Preliminary and Incomplete: Please do not circulate

Using Incentives and Commitments to Overcome Self-Control Problems with Exercise: Evidence from a Workplace Field Experiment

Heather Royer, UC Santa Barbara & NBER
royer@econ.ucsb.edu

Mark Stehr, Drexel University
ms424@drexel.edu

Justin Sydnor, University of Wisconsin
jsydnor@bus.wisc.edu

October 28, 2011

Abstract

Can financial-incentive programs help people improve behavior in health settings where self-control problems are important? We report the results of a large-scale field experiment with employees of a Fortune 500 company on the effect of financial incentives at encouraging exercise. The treatment group for the experiment received a one-month incentive program, during which time they could earn $10 per visit to the on-site company gym for up to 3 visits per week. A secondary treatment offered half of the incentivized group at the end of the incentive period the opportunity to create a self-funded commitment contract for continued exercise at the company gym. We find that the temporary incentive led to lasting increases in exercise even after the incentive was removed. The availability of a commitment contract substantially improved the long-run effects of the incentive program. We discuss the implications of these findings for health-incentive programs and our understanding of time inconsistency and commitment.

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 Rachel Smith, Bert Wagner, and Jon Evans. The comments and suggestions of seminar participants and attendees at the NBER Summer Institute Health Economics Workshop, the Behavioral Economics Annual Meeting, ASHE conference, and the ASSA conference have greatly improved this work.

Can financial-incentive programs help people improve behavior in health settings where self-control problems are important? This paper addresses this question in the context of exercise. There is strong empirical evidence that people exercise less often than they intend (DellaVigna and Malmendier, 2006) and these struggles are a common motivating example for models of time inconsistency (Laibson 1997, O'Donoghue and Rabin, 1999 and 2001). Low levels of exercise are also an important policy issue. The Centers for Disease Control estimates that as of 2009 approximately one third of adults engage in no regular exercise, and this lack of physical activity likely contributes to poor population health and rising health-care costs.[1] In the face of spiraling health costs, policy makers, employers and insurance companies are increasingly interested in using incentive programs to address a range of problematic health behaviors, including exercise. Approximately 40 percent of large employers in the US offer some form of incentive-based wellness program (Hewitt Associates) and recent health-care legislation in the U.S. contains language explicitly authorizing firms to spend up to 30% of the cost of employee insurance on wellness programs. While employers and insurance companies are primarily motivated to offer these programs in attempt to deal with the externality imposed on their bottom line by the poor health behaviors of individuals,[2] models of time inconsistency suggest that incentives may interact with self-control problems in important ways that could improve individual welfare in addition to social welfare.

This study reports the results of a large-scale field experiment examining the effect of financial incentives to encourage exercise among employees at the headquarters of a Fortune 500 company. The treatment group for the experiment received a one-month incentive program, during which time they could earn $10 per visit to the on-site company gym for up to 3 visits per week. A secondary treatment offered half of the incentivized group at the end of the treatment period the opportunity to incentivize themselves by creating a self-funded commitment contract for continued exercise at the company gym. In these contracts, employees put money on the line that was forfeited to charity if they

[1] Regular exercise contributes to a healthy lifestyle, with a wide range of benefits including reductions in heart disease, obesity, diabetes, and cancer (Pate et al. 1995).

[2] Baicker et al. (2010) provide a review of workplace wellness programs and estimate (cross-sectionally) that on average firms see returns of around $5 for every dollar invested in wellness programs.

failed to regularly use the gym. The primary outcomes of interest for the experiment are a) whether the temporary incentive program had lasting effects on exercise at the company gym and b) whether the degree of any such habit formation was stronger for the group offered commitment contracts in addition to their incentives.

Models of time inconsistency suggest a number of reasons that a temporary incentive program could have a lasting effect on behavior. First, there are large start-up costs to beginning an exercise routine, including deciding on the level and type of exercise, gaining familiarity with a gym or other exercise setting, and changing existing daily routines to make time for exercise. In the face of these start-up costs, time-inconsistent people may procrastinate on starting to exercise even though they think the long-run benefits of exercise exceed the costs. Second, the physiology of exercise is such that exercise typically becomes more pleasurable the more one exercises, and as such a person may need to go through an investment period before reaching a steady state where the daily costs of exercise are not too high. Individuals with present-bias may find it particularly difficult to motivate themselves to exercise through this investment period. The incentive program, with its month-long per-use incentives and fixed start date that precludes procrastination, could be expected to help overcome both of these issues. On the other hand, even after establishing an initial exercise routine, people with present bias are more likely on a daily basis to be tempted to avoid the immediate costs of exercising. One-time incentives are unlikely to help with that daily self-control struggle. The availability of a commitment contract, though, may be an effective way to deal with this problem for those who are sophisticated about and self-aware of their time inconsistency.

During the 4 weeks that the incentive program was in place, the incentives led to approximately a doubling of the use of the company gym by those in the treatment group relative to the control group. The subjects in this experiment are working adults in the Midwest, with an average age of 40 and self-reported rates of overweight and obesity (69% combined) matching the national average reported by the Centers for Disease Control. As such, while it is perhaps not surprising to economists that people exercise more while paid to do so, these results nonetheless provide a useful benchmark for understanding how this policy-relevant population responds to incentives for exercise.

The treatment groups continued to use the gym at higher rates than the control group even after the incentive program ended. Among those offered incentives only there was a modest,

statistically-significant increase of 3.5 percentage points in the fraction of employees regularly exercising at the company gym in the weeks after the incentive program ended (a 16% increase relative to the baseline of 22% regularly exercising in the control group).[3] The effect for this group largely dissipates by two months after the program. The post-incentive effects were much stronger for the treatment group offered the ability to create a self-funded commitment contract at the end of the incentive program. On average this group saw an increase of 10 percentage points (45%) relative to control in the fraction regularly exercising over the first two months after the incentive program ended. This long-run effect also dissipates more slowly than for the group only offered incentives.

These results relate most closely to seminal work by Charness and Gneezy (2009) and a recent working paper by Acland and Levy (2010).[4] Charness and Gneezy showed in an experiment with undergraduate students that a one-month incentive program for exercise could lead to habit formation. In the process they revealed that in this context the power of incentives to generate lasting habits by helping people overcome the costs of establishing an exercise routine outweighed longstanding concerns raised by psychologists (Deci 1971) that incentives would have perverse effects by eroding intrinsic motivation. Our findings are consistent with Charness and Gneezy's results, showing little evidence of an erosion of intrinsic motivation. Our results also reveal that a temporary incentive for exercise can have lasting effects even in a population of more sedentary working adults, though the long-run effects for these working adults are more modest than those for undergraduate students. Acland and Levy (2010) present another closely-related study, in which they use a similar experimental design to that of Charness and Gneezy with an undergraduate population and focus on eliciting subjects' expectations of gym use to better understand the nature of time inconsistency. They largely confirm the basic finding that incentives generate habit formation, and they find that subjects systematically overestimate how often they will exercise, suggesting some level of naivety concerning their time inconsistency. Since commitment contracts are only useful to those aware of and trying to overcome self-control problems, our results that the availability of commitment contracts improves

[3] Regular exercise is defined here as using the company gym at least once in a week. The results section shows that the main findings are similar using other outcome measures.

[4] Other related exercise studies 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. On the topic of wellness incentives, Finkelstein et al. (2007) test the effect of cash incentives for weight loss among a group of obese employees of North Carolina universities.

the lasting effects of the temporary incentive suggest that our population has enough sophistication about their time inconsistency in regards to exercise to benefit from commitment.

The setting, design, and large sample size for our experiment provide us with rich information that we can use to understand both the extent to which the incentive program caused people to substitute away from other forms of exercise and the heterogeneity of the response to incentives in ways that the previous studies could not. Our sample includes both employees who were members of the on-site company gym prior to our intervention and employees who had not signed up to use the company gym. This presents an important dimension of heterogeneity and for the members we can observe their exercise behavior at the gym prior to the study period to further delve into the heterogeneity. We supplement this information with surveys in which we collect information from the employees about the amount of exercise they do away from the company gym. We conducted a baseline survey prior to the incentive program and a follow up survey immediately at the end of the incentive period that allow us to test for substitution by looking for changes in these measures.

Understanding the nature of substitution should be an important goal in any study of incentive programs, and is particularly relevant for studies of work-site wellness incentives. To facilitate this analysis we asked subjects in the baseline survey to characterize their current level of exercise relative to how much exercise they wanted themselves to do.[5] Our hypothesis was that those who desired to exercise more had potential scope to exercise at the company gym without substituting away from other activities, whereas any response by those already exercising at their desired level would likely represent substitution. That is the pattern we find. For the 70% of employees who stated they desired to exercise more, the survey data reveals that they did not substitute substantially away from other activities, in part because they had relatively low baseline levels of exercise. This was true both for employees who were and were not existing members of the company gym prior to the study. For the remaining 30% of employees, those who were already exercising their desired amount, we cannot reject that all of their increase in the use of the company gym during the incentive period represented substitution away from other sources of exercise. Because those below their target initially also had

[5] To allow for a comparison of substitution effects across natural dimensions of heterogeneity, the randomization was stratified on the combination of whether the employee was an existing member of the company gym and whether they reported wanting to exercise more than they currently did.

the largest increases in use of the company gym during the incentive, the vast majority (78%) of the increase in exercise generated by the incentive program represents genuinely new exercise.

Turning to the heterogeneity of the response to the availability of incentives and commitment contracts, we find interesting patterns that shed light on how these programs interact with time inconsistency. The long-run (i.e., post-incentive) effects for the group offered only incentives (without the commitment contract option) was concentrated among those who might have faced large start-up costs to establishing an exercise routine at the company gym. That pattern is consistent with the mechanisms by which the temporary incentive might be expected to help people overcome problems of time inconsistency discussed above. The improved long-run effects of the group offered the commitment-contract option are especially pronounced for those who were already occasionally, but not consistently, using the company gym. That result is consistent with the idea that there is a population of people for whom time inconsistency causes an ongoing struggle to maintain an exercise routine and suggests that on average there is enough sophistication among this group about their self-control problem for commitments to help them improve their behavior.

The results in this study add to a small but growing literature on the use of commitment technologies as a mechanism for overcoming problems with time-inconsistency. One of the key insights to come out of work on time-inconsistent preferences is the idea that people might benefit from the ability to commit their future selves to behavior. Previous literature has shown that people will voluntarily place binding constraints on their future behavior. There is evidence that various forms of such commitments can improve test scores for students (Ariely and Wertenbroch, 2002), accumulation of savings (Ashraf, Karlan and Yin, 2006), and smoking cessation rates (Gine, Karlan, and Zinman, 2010).[6] Overall the take-up rates of commitment contracts in our study was 12% but conditional on using the gym at least once during the incentive period, this statistic was 23%. These take up rates are comparable to those found by Ashraf et al. (2006) and Gine et al. (2010). The take-up rate is particularly high among employees previously exercising occasionally, but not consistently, at the gym and we see a strong intention-to-treat effect on long-run exercise of offering commitment

[6] There is also an earlier literature in clinical settings that used deposit contracts in conjunction with other approaches to address smoking cessation and weight loss (Paxton 1979, 1980, 1981, Jeffrey et al., 1990). More recently, Volpp et al. (2008) found that matched deposit contracts were effective in a clinical setting for improving weight loss among the seriously obese, though the matching implies that these are not "pure" commitments and might even have appeal to agents with time-consistent preferences.

contracts among that group. Those effects are consistent with the anticipated benefits commitment contracts could have for those at the margins of creating a successful exercise routine. Interestingly, however, we also find that there is also similar demand for commitment contracts among employees who were consistently exercising at the gym prior to the intervention. The studies by Ashraf et al. and Gine et al. both concentrated on the demand for commitment among those whose behavior was open to change (i.e., non-savers and smokers). The results here are the first to suggest that commitment may be desirable even to those who on the surface are not struggling to overcome time inconsistency. As we discuss below, this result may point to the value of further research exploring the usefulness of overcoming self-control in a broader context than just behavioral-economic models of time inconsistency, for example the ego-depletion literature in psychology (Baumeister et al., 1998, 2000).

Although we find that it is in general difficult to predict take-up of commitments, the one strong predictor of take-up is gender. Female employees made commitments at much higher rates than male employees, even though both groups showed similar responses to the financial incentives during the incentive period. We find that the differences in habit formation between those who received only incentives and those also offered commitments are concentrated on the female employees. Men display significant habit formation whether offered commitment contracts or not. Women on the other hand, have a very low post-treatment response to incentives alone, but when offered commitment contracts on average show essentially the same level of long-run response as the men. Ashraf et al. (2006) also found a relationship between gender and commitment, as the importance of a measure of time inconsistency for predicting take-up of their savings commitment was only significant for women, though they saw similar overall take-up rates across gender. Our study combined with those earlier findings suggests that there may be important gender differences in how commitments can be used to overcome time inconsistency problems, which could be an important direction for future research.

The remainder of the paper proceeds as follows: In Section 2 we describe the experimental design. Section 3 provides summary statistics from our baseline 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 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 might motivate future research. In particular, we touch on what our study has to say about the potential effectiveness of financial incentive programs as part of a company's strategy for reducing health-care costs. Although our study is not designed to serve as program evaluation for a comprehensive wellness program, it can provide some useful insights and benchmarks for the design of those programs.

2. Experimental Design

The experiment was conducted at the headquarters of a Fortune 500 company located in the Midwest. There are approximately 1,300 employees located at the headquarters complex, and employees at this site have a range of job functions, such as finance, administrative support, customer service, legal, sales, etc. There is an on-site wellness center at the headquarters that has the usual amenities of a modern gym. An employee must become a member of the gym in order to use the facilities. Gym members pay a subsidized membership of $12.96 every 2 weeks that is automatically withdrawn from the employee's paycheck.[7,8] Employees log in at a computer terminal when they arrive to use the gym, and the computerized records of these log-ins serve as the primary data for this research. The experiment was described as a university study that was supported by the corporation and the subjects were ensured that none of their individual responses to any surveys or other information would be shared with anyone at the corporation. All respondents agreed to allow us to access their complete records of using the gym both before the study period and for a period of one year from the completion of the survey. Since employees were aware they were participating in a study, this experiment is a "framed field experiment" in the lexicon of List (2009).

The experiment was conducted in 15 cohorts, with the first cohort beginning in February of 2009 and the last cohort starting the study in March of 2011. The cohort design ensures that the gym staff would not be overwhelmed by new-member signups at any point during the experiment and also ensures that the results here are not specific to a specific time of year. We detail the number of participants along each step of the experiment in the Appendix Figure 1. For each cohort a random sample was drawn from the company's full list of employees, excluding high-level executives and members of the human resources team privy to the details of the research. That random sample of

[7] There are no start-up fees or contracts and employees can cancel their membership at any time with no penalty.
[8] The gym is open Monday through Friday from 6:00 a.m. to 7:00 p.m.

employees was then sent an email inviting them to participate in two online wellness surveys (baseline and follow-up) spaced 5 weeks apart (see Appendix for a copy of these surveys). The employees were told they would be compensated with a $25 payment if they completed both surveys. The baseline survey collected a range of information on demographics, self-assessed fitness levels, exercise patterns, and subjective wellbeing. Response rates for this survey averaged 62 percent (see Appendix Figure 1).

The respondents to the survey form the subject pool for the experiment and were then randomly sorted into treatment and control groups. As we thought that there was likely to be heterogeneous responses to the incentives, we stratified the randomization into 4 groups: a cross between a) existing members of the company gym versus non-members of the gym with b) those stating that they were currently exercising less than their personal target versus those currently exercising at least as much as they desired. Since the subject pool consists exclusively of individuals responding to the initial survey, it is important to note that the estimated effects may not be generalizable to that for the broader pool of employees.

Respondents randomly assigned to the incentive treatment were informed of the incentive program via both an email and a letter sent to them via the internal company mail system. This duplication was used to help ensure that all subjects offered incentives were aware of the program. Based on evidence from follow-up surveys, lack of information about the incentive program was not an impediment to participation. These employees could earn a payment of $10 per visit they made to the company gym over a 4 week period for up to 3-visits per week. They could therefore earn up to $120 in per-use incentives during the program. In addition to the per-use incentive, all subjects in the treatment group received a membership reimbursement ($12.96*2=$25.92) for the 4-week incentive program. Finally, because new members who join the gym have to find time (~1 hour) to complete a new-member assessment, employees who were not already members of the gym were additionally offered a $20 enrollment bonus to help compensate for this extra time. Since all treatment groups include both per-use incentives and the membership reimbursements/bonus, while the control group received neither, the incentive program should be thought of as a package of incentives.[9]

[9] In pilot experiments conducted at the company prior to the first main cohort we found essentially zero response to a treatment offering only the $25 membership reimbursement.

Visits for the incentive program were tracked based on log-ins at the company gym, with only one visit in a given day counting toward the program. 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 employees to log out when finished exercising at the gym. As such, it is not possible to know how long the employee exercised during a given visit or the nature of that exercise. In theory, then, there is some scope for employees to cheat on the program by logging in and not exercising. Research assistants discretely monitored the gym during the treatment period for early cohorts and did not observe any behavior consistent with employees logging in but not using the gym. In addition, the gym staff (who were aware of the program but did not know who was offered incentives) reported no increases in employees visiting the gym without exercising and did not notice anything unusual in the patterns of recorded visits relative to who they were observing at the gym. 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 in behavior after the incentive program ends, at which point there is no incentive for this type of cheating.

During the final week of the incentive program, all subjects who responded to the baseline survey (including the control group) were asked to complete an endline survey. This follow up survey largely asked the same questions as the baseline survey (omitting demographics). The response rate to this endline survey was 91.4 percent (see Appendix Figure 1).

At the end of the 4-week incentive period, members of the treatment group were randomized into the second-level treatment, in which roughly half of the incentivized subjects were offered the chance to create a commitment contract. No mention of the commitment option was made before or during the incentive program. In order to ensure balance between the group offered the commitment option and the group not offered the option, 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.[10] The commitment contract for this study was a pledge not to go more than 14 calendar days without attending the company gym. To maintain simplicity of the program – including the ability to describe

[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-plus-commitment-option 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.

the program briefly via email – we decided to set s fixed commitment option for all subjects in the IC group. The level of commitment was set at a low level to ensure that it would not be too ambitious or binding for employees whose work-related travel or vacation would take them away from the office during a particular week. In addition, we hoped to set the commitment at a level that would be potentially attractive to those most on the margin of establishing an effective exercise routine. Naturally, having a fixed contract, in this case one with a modest goal, likely made the contract less desirable to some participants. While we think establishing optimal commitment contract designs is an important goal for future research, it is outside the scope of this study. The subjects could decide whether or not they wanted to make that commitment and if so set an amount of money they wanted to put on the line. They were told that their visits would be tracked over an 8 week period and that if they failed to keep to the commitment that the money they put at stake would be forfeited to the United Way. Subjects could not earn any additional money by making a commitment contract and as such it should only be attractive to those looking to influence their future behavior. Subjects who made commitments could commit the money they earned during the incentive program, or if they wanted to commit more (or did not earn any money during the incentive program) they could write a check made out to the charity that was held for them until the end of the commitment period and returned if they successfully completed the commitment. All payments for the gym-attendance incentive were scheduled to be mailed after this 8-week commitment period, so a subject wanting to create a commitment contract could decide to put their earnings from the incentive program on the line for the commitment without causing an additional delay in receiving their money. In order to ensure that we received an active response to the commitment offer, subjects made the decision of whether or not to create a commitment as part of a form where we collected their mailing address for payments of the survey and gym-use incentives.

Our interest is in assessing the effect of being offered the chance to create a binding financial commitment. However, it is possible that simply suggesting the specific commitment could have an effect on behavior beyond offering the option to put money on the line. In order to control for that potential confounding effect, subjects who received incentives but did not get the commitment contract offer were given a nearly identical email that urged them to commit themselves to not

missing more than 14 days at the gym over the following 8 weeks, but their email did not mention putting money at stake for that goal (See Appendix A).

All subjects in the cohort were invited via email to take a final long-run online survey at the end of the commitment period (10 weeks after the end of the incentive program). That survey re-measured a number of the main variables of interest from the baseline survey. To keep down costs the long-run survey was incentivized with a chance to win one of two $50 payments. Not surprisingly, this weaker incentive resulted in a lower response rate of 67%.

Section 3. Data

Table 1 provides the means for key variables from our initial baseline survey across company gym membership status.[11] Both panels present the overall mean and standard deviations (for non-binary variables in parentheses) along with the means and standard deviations across each of the treatment arms in columns (1)-(4). The last columns in each panel are the p-values from tests of (a) the equivalence of the means across the 3 randomized groups and (b) the equivalence of the means across the incentive and the incentive+commit groups.

Across both groups, our subject pool is on average 40 years old, roughly equally divided across genders, and is well-educated (more than 65 percent have a college degree or more). In comparison, overall in the United States in 2009, just under 30 percent of adults aged 25 and older had at least a college degree. Possible time constraints as measured by marital status, presence of children at home, and commute times are comparable to overall US patterns albeit the commute times are significantly longer.[12] Company employees are on average less unhappy than the US as a whole (14.3 percent report being unhappy in the 2010 General Social Survey).[13] Independent of gym membership status, our subject pool has poor health outcomes – based on self-reports of height and weight, 69 percent are either obese or overweight. These statistics align well with state-level statistics, and moreover with overall national statistics.[14] Interestingly, the gym members have body weights closer to their target weight but have higher rates of being obese or overweight. Indeed if we estimated cross-sectional

[11] Company gym membership is defined as whether or not the individual reports being a member at the time of the baseline survey. If a response to that question is missing, we search for their name in the company gym records.
[12] Based on authors' calculations using the 2010 Census.
[13] Source of statistic is http://sda.berkeley.edu/cgi-bin/hsda?harcsda+gss10.
[14] http://www.cdc.gov/obesity/data/adult.html.

regressions of gym membership status on health outcomes among this population, we would find mixed evidence of important health returns to exercise as members have better outcomes on some dimensions and worse on others. At the bottom of Table 1, it can be seen that rates of self-reported exercise are well below individual's self-reported target levels of exercise; the gap is 1.5 days/week for gym members and 2 days/week for non-gym members. Target levels of exercise are higher for gym members by roughly 1 day/week. While incentives may have the largest impact on the behavior of those already exercising, the likely largest health returns are for the inactive. Even among the members and much more sizable among the non-members, we observe significant shares of individuals reporting no exercise in a typical week. Overall, the groups are fairly well balanced across the different treatments with only tests of the differences in education levels among the non-members imbalanced across the incentive and the incentive+commit groups showing a p-value below 0.05.

Section 4. Results

4.1 Graphical analysis

We begin with a series of simple graphs to display the main results. In each graph we have stacked the 15 cohorts from the experiment and centered them around the incentive-period for the cohort. The graphs run from 20 weeks prior to the intervention through 20 weeks after the start of the intervention, with week 1 corresponding to the first week of the incentive program. We mark the incentive period with solid vertical bars at weeks 1 and 4. Subjects took the baseline survey at week -2 for these graphs. Treated subjects began learning about the incentives at the end of week -1, and week 0 was set aside for non-members to become members of the gym.

In Figure 1 for each week we plot the fraction of subjects who attended the company gym at least once. The green-diamond curve shows the series for subjects in the control group. Approximately 20% of the control group attends the company gym in any given week and the series is quite flat and stable throughout the experimental window. The blue-circle curve shows the series for subjects who were offered the incentive program but not offered the commitment option (incentive only group).[15]

[15] These graphs have been adjusted to normalize the percent of each group who were members of the company gym prior to the intervention, because the control group has a higher fraction of employees who were not members prior to the intervention than the treated groups. The reasons for this is difference is that due to capacity constraints on the gym staff's

The red-square curve show the series for subjects in the incentive + commitment group. As we would expect given random assignment, we see that 20% of each group attended the gym in a typical week prior to the study period. Both incentive groups roughly double their use of the gym during the incentive program. Because the sub-treatment of the commitment contract offer was randomly assigned and not announced until the incentive program ended, we expect to see similar patterns for these groups in the incentive period. The important pattern is that the two groups diverge once the incentive is removed. Those only given incentives saw a sharp reduction in the frequency of their visits following the end of the program. Although their attendance remains elevated relative to the control, the difference is essentially gone 8 weeks after the end of the incentive program. For the group offered commitment contracts, however, attendance in the post-incentive period remains substantially elevated. This effect also dissipates much more slowly. This is the first evidence that the option to create a commitment contract appears to significantly improve the long-run effect of the incentive program.

As we discussed above, we anticipated that there would be substantial heterogeneity in the response to the incentive program between employees who were members of the company gym prior to the study versus those who were not, and we stratified our randomization on this dimension to allow us to explore the patterns for these groups separately. Figure 2a shows that prior to the intervention, around 60% of members attended the gym at some point each week, consistent across control and treatment groups. That fraction rises to around 80% for members of both the IO and IC treatment groups during the incentive period. Among these existing members, however, the effect of the incentive only falls dramatically after the incentive program ends and the IO group shows little difference in attendance rates relative to control in the weeks after the incentive program is removed. For the group offered the chance to commit after the incentive program, in contrast, there are strong lasting increases in attendance.

Figure 2b. shows that the incentive program was effective at generating exercise among employees who were not previously using the company gym. Focusing on the trends for the control

---

ability to process new membership, the treatment probability in the experiment for employees who were not already members of the gym was slightly lower than that of employees who were members. We stratified the randomization on pre-treatment membership status to account for these issues and present most analysis separately for pre-treatment members and non-members as a result.

group, it is clear that in absence of our treatment, employees in this category usually transition into using the company gym at very slow rates, with only a handful of members of the control group having joined the gym within 20 weeks of taking our baseline survey. In contrast, one can see that 20% of subjects offered the incentives decided to join and use the company gym during that treatment period. These levels remain elevated throughout the post-period for both the incentive and incentive + commit groups.

4.2 Regression analysis

In order to better quantify these patterns, we turn now to regression specifications to explore our main effects. We run regression models of the following form:

$$y_{it} = \alpha_0 + \alpha_1(IO) + \alpha_2(IC) + \delta_0 intreatment + \delta_1(IO) \times intreatment + \delta_2(IC) \times intreatment + \beta_0 posttreatment + \beta_1(IO) \times posttreatment + \beta_2(IC) \times posttreatment + \mu_s + \pi_t + \varepsilon_{it},$$

where y_{it} is an outcome measure, such as an indicator for attendance, for subject *i* in week *t.* The specification interacts dummy variables for the two treatment groups incentive only (IO) and incentive with commitment option (IC) with dummies for various time periods, setting the control group as the omitted category. This specification allows us to test for differential use patterns at different periods of the study. The α coefficients measure the level of the outcome in the pre-intervention period and given a balanced randomization, we would expect α_1 and α_2 to be approximately zero in all specifications. The δ coefficients give the estimates of attendance on average over the 4-week incentive period ("intreatment"). We would generally expect $\delta_1$and δ_2 to be significantly positive and similar to each other. Our primary interest is in the β coefficients that measure the average response after the incentive period has ended ("posttreatment"). If the program has lasting effects, β_1 and/or β_2 will be significantly positive, with β_2 greater than β_1 if the availability of commitment contract increases habit formation. In many specifications we include fixed effects for each strata of the randomization x cohort combination (μ_s) to account for level differences between the strata and week fixed effects (π_t) to control for basic time trends. Since we have repeated observations, we adjust the standard errors for clustering at the individual level. Throughout the regression analysis we present

results separately for subjects who were and were not members of the company gym prior to the study.

Table 2 shows the regression results where the outcome measure is an indicator for whether or not the subject attended the gym that week. As such, the coefficients in this table can be naturally interpreted as the change in the fraction of employees using the gym in a given week. We present results using both OLS linear probability models and marginal effects from Probit regressions. The results throughout are not sensitive to the model specification. From the first two rows of Table 2, it is clear that among existing gym members there was little difference (and no statistically significant difference) in attendance rates between either treatment group and the control in the pre-treatment period, which in these regressions we have defined as the 6 weeks prior to the subjects taking the baseline survey and entering our study. By definition, the non-members had no visits during this pre-treatment period. Column (1) shows that among members, the fraction of the Io group attending the gym rose by 23 percentage points and of the IC group by 21 percentage points. In Column (3) the non-members saw increases relative to control of 15 percentage points (IO) and 20 percentage points (IC). The difference in this in-treatment response for the non members is not statistically significant at conventional levels (p-value = 0.18), and the Probit specification shows little difference.

We define the post-treatment period for these regressions as beginning after the incentives are removed (week 5) and runs through the end of the commitment period for the 8-week commitments (week 13).[16] For existing members of the gym, the incentive only group saw a modest and statistically insignificant increase in attendance of 3 percentage points relative to control during this period. Those also offered commitments, however, saw a statistically significant increase of 10 percentage points relative to control. A t-test for the difference between the effects for the Io and Ic groups is statistically significant with a p-value of 0.03. For those who were not members prior to the intervention, the incentive only group showed a statistically significant increase of 4-7 percentage points in the fraction using the gym during this post-treatment period. The Ic group attended at a rate of 9-12 percentage points higher than control over this period. The difference between these effects for the IO group and IC groups is statistically significant (p-value = 0.02 and 0.01 for OLS and Probit respectively).

[16] Subjects were given the opportunity to create the commitment in week 5, which then ran from weeks 6 through 13.

Table 3 shows that all of these patterns hold if the outcome is the number of visits in a week. With this measure average weekly visits for existing members rise by 0.9 visits per week during the incentive period for treated subjects relative to control, which is an increase of 50% relative to the baseline of 1.8 visits-per week for control. For these existing members, the post-intervention effect is 0.16 (not statistically significant) for those offered only incentives and 0.31 (statistically significant) for those offered incentives and commitments, representing lasting increases relative to baseline for the control of 9% and 17% respectively. For non-members the in-treatment effects are between 0.39 and 0.53 for IC and IO respectively and these groups have lasting effects of 0.09 and 0.20 visits per-week in the post-period. Overall, combining members and non-members, we see post-treatment increases for these groups of between 20% and 50% of the baseline. We can compare this to Charness and Gneezy (2009) and Acland and Levy (2010), both of whom report average visit measures. As a fraction of the pre-intervention control mean, both studies find post-intervention effects of approximately 100% increases in the undergraduate population. As a percentage of the in-treatment increase, Charness and Gneezy find a lasting effect of 46% while Acland and Levy find 21%, which is in line with the 20% to 38% increases relative to in-treatment effects we find here.

In Figure 3 we use the results of the Probit estimation for the probability of attending the gym in a given week to explore the longer-run response to the program. This figure shows the estimated difference in the fraction of subjects using the company gym for the two treatment groups relative to the control group in 4-week groups running for 13 months after the start of the incentive program. For those previously members of the company gym (Figure 3a) the treatment group offered incentives only has no lasting increase in exercise relative to the control group past the second month (first month after the incentive program). However, the rates of exercise for treatment group offered commitment contracts is clearly elevated during the commitment period (months 2 and 3) and does not fall off to the control-group levels until approximately month 8. For the non-members (Figure 3b) both the incentive only and incentive commit groups show lasting increases in gym use relative to the control for this entire long-run period. The IC group appears to have higher exercise than the IO group during the commitment period, but the difference falls after the end of the commitment period.

Figure 4 presents another way of looking at the post-treatment effects of the program. The y-axis for this graph gives the fraction of weeks with positive gym visits between the pos-treatment

period (weeks 5-13) and the pre-treatment period, and is a measure of the post-treatment change in exercise behavior. The x-axis shows the change in exercise between the intreatment period and the pretreatment period. If all of the new exercise generated by the incentive program were maintained after the incentives were removed (i.e., full habit formation), then this scatter plot would lie along the 45-degree line. The graph shows (with bubble-size reflecting the relative number of subjects with that combination of changes) that the IC group had a greater long-run response relative to their intreatment changes than did the IO group and that this pattern holds for both the members and non-members prior to the study.

4.3 Take-up of commitment contracts

The results above for the IC group suggest that the offer of commitment contracts improved the lasting effect of the incentive program. Overall among the 346 members of the incentive group offered commitments, 12.4% of subjects chose to make a commitment and on average these committers placed $58 at stake for their commitments. Sixty-three percent of those who created commitments successfully maintained the commitment of not missing more than 14 days at the gym. Among existing members of the company gym, the take-up rate of commitments was 23%. For those who were not previously members of the gym, the overall take-up rate was 6% and it was 21% for this group conditional on making at least one visit to the gym in the incentive period.

Table 4 shows the results of regressions to estimate the correlates of commitment-contract takeup. The table is restricted to the IC group and the dependent variable is an indicator for whether or not the subject decided to create a commitment.[17] For those who were existing members of the gym (Columns 1 and 2), the only strong predictor of commitment take-up is gender, with men making commitments at much lower rates than women. The point estimate for men (relative to women) is similar for those who were not member of the gym (here restricted to those who used the gym at least once during the incentive program), but is not statistically significant in this smaller sample. For non-members of the gym, the only significant predictor of commitment takeup is being overweight. Every 10 lbs a subject reports being over their target weight is associated with a 5 percentage point increase in the probability the subject will create a commitment.

[17] Given the small sample size of subjects who made commitments, we do not present any analysis on the determinants of the dollar amount subjects choose to commit.

This table also includes three measures we anticipated could be used to predict demand for commitment. First, subjects were asked a standard question on intrinsic motivation and we split the sample at the median response to that scale and included an indicator for whether the subject "lacked intrinsic motivation" for exercise. We also asked subjects in our pre-survey to judge their own level of self-control for exercise from low to high. We split that measure at the median and included an indicator for having below-median self-control. Finally, we included a dummy for whether the subject self-reported exercising below her own personal target prior to our study. Interestingly, none of these measures predict the take-up of commitment contracts. Among members, the point estimates suggest that those who rate themselves as having low self-control were somewhat more likely to make commitments, however among non-members that result goes in the opposite direction.

We present two other cuts in this table. First, we asked subjects to give their expectation for how often they would use the gym during the post-treatment period (weeks 5-13) at the end of the incentive program and prior to the commitment group learning about their option to commit. We also asked them to rate their "ideal level of exercise" at the company gym during this period. We anticipated that those who expected to go less than their ideal level were exhibiting sophistication about their self-control issues, however including an indicator for that measure of sophistication has no economic or statistical significance in the regression. Finally, for the members, we can compare the take-up rates as a function of how frequently the subject was using the gym in the pre-treatment period. Interestingly, we find no predictive power of this pre-treatment use and in fact our point estimates reveal that the high-frequency users (those who attended at least once each week in the 6 weeks prior to the study) are actually slightly *more likely* to commit than those who were attending infrequently to make commitments. This is an intriguing finding, suggesting that the demand for commitment contracts is more wide-spread than simple models of time inconsistency would suggest. We return to this topic in the discussion at the end.

4.4 Robustness

Direct comparison of treatment groups. While the results above show that the group offered incentives and commitments had higher post-treatment differences relative to control than the group

offered incentives alone, we can also do a direct comparison of the patterns of use for the treatment groups. In Table 5, we restrict the sample to the two treated groups and focus on the fraction of weeks with positive visits during the initial post-treatment period (weeks 5-13). We find that among members, the estimated difference is 9 percentage points, which is actually a little higher than th estimates from Table 2. Since the randomization into the commitment offer occurs after the incentive program, with this restricted sample, we can also estimate the difference controlling for the number of visits the subject made to the company gym during the treatment period. This ensures that any differences in post-treatment use are not an artifact of random differences in the response to the incentive program for the different groups. Column 2 shows that result for members and we find that controlling for these difference there was a 6 percentage point increase for the IC group relative to IO. Columns 3 and 4 show the same results for the non-members who attended the gym at least once during the treatment program. Without controls for in-treatment visits, the difference is 15 percentage points and with controls is 11 percentage points, though the latter estimate is not statistically significant (p-value = 0.11).

The results above for the IC group are "intention to treat" effects in that they are the results 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 inflate the estimates in Table 5 by the reciprocal of the commitment takeup rate to get an implied IV estimate. Those estimates are given at the bottom of the table and all fall between 0.26 and 0.63 visits per week. 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.

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. In particular, subjects in the IC 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 or focusing on their exercise routine. As such, the intention-to-treat estimates in the main regressions should serve as the primary evidence on the effectiveness of the availability of commitment contracts to generate behavior change.

Substitution. One important issue with any incentive program that targets a particular behavior is to understand the extent to which individuals respond to that incentive by substituting away from un-incentivized activities. This is particularly important in the context of a workplace exercise incentive, since subjects might simply start exercising at the company gym as a substitute of their exercise elsewhere. To test for this possibility of substitution, we can use the initial endline survey to look at how incentives impact self-reported overall rates of exercise. Table 6 presents estimates for this purpose. For each sample in this table, we present 4 estimates: exercise at the company gym as measured by the gym data for all subjects, exercise at the company gym as measured by the gym data for respondents of the first endline survey, self-reported rates of exercise at the company gym, and self-reported rates of overall exercise. A priori, we predicted that the substitution effects would be heterogeneous – in particular, among non-members. The non-member group includes both regular exercisers and non-exercisers. For that reason, we stratified the randomization by whether an individual reported that they were at/above or below their target level of overall exercise. For those at/above their target, the incentive effects are more likely to be substitution.

But before discussing the substitution results, it is important to address two concerns with this analysis. First, there might be a worry that the sample of responders to the post-survey is selected.[18] This concern is lessened when we note that the response rate to this post-survey is particularly high (over 90%). Nonetheless, as a first pass at examining this type of selection, we examine the incentive effects differ for the overall sample and the sample of post-survey responders.[19] These two sets of estimates are very similar and sometimes identical mitigating our concerns about selection bias. Second, the reliability of self-reported data is often questioned. Our best attempt at assessing the

[18] We observe a statistically significant difference in the response rate of the incentive group and the control group. Such selective response may be bias our estimates. In future iterations of this paper, we will address this selection issue following the Lee (2009) bounding procedure.

[19] Of course, if we observe different estimated effects, this divergence could be simply due to heterogenous treatment effects rather than selection bias.

reliability involves a comparison of self-reported incentive effects with the actual incentive effects as measured by the computerized gym data. Although self-reported exercise levels are elevated, the estimated effects are quite consistent across the two data sources with the exception of members at/above their target level of exercise.[20] As such, we may want to interpret the substitution effects for members at/above their target with caution. Excluding this group, the average margin error is fairly low at 16 percent.

To gauge the extent of substitution, we can compare the treatment effects for overall exercise relative to that for company gym exercise. The equality of the estimated coefficients indicates the effects on company gym exercise represent one-for-one increases in overall exercise. For below target level of exercise subjects, as ex-ante predicted, the incentives led to new exercise with little or no substitution. For gym members in the below target group, the effect on overall exercise and company gym exercise are identical, indicating no substitution at play. This finding is not surprising since for most company gym members, their main place of exercise is the company gym. For those above their target level of exercise, according to these results incentives led mainly to substitution of exercise at other venues for exercise at the company gym. Overall, this supplementary analysis fills an important unanswered question in the literature of whether incentives for exercise actually increase exercise rather than just change the location of exercise. For the majority of subjects in our experiment, this workplace exercise incentive generated a real change in exercise behavior.

Heterogeneity of response. We have found that there is modest lasting effect of the program once the incentives are removed and that the lasting effects are considerably stronger if employees are offered the chance to create a commitment contract. In this subsection, we discuss the heterogeneity of the response to the incentive and incentive-plus-commitment treatments in order to gain more insight into how financial incentives interact with time inconsistency in this context.

The primary mechanism by which we might expect the temporary incentive to have a lasting effect on behavior is by getting people over the initial "start-up" costs of using the gym. The post-

[20] Eighty-seven percent of members at/above their target level of exercise report going to the company gym 3 or more times per week, so we would not expect that the incentives have much of a behavioral effect. Moreover, for this group, how often they go (2 versus 3 times) may be less salient than a less regular user. Also, it appears that measurement error of the self-reported data is high; the average visits for the control is 3.51 when in fact it is 2.68 as determined from the gym data. Thus, it is more difficult to detect an incentive effect among this population in both the survey and administrative gym data.

treatment effects for the incentive only group appear stronger for those who were not members of the gym prior to the experiment. That pattern is largely consistent with the anticipated habit-formation mechanism of a temporary incentive, since many existing members would already have overcome the start-up costs associated with establishing an exercise routine at the company gym, while the non-members clearly have not. In contrast, the commitment contract could in theory help generate lasting behavior change even for those who have overcome the start-up costs of exercising, if it helps them overcome day-to-day struggles with self control. Consistent with that idea, there is a strong lasting effect for the group offered the commitments relative to control for both members and non-members.

In the top section of Table 7 we present a heterogeneity cut to investigate how these different treatments play out as a function of the exercise behavior of employees prior to our study. It is, of course, important to recognize that this type of heterogeneity cut is not based on random assignment and as such, 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 we observe. Nonetheless, we feel this analysis can provide useful suggestive evidence about the mechanisms behind the main effects. 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 in 33% or less of the weeks, the "middle" group 50%-83% of the weeks and the "high" group every week in the pre-period. The "low" group here is analogous to the population identified as driven by present bias in DellaVigna and Malmendier's (2006) study of gym-going behavior. 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 posttreatment effects. Concentrating on columns 3 and 4, where we account for any imbalance in these subgroups by controlling for observables, we find that the long-run effects of the incentive only are statistically insignificant for all groups, but have a fairly high point estimate of 0.09 for those with low exercise prior to the study. The somewhat higher response for that group could be consistent with the incentive helping them overcome "startup costs" associated with re-initiating their use of the gym. The largest difference between the IC and the IO group (0.16 vs. 0.02) is found for those who were exercising, but not consistently (the "middle" group) in the pre-period. Finally, and unsurprisingly we

find no estimated increase for either group among those who were exercising regularly prior to the study.

Among non-members prior to our study we conduct a similar tercile cut based on their self-reported frequency of overall exercise in the pre-treatment period. We are especially cautious about over-interpreting these cuts based on the potentially noisy and unverifiable self-reports. Each of these sub-groups responded fairly similarly to the incentive program during the treatment period. We also find little meaningful differences among these groups looking at the post-treatment period, though the low-exercisers appear not to have had any lasting effect from the incentive alone. For each sub-group, the point estimates suggest a higher post-treatment effect for the group offered incentives and commitments than those only given incentives. Given potential concerns about substitution, it is useful to note that the main patterns of effects for non-members observed earlier hold for the two lower groups of self-reported exercise and are not driven solely by those who reported exercising frequently prior to our study.

Given our findings that men make commitment contracts at much higher rates than women, a second natural heterogeneity cuts is to split the samples by men and women. Among both existing members and non-members there were roughly equal fractions of men and women in each group. The middle rows in Table 7 presents this heterogeneity cut for both members and non-members. Among male members, both the IO and IC groups showed increased post-intervention frequency of exercise of over 12 percentage points, with no difference between IO and IC. For the female members, however, the point estimate for the initial post-intervention period for the Io group is actually negative (-0.10), and statistically significant when controls are included, and could be consistent with a drop in intrinsic motivation for women who receive financial incentives and then have those incentives removed. The availability of commitment contracts substantially improves the long-run response for women to a positive, though statistically insignificant, 0.08. Taken together and combined with the results on the take-up of commitments, these results suggest that for members, men have habit formation from the incentive alone and neither see a need for or appear to benefit from the availability of commitments. Women in contrast, do not show lasting effects of the incentives alone, but recognize a need for commitment and benefit substantially from the availability of commitments. Among non-members there are smaller differences between the genders. In both cases, the point estimates on the post-

intervention period are higher for the IC group than the IO group, though there are also in-treatment differences between these groups for the female employees who were non members prior to the study.

Section 5. Discussion and Conclusion

In this paper we have shown that working adults respond strongly to a financial incentive to use their company gym over a 4 week period and that this temporary incentive program results in increased exercise even after the incentive is removed. The long-run effect of the incentive program appears to be concentrated on employees who were not using the gym prior to the stud. This suggests that the high start-up costs of exercise are a barrier to people with time inconsistency and that that incentive programs could be useful at helping people overcome that barrier. However, the lasting effect of the financial incentive alone is also quite modest and fades out within 2 months of the end of the program. 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.

We find, though, that coupling the incentive program with an option for subjects to create their own self-funded commitment contracts substantially improves the long-run effects of the program. The take-up and effect of commitment contracts on exercise is direct evidence that people recognize their problems with self-control in this domain and can benefit from binding commitment technologies. 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 idea that there are important time-inconsistency problems that limit peoples' 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 interventions. We are optimistic that recognizing the potential value of commitment technologies could help improve the long-run impact of a range of health-behavior interventions.

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 that call for

further research. First, the take-up rates of commitment contracts below 20% in our study are in line with previous studies of commitment devices. That basic level of take-up holds even among sub-populations who state that they exercise less than they would like, for whom theory would suggest these commitments should be attractive. While it is perhaps not surprising that there are subjects who were not interested in the very specific commitment we offered in this study, the general pattern across this literature of positive, but modest take-up of commitment, leaves open a question of why more people do not commit. Some questions that could be addressed in future research on this topic include: a) what fraction of people are naïve about their self-control problems and think they could not benefit from commitment?, b) what other means of overcoming self-control do people use in these settings and to what extent are commitment contracts offered in research studies competing with other ways people attempt to overcome self-control? and c) how should commitment devices, especially self-funded financial commitment contracts, be designed to improve their appeal and effectiveness?

A second open question arises when we instead focus on commitment-contract demand among employees who already exercise regularly and hence do not on the surface appear to be struggling with self-control problems. We find take-up rates of commitment contracts above 20% for these individuals. Why do these employees make commitments? One possibility is that these individuals have self-control problems that they have previously found ways of overcoming, but that they find the financial commitment more attractive than these other approaches. If financial commitments are attractive alternatives to other forms of self-control, it suggests that they could help improve individual welfare even when they do not result in observable changes in behavior. One possibility that may bear further exploration in future research is that financial commitments can substitute for costly internal self-control. While economists tend to model self-control as a struggle between different time-period "selves", psychologists often think about self-control as a constant mental battle to overcome temptation. 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 overcome self-control. The fact that many subjects who make commitments currently exercise frequently enough that they are not near the margin where the commitment binds (attending once every 2 weeks) would be consistent with the idea that

commitments are lowering the costs of maintaining self-control rather than controlling behavior at the margin. Since our study was the first to offer commitments broadly to populations with no apparent need for commitments, little is known about this issue and more research is needed to better understand the boundaries of the use of commitment.

A third open question about commitment contracts is the extent to which the gender differences we find here in both the apparent need for and demand for commitments is a general effect. Ashraf et al (2006) found that measures of present-bias were predictive of commitment-contract takeup only for women. This might suggest that women tend to have different degrees of sophisticated time inconsistency than men. However, Ashraf et al (2006) did not find any difference in the takeup rate of their program between men and women. Taken together, our results suggest that there may be important gender differences in time inconsistency and the value of commitment, but more work is needed to understand those differences fully.

Finally, although our study was not designed as a program evaluation of a comprehensive corporate wellness program, there are a number of insights from our research that might help inform the design of wellness programs that incorporate incentives for exercise. The effects of our program on behavior during the incentive program can provide a benchmark for how exercise behavior might respond to a financial incentive. We observe an approximate doubling of exercise at the company gym during our treatment period. On average, the group offered incentives attended the gym 4.8 times in the 4-week incentive period for a cost per-participant when we include the gym-membership reimbursement of just under $75. If one were willing to assume that the in-treatment responses we observe during the study would hold if the program was run on an annual basis, this suggests a back-of-the-envelope figure of $900 per year to double the exercise of the employee population. That represents 18% of the average individual health-insurance premium of $5,049 reported by the Kaiser Family Foundation (kff.org) for large employers. So the question becomes whether a doubling of exercise frequency could generate an 18% reduction in health-care costs. Our study is not designed to provide the answer to that question but does provide this rough guideline as to the effect that might be needed to make an incentive program cost effective in terms of the reductions in health-care costs. Of course, one needs to be cautious before extrapolating from our 4-week program to the effects of a year-long incentive program. We do see slight declines in use of the gym even during our 4-week

program, so it is possible that over the course of the year the effects of an incentive might be lower. On the other hand, the calculation above only considers health-care costs and increased exercise might improve both absenteeism and worker productivity.

Our findings also show that a substantial fraction of the money paid out in an incentive program of this type goes to reward employees for exercise they already do. Since the program doubles use of the company gym, half of the exercise during the incentive period would have been expected even in absence of the program, so a full 50% of the money rewards exercise that would have occurred at the company gym anyway. Our substitution analysis also suggests that approximately 13% of the new exercise generated at the company gym may be substituting from other forms of the exercise the employees do elsewhere. Taken together, this suggests that only 37% of the cost of the program was spent on generating new exercise, and highlights the potential value of finding ways to target wellness programs on new exercise if possible.

The more general insight that corporations and policymakers could take away from this paper is that because a temporary incentive, especially when coupled with commitment devices, has lasting effects on exercise behavior. This habit formation could be exploited to improve the cost effectiveness of wellness incentives. For the group offered incentives coupled with commitments, we found an increase of 100% in the fraction of regular exercisers during the incentive and a lasting increase of 50% over the 2-months after the program ended. This suggests that a periodic incentive program with 1 months of incentives followed by 2-months of commitments with no incentives could potentially generate 66% of the effect on exercise for one-third of the cost of a program where incentives were in place at all times. Of course, to date we know little about either the long-run response to a continuing incentive program or the response to programs that are run periodically. The results in this paper suggest that there could be large value to research that explores these issues further.

References:

Acland, Dan and Matthew Levy. 2010. "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.

Bandura, Albert. 2006. "Guide for Constructing Self-Efficacy Scales." In F. Pajares & T. Urdan (Eds.). *Self-efficacy Beliefs of Adolescents*, (Vol. 5., pp. 307-337). Greenwich, CT: Information Age Publishing.

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.

Benabou, R. and J. Tirole. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies*, 70; 489-520.

Benabou, R. and J Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5); 1652-1678.

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

Berry, J.D., B. Willis, S. Gupta, C. Barlow, S.G. Lakoski, A. Khera, A. Rohatgi, J.A. de Lemos, W. Haskell, and D.M. Lloyd-Jones. 2011. "Lifetime Risks for Cardiovascular Disease Mortality by Cardiorespiratory Fitness Levels Measured at Ages 45, 55, and 65 Years in Men." *Journal of the American College of Cardiology,* 57: 1604-1610.

Buchman, A., P. Boyle, R. Wilson, et al. 2007. "Physical Activity Decline in Older Persons." *Muscle Nerve,* 35(3); 354-362.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. 2011. "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction." NBER Working Paper.

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

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

Colcombe, S., and A. Kramer. 2003. "Fitness Effects on the Cognitive Function of Older Adults: a Metaanalytic Study" *Psychological Science,* 14(2);125-130.

Deci, E. 1971. "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology*, 18; 105-115.

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

Fox, K. 1999. "The Influence of Physical Activity on Mental Well-Being." *Public Health Nutrition,* 2(3A); 411-18.

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, U. and A. Rustichini. 2000. "Pay Enough-Or Don't Pay At All." *The Quarterly Journal of Economics*, 115 (3); 791-810.

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

Gupta, S., A. Rohatgi, C.R. Ayers, B.L. Willis, W.L. Haskell, A. Khera, M.H. Drazner, J.A. de Lemos, and J.D. Berry. 2011. "Cardiorespiratory Fitness and Classification of Risk of Cardiovascular Disease Mortality." *Circulation,* 123:1377-1383**.**

Hu, F., T. Li, G. Colditz, et al. 2003. "Television Watching and Other Sedentary Behaviors in Relation to Risk of Obesity and Type 2 Diabetes Mellitus in Women" *JAMA,* 289(14); 1785-1791.

Jeffery, Robert, Paul Thompson and Rena Wing. 1978. "Effects on Weight Reduction of Strong Monetary contracts on Calorie Restriction or Weight Loss." *Behaviour Research and Therapy*, 16; 363-9.

Jeffery, Robert W., Wendy Gerber, and Barbara Rosenthal, Ruth Lindquist. 1983. "Monetary Contracts in Weight Control: Effectiveness of Group and Individual Contracts of Varying Size." *Journal of Consulting and Clinical Psychology*, 51(2); 242-248.

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.

Katz, Lawrence, J. Kling, and J. Liebman. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1); 83-119.

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

Lee, David. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects" *Review of Economic Studies* 76(3): 1071-1102.

Li, T., J. Rana, J Manson, W. Willett, M. Stampfer, G. Colditz, K. Rexrode, and F. Hu. 2006. "Obesity as Compared With Physical Activity in Predicting Risk of Coronary Heart Disease in Women." *Journal of the American Heart Association,* 113: 449-506.

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

Littman, A., A. Kristal, and E. White. 2005. "Effects of Physical Activity Intensity, Frequency, and Activity Type on 10-y Weight Change in Middle-aged Men and Women." *International Journal of Obesity,* 29(5); 524-533.

Markland, David. 1999. "Self-determination Moderates the Effects of Perceived Competence on Intrinsic Motivation in an Exercise Setting." *Journal of Sport and Exercise Psychology*, 21; 351-361.

Matthews, C, A. Jurj, X. Shu, et al. 2007. "Influence of Exercise, Walking, Cycling, and Overall Nonexercise Physical Activity on Mortality in Chinese Women" *American Journal of Epidemiology,* 165(12); 1351-1353.

Melov, S., M. Tarnopolsky, K. Beckman, et al. 2007. "Resistance Exercise Reverses Aging in Human Skeletal Muscle." *PLoS ONE* 2(5): e465. doi:10.1371/journal.pone.0000465

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.

Pate P., M. Pratt M, SN Steven, W.L. Haskell, C.A. Macera, and C. Bouchard et al. 1995. "Exercise and Public Health: A Recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine." *JAMA,* 273:402–7.

Paxton, Roger. 1979. "Do Methods Matter in Smoking Clinic?" *Health Bulletin*, 37**;** 164–167.

Paxton, Roger. 1980. "The Effects of a Deposit Contract as a Component in a Behavioural Programme for Stopping Smoking." Behaviour Research and Therapy, 18(1); 45-50.

Paxton, Roger. 1981. "Deposit Contracts with Smokers: Varying Frequency and Amounts of Repayments." *Behaviour Research and Therapy*, 19;117-123.

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.

Tangney, J. P., R. F. Baumeister, and A. L. Boone. 2004. "High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success." *Journal of Personality*, 72 (2); 271-324.

Volpp et al. 2008. "Financial Incentive-Based Approaches for Weight Loss: A Randomized Trial." *JAMA,* 300(22): 2631-2637.

Zhang F. 2010. *Incentive Experiments: NSF Experiences.* Working Paper SRS 11-200. Arlington, VA: National Science Foundation, Division of Science Resources Statistics.

Figure 1. Fraction with positive gym visits by treatment status

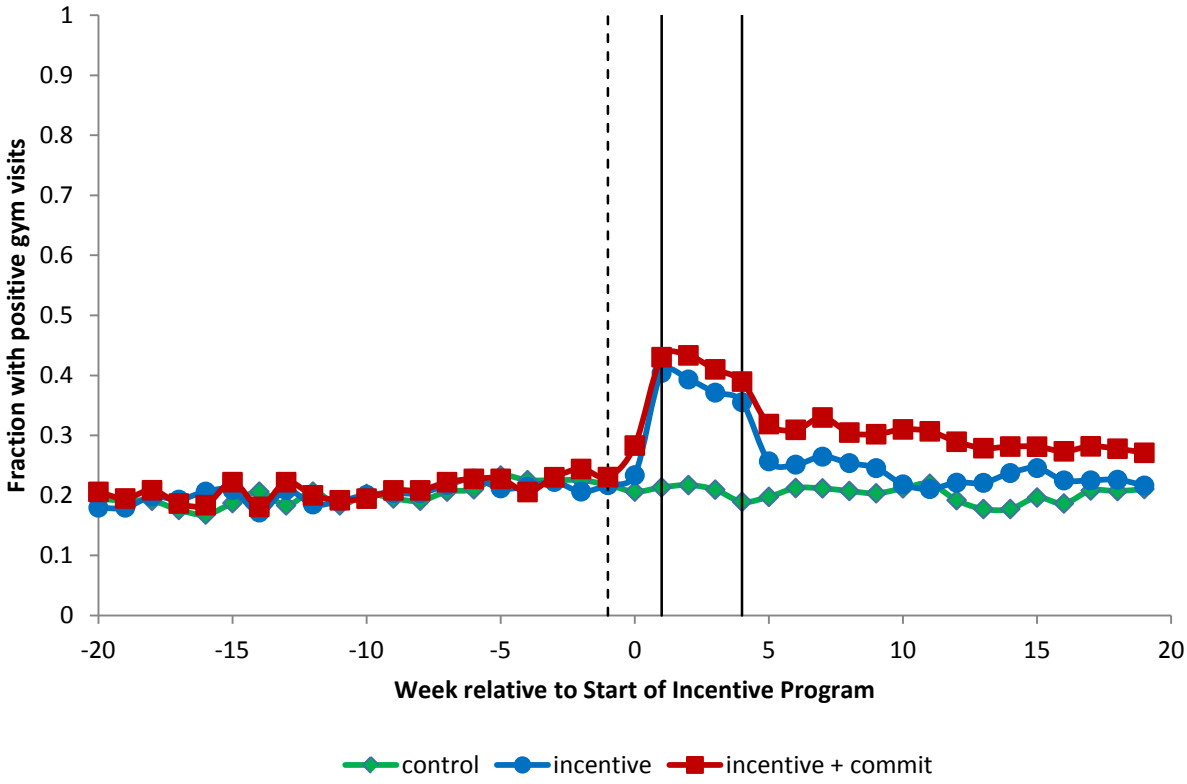


Figure 2a. Fraction with positive gym visits by treatment status (members only)

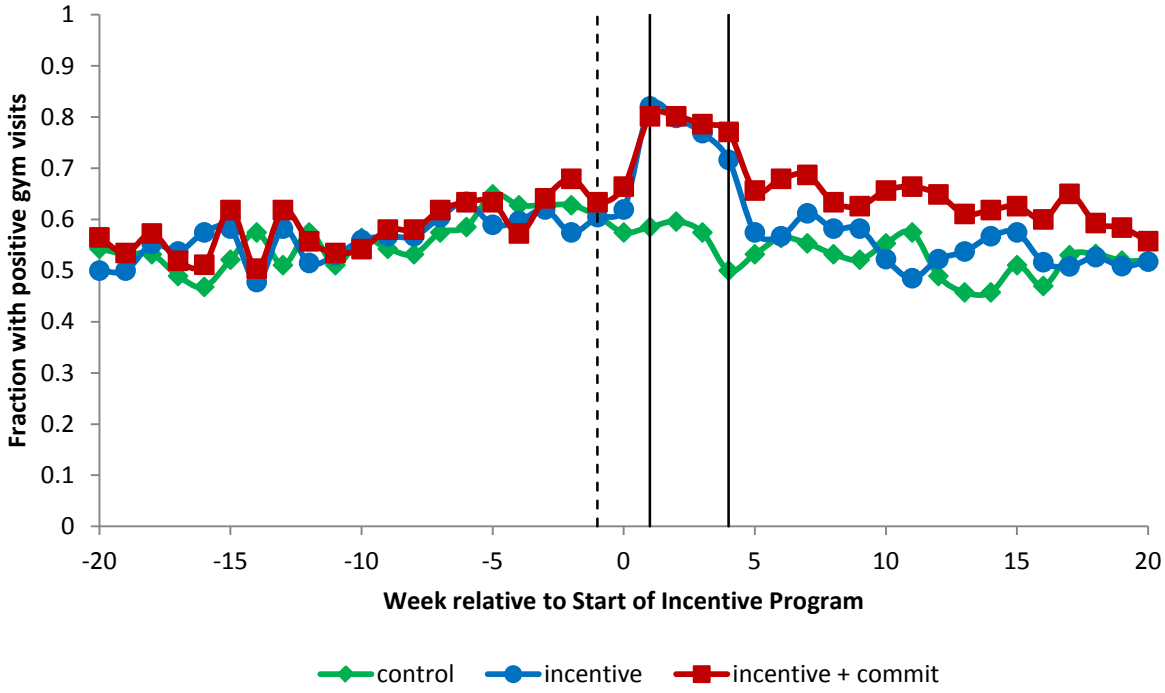


Figure 2b. Fraction with positive gym visits by treatment status (non-members only)

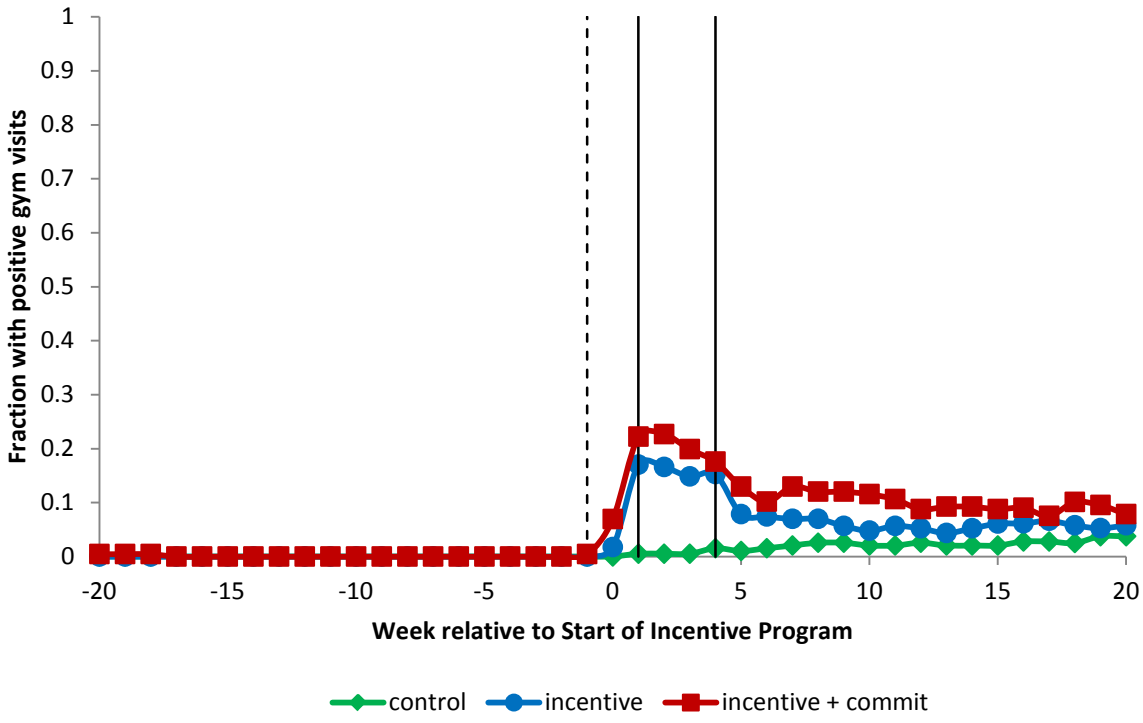


Figure 3 a. Long Run Treatment Effects (members only)

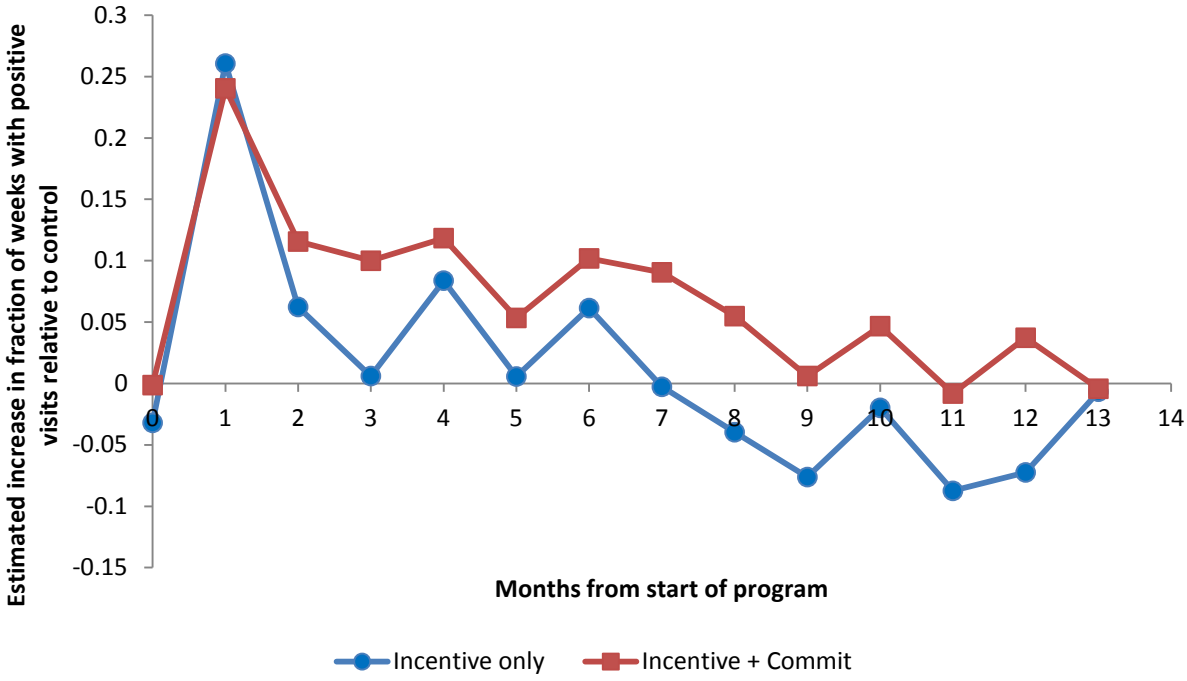


Figure 3 b. Long Run Treatment Effects (non-members only)

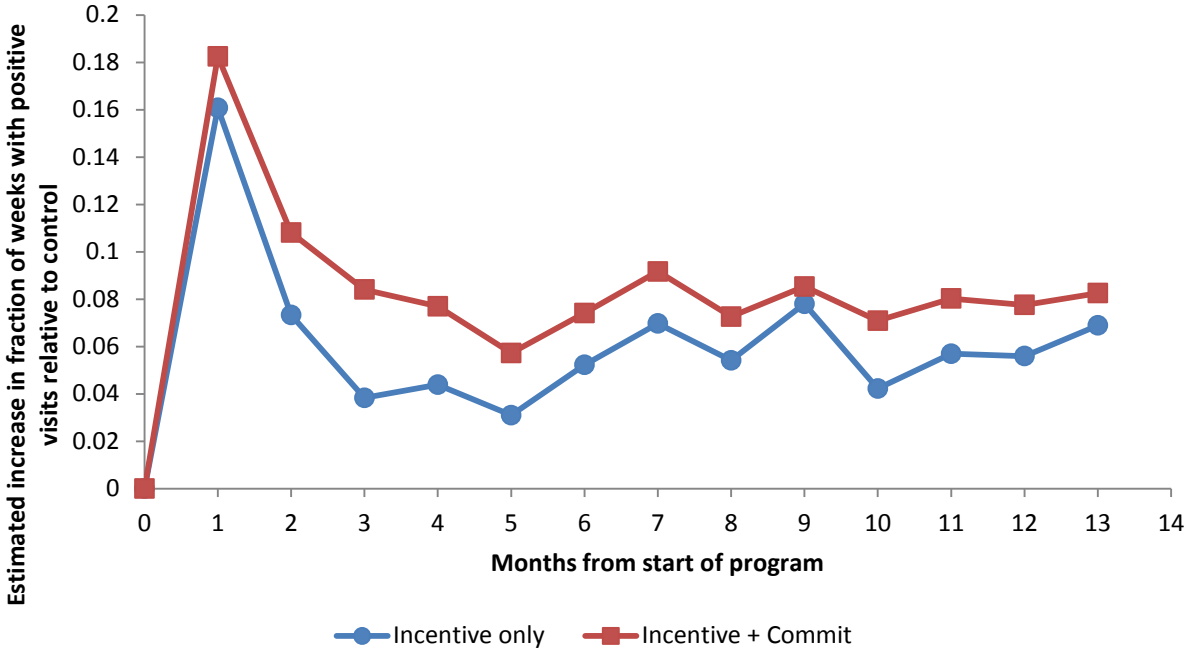


Figure 4 a. Post-Treatment Changes versus In-Treatment Changes in Exercise (members only)

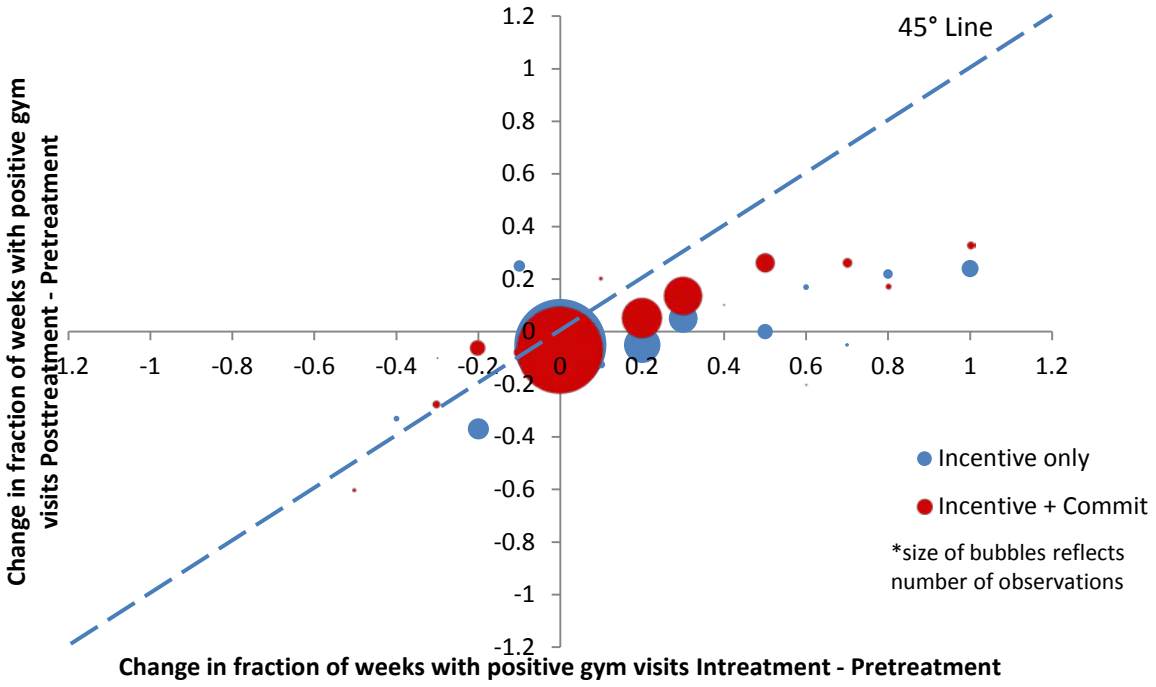


Figure 4 b. Post-Treatment Changes versus In-Treatment Changes in Exercise (non-members only)

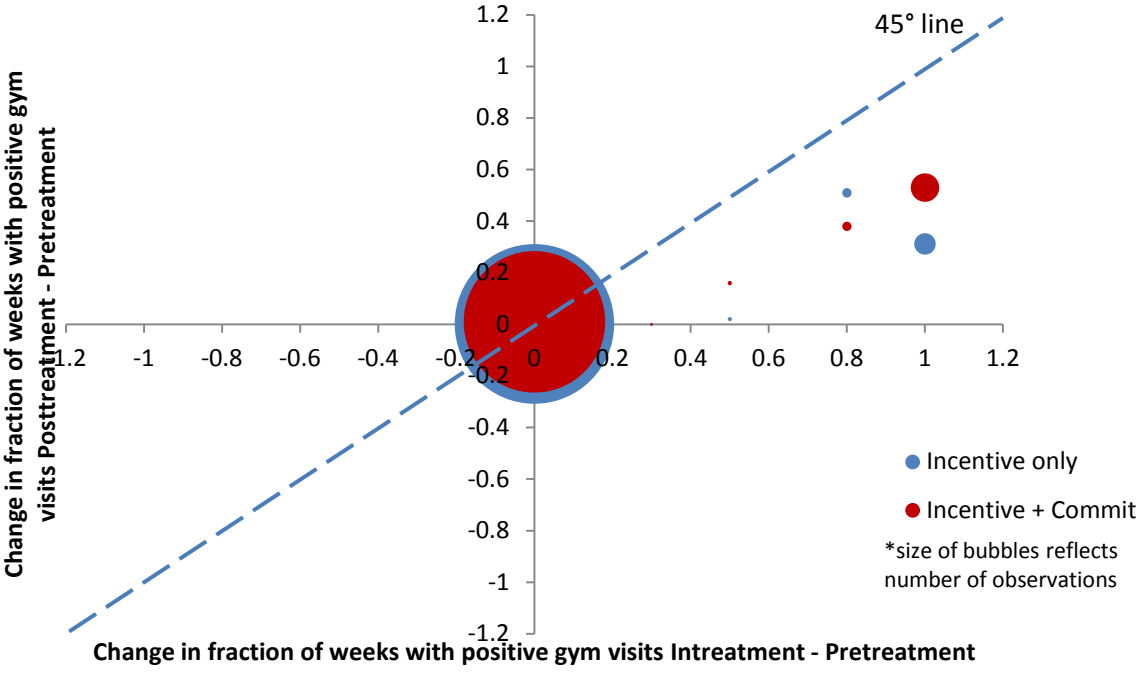


Table 1. Descriptive Statistics

Panel A. Pre-Treatment Survey Variables for Members

	(1) Overall	(2) Control	(3) Incentive	(4) Incentive+ Commit	p-value (2)=(3)=(4)	p-value (3)=(4)
Basic Demographics						
Age	39.55	40.12	38.71	40.00	0.61	0.57
	(10.73)	(10.63)	(10.16)	(11.39)		
Male	0.51	0.46	0.51	0.55	0.53	0.35
College degree or more	0.66	0.61	0.71	0.66	0.37	0.96
Living Situation						
Married	0.71	0.68	0.71	0.74	0.78	0.53
Has Children at Home	0.50	0.45	0.52	0.52	0.47	0.64
One-Way Commute (Minutes)	36.78	37.82	35.29	37.54	0.48	0.53
	(17.53)	(21.23)	(15.51)	(16.54)		
Subjective Wellbeing						
Unhappy with Life	0.08	0.07	0.07	0.09	0.89	0.63
Unhappy with Fitness	0.36	0.34	0.38	0.34	0.64	0.72
Unhappy with Weight	0.53	0.58	0.54	0.48	0.33	0.14
Health and Fitness						
Pounds over Target Weight	19.49	20.28	17.72	20.73	0.69	0.44
	(26.85)	(18.68)	(30.21)	(28.24)		
BMI	27.84	28.31	27.51	27.85	0.68	0.92
	(5.53)	(5.52)	(5.16)	(5.90)		
Overweight	0.43	0.43	0.39	0.47	0.50	0.31
Obese	0.27	0.30	0.28	0.23	0.47	0.22
Takes blood pressure meds	0.14	0.12	0.16	0.13	0.78	0.72
Exercise						
Average days of overall exercise	3.39	3.36	3.32	3.49	0.34	0.16
	(1.61)	(1.65)	(1.72)	(1.47)		
0 days of overall exercise	0.06	0.05	0.08	0.03	0.16	0.07
Target days of exercise	4.86	4.79	4.81	4.96	0.26	0.10
	(1.04)	(1.08)	(1.08)	(0.97)		
Number of Observations	359	94	134	131		

Notes: Means for selected variables are presented without parentheses. Standard deviations for selected variables are presented with parentheses.

Table 1. Descriptive Statistics

Panel B. Pre-Treatment Survey Variables for Non-Gym Members

	(1) Overall	(2) Control	(3) Incentive	(4) Incentive+ Commit	p-value (2)=(3)=(4)	p-value (3)=(4)
Basic Demographics						
Age	39.55	39.62	39.11	39.96	0.65	0.52
	(10.99)	(10.82)	(11.77)	(10.30)		
Male	0.53	0.52	0.54	0.52	0.90	0.69
College degree or more	0.68	0.64	0.66	0.73	0.06	0.02
Living Situation						
Married	0.66	0.67	0.64	0.68	0.65	0.47
Has Children at Home	0.47	0.48	0.43	0.52	0.18	0.10
One-Way Commute (Minutes)	38.18	38.03	39.10	37.35	0.69	0.41
	(20.48)	(20.54)	(24.29)	(15.41)		
Subjective Wellbeing						
Unhappy with Life	0.12	0.11	0.12	0.13	0.91	0.74
Unhappy with Fitness	0.48	0.54	0.44	0.46	0.08	0.48
Unhappy with Weight	0.52	0.55	0.51	0.51	0.61	0.53
Health and Fitness						
Pounds over Target Weight	23.40	22.72	22.09	25.40	0.58	0.31
	(30.00)	(28.59)	(31.36)	(29.82)		
BMI	28.31	28.22	27.88	28.85	0.36	0.23
	(6.54)	(6.51)	(6.50)	(6.61)		
Overweight	0.38	0.42	0.36	0.38	0.44	0.86
Obese	0.31	0.30	0.28	0.34	0.43	0.25
Takes blood pressure meds	0.12	0.13	0.13	0.10	0.75	0.45
Exercise						
Average days of exercise	1.91	1.98	1.90	1.86	0.54	0.87
	(1.72)	(1.73)	(1.71)	(1.72)		
0 days of exercise	0.26	0.24	0.25	0.27	0.86	0.65
Target days of exercise	3.99	4.05	3.89	4.05	0.30	0.40
	(1.42)	(1.33)	(1.41)	(1.50)		
Number of Observations	638	195	228	215		

Notes: Means for selected variables are presented without parentheses. Standard deviations for selected variables are presented with parentheses.

Table 2. Regression results -- Any Visits to the Gym

Dependent variable: Indicator for whether the subject attended the company gym that week

	Members		Non-Members	
	(OLS)	(Probit)	(OLS)	(Probit)
Control mean of dep var in pre-period	0.62	0.62	0	0
Incentive only	-0.02	-0.02	-	-
	(0.05)	(0.04)		
Incentive + Commit	0.01	0.01	-	-
	(0.05)	(0.04)		
Intreatment period (weeks 1-4)	0.03	0.03	-	-
	(0.04)	(0.04)		
(Incentive only) x (Intreatment)	0.23***	0.24***	0.15***	0.21***
	(0.04)	(0.04)	(0.02)	(0.03)
(Incentive + Commit) x (Intreatment)	0.21***	0.22***	0.20***	0.23***
	(0.04)	(0.04)	(0.03)	(0.03)
Posttreatment (weeks 5-13)	0.03	0.04	0.03**	0.07***
	(0.05)	(0.05)	(0.01)	(0.02)
(Incentive only) x (Posttreatment)	0.03	0.03	0.04**	0.07**
	(0.03)	(0.03)	(0.02)	(0.03)
(Incentive + Commit) x (Posttreatment)	0.10***	0.10***	0.09***	0.12***
	(0.03)	(0.03)	(0.02)	(0.03)
Strata and Week Fixed Effects	X	X	X	X
Subject-week observations	6,821	6,821	8,333	8,138
Number of subjects	359	359	641	626
Adjusted R-squared	0.17	NA	0.07	NA
Pvalues test of equal effects -- incentive vs. incentive + commit:				
Pretreatment	0.46	0.46	NA	NA
Intreatment (weeks 1-4)	0.72	0.65	0.15	0.13
Posttreatment (weeks 5-13)	0.03	0.03	0.02	0.01

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

Estimates in columns 2 and 4 for Probit model give marginal effects.

Table 3. Regression results -- Total Weekly Visits to the Gym

Dependent variable: Number of visits subject made to the gym that week

	Members		Non-Members	
	(OLS)	(Tobit)	(OLS)	(Tobit)
Control mean of dep var in pre-period	1.80	1.80	0	0
Incentive only	-0.19	-0.19	-	-
	(0.16)	(0.16)		
Incentive + Commit	-0.03	-0.02	-	-
	(0.16)	(0.16)		
Intreatment period (weeks 1-4)	0.14	0.11	-	-
	(0.13)	(0.13)		
(Incentive only) x (Intreatment)	0.86***	0.85***	0.38***	0.52***
	(0.13)	(0.13)	(0.07)	(0.09)
(Incentive + Commit) x (Intreatment)	0.94***	0.88***	0.53***	0.59***
	(0.13)	(0.12)	(0.08)	(0.09)
Posttreatment (weeks 5-13)	0.20	0.18	0.06*	0.18***
	(0.14)	(0.14)	(0.03)	(0.06)
(Incentive only) x (Posttreatment)	0.15	0.15	0.09*	0.17**
	(0.11)	(0.11)	(0.05)	(0.08)
(Incentive + Commit) x (Posttreatment)	0.30***	0.33***	0.20***	0.30***
	(0.10)	(0.10)	(0.06)	(0.08)
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
Adjusted R-squared	0.25	NA	0.06	NA
Pvalues test of equal effects -- incentive vs. incentive + commit:				
Pretreatment	0.25	0.22	na	na
Intreatment (weeks 1-4)	0.53	0.84	0.13	0.11
Initial posttreatment (weeks 5-13)	0.16	0.09	0.08	0.02

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

Tobit models in columns 2 and 4 account for weekly visits observations censored between 0 and 5. For these models the table gives marginal effects on the predicted number of visits.

Table 4: Predicting Uptake of Commitment Contracts

Dependent variable: indicator for whether subject made a commitment contract

	Members		**Non-members with at least 1 visits during treatment**	
mean commitment takeup rate:	0.25		0.21	
	(OLS)	(Probit)	(OLS)	(Probit)
Male	-0.19**	-0.18**	-0.15	-0.15
	(0.08)	(0.07)	(0.14)	(0.13)
Married	-0.12	-0.11	-0.08	-0.10
	(0.10)	(0.09)	(0.15)	(0.14)
Has children	0.02	0.03	0.06	0.13
	(0.09)	(0.08)	(0.14)	(0.14)
Lbs over target weight (in 10s)	0.02	0.02	0.05**	0.04**
	(0.02)	(0.02)	(0.02)	(0.02)
College degree	0.06	0.05	0.03	0.03
	(0.08)	(0.08)	(0.13)	(0.12)
Lacks "intrinsic motivation"	0.03	0.03	-0.13	-0.14
	(0.09)	(0.08)	(0.14)	(0.11)
Low self-control for exercise	0.06	0.05	-0.16	-0.19
	(0.12)	(0.11)	(0.15)	(0.14)
Exercise below target in pre-period	-0.04	-0.02	0.07	0.06
	(0.10)	(0.10)	(0.18)	(0.15)
Expected exercise post < ideal	-0.003	-0.01	-0.02	-0.04
	(0.09)	(0.09)	(0.15)	(0.13)
Expected exercise post > pretreatment	0.01	0.003		
	(0.10)	(0.09)		
Medium frequency user pretreatment	0.06	0.05		
	(0.11)	(0.11)		
High frequency user pretreatment	0.11	0.10		
	(0.13)	(0.12)		
Observations	117	117	48	48
R-squared	0.09		0.20	

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Estimates for Probit models give predicted marginal effects of the independent variables on probability of making a commitment. See the text for further description of the variables used in this table

Table 5. Estimated posttreatment differences between incentive groups

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

	Members		Non-Members with at least 1 visits during treatment	
	(Probit)	(Probit)	(Probit)	(Probit)
Mean for incentive-only in weeks 5-13	0.55	0.55	0.31	0.31
Incentive + Commit	0.09**	0.06*	0.15**	0.11
	(0.04)	(0.03)	(0.07)	(0.07)
Number visits during treatment control		X		X
Strata Fixed Effects	X	X	X	X
Subject-week observations	2,376	2,376	747	747
Number of subjects	264	264	83	83
Commitment contract takeup rate	0.23	0.23	0.24	0.24
Implied IV estimate	0.39	0.26	0.63	0.46

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

Probit estimates give marginal effects of the independent variable on the probability of attending the gym in a given week. The implied IV estimate is simply a scaling that divides the treatment effect by the commitment contract takeup rate.

Table 6a. Substitution Analysis for Members

Panel A: All subjects

Dependent Variable:	Weekly visits Gym data	Weekly visits Gym data†	Weekly visits Survey data	Overall exercise Survey data
Incentive or inc + commit	0.79***	0.79***	0.49***	0.40**
	(0.15)	(0.16)	(0.17)	(0.18)
Observations	359	335	335	337
Mean control	1.59	1.68	2.26	3.25

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

Dependent Variable:	Weekly visits Gym data	Weekly visits Gym data†	Weekly visits Survey data	Overall exercise Survey data
Incentive or inc + commit	0.88***	0.93***	0.76***	0.76***
	(0.18)	(0.19)	(0.21)	(0.22)
Observations	209	190	190	192
Mean control	0.97	1.03	1.46	2.32

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

Dependent Variable:	Weekly visits Gym data	Weekly visits Gym data†	Weekly visits Survey data	Overall exercise Survey data
Incentive or inc + commit	0.60**	0.52*	0.03	-0.16
	(0.26)	(0.27)	(0.29)	(0.29)
Observations	150	145	145	145
Mean control	2.58	2.68	3.51	4.71

Table 6b. Substitution Analysis for Non-Members

Panel A: All subjects

Dependent Variable:	Weekly visits Gym data	Weekly visits Gym data†	Weekly visits Survey data	Overall exercise Survey data
Incentive or inc + commit	0.45***	0.53***	0.60***	0.38***
	(0.05)	(0.06)	(0.07)	(0.13)
Observations	641	571	571	572
Mean control	0.03	0.03	0.066	2.09

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

Dependent Variable:	Weekly visits Gym data	Weekly visits Gym data†	Weekly visits Survey data	Overall exercise Survey data
Incentive or inc + commit	0.48***	0.55***	0.65***	0.45***
	(0.06)	(0.06)	(0.08)	(0.15)
Observations	499	446	446	447
Mean control	0.003	0.003	0.030	1.58

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

Dependent Variable:	Weekly visits Gym data	Weekly visits Gym data†	Weekly visits Survey data	Overall exercise Survey data
Incentive or inc + commit	0.37***	0.46***	0.43**	0.11
	(0.14)	(0.16)	(0.19)	(0.31)
Observations	142	125	125	125
Mean control	0.11	0.11	0.20	3.96

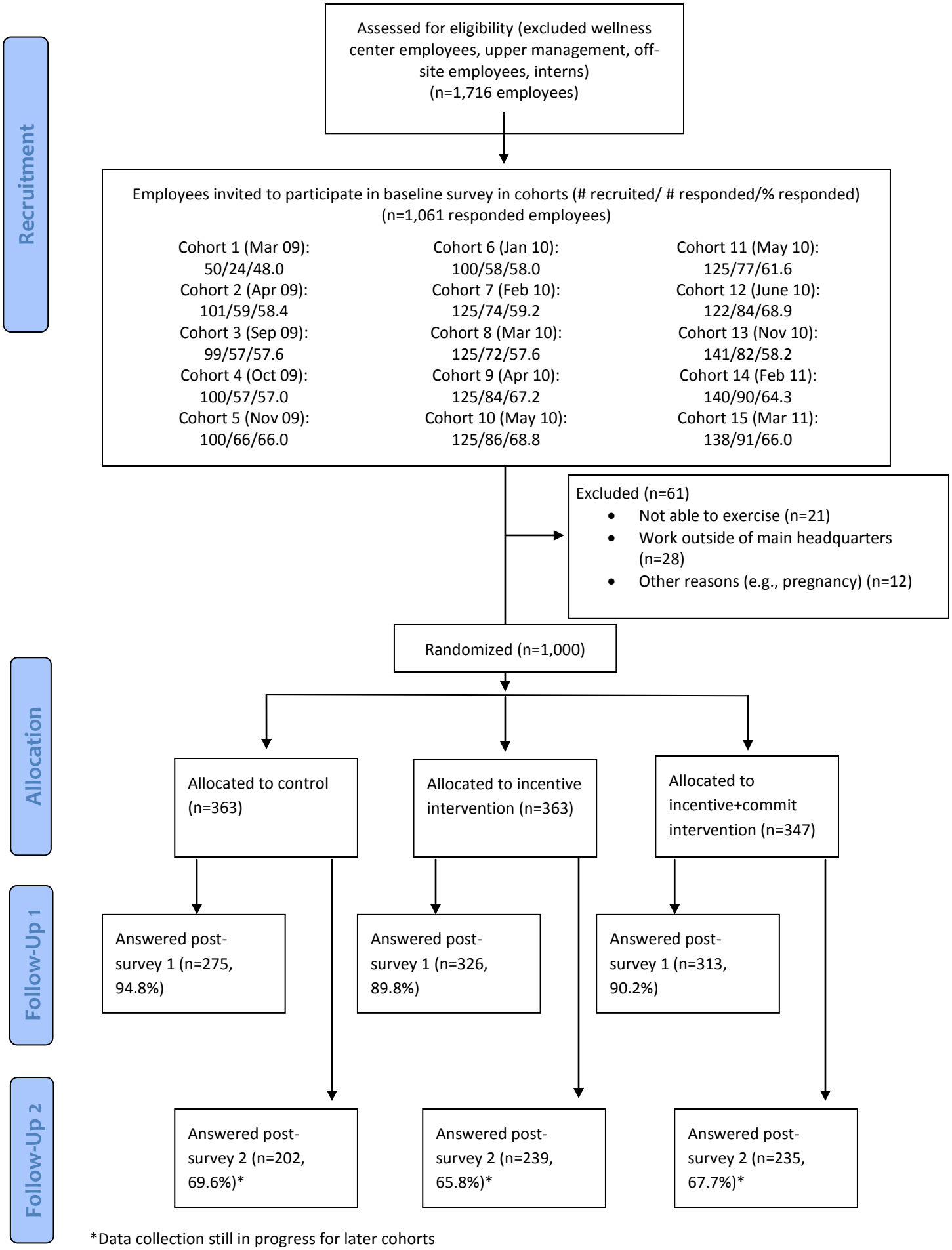
Notes: Dependent variable is average of weekly visits or exercise over the treatment period.
Includes dummies for week of the year. SE's clustered by subject.
† Just those subjects who answered weekly visits question in survey.

Table 7. Heterogeneity Cuts on Treatment Effects

		Members				Non-Members				
		W/O Controls		W/ Controls			W/O Controls		W/ Controls	
Heterogeneity Cut		**Incentive only**	**Incentive + Commit**	**Incentive only**	**Incentive + Commit**		**Incentive only**	**Incentive + Commit**	**Incentive only**	**Incentive + Commit**
Low preperiod exercise	**Intreatment period**	0.32***	0.28***	0.31***	0.30***	**Intreatment period**	0.14***	0.18***	0.12***	0.17***
	Control Mean = 0.07	(0.09)	(0.09)	(0.11)	(0.10)	Control Mean = 0.004	(0.04)	(0.04)	(0.04)	(0.04)
	Posttreatment period	0.07	0.16**	0.09	0.17**	**Posttreatment period**	0.003	0.07*	-0.005	0.06
	Control Mean = 0.07	(0.05)	(0.06)	(0.06)	(0.07)	Control Mean = 0.03	(0.03)	(0.04)	(0.03)	(0.04)
Mid preperiod exercise	**Intreatment period**	0.20***	0.22***	0.21***	0.22***	**Intreatment period**	0.15***	0.23***	0.15***	0.25***
	Control Mean = 0.67	(0.05)	(0.05)	(0.06)	(0.05)	Control Mean = 0.00	(0.04)	(0.05)	(0.05)	(0.06)
	Posttreatment period	0.05	0.18***	0.02	0.16**	**Posttreatment period**	0.04*	0.11***	0.05*	0.11***
	Control Mean = 0.58	(0.07)	(0.06)	(0.07)	(0.07)	Control Mean = 0.00	(0.02)	(0.04)	(0.03)	(0.04)
High preperiod exercise	**Intreatment period**	0.09*	0.10**	0.09*	0.10**	**Intreatment period**	0.14***	0.18***	0.15***	0.21***
	Control Mean = 0.88	(0.05)	(0.05)	(0.05)	(0.04)	Control Mean = 0.02	(0.04)	(0.05)	(0.04)	(0.05)
	Posttreatment period	-0.02	0.002	-0.02	-0.003	**Posttreatment period**	0.06	0.08**	0.06	0.09**
	Control Mean = 0.87	(0.05)	(0.04)	(0.05)	(0.04)	Control Mean = 0.03	(0.04)	(0.04)	(0.04)	(0.04)
Men	**Intreatment period**	0.28***	0.28***	0.26***	0.23***	**Intreatment period**	0.16***	0.17***	0.17***	0.17***
	Control Mean = 0.51	(0.07)	(0.07)	(0.05)	(0.06)	Control Mean = 0.01	(0.03)	(0.04)	(0.04)	(0.04)
	Posttreatment period	0.12*	0.15**	0.10*	0.07	**Posttreatment period**	0.04*	0.10***	0.05**	0.09***
	Control Mean = 0.50	(0.07)	(0.07)	(0.05)	(0.05)	Control Mean = 0.01	(0.02)	(0.03)	(0.02)	(0.03)
Women	**Intreatment period**	0.14**	0.16**	0.13***	0.12**	**Intreatment period**	0.13***	0.21***	0.13***	0.22**
	Control Mean = 0.61	(0.07)	(0.07)	(0.05)	(0.05)	Control Mean = 0.00	(0.03)	(0.04)	(0.03)	(0.04)
	Posttreatment period	-0.11	0.11	-0.10**	0.08	**Posttreatment period**	0.03	0.07**	0.02	0.07**
	Control Mean = 0.56	(0.06)	(0.07)	(0.05)	(0.05)	Control Mean = 0.03	(0.03)	(0.03)	(0.02)	(0.03)
No college degree	**Intreatment period**	0.23***	0.28***	0.22***	0.21***	**Intreatment period**	0.15***	0.26***	0.15***	0.26***
	Control Mean = 0.48	(0.08)	(0.08)	(0.06)	(0.06)	Control Mean = 0.01	(0.04)	(0.05)	(0.04)	(0.05)
	Posttreatment period	0.04	0.17***	0.06	0.10**	**Posttreatment period**	0.06**	0.17***	0.06**	0.16***
	Control Mean = 0.45	(0.07)	(0.06)	(0.05)	(0.04)	Control Mean = 0.01	(0.02)	(0.04)	(0.02)	(0.04)
College degree	**Intreatment period**	0.15***	0.16**	0.19***	0.19***	**Intreatment period**	0.15***	0.16***	0.16***	0.16***
	Control Mean = 0.65	(0.06)	(0.07)	(0.05)	(0.05)	Control Mean = 0.01	(0.03)	(0.03)	(0.03)	(0.03)
	Posttreatment period	-0.03	0.09	0.02	0.13***	**Posttreatment period**	0.01	0.04	0.02	0.03
	Control Mean = 0.62	(0.06)	(0.06)	(0.04)	(0.04)	Control Mean = 0.04	(0.02)	(0.03)	(0.03)	(0.03)

Notes: Cells in the table give coefficients from linear regression models of treatment effects for indicated treatment with standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Intreatment period is weeks 1-4 with incentive in place. The posttreatment 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.

Appendix Figure 1. Flow Diagram



*Data collection still in progress for later cohorts