

NBER WORKING PAPER SERIES

THE STRUCTURE OF HEALTH INCENTIVES:
EVIDENCE FROM A FIELD EXPERIMENT

Mariana Carrera
Heather Royer
Mark F. Stehr
Justin R. Sydnor

Working Paper 23188
<http://www.nber.org/papers/w23188>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2017

This study was funded by a grant from the Robert Wood Johnson Foundation through the Applying Behavioral Economics to Perplexing Problems in Health and Health Care initiative via the Center for Health Incentives and Behavioral Economics at the University of Pennsylvania and from a UCSB Faculty Senate Grant. We greatly appreciate the comments and suggestions of John Cawley and Gary Charness. We thank Chang Lee, Ben Marks, Alan Thomas, Sarah Mattson, Steven Cramer, Leah Neustadt, and Jacob Rubin for excellent research assistance, and various seminar and conference participants for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Mariana Carrera, Heather Royer, Mark F. Stehr, and Justin R. Sydnor. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Structure of Health Incentives: Evidence from a Field Experiment
Mariana Carrera, Heather Royer, Mark F. Stehr, and Justin R. Sydnor
NBER Working Paper No. 23188
February 2017
JEL No. D03,I10,J30

ABSTRACT

The use of incentives to encourage healthy behaviors is increasingly widespread, but we have little evidence about how best to structure these programs. We explore how different incentive designs affect behavior on the extensive and intensive margins through an experiment offering incentives to employees of a Fortune 500 company to use their workplace gym. Overall the likelihood of joining the gym was not strongly affected by the incentive design. Notably, front-loading incentives to encourage initial participation was not more effective than an incentive kept constant over time. For those who were already at least occasional users of the gym, however, we find more evidence that the design of incentives matters. For this group, front-loading incentives appears to be detrimental relative to a constant incentive, but a novel design that spreads out the incentive budget by turning incentives on and off over a longer period of time is effective.

Mariana Carrera
Case Western Reserve University
2452 Kenilworth Rd.
Cleveland Heights, OH 44106
mpc67@case.edu

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

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

Justin R. Sydnor
Wisconsin School of Business,
ASRMI Department
University of Wisconsin at Madison
975 University Avenue, Room 5287
Madison, WI 53726
and NBER
jsydnor@bus.wisc.edu

Firms are increasingly using targeted financial incentives to encourage workers to improve their health behaviors. According to the Kaiser Family Foundation Employer Benefits Survey, in 2015 81% of large firms offered a wellness program, and 38% of those programs offered their employees monetary incentives for participation.¹ The Affordable Care Act also encourages the use of health incentives by allowing up to 30 percent of the cost of health care coverage to be allocated to wellness program rewards.²

A burgeoning academic literature explores how people respond to financial incentives to quit smoking, lose weight, eat healthier or exercise more (Finkelstein et al, 2007; Volpp et al., 2008; Charness and Gneezy, 2009; Volpp et al., 2009; Babcock and Hartman, 2010; Gneezy, Meier, and Rey-Biel, 2011; John et al., 2011; Cawley and Price, 2013; Just and Price, 2013; Milkman, Minson and Volpp, 2013; Acland and Levy 2015; Babcock et al., 2015; Halpern et al., 2015; Royer, Stehr and Sydnor, 2015; Loewenstein, Price and Volpp, 2016; Patel et al 2016; Mochon et al., 2017). These studies show that incentives can meaningfully affect behavior while they are in place and in a few cases have even found persistent effects once incentives are removed (e.g., Charness and Gneezy, 2009; Loewenstein, Price and Volpp, 2016; Mochon et al., 2017).

Relatively little research, though, has explored how, given a fixed budget, incentives should be structured to best promote behavior change. In particular, an important open question is what sort of tradeoffs different incentive designs create for affecting behavior on the intensive and extensive margins. For example, for behaviors like exercise, motivating people who are not engaged in the behavior at all (i.e., extensive margin) may involve overcoming high start-up costs. Models of present bias suggest that in these situations it might be beneficial to combat

¹ Web link: <http://kff.org/report-section/ehbs-2015-summary-of-findings/>

² See Federal Register Doc No: 2013-12916.

procrastination by using sharp incentives for immediate action (O'Donoghue and Rabin, 1999). However, for people who are already engaging in an activity but perhaps less than desired, it may be more effective to focus on maintaining the power and salience of incentives over time. This may be particularly true if a period of sustained activity leads to habits that promote continued behavior change even once incentives are removed.

We explore these issues using a randomized field experiment incentivizing exercise at the on-site company gym for 980 employees of a Fortune 500 company. A number of features of this setting are useful for exploring incentive design for motivating behavior on the intensive and extensive margins. First, at this company employees must join the fitness facility, creating a natural division between employees who need to be motivated on the extensive margin (non-members) versus those for whom an incentive program targets the intensive margin (members).³ Second, there is substantial variation in ex-ante gym use among members, and a large fraction of the existing members state in surveys that they wish they visited more often. Finally, there are computerized records of the visits employees make to the gym, which allow us to track visit patterns prior to, during and after the incentive period.

The two primary treatment arms for the experiment contrast a front-loaded incentive versus a consistent incentive. These two treatments, which we refer to as the “kickstart” and “constant” treatments respectively, offered employees the same maximum possible earnings (\$160) over an 8-week incentive period but differed in how the earnings were allocated over time. The constant treatment offered an incentive of \$10/visit for up to two visits each week throughout the incentive period. In contrast, the kickstart treatment front-loaded incentives,

³ Of course, several non-members (at least 20 percent) belong to other gyms and thus, likely do not need motivation on the extensive margin.

offering \$25/visit in the first two weeks followed by \$5/visit in the remaining six weeks.⁴ We anticipated that the front-loaded kickstart design might be effective at overcoming procrastination problems and hence might help encourage participation at the extensive margin for non-members. On the other hand, large incentives early on with smaller incentives later might be less effective at sustaining motivation for existing members over the course of the incentive period. A comparison of the kickstart and constant treatments allows us to better understand the nature of these tradeoffs.

Our primary finding is that the front-loaded kickstart program failed to deliver stronger effects on the extensive margin while also delivering somewhat weaker effects on the intensive margin. The kickstart treatment increased join rates for non-members by ten percentage points relative to control, but the constant incentive was equally effective. Among those who were already gym members, both the kickstart and constant incentives generated economically and statistically significant increases in visit rates, but the effects for kickstart were somewhat weaker. Beyond the incentive period, those assigned to the constant treatment continued to have a meaningful increase in gym attendance relative to the pre-intervention period, but the kickstart-incentive group exhibited no persistent effect. Overall, the tradeoff we anticipated between these programs did not emerge: as expected, the effects of the constant incentive on the intensive margin were somewhat stronger, but more surprisingly, there was no benefit, even in the two weeks with high incentives, from the front-loading of the kickstart design on extensive margin behavior.

These results contribute to a small set of studies exploring how behavioral factors interact with the design of incentives for health-behavior change. In particular, there is some evidence that using lottery and deposit-based designs that target behavioral phenomena such as probability

⁴ As in the constant incentive, employees in the kickstart could earn the incentives for up to two visits each week.

weighting and loss aversion can be effective (Volpp et al., 2008; Haisley et al., 2012; Halpern et al., 2015; Patel et al., 2016). Our results contrast somewhat with these positive findings by showing that a front-loaded design to create a sharp incentive to overcome procrastination is not, in fact, more effective at encouraging participation. Our results are, however, broadly consistent with a concurrent working paper exploring the structure of incentives by Bachireddy et al. (2017), which finds that a constant incentive more effectively encouraged steps by participants with pedometers than continuously increasing or decreasing incentives with the same overall possible earnings.

Our primary findings suggest that it may be more effective to structure incentive programs with an eye toward sustaining behavior change over longer periods and generating habit formation than it is to use sharp incentives for generating initial behavior change. That takeaway message is bolstered by the findings of two secondary treatment arms in the study.

For the non-members, we offered a secondary treatment arm (“constant-short”) that used the same \$10/visit structure as the constant incentive but over 4 weeks instead of 8 weeks. We find that more non-members joined the company gym during the 8-week program than during the 4-week program. Hence, while front-loading the incentives (kickstart vs constant 8-week programs) did not increase participation on the extensive margin, the length of the program did have an effect. This tells us that non-members are not responding only to spot incentives. It further shows that longer programs not only lead to longer treatment effects but also increase the fraction of employees participating on the extensive margin.

The secondary treatment for the existing members offered the same \$10/visit structure for 8 weeks as in the constant treatment, but participants were incentivized for 8 randomly-selected weeks out of 16 weeks. The purpose of this novel “extended-sporadic” design was to test

whether short-run habit formation could be leveraged to generate consistent increases in activity over a longer period without increasing the budget for the program. Participants in this treatment group were informed each week whether the following week was incentivized, which created an incentive program where the incentives turn on and off sporadically over the course of 16 weeks.⁵ The extended-sporadic incentive generated a treatment effect between 60%-100% of the incentive-period effect of the constant incentive but spread over 16 weeks instead of 8 weeks. The results for this incentive design suggest it may be possible to leverage short-run habit formation to encourage prolonged periods of elevated exercise without needing to increase the budget for incentive programs.

The design of this study is closely related to a series of experiments on incentives for attending a gym that began with the pioneering work of Charness and Gneezy (2009) and was followed by closely related studies by Acland and Levy (2015) and Royer, Stehr, and Sydnor (2015). Both the studies by Charness and Gneezy and Royer et al., examine effects of a single primary incentive treatment separately for those who were and were not previously using the fitness facility (i.e., intensive vs extensive margins). The primary contribution of this study is to explore the effects of different incentive designs, including the kickstart program that was expressly designed to test the benefits of front-loading incentives for motivating participation on the extensive margin.

Our study also adds additional evidence consistent with original findings of Charness and Gneezy (2009) on a couple of dimensions. First, we find that providing financial incentives does not appear to lead to a crowd out of intrinsic motivation once incentives are removed. Moreover, for those who already exceeded our two-visit-per-week target, we find no detrimental effect on

⁵ This is distinct from the lottery design of Volpp et al. (2008) where meeting the target activity (daily weight loss) made the participant eligible for a lottery with a 1 in 100 chance of a \$100 reward and a 1 in 5 chance of a \$10 reward.

behavior. Thus, there were no unintended consequences for this population related to setting a low norm or target. Finally, we find meaningful habit-formation effects after the end of the temporary 8-week constant incentive program among existing members of the gym. Royer, Stehr, and Sydnor (2015) found weaker post-treatment effects from a similar 4-week program offered to this same population and it seems plausible that the difference may stem from the longer incentive program used in this study. Overall, both this finding of habit formation after the 8-week program and the inter-week habit formation seen in the extended-sporadic treatment help to underscore the potential value of designing incentive programs in ways that both promote and leverage the establishment of habits.

2. Experimental Design

2.1 Setting

Our experiment took place at the headquarters of a Fortune 500 company in the Midwest. Roughly 2000 employees work at this location, ranging from the call center to high-level management. The experiment operated in coordination with an onsite fitness facility, which is open 6 am to 6 pm Monday through Friday. The gym offers classes and standard aerobic and strength training machines. Gym membership is subsidized and costs employees \$12.96 each 2-week pay period.

When employees enter the gym, they type their company identification number into one of two computer terminals inside the gym's main door entrance. These computerized login records provide high-quality data on visits made to the gym and provide both the data on which program incentives are based and our primary outcome measures. Importantly, participation in the study did not require employees to do anything different to log their visits. As is standard in

many gyms, however, employees are not required to log out when they leave. As such, we do not have data on the length of time employees spend at the gym during a visit or on precisely what they do when they are there. We did not attempt to alter the check-in/out process at the gym because we wanted to keep the environment natural and worried that any new procedures might lead to differential compliance based on treatment status.

The reliance on check-in records may raise concern that gym visits do not necessarily correspond to exercise since employees could earn incentives by visiting the gym without working out. We are not very concerned about this type of behavior for several reasons. First, there is a full-time staff at the gym who interact frequently with employees and monitor activity at the gym. This staff reported no changes in these types of activities or a rise in employees entering the gym without exercising. Second, Royer, Stehr, and Sydnor (2015) offered similar-sized incentives at this same site two years prior to this study and employed research assistants to unobtrusively monitor the gym on multiple days. The research assistants recorded no instances of employees checking into the gym without exercising. Finally, analysis of our follow-up surveys indicates that a visit to the company gym was associated with around 45 minutes of exercise at the gym. A full description of this analysis appears in the Appendix.

2.2 Recruitment and Treatment Assignment

We recruited subjects from a list of all employees at the firm's headquarters by sending an email that contained a link to an online survey on wellness.⁶ The email informed participants that they could earn a \$25 gift card for completing both this initial survey and a short follow-up

⁶ We exclude gym staff, executives, and human-resource staff who might have been aware of the details of this program. A copy of the survey is available upon request and will be included in an online appendix.

survey to be administered 8 weeks later. Gift cards were used throughout the study to simplify the distribution of incentives. We gave subjects a choice of gift cards to make them as fungible and close to cash as possible; the options included Amazon, Target, Wal-Mart and several local gas stations.

Respondents to our initial survey formed our subject pool for randomization, provided that they worked at the corporate headquarters and were physically able to exercise. We ran the experiment in 15 cohorts to ensure that it did not overly tax the gym and gym staff. We enrolled the first cohort in June 2013 and the last cohort in April 2014. The closed nature of the corporate setting and the overlapping cohorts raise the possibility for cross-talk between participants. We discuss this issue in Section 4 and provide evidence that suggests communication between employees did not bias treatment effects.

Individuals were randomized into the treatments detailed in Table 1 and the randomization was stratified by whether they belonged to the company gym at the time of the initial survey. Both members and non-members were randomized into control, constant treatment and kickstart treatment. We chose to incentivize up to 2 visits per week with only 1 visit per day eligible for incentives. Setting a cap higher than this would have shifted much of our incentive costs to inframarginal individuals (e.g., regular exercisers). Two visits per week however still allowed for incentivizing substantial changes in exercise given that among existing members, 65 percent attended 2 or fewer days per week prior to the study.

In addition to these two primary treatments, we offered two other targeted incentives to members and non-members separately. For the existing members this additional treatment, extended-sporadic, offered the same \$10/visit incentive as the constant was for 8 weeks but randomized the 8 weeks in which the incentive was received over a 16-week period. Each week,

participants in this group received an email telling them whether the following week would be an incentivized week. The secondary treatment for the non-members offered them a \$10/visit incentive but only over a 4-week period.⁷ This constant-short incentive is the only one of the incentive schemes to have smaller total possible earnings – specifically half of the \$160 maximum incentive earnings for the other programs. Finally, the non-members received an additional \$25 payment if they decided to join the gym during the incentive period, which was designed to offset the additional burdens they face in participating, including the need for a fitness assessment taking place at the gym as part of the gym-joining process.

This experiment took place in the same location as Royer, Stehr, and Sydnor (2015), which offered 4-week incentives to employees between February 2009 and March 2011. Roughly 78 percent of the employees in the company at the start of this experiment were employed at the company at the end of that earlier study, and although IRB regulations restrict us from matching the subject pools across experiments, some of the participants in the earlier study surely enrolled in this study. We think, however, that the effects of the earlier study on our findings here are likely minimal as the gap between the two experiments is two years. Additionally, the internal validity of our study is not compromised because treatment assignment should be orthogonal to previous participation because of randomization.

The online follow-up survey collected self-reported measures of exercise type and duration to help us understand exercise patterns at and away from the company gym. Additionally, when picking up their earnings, individuals could earn \$25 for a brief self-

⁷ We began the experiment assigning non gym-members to the extended sporadic treatment. After three cohorts, we received anecdotal reports that non-members were reluctant to join under this program because the incentives stretched over 16 weeks, requiring more gym-membership fees, while only incentivizing 8 of those weeks. Because of that difference in the nature of the effective treatments between the constant and extended sporadic incentive for the non-members, we discontinued allocating non-members to this treatment. We instead switched our secondary treatment for this group to be a 4-week incentive at the \$10/visit rate.

administered health assessment. Health measures included blood pressure, resting pulse, percent body fat and weight.

Appendix Figure A1 summarizes the flow of subjects by treatment status. Of the 1759 employees we contacted, 1075 or 61% took the first survey and consented to participate in the study.⁸ Of those, 95% or 1024 were randomized into treatment. Of the 1024 subjects who were randomized to treatment, the final analysis sample includes 980 subjects with useable data.⁹ Of those 980, 845 or 86% filled out the follow-up survey and 702 or 72% recorded their health measures. As we discuss in the Appendix, the lower response rate to the health measures creates scope for selection problems. The analysis in the primary paper, however, is focused on gym-attendance records, for which selection and attrition are not concerns.

Our final sample of subjects, 980, is fairly sizable. We based our power calculations on detecting in-treatment effects on gym attendance. Since ex-ante we expected that the differences between the control group and treatment groups would exceed those between the different treatment groups, we assigned more individuals to each of the treatment groups than to the control group. At 80 percent power, ex-ante, we had sufficient power to detect a 0.12 difference in the probability of visiting the company gym in a given week between any two treatment groups and a 0.09 difference in this probability between the control group and any treatment group given baseline attendance rates and results from Royer, Stehr, and Sydnor (2015). We anticipated being underpowered to detect any differences in health measures based on treatment

⁸ This includes the individuals randomized plus those filling out the survey but not meeting the inclusion criteria.

⁹ The number of subjects decreased to 1024 when we excluded those not medically able to exercise or who did not work at the site where the company gym is located. The number of subjects drops to 980 when we exclude those non gym-members in cohorts 1-3 who were assigned to the extended-sporadic treatment, which as noted in the preceding footnote was replaced with the constant-short treatment beginning with cohort 4.

status and focus our analysis here on differences in visit patterns. We discuss the health measures in the Appendix.

2.3 Descriptive Statistics

Table 2 presents descriptive statistics on the sample population, showing the overall mean and the mean for the control group for a number of characteristics measured with the initial survey. The table also provides a check that the randomization resulted in balance by providing the difference in these values for each treatment group relative to control and the p-value for a joint test that the mean of each variable is the same for all groups.¹⁰ The table is split into two panels to separate participants by their ex-ante company gym membership status, which we stratify our randomization on. Overall, the randomization is balanced on observable characteristics. Only one of 26 p-values testing whether the pre-treatment means are equivalent across groups is at or below 0.05.

Both the samples of members and non-members are roughly equally split along gender lines. The sample is well-educated; over 70 percent have college degrees. Rates of overweight and obesity are high, although our overall rate of 62 percent is a bit below the overall national average of 70.7 percent. The slightly lower rate in our sample is not surprising given the negative correlation between education and rates of overweight/obesity. Interestingly, despite different rates of exercise, members and non-members have similar rates of overweight/obesity.

¹⁰ The differences are regression adjusted to control for cohort fixed effects.

In the initial recruitment survey, we inquired about individual exercise patterns – overall, at the company gym, at any other gym, and outside of the gym.¹¹ Such data are useful for assessing how well a company gym based intervention might work for changing exercise behaviors. For members, most of their exercise takes place at the company gym (average of 2.3 days/week out of 3.57 days/week). Non-members exercise less: 2.24 days/week, with 74% of this exercise occurring outside of a gym and 26% occurring at other gyms, outside the workplace. Non-members and members alike appear to have a desire to increase their exercise as their target level of exercise exceeds their actual exercise, with non-members reporting a lower subjective probability of reaching their self-chosen targets.

3. Results

3.1 Extensive Margin: The Participation Decision of Non-Members

We begin the discussion of our empirical results by examining incentive effects among non-gym members. As described above, these employees were randomized into a control group, a constant incentive of 8 weeks' duration, a kickstart incentive of 8 weeks' duration, or a constant-short incentive of 4 weeks' duration.

To measure the effects of these incentives on the decision to join the gym and begin exercising there, we define *Joined* as a binary variable equal to 1 if an employee appears in the gym visit records at least once in the eight-week incentive period. We estimate OLS regressions that include treatment-group dummies and controls for observable characteristics. Since subjects

¹¹ The overall exercise question was asked separately of the other three (i.e., it is not a sum of the other three).

were recruited and randomized over 15 different cohorts, we also include dummies for cohorts indexed by j . The regression equation is:

$$y_{it} = \alpha + \delta_1 D_i^{constant} + \delta_2 D_i^{kickstart} + \delta_3 D_i^{constant-short} + X_i B + \sum_{j=1}^{14} \mu_j I_i^j + \epsilon_{it}, \quad (1)$$

where D_i^x are the 3 treatment indicators (constant, kickstart, and constant-short) and I_i^j are binary indicators for the cohort. The controls included in X_i are male, age, married, a set of dummies for educational attainment, and a set of dummies for level of interest in joining the company gym. All of these variables were collected in the initial survey, prior to treatment assignment.¹²

Table 3 displays the results. Column 1 shows that in the control group 4.6% of employees became members during the study period despite the fact that they received no financial incentives to do so. This is not surprising since some non-members naturally join the gym even in the absence of the experiment. The constant and kickstart effects on the probability of joining are approximately equal effects of increasing the probability of joining the gym by 10 percentage points relative to the control group. In levels, this implies that the join rate was approximately 3 times higher for these two treatment groups relative to control. The constant-short incentive works less well; the point estimate is 0.032 and statistically insignificant. The p-value of a Wald test for the equality of the effects of the constant and constant-short incentives is 0.059.

Since the constant-short incentive offered the same week-by-week incentives as the constant incentive, but for only four rather than eight weeks, these results suggest that non-members are considering the value of the incentive program as a whole when deciding whether

¹² We include the controls to help improve the precision of the estimates. Due to the randomization of treatment, we do not expect controls to affect the estimated treatment effects, and indeed we find only small differences in treatment effects if we instead exclude the additional controls.

to participate. This is consistent with the notion that fixed costs pose a barrier that makes the extensive margin participation decision differ from an intensive margin response to spot incentives.

Interestingly, though, we do not find evidence consistent with the possibility that naïve present bias limits participation in these programs due to procrastination. The similar effects of the constant and kickstart incentives, which offer the same total possible earnings but very different incentives in the first two weeks, suggest that potential participants are not heavily discounting over the eight-week program period when considering the likely gains from participating. Taken together with the findings on the constant-short program, this suggests that the decision to participate is based on a consideration of the full value of the programs without strong discounting that would make the exact timing of incentives important to the participation decision.

Of course, the overall effect of the program among non-members may not be the only relevant concern. In particular, companies might be especially interested in how different incentive programs motivate participation for those who are both not using the company gym and not exercising much elsewhere. In columns (2) and (3) of Table 3, we show the estimates of Equation 1 for two subsamples of non-members: 1) those who reported exercising two or fewer days in the week prior to the initial survey and reported that they typically exercise two or fewer days per week (“low exercisers”), and 2) those who reported three or more days in the week prior to the survey and reported that they typically exercise three or more days per week (“high exercisers”). This split was not pre-specified, so it is important to have some caution in

interpreting these results. However, we believe this cut of the data is natural for heterogeneity considerations and can provide useful suggestive evidence.¹³

For the low exercisers, we can clearly reject the hypothesis that constant and constant-short incentives had the same effect ($p=0.01$). Among those who reported exercising three or more days, the treatment effect estimates are remarkably similar for all three incentive groups, and we cannot reject the hypothesis that constant and constant-short incentives had the same effect ($p=0.77$). In both subsamples, there is no notable difference in the estimated treatment effects of the kickstart and constant incentives.

To summarize, we find that the difference in the extensive margin effects of the constant and constant-short treatments is driven by low exercisers. This implies that the program duration and the number of incentivized visits are important factors for motivating the participation of people who are interested in starting a new exercise regimen but typically find it difficult to do so. Front-loading the earnings did not increase participation rates among this subpopulation, which is surprising because they score lower on a self-control scale in the pre-survey and might have been expected then to be more responsive to the immediate rewards of the kickstart treatment.¹⁴ In contrast, those who exercise frequently and without difficulty appear less concerned with the total value and duration of an incentive program, perhaps because all of our incentive programs appeared to them to offer “easy money.”

¹³ Our categorization of low and high exercisers takes advantage of two different measures of self-reported exercise collected in the initial survey, one about days of exercise last week, and one about exercise frequencies in a typical week. For both measures, the median value reported is 2. Our findings are very similar if we instead use either measure alone to categorize exercise frequency.

¹⁴ Daly, Harmon and Delaney (2009) find the Tangney, Baumeister and Boone (2004) self-control scale to be correlated with experimentally measured discount rates of financial rewards. Based on the 9-question version of this self-control scale that we included in the survey, non-members in the low-exercise group scored on average 0.63 standard deviations lower than non-members in the high-exercise group ($p<0.001$).

3.2 Intensive Margin: Treatment Effects for Members

We now compare the effects of the three incentive structures given to existing gym members, which is the group for which incentives were expected to act primarily on the intensive margin. For this group, while all three incentive designs had the same budget, the possible earnings were spread out in three different ways as described in Table 1.

We begin our analysis of the treatment effects along the intensive margin by examining the treatment effects on two measures of visit rates: *Any Visit*, a binary variable indicating at least one visit in a given week, and *Number of Visits*, the number of days with a visit in a given week. We organize the analysis into two periods – weeks 1-8 when all incentive treatments were in play and weeks 9-16 when only the extended-sporadic group was eligible for incentives. Our regressions to estimate the treatment effects take the following form:

$$y_{it} = \mu_i + \sum_{k=1}^2 P_{i,t}^k (\gamma_k + \delta_{1,k} D_i^{constant} + \delta_{2,k} D_i^{kickstart} + \delta_{3,k} D_i^{extended-sporadic}) + \varepsilon_{it} \quad (2)$$

where y_{it} is the person-by-week visit measure, D_i^x are the 3 treatment indicators (constant, kickstart, and extended-sporadic), and $P_{i,t}^k$ are two time-period indicators (weeks 1-8 and weeks 9-16, respectively). The parameters $\delta_{1,k}, \delta_{2,k}, \delta_{3,k}$, our coefficients of interest, estimate the treatment differences for each of the treatment groups (constant, kickstart, and extended-sporadic respectively) relative to control in each of the time periods. The parameters γ_k estimate the visits rate differences for the control group in these two different time periods relative to their pre-treatment visit rates. As such, our main treatment effects ($\delta_{1,k}, \delta_{2,k}, \delta_{3,k}$) are essentially

difference-in-difference estimates. The μ_i are individual fixed effects, which are included to improve the precision of the estimates and control for subject-level differences in pre-treatment visit rates. To give a sense of the underlying data, we also provide smoothed-averages of *Any Visit* for each group over time in Figure 1. The averages come from LOESS estimation.

Table 4 displays estimates of this regression using the outcomes *Any Visit* and *Number of Visits* in columns (1) and (3) respectively. All three of the incentive schemes have strong effects on gym attendance among members, albeit with different levels of success. In weeks 1-8, the constant incentive elevated attendance rates by 18 percentage points from a base of 62 percent (a 29 percent increase). The point estimates for kickstart and extended sporadic are smaller, with increases in *Any Visit* of 11-12 percentage points. All of the incentive programs show virtually identical average treatment effects when the outcome is the number of visits (around 23% increases in average visit rate).

The results from weeks 9-16 provide some evidence of habit formation in the constant group, at least for the any-visit measure. This group showed a statistically significant 9 percentage point increase in *Any Visit* which is half the size of the treatment incentive effect. The kickstart group, however, exhibits no long-run effects.

Overall, taking into account both the in-treatment and post-treatment effects, the constant incentive scheme outperforms the kickstart incentive. The average difference in the treatment effects on any visits between constant and kickstart across weeks 1-16 is 0.075 (p-value 0.025). Even in weeks 1 and 2, when the kickstart incentive (\$25/visit) exceeds that for the constant (\$10/visit), the patterns of usage are similar (see Figure 1).

The extended sporadic exhibits a more stable and consistent effect over the 16 week incentive period than the other two incentive treatments. The rate of visits for the extended sporadic suggests that this scheme may effectively leverage some combination of inter-week habit formation and increased salience of incentivized weeks. Specifically, if the treatment effect for the extended sporadic were zero in non-incentivized weeks and equal to that of the constant group in incentivized weeks, the extended sporadic treatment effect would be one-half of the constant treatment effect. However, the extended sporadic treatment generates a stronger treatment effect than this benchmark of half of the effect for twice as long. In weeks 1-8, the extended sporadic effect is 60 percent to 100 percent the size of the constant effect for *Any Visit* and *Number of Visits* respectively. In weeks 9-16, when the constant and kickstart groups no longer earn incentives but the extended sporadic group continues to be incentivized, gym use remains elevated in the extended sporadic. In fact, the extended sporadic group's attendance is higher relative to the control during weeks 9-16 than during weeks 1-8, indicating little if any depreciation of the effect of incentives.

Columns (2) and (4) of Table 4 delve more deeply into the timing of the sporadic treatment effects by adding to the regression model interactions between the sporadic incentive indicator and dummy variables for whether or not week t was incentivized for member i . The incentivized-week effects for the extended sporadic group during both weeks 1-8 and 9-16 are roughly as strong as those for the constant group during its incentive period of weeks 1-8. For the number of visits measure, the extended sporadic effects actually exceed the constant effects. This dampens any concern that the sporadic nature of the incentives reduced their salience and if anything, suggests instead that sporadic incentives may slightly strengthen the response to incentivized weeks. During the non-incentivized weeks for the extended sporadic, there is a

positive treatment effect of 0.065 to 0.11 on *Any Visit*, depending on the time period, and for weeks 9-16, this effect is statistically significant. This treatment effect is roughly the size of the constant incentive's effect for weeks 9-16, which we attribute to habit formation.

Comparing across 16 weeks, the constant incentive increases *Any Visit* by 0.17 and *Number of Visits* by 0.48 whereas the similar figures for the extended sporadic are 0.16 and 0.55, respectively.¹⁵ Although the point estimates for the sporadic appear higher than those for the constant in many comparisons, these differences are never statistically significant. Thus, overall, we cannot reject that the two incentive schemes perform equally well. Taken together, the results for the sporadic incentive suggest that this treatment can mostly be thought of as spreading out the short-run habit formation we see for the constant incentive.

3.3 Heterogeneity for Members by Ex-ante Attendance Level

In this subsection, we explore how the incentives affected existing members based on their pre-intervention visit frequency. In particular, for those regularly attending the company gym twice or more in a week, the incentive program may not lead to increases in gym attendance because the incentives paid for at most two visits per week. In fact, the incentive effect could even be negative if setting the incentivized level below their usual attendance sets an implicit target and reduces their intrinsic motivation to exercise beyond the targeted level. On the other hand, as was seen in Charness and Gneezy (2009), less frequent exercisers – those attending less than twice a week – may increase their gym-attendance more dramatically than seen for the overall member population in Table 4.

¹⁵ These figures come from separate regressions when combine the week 1-8 and 9-16 periods.

We test these hypotheses in Table 5 where we divide members into two groups: 1) those who exercised 2 or more times per week for at least half of the weeks from week -9 to week -4, and 2) those who exercised 2 or more times per week for less than half of the weeks from weeks -9 to -4. In this sense, the “high-exercise” group consists mainly of those who are already exercising 2 or more days a week and for whom the incentives should have little impact on their behavior. Columns (1) and (2) provide the estimates with *Any Visits* as the dependent variable and columns (3) and (4) use the *Number of Visits* as the dependent variable.

As expected, the ex-ante low-use members are responsible for the significant treatment effects in Table 4. For the low-use members, the effects of all incentive schemes for weeks 1-8 are large and statistically significant at the 5 percent level. In contrast, fewer of the effects (4 out of 6) are statistically significant at the 5 percent level for high-use members. For the high-use members, the concern that the treatment effects might be negative if financial incentives crowd out intrinsic motivation turns out to be unjustified.¹⁶ Consistent with the findings of Charness and Gneezy (2009) and Sen et al. (2016), financial incentives do not appear to harm intrinsic motivation in this setting or to otherwise create a negative target effect for those who are already heavily engaged in exercise.

3.4 Intensive Margin: Treatment Effects for Non-Members

Like we have done for members, we can also examine the intensive-margin responses for ex-ante non-members. This analysis is naturally weaker among the non-members since the

¹⁶ The only coefficients that appear significantly negative are those for number of visits for all time periods. This simply reflects mean reversion (recall that we have defined this subgroup as employees with above-median visits in the pre-period, which is the omitted period from the regression). Mean reversion can also explain why the trend in the control group’s gym visits was downward for members as a whole, since some current members might have quit their memberships during our study period but remained classified as members within our study.

majority of them did not join the gym (i.e., had no extensive-margin response) and hence by definition will have zero visit rates.¹⁷ Nonetheless, it is instructive to consider how the visit patterns evolved for non-members through the incentive periods.

Our regressions follow the same basic format as those for members except our division of time periods is different (3 time periods: weeks 1-4, 5-8, and 9-16) and we include the controls we used earlier for the extensive margin responses for non-members. Figure 2 shows the time trend averages in any visit measure for the four groups. The averages again come from LOESS estimation.

We display the regression results in Table 6. Column (1) provides the results with any visit as the dependent variable whereas column (2) shows the results with the number of visits as the dependent variable. The column (1) estimates are not remarkably different from those seen earlier for the extensive margin participation decisions. In weeks 1-4, kickstart and constant work equally well – increasing any visits margin by 0.11 and the number of visits by 0.25-0.28. After that period, the effects begin to diverge slightly. From Figure 2, it is evident that kickstart has a declining effect over time; the constant group’s patterns are flatter. The constant effect in weeks 5-8 is 0.10 whereas for the kickstart it is 0.07 for any visits. Unlike for members, the effects during the post-incentive period using the any visits measure for these two are roughly equal in magnitude; 0.036 for constant and 0.030 for kickstart.

Perhaps the most interesting comparison in this table is constant versus constant-short. Despite having the exact same incentive scheme in weeks 1-4, the treatment effects for constant are nearly twice as large as those for constant-short in these weeks. Although not statistically

¹⁷ An alternative approach would be to compare visit rates for non-members conditional on joining. However, given that the join decision is endogenous, this creates selection concerns in interpreting the comparison of visit rates for joiners across treatment and control.

significant, these differences are in line with the more pronounced differences for the participation decision discussed earlier. These results suggest it may be challenging to extrapolate findings from a shorter study to a longer study. It would be natural to surmise that doubling the length of a 5-week intervention might extend the 5-week effects over a 10-week period. Our findings caution against such a conjecture and suggest instead that the 10-week intervention would likely induce additional subjects to participate. Furthermore, our subgroup analysis in section 3.1 suggests that the marginal participants induced by a longer intervention might be those who find the incentivized behavior more challenging, and thus, may benefit more from the intervention.

3.5 Treatment Effects and Long-Run Effects Compared to Prior Estimates

It is useful to compare our estimates to others in the literature – particularly Charness and Gneezy (2009), Acland and Levy (2015), and Royer, Stehr, and Sydnor (2015).¹⁸ Both Charness and Gneezy (2009) and Acland and Levy (2015) use samples of mainly university students while the subject pool for Royer, Stehr, and Sydnor (2015) comes from the firm covered in this study.

The main treatment arm of Charness and Gneezy (2009) offers participants \$100 to attend the gym 8 times over a 4-week period. Their in-treatment effects imply an increase in weekly gym visits of 1.2. However, the main point of their paper is to assess the degree to which the incentives affect behavior following the incentive period. The post-treatment effect, measured at 7 or 13 weeks after the intervention across their two studies, was 0.59 visits per week, or roughly half of the in-treatment effect. Moreover, they find that all of this habit-

¹⁸ Two other field experiments of exercise incentives, Babcock and Hartman (2010) and Babcock et al. (2015), do not discuss their long-run effects. Thus, we exclude them from discussion here.

formation effect was concentrated on the subset of participants who were not regular users of the gym prior to the intervention.

Acland and Levy (2015) follow the same general design except they require 2 visits per week for 4 weeks to be eligible to earn \$100. Their focus is on the extent to which individuals can predict their post-intervention behaviors. Their in-treatment effects are similar to Charness and Gneezy but their post-intervention effects are smaller at 0.26 visits per week.

Royer, Stehr, Sydnor (2015) offer participants \$10 per visit up to 3 visits per week for a 4-week period. Their contribution is the combining of an incentive program with a commitment contract following the incentive period to see if the commitment contract improves habit formation. In their full sample (pooling members and non-members) the incentive without an additional commitment option raised gym attendance by 0.56 visits per week in-treatment and by 0.12 in the 9-week period following the incentives. Measuring habit formation as the ratio of the two effects, they found 21% habit formation compared to Charness and Gneezy's estimate of 50%.

If we pool our sample (members and non-members) and focus on the constant-incentive group for the purpose of comparison, we find that the in-treatment effect for weeks 1-8 for this group is 0.33 visits per week and the post-treatment effect for weeks 9-16 is 0.15 visits per week. Compared to Royer, Stehr and Sydnor (2015), which took place in the same setting, the in-treatment effects are somewhat smaller in the current study, perhaps because the incentives were capped at 2 visits per week whereas, in the earlier study, subjects could earn money for up to 3 visits per week. Our habit formation effect, the ratio of post-intervention to intervention effect size, in contrast, is roughly double that of Royer, Stehr, and Sydnor (2015) and is more similar to that of Charness and Gneezy (2009). A plausible, though untestable, reason for the larger habit-

formation effect in this study is the longer length of the intervention – 8 weeks versus 4 weeks in Royer, Stehr, and Sydnor (2015).

4. Robustness Checks

4.1 Cross-contamination concerns

Participants in an experiment conducted in a closed environment like a workplace may become aware of treatments to which they were not assigned, and knowledge of the treatment assignment of a peer could make one's own treatment assignment seem less or more desirable.¹⁹ Such cross-contamination would bias our results and while we certainly cannot eliminate the possibility that cross-contamination occurred, we can point to evidence that cross-contamination was minimal.

First, since cross-contamination is most likely to occur between employees who work together, we can examine the overlap within cohorts of employees in the same organizational unit.²⁰ Across all cohorts, 73 percent of subjects were the lone employee from their unit in their cohort in the experiment. Second, the treatment effects do not vary with the number of employees from one's unit in the experiment. If the incentives were more effective when one's co-worker was also in the experiment, we would expect that the treatment effects would have positive interactions with the number of co-workers in the experiment. Third, the introduction of incentives does not appear to affect the behavior of the control group. In the main treatment

¹⁹ Cross-contamination of this type would violate SUTVA (Stable Unit Treatment Value Assumption), one of the fundamental assumptions of identifying causal effects in randomized-controlled trials.

²⁰ There are 268 different units within the organization.

periods, we observe no decline in attendance among the control group.²¹ Overall, these three pieces of evidence suggest that any bias from cross-contamination is likely to be small.

4.2 Substitution effects

Our results show that incentives were successful in getting workers to use the company gym, but to fully understand their effects on exercise, we need to know whether the exercise was “new” or merely shifted (substituted) from other locations. To test for substitution of workplace exercise for other exercise, we combine computerized gym visit data with self-reported exercise data from our follow-up survey. The follow-up survey asked subjects both how many days they exercised at the company gym and how many total days they exercised in the previous week.²² We then pooled the treatments and estimated treatment effects with three different dependent variables: company gym days from the computerized login records, company gym days from the survey and total exercise days from the survey. If there is a treatment effect on the total days of exercise, then we know that the experiment-induced exercise at the gym is not entirely due to substitution from exercise elsewhere. Moreover, the degree to which the treatment effect on total days of exercise is smaller than that for company gym exercise reflects the degree to which there was substitution. We present further details of this analysis in the Appendix and only discuss the results here.

²¹ This result refers to pooled results among non-members and members. That is, if we pool members and non-members, we observe no significant trends in gym use in the control group from the pre-treatment period to the main treatment periods (weeks 1-8 and 9-16). If, however, we looked only at members during weeks 9 through 16, there is a statistically significant downward trend among control-group members. That result, however, is expected because there will naturally be some rate of gym quitting over time in a group of pre-identified members.

²² By inspecting the date each subject completed the survey, we were able to determine the week of the experiment for which each subject was reporting his exercise behavior. The reporting week varied both for administrative reasons and because subjects sometimes waited before filling out the survey. 6.6% reported their exercise for week 4, 6.4% for week 5, 14.6% for week 6, 66.7% for week 7, 5.5% for week 8, and 0.2% for week 9.

This analysis points to two findings. First, although self-reported, the survey measures of overall exercise may be useful to assess treatment effects, at least for non-members. For existing members, there appears to be larger measurement error in self-reported exercise. However, since existing members mostly exercise at the company gym, there is less concern that increases in the use of the gym represent substitution for this group. Second and most important, for ex-ante non-members, approximately 79% of the gym days induced by the incentives represent new exercise.

5. Conclusion

The use of incentives is pervasive in many domains, including exercise. While much literature has discussed whether such incentives work, less attention has been spent on understanding how the temporal allocation of the incentives affects their efficacy. Moreover, there has been little focus on how well incentives work at inducing behavioral change across different sub-populations that may face behavioral challenges on different margins (i.e., the initiation of any physical activity versus increasing and sustaining physical activity). Our main conclusion regarding incentives for motivating exercise on the extensive margin are that a) the timing of incentive payouts, and in particular front-loaded designs, does not appear to have a meaningful effect on behavior but b) the overall length and hence earnings available in the program are important for how many employees respond. For motivating gym use on the intensive margin our primary findings are that a) constant incentives are somewhat more effective than front-loaded incentives and b) it is possible to generate and leverage short run habit formation to smooth out incentive effects using a sporadic incentive. We also find suggestive evidence that habit formation is stronger in longer programs. The post-treatment effects for our eight-week constant incentive program were stronger than those for its four-week

counterpart and also stronger than those for a similar 4-week incentive program in Royer, Stehr, and Sydnor (2015) who use a very similar subject pool.

These results suggest a few natural areas for further research in future studies. For example, it would be helpful to know more about how different variations in the length of incentive programs affect both participation on the extensive margin and habit formation. Similarly, the sporadic incentive design shows enough promise in this setting to suggest value in future studies exploring this idea more fully, including examining issues like how often incentives need to be in place to maintain elevated behavior. The sporadic design is also unique in that it allows the program to be longer for a fixed budget without shrinking the size of the per-activity incentive. Future studies could explore the tradeoff between incentive-per-activity size and program length more explicitly. Our heterogeneity analyses also suggest that there will surely be value to future studies that leverage new techniques, such as predictive models using machine learning, to tailor incentives to specific subpopulations. Finally, better understanding how to motivate behavior on the extensive margin is ultimately important in a range of settings beyond exercise, including for example motivating the use of public transport, eating new types of healthy foods, biking and walking to work, meditation and mindfulness, and accessing primary care services. The somewhat surprising fact that front-loading incentives did not increase participation on the extensive margin suggests that there may be value in further studies that explore the nature of the fixed costs people face when they must engage in new behaviors and how incentive designs can best help overcome those barriers.

References

- Acland, Dan and Matthew Levy. 2015. "Naiveté, Projection Bias, and Habit Formation in Gym Attendance." *Management Science*, 61(1): 146-160.
- Babcock, Philip and John Hartman. 2010. "Networks and Workouts: Treatment Size and Status Specific Peer Effects in a Randomized Field Experiment." NBER Working Paper 16581.
- Babcock, Philip, Kelly Bedard, Gary Charness, John Hartman, and Heather Royer. 2015. "Letting Down the Team? Evidence of Social Effects of Team Incentives." *Journal of European Economic Association*, 13(5): 841-870.
- Bachireddy, Chethan, Leslie John, Katherine Milkman, Francesca Gino, and Bradford Tuckfield. 2017. "How Can We Optimally Reward Exercise and Build Lasting Habits? A Field Experiment." Working paper.
- Cawley, John and Joshua Price. 2013. "A Case Study of a Workplace Wellness Program That Offers Financial Incentives for Weight Loss." *Journal of Health Economics*, 32(5): 794-803.
- Charness, Gary, and Uri Gneezy. 2009. "Incentives to Exercise." *Econometrica*, 77(3): 909–931.
- Daly, Michael, Colm P. Harmon, and Laim Delaney. 2009. "Psychological and Biological Foundations of Time Preference." *Journal of the European Economic Association*. 7(2-3): 659-669.
- Finkelstein, Eric, Laura Linnan, Deborah Tate, and Ben Birken. 2007. "A Pilot Study Testing the Effect of Different Levels of Financial Incentives on Weight Loss among Overweight Employees." *Journal of Occupational and Environmental Medicine* 49(9): 981-989.
- Gneezy, Uri, Stephen Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *The Journal of Economic Perspectives* 25(4): 191–209.
- Haisley E, Volpp KG, Pellathy T, Loewenstein G. 2012. "The impact of alternative incentive schemes on completion of health risk assessments." *American Journal of Health Promotion*, 26(3): 184–188.
- Halpern, SD, B French, DS Small, K Saulsgiver, MO Harhay, J Audrain-McGovern, G Loewenstein, TA Brennan, DA Asch, KG Volpp. 2015. "Randomized Trial of Four Financial-Incentive Programs for Smoking Cessation." *New England Journal of Medicine*, 372(22): 2108-2117.
- John, Leslie K, Loewenstein, George, Troxel, Andrea B, Norton, Laurie, Fassbender, Jennifer E., and Kevin G. Volpp. 2011. "Financial Incentives for Extended Weight Loss: A Randomized, Controlled Trial." *Journal of General Internal Medicine*, 26(6): 621-626.

- Just, David. R., and Joseph Price. 2013. "Using Incentives to Encourage Healthy Eating in Children." *Journal of Human Resources* 48(4):855-872.
- Loewenstein, George, Joseph Price and Kevin G.M. Volpp. 2016. "Habit formation in children: Evidence from Incentives for Healthy Eating." *Journal of Health Economics*, 45: 47-54.
- Milkman, Katherine L, Julia A Minson, and Kevin G.M. Volpp. 2014. "Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling." *Management Science* 60(2): 283-299.
- Mochon, Daniel, Janet Schwartz, Josiase Maroba, Deepak Patel, and Dan Ariely. 2017. "Gain Without Pain: The Extended Effects of a Behavioral Health Intervention." *Management Science*, 63(1): 58-72.
- O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review*, 89(1): 103-124.
- Patel MS, Asch DA, Rosin R, et al. 2016. "Framing financial incentives to increase physical activity among overweight and obese adults: a randomized, controlled trial." *Annals of Internal Medicine*, 164(6): 385-394.
- Royer, Heather, Mark Stehr, and Justin Sydnor. 2015. "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company." *American Economic Journal: Applied Economics*, 7(3): 51-84.
- Sen, Aditi, David Huffman, George Loewenstein, David A. Asch, Jeffrey T. Kullgren, Kevin Volpp, "Do Financial Incentives Reduce Intrinsic Motivation for Weight Loss?: Evidence from Two Tests of Crowding Out" (2016) *Nudging Health: Health Law and Behavioral Economics*, edited by I. Glenn Cohen, Holly Fernandez Lynch, and Christopher T. Robertson.
- Tagney, June P., Roy F. Baumeister, and Angie Luzio Boone. 2004. "High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success." *Journal of Personality*, 72: 271-324.
- Volpp, Kevin G., Leslie John, Andrea Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein. 2008. "Financial Incentive-Based Approaches for Weight Loss: A Randomized Trial." *Journal of American Medical Association* 300(22): 2631-2637.
- Volpp, Kevin G., Andrea B. Troxel, Mark V. Pauly, Henry A. Glick, Andrea Puig, David A. Asch, Robert Galvin et al. 2009. "A randomized, controlled trial of financial incentives for smoking cessation." *New England Journal of Medicine* 360(7): 699-709.

Figure 1.
Fraction of Existing Members with Gym Visits by Treatment Group and Week

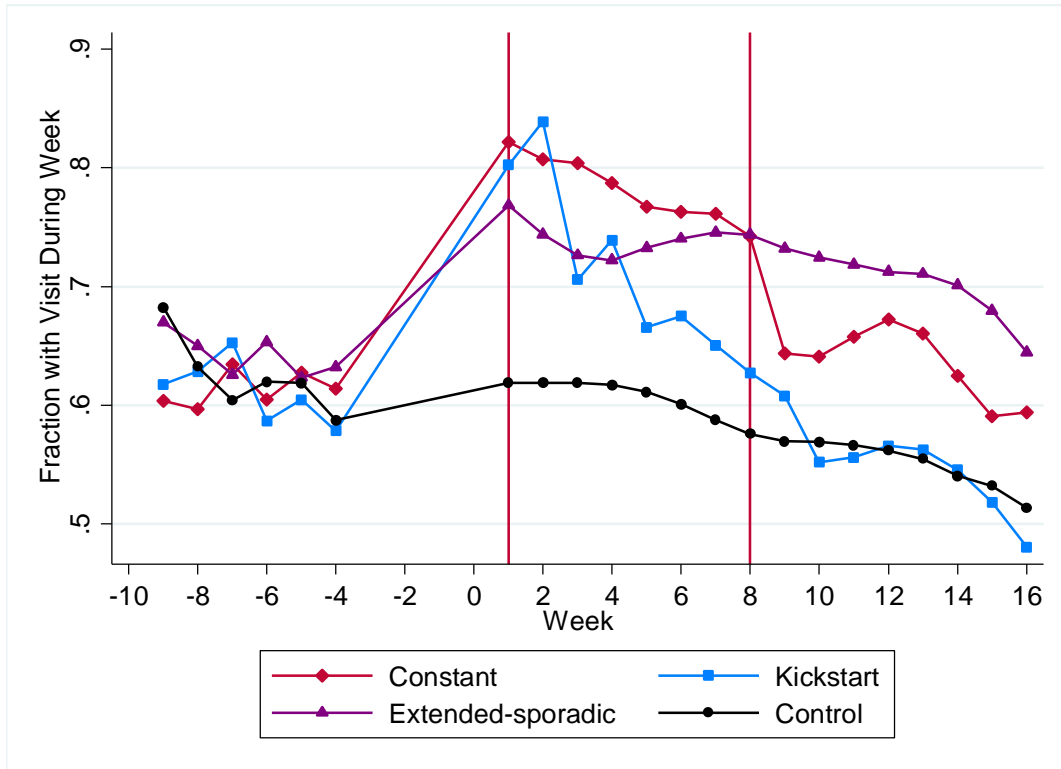


Figure notes: This figure presents LOESS smoothed plots with bandwidth set at 0.4. Outcome variable is an indicator for making at least one visit to the company gym during that week. Week 1 (marked with a vertical line) denotes the first week of the incentive program. Pre-treatment visits rates for weeks -9 through -4 are included in the graph. Weeks -3 through 0 are omitted because these were weeks in which the initial survey and/or information about the treatment assignment was sent to subjects.

Figure 2.

Fraction of Non-Members with Gym Visits by Treatment Group and Week

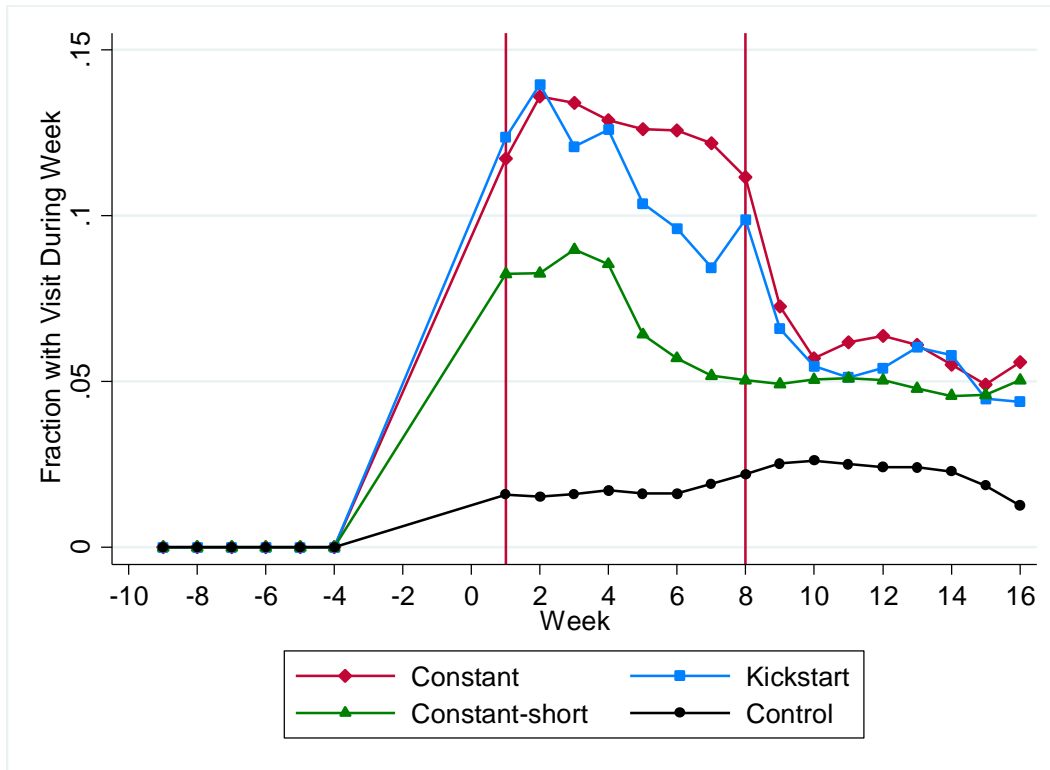


Figure notes: This figure presents LOESS smoothed plots with bandwidth set at 0.4. Outcome variable is an indicator for making at least one visit to the company gym during that week. Week 1 (marked with a vertical line) denotes the first week of the incentive program. Pre-treatment visits rates for weeks -9 through -4 are included in the graph. Weeks -3 through 0 are omitted because these were weeks in which the initial survey and/or information about the treatment assignment was sent to subjects.

Table 1. Treatments

Treatment	Sample	Length of Incentive Period	Length of Program	Maximum Earnings	Description of Incentives
Constant	Members & Non-Members	8 weeks	8 weeks	\$160	\$10/visit for 8 weeks (capped at 2 visits/week)
Kickstart	Members & Non-Members	8 weeks	8 weeks	\$160	\$25/visit for 1st 2 weeks of incentive period; \$5/visit the last 6 weeks of incentive period (capped at 2 visits/week)
Constant Short	Non-Members	4 weeks	4 weeks	\$80	\$10/visit for 4 weeks (capped at 2 visits/week)
Extended-Sporadic	Members	8 weeks	16 weeks	\$160	\$10/visit for a random 8 weeks out of a 16 week period (capped at 2 visits/week)
Control	Members & Non-Members	NA	NA	\$0	NA

Table 2. Summary Statistics

Panel A. Non-member participants

Variable	Overall Mean	Control Mean	Constant Difference	Kickstart Difference	Constant-Short Difference	P-value of All Treatments=0
Age	41.29 [11.43]	41.24 [12.1]	-0.25	-0.22	-0.28	1.00
Male	0.52	0.49	0.05	0.05	0.06	0.84
College or More	0.7	0.7	0.03	-0.02	0.03	0.8
Interest in Joining Gym (-2 to 2, most interested)	-0.83 [1.35]	-0.77 [1.32]	0.07	0.06	-0.02	0.94
Married	0.67	0.69	-0.01	-0.03	-0.04	0.89
One or More Children at Home	0.5	0.48	0	0.08	-0.03	0.3
Days of Exercise Last Week	2.24 [1.81]	2.31 [1.75]	-0.02	-0.28	-0.17	0.58
Target Days of Exercise	3.6 [1.54]	3.59 [1.38]	0.01	-0.1	-0.16	0.77
Subjective Probability of Meeting Target	58.91 [32.4]	61.61 [30.22]	0.59	-6.76	-3.29	0.23
Days Exercised at Different Gym in Week	0.6 [1.35]	0.65 [1.35]	-0.05	-0.03	0.03	0.97
Days Exercised not at Gym in Week	1.85 [1.79]	1.82 [1.65]	0.08	0.01	-0.14	0.80
Body Mass Index	28.16 [6.18]	27.57 [5.74]	1.04	0.66	1.09	0.54
Fraction Overweight or Obese	0.66	0.61	0.13	0.06	0.03	0.18
Number of Observations	609	109	180	182	138	

Table notes: The overall mean column is the overall mean. The control mean column is the control group mean. The constant difference column is the mean difference between the constant group and the control group; they come from regressions that include cohort fixed effects. The kickstart difference column is the mean difference between the kickstart group and the control group; they come from regressions that include cohort fixed effects. The constant-short difference column is the mean difference between the constant-short group and the control group; they come from regressions that include cohort fixed effects. The p-value column display the p-values testing that the means of all 4 groups (3 treatment groups + 1 control) are equal. For the non-dichotomous variables, the numbers in brackets represent the standard deviations.

Table 2. Summary Statistics continued

Panel B. Member participants

Variable	Overall Mean	Control Mean	Constant Difference	Kickstart Difference	Extended Sporadic Difference	P-value of All Treatments=0
Age	40.72 [10.46]	41.39 [11.43]	-1.56	-1.16	0.18	0.61
Male	0.48	0.42	0.07	0.11	0.04	0.55
College or More	0.72	0.79	-0.09	-0.04	-0.12	0.36
Married	0.71	0.75	0.01	-0.15	-0.01	0.05
One or More Children at Home	0.56	0.59	-0.03	-0.01	-0.03	0.98
Days of Exercise Last Week	3.57 [1.81]	3.79 [1.81]	-0.31	-0.24	-0.23	0.75
Target Days of Exercise	4.52 [1.15]	4.67 [1.14]	-0.13	-0.24	-0.19	0.61
Subjective Probability of Meeting Target	74.29 [24.98]	74.05 [25.65]	0.73	-0.05	1.42	0.97
Days Exercised at Company Gym in Week	2.3 [1.84]	2.73 [1.93]	-0.61	-0.57	-0.35	0.20
Days Exercised at a Different Gym in Week	0.43 [.97]	0.55 [1.12]	-0.16	-0.2	-0.05	0.56
Days Exercised not at Gym in Week	1.57 [1.46]	1.26 [1.39]	0.34	0.43	0.35	0.32
Body Mass Index	27.32 [7.77]	27.01 [6.22]	0.27	-0.5	1.41	0.48
Fraction Overweight or Obese	0.62	0.53	0.06	0.15	0.12	0.21
Number of Observations	371	63	100	102	106	

Table notes: The overall mean column is the overall mean. The control mean column is the control group mean. The constant difference column is the mean difference between the constant group and the control group; they come from regressions that include cohort fixed effects. The kickstart difference column is the mean difference between the kickstart group and the control group; they come from regressions that include cohort fixed effects. The constant-short difference column is the mean difference between the constant-short group and the control group; they come from regressions that include cohort fixed effects. The p-value column display the p-values testing that the means of all 4 groups (3 treatment groups + 1 control) are equal. For the non-dichotomous variables, the numbers in brackets represent the standard deviations.

Table 3. OLS Regression of Non-Member Participation Decision

Dependent variable: <i>Subgroup (if any):</i>	(1) Joined <i>All</i>	(2) Joined <i>Low exercisers</i>	(3) Joined <i>High exercisers</i>
Constant-Short	0.032 (0.033)	-0.000086 (0.046)	0.11* (0.055)
Constant	0.10*** (0.034)	0.13** (0.053)	0.13*** (0.045)
Kickstart	0.10*** (0.035)	0.12** (0.053)	0.11** (0.043)
Observations	596	318	253
Mean of dep var in control group	0.046	0.054	0

Table notes: Each column represents a separate regression. Heteroskedasticity-robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1 "Low exercisers" are non-members reporting 2 or fewer days of exercise in the week prior to the initial survey as well as in their typical week. "High exercisers" are non-members reporting 3 or more days of exercise in the week prior to the initial survey as well as in their typical week. We drop from columns (2) and (3) the 25 subjects whose prior week exercise was at odds with their typical exercise.

Table 4. OLS Regression of Visit Rates for Existing Members

Dependent Variable:	(1) Any Visit	(2) Any Visit	(3) Number of Visits	(4) Number of Visits
Weeks 1-8	-0.017 (0.035)	-0.017 (0.035)	-0.091 (0.093)	-0.091 (0.093)
Weeks 1-8 * Constant	0.184*** (0.048)	0.184*** (0.048)	0.433*** (0.135)	0.433*** (0.135)
Weeks 1-8 * Kickstart	0.119*** (0.044)	0.119*** (0.044)	0.436*** (0.121)	0.436*** (0.121)
Weeks 1-8 * Extended Sporadic	0.111** (0.044)		0.434*** (0.134)	
Weeks 1-8 * Incentive ON for Extended Sporadic		0.158*** (0.045)		0.562*** (0.137)
Weeks 1-8 * Incentive OFF for Extended Sporadic		0.065 (0.046)		0.311** (0.141)
Weeks 9-16	-0.073** (0.036)	-0.073* (0.037)	-0.304** (0.123)	-0.304** (0.126)
Weeks 9-16 * Constant	0.093** (0.047)	0.093* (0.048)	0.233 (0.157)	0.233 (0.161)
Weeks 9-16 * Kickstart	0.009 (0.043)	0.009 (0.044)	0.198 (0.140)	0.198 (0.144)
Weeks 9-16 * Extended Sporadic	0.137*** (0.046)		0.491*** (0.161)	
Weeks 9-16 * Incentive ON for Extended Sporadic		0.165*** (0.048)		0.598*** (0.169)
Weeks 9-16 * Incentive OFF for Extended Sporadic		0.109** (0.047)		0.382** (0.163)
Observations	8,184	8,184	8,184	8,184
Number of persons (cluster level)	372	372	372	372
Control pre-period mean of dep var	0.624	0.624	1.897	1.897

Table notes: Each column represents a separate regression. All regressions include individual fixed effects. "Incentive ON for Extended Sporadic" represents the weeks in which members of the extended sporadic group were incentivized to visit. These 8 weeks were randomly allocated within weeks 1-16, separately for each participant. Robust standard errors clustered on individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table 5. OLS Regression of Visit Rates for Existing Members
by Ex-Ante Attendance Rates**

Dependent variable:	(1)	(2)	(3)	(4)
	Any Visit	Any Visit	Number of Visits	Number of Visits
Subgroup:	Less than 2 visits	2 or more visits	Less than 2 visits in	2 or more visits in
	in pre period	in pre period	pre period	pre period
Weeks 1-8	0.045 (0.067)	-0.058 (0.037)	0.108 (0.116)	-0.223* (0.130)
Weeks 1-8 * Constant	0.325*** (0.089)	0.080* (0.044)	0.774*** (0.189)	0.180 (0.166)
Weeks 1-8 * Kickstart	0.172** (0.082)	0.062 (0.043)	0.485*** (0.168)	0.354** (0.160)
Weeks 1-8 * Extended Sporadic	0.217** (0.085)	0.054 (0.044)	0.692*** (0.208)	0.300* (0.164)
Weeks 9-16	-0.005 (0.059)	-0.117*** (0.045)	0.068 (0.119)	-0.548*** (0.178)
Weeks 9-16 * Constant	0.116 (0.081)	0.074 (0.053)	0.293 (0.182)	0.170 (0.215)
Weeks 9-16 * Kickstart	-0.038 (0.070)	0.037 (0.054)	-0.082 (0.136)	0.364* (0.209)
Weeks 9-16 * Extended Sporadic	0.235*** (0.078)	0.086 (0.053)	0.729*** (0.213)	0.381* (0.207)
Observations	3,366	4,818	3,366	4,818
Number of persons (cluster level)	153	219	153	219
Control pre period mean of dep var	0.220	0.890	0.387	2.890

Table notes: Each column represents a separate regression. All regressions include individual fixed effects. Robust standard errors clustered on individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Members visiting less than 2 visits in pre-period are those who in weeks -4 to -9 went to the gym 2 or more days fewer than half of the weeks. Members visiting 2 or more visits in pre-period are the remaining members.

Table 6. OLS Regression of Visit Rates for Non-Members

Dependent variable:	(1) Any Visit	(2) Number of Visits
Weeks 1-4 * Constant	0.116*** (0.027)	0.275*** (0.064)
Weeks 1-4 * Kickstart	0.113*** (0.027)	0.248*** (0.067)
Weeks 1-4 * Constant-Short	0.064** (0.026)	0.167** (0.067)
Weeks 5-8	0.002 (0.005)	-0.016 (0.019)
Weeks 5-8 * Constant	0.106*** (0.025)	0.265*** (0.057)
Weeks 5-8 * Kickstart	0.074*** (0.023)	0.155*** (0.049)
Weeks 5-8 * Constant-Short	0.027 (0.020)	0.109** (0.054)
Weeks 9-16	0.007 (0.013)	0.001 (0.030)
Weeks 9-16 * Constant	0.036** (0.018)	0.104** (0.047)
Weeks 9-16 * Kickstart	0.030 (0.018)	0.072* (0.043)
Weeks 9-16 * Constant-Short	0.017 (0.018)	0.070 (0.045)
Observations	9,536	9,536
Number of persons (cluster level)	596	596
Control pre period mean of dep var	0	0

Table notes: All regressions control for male, education degree fixed effects, age, interest in the company gym fixed effects, whether or not the individual is married, and cohort fixed effects. Each column represents a separate regression. Robust standard errors clustered on individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

APPENDIX

A.1 Substitution Analysis

In this section of the appendix, we provide further details of the substitution analysis described in the main text. Our substitution analysis takes two steps: 1) assess how well self-reported measures work in calculating the treatment effects on company gym exercise and 2) then based on the findings of 1), proceed to estimate the treatment effects on overall exercise.

A slight complication arises when comparing treatment effects estimated for computerized logins with those estimated for total days of exercise from the survey. The computerized records should have little or no measurement error whereas the overall exercise measures are self-reported, and thus may be prone to recall error. To assess the validity of the self-reported measure, we estimate and then compare treatment effects from the computerized records with those from the self-reported attendance. If those treatment effects are similar, then it may be reasonable to believe that the treatment effects on overall exercise days reported in the survey are also reliable.

Appendix Table A1 presents the effects of the incentive treatments on attendance from computerized company gym swipes, self-reported company gym days, and self-reported total exercise days separately for members and non-members. Note we combine all treatments (i.e., *Treated* is a binary indicator equal to 1 if one belongs to one of the incentivized groups and 0 otherwise) to increase precision in this analysis.

For members, the self-reported company gym attendance effects are significantly smaller than the company gym swipes effects. Thus, for this group, using the self-reported overall exercise days is unlikely to be fruitful. But, we worry less about substitution for this group

because they are already members of the gym, and the primary location of exercise for this group is the company gym.

For the non-members, the two estimates measuring company gym attendance effects are quite similar, 0.218 for company gym logins versus 0.186 for self-reported company gym days. Thus, the size of the total exercise days effect relative to the company gym days effect may be a reasonable measure of substitution. If we take the ratio of coefficients in columns (5) and (6), we calculate that 79% of the gym days induced by the incentives represented new exercise.²³ These results are similar to those in Royer, Stehr and Sydnor (2015) who followed a similar methodology and found that, again depending on the measure of company gym visits, either 63% (from self-reports) or 84% (from logins) of incentive-induced exercise at a gym represented new exercise for those who were not previously members of the gym. In summary, our results are consistent with the notion that for non-members, the incentive programs induced substantial new exercise.

A.2 Understanding the Quantity and Type of Exercise Induced by the Incentives

In this study, subjects in the treatment groups received incentives if the electronic records at the company gym showed that they made the required visits. Although this outcome measure has been used in the literature, it faces several shortcomings as a measure of new exercise. First, subjects may “game the system” by logging in and out of the gym without exercising and for the sole purpose of earning incentives. Second, as described in the body of the paper, subjects may substitute exercise at the company gym for other forms of exercise so that only a fraction of the exercise at the company gym is in fact new exercise. We address these concerns in two ways.

²³ If we were to instead rely on the ratio of the coefficients in columns (4) and (6), this figure would fall to 67%.

The first relies on surveying subjects about their exercise behavior at and away from the company gym while the second relies on analyzing health outcomes that are plausibly affected by changes in exercise.

Our follow-up survey asked subjects how many minutes they spent exercising at the company gym, at other gyms and at places other than a gym during the calendar week before they took the survey. Controlling for exercise at these other venues, a regression of total minutes spent exercising on visits to the company gym shows that each visit to the company gym was associated with 45 additional minutes of exercise ($p = 0.01$). This relationship between minutes and visits was largely driven by exercise the subjects characterized as moderate to vigorous cardio and strength training rather than light cardio suggesting that the exercise was, in fact, substantive.

It is possible that subjects in the treatment groups might have over reported their exercise minutes relative to the control group if they thought their answers would influence their incentive payments, though they were instructed that payments were based on gym login records. To test this hypothesis, we included in the regression an interaction term between a control group dummy and the number of visits to the company gym. The coefficient on this interaction term was very small (0.28 minutes) and statistically insignificant ($p = 0.97$), so we fail to reject that the relationship between visits to the company gym and minutes of exercise is different for the control group than for the treatment groups. As such, receipt of incentives is an unlikely explanation for the association between company gym visits and exercise minutes. It's also possible, perhaps even likely, that all subjects were overstating their exercise minutes to a certain extent, however, even if subjects were overstating their exercise by 50 percent, a visit to the company gym would still be associated with half an hour of exercise.

A.3 Estimating the Health Effects of the Incentives

As part of our study, we offered participants an additional incentive to provide health measurements at the time that they collected their survey payments.²⁴ While we anticipated ex-ante that we would be underpowered to detect health effects of the exercise induced by our intervention, the data we collected can aid future researchers in making power calculations for similar employee populations. Thus, we present and discuss our health measure results, along with their caveats, in this section.

The ideal way to test for health effects of exercise induced by an intervention would be to collect all subjects' health measures before and after the intervention period, as done by Charness and Gneezy (2009) with a college student population. In corporate settings like ours, however, requiring multiple in-person visits for health measures could greatly reduce employees' interest in participating in any study or wellness program. Thus, for some of the health measures we collected, our analysis is limited to a comparison of ex-post measurements between treated and control groups, rather than a comparison of individual-level changes between pre- and post-intervention measurements. For these measures (blood pressure, pulse), any initial differences between the treatment and control groups cannot be observed, making our identification of a treatment effect rely on the assumption that initial (ex-ante levels) were equal on average. For weight and BMI, on the other hand, we can calculate individual-level *differences*

²⁴ We took advantage of the fact that all study participants had to visit the gym in person to pick up their earnings (including compensation for taking the survey as well as any earned incentives) at the end of the study, and asked them at this time to record their health measures (including weight) for an added payment of \$25.

between self-reported values provided in the initial survey and ex-post measurements.²⁵ For all comparisons, we pool together the different incentive programs in order to increase power, comparing them against the control group alone.

Regardless of whether the comparison between treated and control groups uses only ex-post values or individual-level differences, the analysis necessarily excludes all subjects who did not provide their health measures at the end of the 8-week study period. This raises the important concern of self-selection into the subsample providing health measures. We obtained ex-post health measures for 77% of the control group and 70% of the incentivized groups.²⁶ If employees who exercised more in response to the incentives were more likely to self-select into showing up and providing health measures, but there was no similar selection within the control group, this would bias us towards overestimating the treatment effect of the incentives on weight. Interestingly, there appears to be far less selection in our sample of gym members, of whom 88.9% of the control group and 86.7% of the treated participants gave health measures, as opposed to the non-members.

In Appendix Table A2, we report the difference-in-difference estimates of changes in weight caused by our intervention in three populations: all study participants, non-members and members, and show p-values based on the Wilcoxon rank-sum test of the difference-in-differences. In all populations, the incentivized groups appear to have a smaller increase in weight between the self-reported ex-ante measure and the objective ex-post measure, although

²⁵ While there is sure to be measurement error in the initial self-reported weight and height, most likely a systematic underreporting of weight, this should be orthogonal to the assignment of incentives given that the initial survey preceded subjects' assignment to incentive programs.

²⁶ Perhaps because they had fewer opportunities to earn money throughout the study, employees in the control group were also more likely than those receiving incentives to complete the follow-up survey (94.8% vs. 84.0%), more likely to pick up their earnings, and therefore more likely to provide their health measures at the time of earnings pickup.

the size and significance varies.²⁷ The difference-in-differences suggests that our incentives led to a statistically significant average weight reduction of 1.7 pounds among non-members, while the estimate for members was smaller (0.93 pounds) and statistically insignificant. Since the estimated treatment effects of our incentive programs on number of gym visits were much larger for members versus non-members, this contrast appears somewhat puzzling. In the absence of selection concerns, the larger impact on weight for non-members despite a smaller effect on gym visits could potentially be explained by stronger weight-loss induced by changes in exercise for those who had lower baseline levels of exercise. However, this pattern could easily be explained by different rates of self-selection into providing health measures between members and non-members, described in the previous paragraph.

The self-selection concern is amplified by the fact that the estimated reductions in weight associated with receiving incentives are implausibly large considering the small number of gym visits (on average) induced by the incentives. For the pooled sample of members and nonmembers, the estimates suggest that one additional gym visit per week, over 8 weeks, is associated with weight loss of 4.53 pounds, implying that the average induced gym visit burned 0.57 pounds or 1981 calories based on the commonly accepted “3,500 calories per pound” rule.²⁸ Based on the average body weight in our sample, burning 1981 calories would require 2.4 hours of vigorous exercise such as running at 6mph or cycling 14-15.9mph.²⁹ This is far more than the 45-minute average duration of a visit reported by participants (see Section A.1). Thus, we suspect that self-selection may be biasing our estimates of changes in weight.

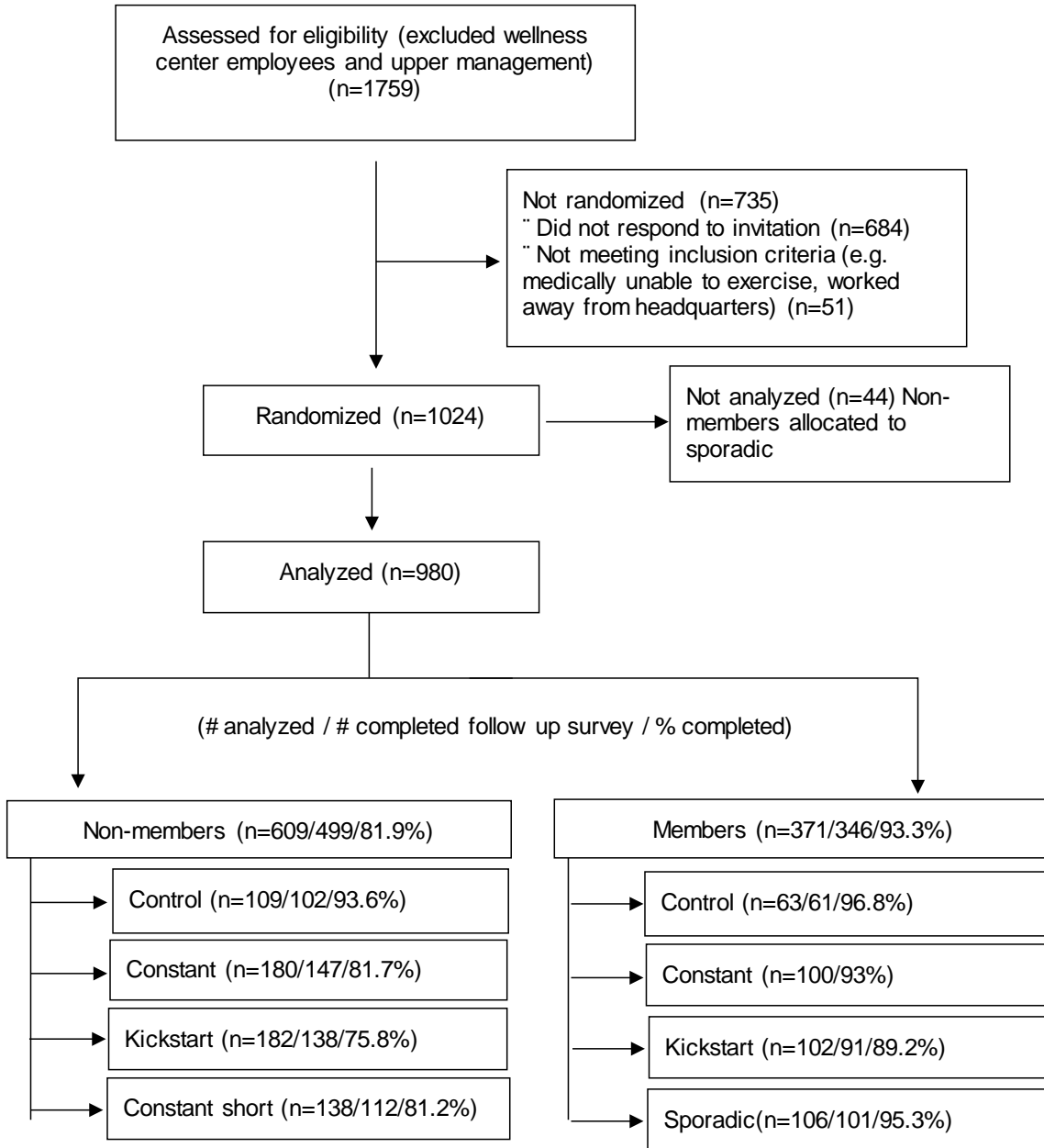
²⁷ The fact that the “change” we measure is generally positive reflects individuals’ tendencies to underreport their weight. If we assume that weight in the control group was constant between the time of the initial survey and the date when health measures were taken, our results suggest that individuals underreported their weight by an average of 7.4 pounds.

²⁸ JAMA Patient Page “Healthy Weight Loss” *JAMA*. 2014;312(9):974. doi:10.1001/jama.2014.10929

²⁹ Source: NutriStrategy. Calculations are based on research data from *Medicine and Science in Sports and Exercise*, the official journal of the American College of Sports Medicine. <http://www.nutristrategy.com/caloriesburned.htm>

For the remaining health measures, in which we only have ex-post measurements, we report averages for the treated and control groups in Appendix Table A3, again pooling all incentive programs together. Results are extremely noisy, with no t-tests detecting differences between the groups at conventional significance levels. Although insignificant, the differences indicate that the treated employees were 3.1 percentage points more likely to have high blood pressure readings and had slightly higher pulse readings (1.07 more beats per minute). This is likely related to the fact that within our sample providing health measures, the average self-reported weight was higher in the treated than control groups (180.4 vs. 176.1 pounds). However, to statistically detect this small treatment difference at a p-value of 0.05 we would need a sample approximately 10 times larger than what we have here in this study. For future studies interested in assessing the health effects, we suggest narrowing the number of treatments and increasing the sample size considerably to get more power for these health outcomes.

Appendix Figure A1. Experimental Participation Flowchart



Appendix Table A1. Substitution Analysis for In-Treatment Effects

	Members			Non-members		
	Company gym swipes	Company gym days from survey	Total exercise days from survey	Company gym swipes	Company gym days from survey	Total exercise days from survey
Treated	0.265	-0.03	-0.162	0.218***	0.186***	0.147
	(0.251)	(0.256)	(0.245)	(0.044)	(0.065)	(0.201)
Mean dependent var in control group	1.93	2.28	3.61	0.00	0.10	2.25
Observations	343	343	343	490	490	490

Table notes: Heteroskedasticity-robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Analysis is based on gym data and data from follow up survey. The dependent variable is days of exercise per week as measured by gym swipes or self-reported survey data. Treated includes treatment groups constant, kickstart and extended-sporadic for members and constant, kickstart and constant-short for non-members.

Appendix Table A2. Changes in Weight

	Control Group			Incentivized Groups			Difference-in-differences	
	N	Ex-ante (self-report)	Ex-post minus ex-ante	N	Ex-ante (self-report)	Ex-post minus ex-ante	$D_{\text{incentivized}} - D_{\text{control}}$	Change associated with an increase of one visit per week
<i>Weight (lbs.)</i>								
Full sample	130	176.14 [3.65]	7.45 [1.14]	560	180.4 [1.52]	6.05 [0.405]	-1.36* [0.051]	-4.53
Non-members	75	181 [5.36]	7.50 [1.74]	297	183.8 [2.41]	5.72 [0.56]	-1.78** [0.029]	-8.77
Members	55	169.6 [6.32]	7.39 [1.22]	263	176.4 [2.36]	6.42 [0.55]	-0.971 [0.65]	-2.24

Table notes: Participants gave self-reported weight in the initial survey ("Ex-ante") and provided objective weight measurements in Weeks 9-10 ("Ex-post"). The sample sizes shown are smaller than the full sample in the rest of the paper because we exclude participants who did not opt to provide health measures at the conclusion of the study. For the differences in differences column, p-values of a Wilcoxon rank-sum test are shown below the difference-in-differences. In the other columns, standard errors are shown below the mean values. In the last column, the difference-in-differences is rescaled based on the average increase in visits per week between control and incentivized groups in each of the three populations. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Table A3. Other Health Measures

	Control Group		Incentivized Groups		Difference
	N	Mean [Std. Err.]	N	Mean [Std. Err.]	Incentivized mean - Control group mean
<i>High blood pressure (0/1)</i>					
Full sample	130	0.331 [0.041]	547	0.362 [0.021]	0.031 [.047]
Non-members	75	0.373 [0.056]	289	0.419 [0.029]	0.045 [.064]
Members	55	0.273 [0.061]	258	0.298 [0.029]	0.026 [.068]
Members with <2 visits per week, ex-ante	30	0.267 [0.082]	134	0.269 [0.038]	0.002 [.090]
<i>Pulse (beats per minute)</i>					
Full sample	129	73.24 [1.07]	551	74.31 [.538]	1.07 [1.23]
Non-members	74	75.62 [1.36]	291	76.59 [0.75]	0.97 [1.63]
Members	55	70.04 [1.65]	260	71.76 [0.74]	1.73 [1.79]
Members with <2 visits per week, ex-ante	30	72.7 [1.68]	135	73.59 [0.96]	0.89 [2.19]

Table notes: For this analysis, our sample is limited to participants who opted to provide their health measures, which were collected after the post-survey following the first 8 weeks of the study. Following the CDC definition, we classify an individual as having high blood pressure if their systolic reading is 140mmHg or higher or if their diastolic reading is 90mmHg or higher. Standard errors are shown below the mean values and mean differences. *** p<0.01, ** p<0.05, * p<0.1.