**	**Post-treatment period**	-0.005	0.07*
	Control Mean = 0.07	(0.05)	(0.06)	Control Mean = 0.03	(0.03)	(0.04)
Mid preperiod exercise	**In-treatment period**	0.20***	0.22***	**In-treatment period**	0.15***	0.23***
	Control Mean = 0.67	(0.05)	(0.05)	Control Mean = 0.00	(0.04)	(0.05)
	Post-treatment period	0.05	0.18***	**Post-treatment period**	0.04*	0.11***
	Control Mean = 0.58	(0.07)	(0.06)	Control Mean = 0.00	(0.02)	(0.04)
High preperiod exercise	**In-treatment period**	0.09*	0.10**	**In-treatment period**	0.14***	0.18***
	Control Mean = 0.88	(0.05)	(0.05)	Control Mean = 0.02	(0.04)	(0.05)
	Post-treatment period	-0.02	-0.003	**Post-treatment period**	0.06	0.08**
	Control Mean = 0.87	(0.05)	(0.04)	Control Mean = 0.03	(0.04)	(0.04)
Men	**In-treatment period**	0.26***	0.28***	**In-treatment period**	0.16***	0.17***
	Control Mean = 0.51	(0.07)	(0.07)	Control Mean = 0.01	(0.03)	(0.04)
	Post-treatment period	0.12*	0.15**	**Post-treatment period**	0.04*	0.10***
	Control Mean = 0.50	(0.07)	(0.07)	Control Mean = 0.01	(0.02)	(0.03)
Women	**In-treatment period**	0.14**	0.16**	**In-treatment period**	0.13***	0.21***
	Control Mean = 0.61	(0.07)	(0.07)	Control Mean = 0.00	(0.03)	(0.04)
	Post-treatment period	-0.11	0.11	**Post-treatment period**	0.03	0.07**
	Control Mean = 0.56	(0.06)	(0.07)	Control Mean = 0.03	(0.03)	(0.03)

Notes: Cells in the table give coefficients from linear regression models of treatment effects for indicated treatment with clustered standard errors on individual in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In-treatment period is weeks 1-4 with incentive in place. The Post-treatment period for these regressions is weeks 5 - 13, while the commitment contract was in place. Heterogeneity cuts on preperiod exercise are based on tercile splits. For the members this is based on fraction of weeks with positive visits to the company gym in the 6 weeks prior to our intervention. For non-members it is based on self-reports of average days of exercise (at all locations) from the baseline survey. The ranges for members are low = 0 - 0.33, mid = 0.5-0.83, high = 1 and for non-members are low = 0 - 0.5, mid = 1-2.5, high = 3 - 7.

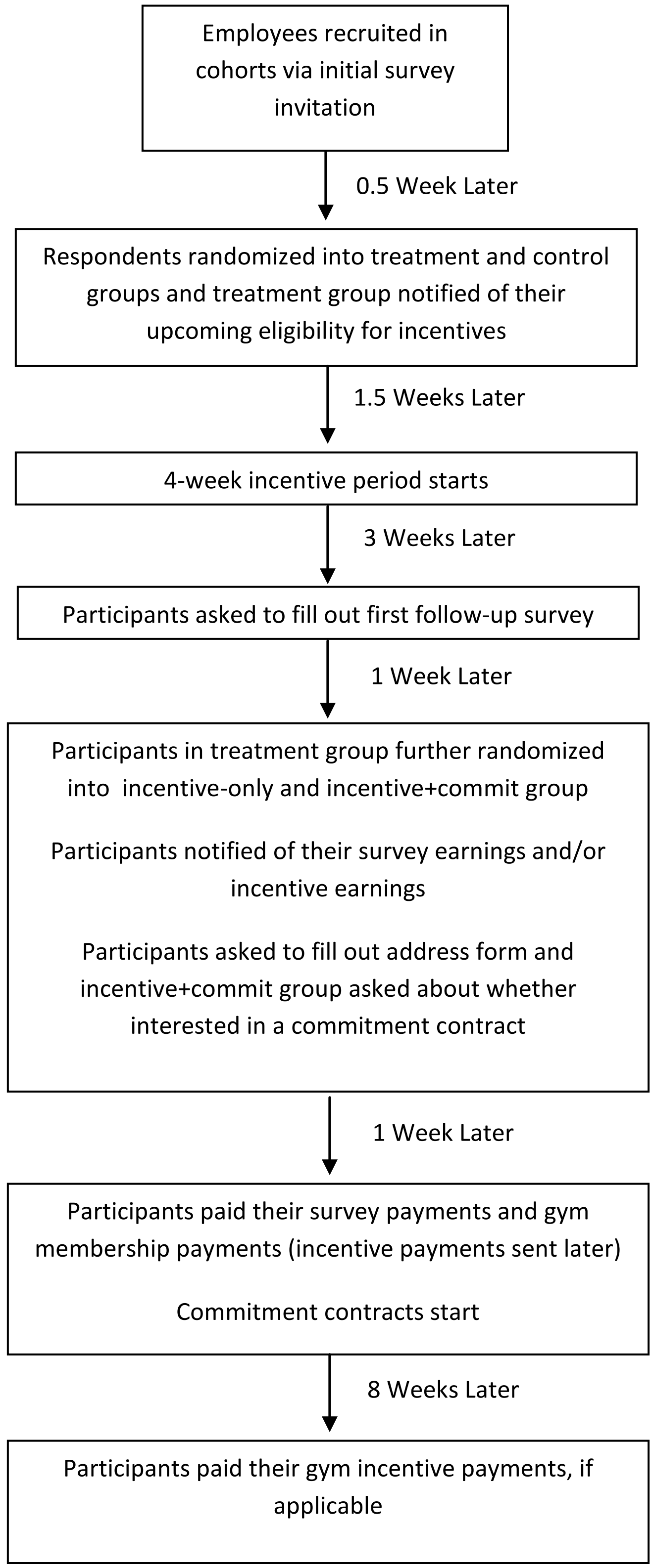
Table 6: OLS Regressions Predicting Uptake of Commitment Contracts

Dependent variable: indicator for whether subject made a commitment contract
Sample restricted to those offered incentives who desired exercise in post period (see note).

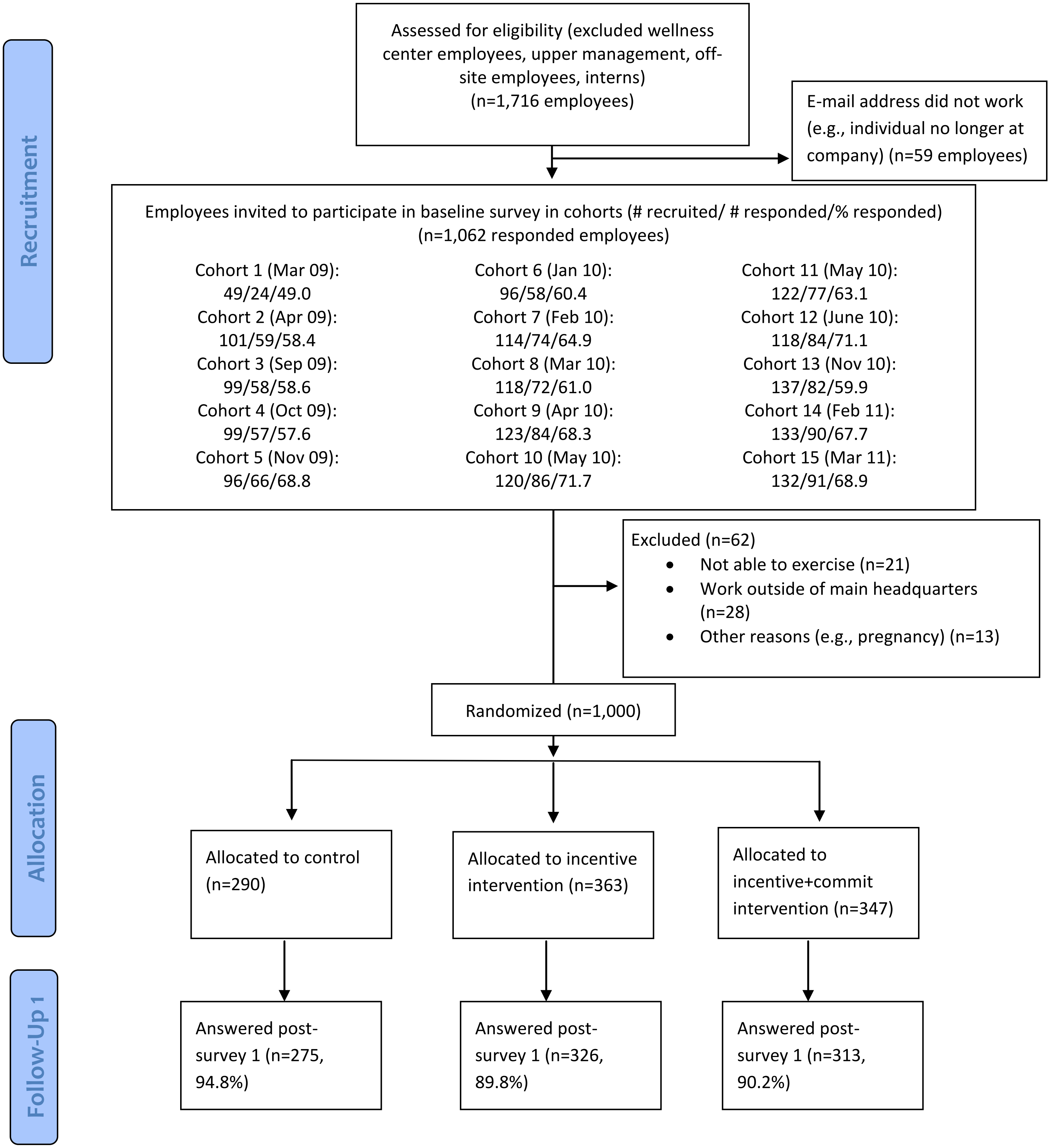
Mean commitment takeup rate: 0.19	(1)	(2)	(3)	(4)	(5)	(6)
Member of gym prior to study	0.06	0.05	0.06	0.07	0.05	0.04
	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)
Avg visits in-treatment: (0 < avg < 2)	0.16**	0.16**	0.15**	0.20**	0.13	0.13
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
Avg visits in-treatment: (2 ≤ avg < 3)	0.31***	0.31***	0.31***	0.31***	0.28***	0.26**
	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)
Avg visits in-treatment: (3 ≤ avg)	0.11	0.11	0.12	0.11	0.10	0.08
	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)
Male	-0.17***	-0.17***	-0.17***	-0.18***	-0.17***	-0.18***
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Young (age ≤ 32)	-0.15***	-0.16***	-0.16***	-0.18***	-0.20***	-0.21***
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Has children	-0.02	-0.02	-0.02	-0.02	-0.03	-0.03
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
College degree	0.02	0.02	0.02	0.03	0.01	0.02
	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)
Below target exercise prior to study	-0.03					-0.02
	(0.06)					(0.08)
Self-reports low self-control for exercise		-0.05				-0.03
		(0.05)				(0.07)
Self-reports low intrinsic motivation for exercise			-0.03			-0.02
			(0.05)			(0.06)
"Aware of self-control problem": expects to go less than ideal in post period				-0.02		-0.02
				(0.07)		(0.07)
"Extreme optimist": expects to go more in post period than during incentive period					0.12*	0.12*
					(0.07)	(0.07)
Number of subjects	224	223	220	188	186	181

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Sample restricted to those in treatment group who were offered commitment option and further restricted to those who expressed a desire to exercise at the company gym in the post treatment period (231 of 347). The takeup rate among those with no desire was zero. See text for descriptions of variables. Sample size is lower in columns 1, 2 and 6 due to missing values of exercise expectations among those who did not respond fully to second survey. The variable "young" is an indicator for being in the bottom quartile of age.

Appendix Figure 1. Timeline



Appendix Figure 2. Flow Diagram



Appendix Table 1. Post-survey response rates as a function of treatment

Dependent variable: Indicator for whether subject responded to the post-survey

	Members			Non-Members		
	All	Below Target	Above Target	All	Below Target	Above Target
Mean for control group	0.93	0.91	0.94	0.96	0.96	0.95
Incentive-only or inc + commit	0.02	0.01	0.03	-0.09***	-0.09***	-0.09*
	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.05)
Number of subjects	359	209	150	641	499	142

Robust standard errors in parentheses. All regression estimates control for strata. *** p<0.01, ** p<0.05, * p<0.1

Appendix A: The Substitution Effect after Accounting for Non-Response Bias

First, it is imperative to define the substitution effect. We define the substitution effect as the ratio of the effect of the incentives on overall days of weekly exercise to the effect of the incentives on days of weekly company gym exercise. A ratio of 1 indicates that there is no substitution whereas a ratio of 0 reflects complete substitution. Mathematically, we define the true substitution effect as the following:

$$\text{True substitution} = \frac{\overline{Y}_{treated}^{overall} - \overline{Y}_{control}^{overall}}{\overline{Y}_{treated}^{gym} - \overline{Y}_{control}^{gym}}$$

where $\overline{Y}_{treated}^{overall}$ is the average overall weekly days of exercise for the treated group (i.e., the incentive-only and incentive+commit group combined) during the incentive period, $\overline{Y}_{control}^{overall}$ is the average overall weekly days of exercise for the control group, $\overline{Y}_{treated}^{gym}$ is the average weekly days of company gym exercise for the treated group during the incentive period, and $\overline{Y}_{control}^{gym}$ is the average weekly days of company gym exercise for the control group. In contrast, the measured substitution effect from the post-survey data:

$$\text{Measured substitution} = \frac{\overline{Y}_{treated,r}^{overall} - \overline{Y}_{control,r}^{overall}}{\overline{Y}_{treated,r}^{gym} - \overline{Y}_{control,r}^{gym}}$$

where the means pertain to the group of survey responders (i.e., $\overline{Y}_{treated,r}^{overall}$ is the average weekly days of overall exercise for the treated group survey responders).

The measure of true substitution above can be expanded as follows:

$$\text{True Substitution} = \frac{p_{treated,r}\overline{Y}_{treated,r}^{overall} + (1-p_{treated,r})\overline{Y}_{treated,nr}^{overall} - (p_{control,r}\overline{Y}_{control,r}^{overall} + (1-p_{control,r})\overline{Y}_{control,nr}^{overall})}{p_{treated,r}\overline{Y}_{treated,r}^{gym} + (1-p_{treated,r})\overline{Y}_{treated,nr}^{gym} - (p_{control,r}\overline{Y}_{control,r}^{gym} + (1-p_{control,r})\overline{Y}_{control,nr}^{gym})}$$

where $p_{treated,r}$ is the probability of a treated individual responding to the survey and $p_{control,r}$ is the probability of a control individual responding to the survey. The nr subscripts on the averages denote non-responders (e.g., $\overline{Y}^{overall}_{control,nr}$ is the average weekly days of overall exercise for the control group non-responders). We can simplify this expression as $\overline{Y}^{gym}_{treated,nr} = 0$ and $\overline{Y}^{gym}_{control,nr} = 0$ because in our data, we observe that all non-responders did not attend the gym during the incentive period. Therefore, the true substitution effect simplifies to

$$\text{True Substitution} = \frac{p_{treated,r}\overline{Y}^{overall}_{treated,r} + (1-p_{treated,r})\overline{Y}^{overall}_{treated,nr} - (p_{control,r}\overline{Y}^{overall}_{control,r} + (1-p_{control,r})\overline{Y}^{overall}_{control,nr})}{p_{treated,r}\overline{Y}^{gym}_{treated,r} - p_{control,r}\overline{Y}^{gym}_{control,r}}$$

$$= \frac{\overline{Y}^{overall}_{treated,nr} - \overline{Y}^{overall}_{control,nr} + p_{treated,r}(\overline{Y}^{overall}_{treated,r} - \overline{Y}^{overall}_{treated,nr}) - p_{control,r}(\overline{Y}^{overall}_{control,r} - \overline{Y}^{overall}_{control,nr})}{p_{treated,r}\overline{Y}^{gym}_{treated,r} - p_{control,r}\overline{Y}^{gym}_{control,r}}$$

We have data on all objects in this formula except for $\overline{Y}^{overall}_{treated,nr}$ and $\overline{Y}^{overall}_{control,nr}$. One might consider a reasonable approximation for these objects to be their pre-incentive levels from the initial survey. Using these values, we can show that our estimates of true substitution and measured substitution are close to one another; they differ by at most 10%.