The impacts of social-comparison information on physical activity*

Lianjun Li†

April 2021

Abstract

Existing evidence shows people are likely to change their behavior for financially or socially related reasons. This paper aims to uncover how patterns in the daily physical activity of adults are affected by information that compares one with unknown peers. We conducted a field experiment that used fitness trackers to collect data on the total number of steps an individual takes in a day. Participants were randomized into a group that was provided with comparison information and a group that did not receive such information. We examined whether individuals in the two groups behaved differently during and after the intervention period. We find no clear evidence of an impact of social norms on the daily number of steps taken. We are also not able to conclude heterogeneous effects across groups with different baseline activity levels, marital status, gender, or procrastinating level. We find slight evidence of an increase in the high activity days per week, using number of days of a week such that the number of daily steps exceeds 8000. Using the quantile treatment method, greater treatment effects may exist among individuals whose number of steps are at the tails of the distribution curve. Unlike altruistic behaviors, physical activity is more likely to be driven by positive social motivators based on the follow-up survey results.

*I am very grateful to David Frisvold, Linnea Polgreen, Lucas Carr, Julia Garlick, Suyong Song, Nick Street, Lester Lusher, Erez Yoeli, and seminar participants at the University of Iowa, and at the Advances with Field Experiments Conference, who have provided helpful comments. I am also thankful for funding from the Department of Economics at Tippie School of Business, University of Iowa, Fahs-Beck Fund for Research and Experimentation, and the Graduate & Professional Student Government research grants of the University of Iowa.

†Contact information: lianjun.li@nus.edu.sg

1 Introduction

Peers are a powerful influence in human life. High school students with disruptive classmates are prone to get lower test scores (Lazear, 2001), whereas those with peers of comparable abilities perform better (Sacerdote, 2011).[1] A teenage girl is more likely to exhibit better health and less likely to encounter domestic violence when her family moves from a higher-poverty neighborhood to a lower-poverty one (Kling et al., 2007). Even fitness levels are subject to the "bad apple" effect: an individual with more unfit friends is more likely to fail a fitness exam at the U.S. Air Force Academy (Carrell et al., 2011).

Individual behaviors and outcomes are subject to the influences of peers because of direct social interactions or social pressure. However, other, anonymous persons can also play a role in forming habits or changing behaviors. This could be driven by disparities between one's attitudes and her perceived norms. A college student drinks more alcohol if she believes her personal feelings about drinking are similar to general campus attitudes about drinking (Perkins and Berkowitz, 1986). This finding implies intervention strategies can be designed to correct perceptions and thus alter behaviors. In particular, when people are presented with information about what others do (descriptive norms) or what others think (injunctive norms), they might be nudged to change their own behaviors. This information-based strategy has been increasingly attractive to policymakers due to its cost-effectiveness and feasibility of implementation.

To understand the impact of comparative information, we studied one specific behavior – daily physical activity – from the results of a field experiment conducted in a university setting. We examined whether individuals provided with information about the physical activity levels of their peers changed their own daily physical activity relative to individuals who were not provided with such information. We also documented the persistence, or lack thereof, of changes in physical activity after individuals no longer receive such information.

We focused on physical activity because of its inherent correlation with health outcomes. In

[1]Note that according to a recent criticism by Angrist (2014), the "tautological nature" of the linear in mean regressions, even if they are elaborate, leads to a fact that the parameters of interest in some peer effects literature is rather an econometrics model behavior than individual social behaviors. The social returns and the social multiplier effects are derived from the divergence between the OLS and the IV estimates.

the industrialized world, chronic diseases are major causes of mortality and morbidity. Sedentary behavior is known to increase one's risk of cardiovascular disease, Type II diabetes, hypertension, and obesity (Carr et al., 2013). High blood pressure and high BMI accounted for over 600,000 U.S. deaths in 2005 (Danaei et al., 2009). In the same year, physical inactivity was one of the top modifiable risk factors that contributed to 191,000 U.S. deaths (Danaei et al., 2009). The World Health Organization (World Health Organization, 2009) estimates that physical inactivity is a contributing factor in 7.7% of deaths, accounting for the loss of 4.1% of disability-adjusted life-years (DALY[2]) for the citizens of almost two thirds of countries in 2004.

Another reason for focusing on physical activity is the fact that most American adults are insufficiently active. In 2014, more than half of adults reported levels of physical activity not meeting the adult aerobic guideline (CDC, n.d.). Furthermore, inadequate physical activity exerts non-negligible externalities on the entire population. Between 2006 and 2011, the burden of physical inactivity on the U.S. population accounted for 8.8% of health care expenditures, or about $79 billion per year; when adults who report walking difficulty are excluded, this figure drops to 6.6%, or about $59 billion (Carlson et al., 2015).[3]

Accordingly, we conducted a randomized field experiment to test how social comparison information influences behavior. We invited students and staff at a public university to participate in the study. Physical activity data was collected through individually distributed fitness trackers and comparison-based information was sent through text messages. The fitness tracker we used in the experiment functions as a pedometer, which tracks participants' number of steps taken per minute and per day as well as other daily statistics (sleeping, distanced walked, and calories burned). It also has a corresponding phone app that participants needed to install on their smart phones. Participants could use the app to view their own data tracked by the device, set goals of daily steps, and also set alarms.[4] We randomly selected participants for a control and treatment group. The treatment period spanned four weeks, during which each

[2]DALY, according to WHO (`http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/`), is the sum of the years of life lost due to premature mortality in the population and they years lost due to disability for people living with a health condition or its consequences.

[3]Data from the National Health Interview Survey (NHIS) and the Medical Expenditure Panel Survey (MEP).

[4]Once a participant meets her step goal, the fitness tracker would vibrate.

individual in both groups received a daily text message. A participant in the treatment group received text messages that contained information on the number of steps they took the previous day and how that compared to her peers within the group.[5] Each participant in the control group received a simple daily text message informing her of the number of steps taken the previous day and reminding her to sync the tracker with the corresponding phone app.[6] Since participants were able to view their number of steps, the control-group participants received no new information about their peers. An initial survey and a follow-up survey were collected two weeks before and after the period when the daily text messages were sent.

Our primary focus is to compare patterns of the daily number of steps taken by the treatment group, who was provided with peer comparison information, versus the control group, who did not receive such information. No cash prize was linked to participants' performance, no tips about how to exercise or what to do were offered, and no benefits of walking or being physically active were emphasized. The goal was to extract the effects from the information about the unknown peers only, controlling for any confounding effects as much as possible. Individual decision making is a cost-benefit analysis. Generally, adults are motivated to participate in physical activity for health benefits, body image, stress management, and enjoyment (Aaltonen et al., 2014). To hold "health capital," people also need to incur an opportunity cost of time and money that must be withdrawn from alternative uses. By randomly selecting members of each group, we ensure similar aggregate characteristics for controlled variables, like their values of health and the costs of being physically active. The only difference between the groups is that one group is provided with information about their performance compared to their peers.

Exposing people to information may make them realize the discrepancy between their beliefs about their own activity level and the perceived norm, which instigates a desire to change their level of exercise. Information can have a positive or negative effect on motivation. Self-affirming information may create a feeling of competence, motivating an originally-active per-

[5]Using steps as the outcome does not lose the purpose of motivating health. Walking reduces the risk of cardiovascular events by 31% and the risk of dying by 32% for both men and women. The benefits are noticeable even at a distance of 5.5 miles per week and a pace of 2 miles per hour (around 60 steps per minute). (Harvard Health Publications, n.d.)

[6]Syncing the app and the device is critical for me to obtain the data. Without doing so, the activities tracked by the device would not be updated in the app, and we would not be able to get the data tracked by the device.

son to sustain or improve her current status. Finally, because forgetfulness is a common barrier to individuals engaging in regular behaviors (McKenzie-Mohr and Schultz, 2014), a consistent provision of information may also function as a prompt to remind people to stand up and move. Therefore, we designed an experiment with information supplied that helps to detect behavioral change on both the intensive (those who are already active) and extensive (those who are currently sedentary) margin.

2 Literature and background

In promoting health behaviors, researchers have been interested in studying different incentives to motivate individuals to form positive habits or to quit negative ones. Many researchers attempt to understand the incentive structure using financially-motivated strategies. Strong positive effects of financial incentives (Volpp et al., 2008, Charness and Gneezy, 2009, John et al., 2011) and commitment devices (Royer et al., 2015) have been found for outcomes like gym attendance and weight loss among college students, adults with high BMI, and workers in a company setting.[7] Financial incentives reward a certain behavior directly by cash, and commitment devices exert a financial cost of a future behavior. Financial incentives can be costly, risk harming intrinsic motivation (Kamenica, 2012), and are associated with high attrition rates (Cawley and Price, 2013). It is also difficult to calibrate the correct monetary amount. The take-up rate of commitment devices is low,[8] possibly because people are not precisely aware of their time preferences. Understanding how human behavior is driven by financially-related incentives is important, but are people only driven by money? A cost-effective program to increase physical activity will help both policy makers and individuals understand how to improve healthy behavior.

The design of this study is closely related to other studies and programs delivering normative information to "nudge" actions. One of these programs is the "home energy report"

[7]For instance, Charness and Gneezy (2009) find financial incentives have successfully helped ex-ante non-regular gym attendees form a habit of exercising more both during the intervention period and after the stimulus is removed, but there is a slight crowding out effect for ex-ante regular gym attendees. Volpp et al. (2008) document positive effects of lottery-based financial incentives and deposit contracts on weight loss among obese adults.

[8]In the setting of a large private company, Royer et al. (2015) find positive and lasting effects in gym attendance, but with a low take-up rate, a little over 20%.

generated by the company Opower, who cooperates with utility companies to send customized reports for the purpose of energy conservation. The energy reports include a comparison of energy use with households' neighbors and energy saving tips. Households who received the reports that compare them to their neighbors' significantly drop their energy use and reach a stable status in the long run (Allcott and Rogers, 2014). Similar desirable results are also found in household water use (Ferraro and Price, 2013, Brent et al., 2015). However, the utility bills offered to households in prior studies offer extra information like tips to reduce energy or water use, and the amount of money that can be saved, in addition to the social comparison information. The effects of the comparison information are confounded with either the additional information provided or the report itself (regardless of what is included). Our goal is to investigate whether individuals are affected by only receiving information about their peers, even if no financial incentive factors are involved. It is also worth to distinguish the potential differences in altruistic and health-related behaviors.

Among other studies that make use of intervention messages, Beshears et al. (2015) find that when employees are presented information about their co-workers' savings decisions, those who are originally not enrolled in the company's 401(k) plan decrease their contribution and those who chose a low contribution rate show a positive reaction to the information provision. Savings behavior, in some aspects, is similar to health behaviors. To save, one has to give up current consumption for a return accumulated over a long period of time; people have to incur consistent time cost to be physically active to see noticeable effects in the distant future. So, similar results might be expected in studies of physical activity and saving. However, being physically active can also generate shorter-run benefits like better sleep quality or relaxation, and hence might have higher marginal benefits than savings do.

Yun and Silk (2011) conduct a survey study to identify how different types of social norms and individual differences affect intrinsically beneficial behaviors. They refer "most xx university students" and "the majority of xx university students" for measures of distal norms and use "friends" (and others) to measure proximal norms. They find that proximal norms have greater effects on intention to exercise compared to distal norms, and distal and proximal norms have similar effects on intention to maintain a healthy diet. Beatty and Katare (2018) analyze results from a field experiment and find a significant positive effect from lottery-based financial incentives with high prizes, and little effect from social norms among college first-year students, using gym attendance as outcomes. However, because the information was provided through emails and their recruitment method was opt-out, it is likely students were not aware of the messages and the social norm information was ignored.

Our study combines the practice of text messaging and comparison-based information and examines the impact of information provision on number of daily steps taken. We also test for the existence of heterogeneity.[9]

The participants of our experiment were not restricted to undergraduate students. In fact, about half of the participants were university staff.[10] The outcomes of interest in our study use daily activity, rather than gym attendance as in most previous literature, allowing us to analyze behavior at a finer level of detail. We used high activity days per week as a secondary outcome. We use the standard difference-in-difference model as our primary analysis method. To investigate heterogeneity, stratified analysis were conducted according to gender, marital status, ex-ante level of daily number of steps etc. We also applied the unconditional exogenous quantile treatment effect method (QTE) to examine distributional effects of treatment.

To preview the results, we find that although a slight increase is detected during the intervention and shortly after the intervention period, it fades away and becomes negative three weeks after the incentive is removed. The effects are not significant in most of the observed time periods. There is also no clear evidence of the boomerang effect, which refers to an unintended consequence of the intervention resulting in an opposite reaction.[11] These results add additional evidence consistent with findings in Beatty and Katare (2018). The null results may be due to behavioral inertia, for both the groups whose baseline daily steps were above

[9] We were motivated to use text messages because of its application by health providers and medical services. Text messaging has been variously used to assist patients with management of chronic diseases (Liang et al., 2011), to remind people to comply with pharmacological procedures, and to support smoking cessation and weight loss maintenance (Gerber et al., 2009). Only a small group of researchers combine such technology with promotion of physical activity; although significant changes in physical activity have been detected, they are confounded by a lack of an attention control group (Buchholz et al., 2013).

[10] Using college students as experimental subjects should not cause problems for the external validity of a study (Druckman and Kam, 2009), particularly in the area of health (as compared to consumer behavioral research). However, it is of interest to study a more diverse population.

[11] For instance, individuals whose activity levels are originally above the average reduce their activity level.

and below the median level. For a habitually active individual, the marginal benefits are little, because intrinsic motivation is more important for maintaining such activity (Aaltonen et al., 2014). Marginal costs are high for those who are not active due to the discomfort of starting an exercise routine. We also failed to find significantly heterogeneous treatment effects.

3 The experiment

The experiment began in August 2018. To recruit participants, a screening survey was sent to all University students and staff. Those eligible were persons between the ages of 18 and 60 who did not have any history of eating disorders and had not been advised against exercising.[12] We met with eligible respondents to explain the consent process, help them complete an initial survey, give them the tracking device (hereafter, fitness tracker), and assist them in setting up an account. All participants were informed that they could use the fitness tracker as much as they wanted once the study began.[13][14] Participants were not required to return the fitness tracker at the end of the experiment.

We emailed some 750 respondents to schedule individual meetings. Approximately 280 responded, 155 showed up to the meeting, and 152 signed up to participate in the study. Participants did not represent a random sampling of students and stuff, as they consisted of a self-selecting group that volunteered to have their activity levels tracked and studied.

After agreeing to participate in the study, participants were randomly assigned to either a control group or a treatment group. Of the 152 participants that began the experiment, one dropped out soon after it began, and four exhibited negligible levels of participation. In addition, a second survey revealed that four participants did not receive text messages relevant to the experiment during the study period. Thus, the number of participants whose results were

[12] The recruitment email mentioned that the study would use fitness trackers and text messages.

[13] Financial incentives were not the main interest of this experiment. Therefore, there is no reward tied to the performance of the participants. However, the amount of time the participants used the fitness trackers did affect the quality of the data. We ran separate cash-prize drawings to encourage participants to wear the fitness trackers and regularly sync them with the corresponding phone app. Specifically, for each drawing period, participants who wore the tracker for at least two hours per day were eligible to be entered in the drawing pool and one winner was chosen to receive $100.

[14] Since recruitment took three weeks, some participants who signed up early started using the tracker early.

analyzed is 143.[15] The initial survey collected information about participants' demographics, physical activity levels, height and weight, time discount factors, hyperbolic discount factors, level of procrastination, and resistance to peer influences.[16]

The summary statistics for individual characteristics are in Table 1.[17] Panel B of Table 1 summarize some baseline (pre-treatment period) statistics. The final columns show the calculated difference between the two groups. The full list of questions in the initial survey and variables measured from it are shown in the appendix tables A.13 to A.19.

The groups were largely similar, although some differences were observed. For instance, participants in the experimental group were on average five years younger and more physically active. They incurred 139 more minutes of moderate activity and 20 more minutes of vigorous activity each week. These participants also walked 694 fewer steps each day than those in the control group. Nevertheless, differences in physical activity level and average daily steps are not statistically significant. By conducting a joint significance test for the discrepancies in these variables, we failed to reject the null hypothesis that the groups were the same.

Table 2 details the timeline of the experiment. Participants received two text messages reminding them of the study. The first was sent out one week before the study began, and the second the day before it began. Data collected over the first two weeks served as a baseline observation. This was followed by a four-week treatment period during which each participant received a text message at the same time each morning. This message contained information specific to that participant, and read as follows: *"Yesterday, you took 4,663 steps. You were ahead of 30 (out of 75) (40%) of your peers. If you walked 1,432 more steps, you would be ahead of 37 (50%) of your peers."* for treatment group participants, and " *Yesterday, you walked* 677 *steps. Please remember to sync your fitness tracker."* for control group participants. A second survey

[15]The null hypothesis is that there is no treatment effect. The alternative hypothesis is that the minimum expected treatment effect is roughly a ten-minute walk (approximately 1000 steps). The variance of outcomes across the population, derived through NHANES data, was approximately 19.55. We used the formula in List et al. (2011) and calculated the minimum statistically significant sample size for one group be about 60. Therefore, a sample size of 143 should be adequate.

[16]We used IPAQ (2002) and GPAQ (2002) to get estimates of minutes of self-reported physical activity. We intended to get individual levels of time discount factors and hyperbolic discount factors by selecting questions from Hardisty et al. (2013).

[17]Resistance to Peer Influence (RPI) is measured using a series of questions constructed in Steinberg and Monahan (2007) with scales from 1 to 4. A low score indicates that an individual is more susceptible to the influence of peers, and a higher one indicates greater resistance to such influence. The average score was 2.9.

was emailed to participants two weeks after the text message were no longer sent. Participants who completed the survey were mailed $10.

4 Results

4.1 Wearing behaviors

Participants were encouraged to wear the fitness tracker as much as they can so that sufficient data would be available for us. We can safely assume that most of the participants' steps were walked while they were wearing the trackers. After all, it is unlikely that they went for a walk without a tracker. We show descriptive statistics of wearing time for the two groups at different time periods in Table 3 and trend of wearing time by week in appendix figure A.1 to ensure that participants have worn an adequate amount of time for data analysis, and no significant difference between the two groups. The time trend of wearing time shows a significant drop in the time period after the second survey was sent, and lower than 5 hours after November. Therefore, even though we continued to collect the data till the end of December, our data analysis restricts to the end of week 16 (November 28th).

4.2 Baseline results

The key outcomes of interest are the number of steps taken per day, and number of days per week such that the daily steps exceeded 8000 steps.[18] Figure 1 graphically presents the outcomes of interest in different study periods for treatment and control groups.

The average daily number of steps taken for the control group slightly decreased from the pre-treatment period to the treatment period, from 9683 steps to 9343 steps, then increased slightly to 9438 steps in the post-treatment period, when no daily text messages were sent to the participants. After the second survey (follow-up survey) was taken, the average daily number of steps taken dropped to 8424, which was below the baseline level. The corresponding changes for the treatment group were from 9070 steps in the pre-treatment period to 9193

[18] 8000 steps was picked because it appeared to be the goal of the daily number of steps on the app of the fitness tracker.

steps in the treatment period, and increased further to 9981 steps during the post-treatment period. After the follow-up survey was taken, the average number of steps taken dropped to 9153 steps, which was almost the same as the pre-treatment level.

Figure 2 shows the trend of daily number of steps taken by week. It actually shows some slight evidence of persistent effects after the incentive was removed.[19]

We observe greater changes when using the number of days per week such that the daily number of steps was more than 8000 steps (Figure 2 (b)). FOr the control group, the value went down slightly from 3.9 days per week in the pre-treatment period to 3.6 days, and then 3.1 days in the post-treatment period and 1.6 days after the follow-up survey was taken. Correspondingly, for the treatment group, the average number of days per week such that the daily number of steps taken exceeded 8000 steps went up from 3.4 days in the pre-treatment period to 3.6 days in the treatment period, and went down to the baseline level — 3.4 days in the post-treatment period, and finally 1.5 days after the second survey was taken.

The first column of Table 4 shows the primary regression results for average treatment effects for total number of steps taken daily. [20] We used a standard difference-in-difference model to estimate the treatment effects. We show the average effects for the treatment, post-treatment periods, and the time period that was after the follow-up survey was taken. On average, the difference between the treatment group and the control group is an insignificant 324 steps total during the treatment period. The results for steps due to walking or running were just above 100 steps. The differences became greater for the post-treatment period (an insignificant increase of 629 steps), yet negative for the following 8 weeks that were after the second survey was taken. Overall, the effects decay after the intervention is removed. In terms of the steps taken due to brisk walking, the magnitudes are small: a close to 300 steps of brisk walking per day that represents about 3 minutes of moderate intensity activity per day or about

[19] The incentive refers to the provision of the social comparison information.

[20] Results using number of steps taken due to brisk walking, walking, or running are shown in Table A.8. In the minute-by-minute data, the column "mode" indicates different status for each minute of a user, like "shallow sleep," "walking," "running," etc. We use this information to obtain the number of steps/minutes due to walking, running, or brisk walking. The data includes a column "activeness" that indicates the active level for each minute of a user. The activeness ranges from 0 to about 150. The "activeness" of walking at a pace of 105 steps/minute (approximately the pace of brisk walking) is about 85, so we use the minutes when the active level is greater than 85 to get steps due to brisk walking.

20 minutes per week. [21]

Figure 3 displays graphical demonstration of the weekly treatment effects.. The weekly effects are consistently small and insignificant for the entire 16 weeks. The difference in daily average steps between the two groups for each week ranges from a little above 0 zeros to somewhere close to 1000 steps (week 5)After the intervention, the effects do not fade away right after the incentive is removed. The values of the effects remained at a moderate level then it dropped to below 0. All the magnitudes are insignificant. Overall, there is no consistent pattern that shows evidence of any significant treatment effects during the intervention period or after.

For the number of days per week when the number of daily steps taken exceeded 8000 steps, there is a weak evidence of positive treatment effects, which is shown in the second columns of Table 4. We used the same model specification, removing the duplicated observations as we used an outcome variable indicating weekly attributes. We see an increase of 0.5 high activity days per week. COnsistent with the pattern of the effects for daily number of steps taken, this difference between the became greater for the post-treatment period, which was 0.8 days (at 5% significance level). The effect dropped to an insignificant 0.48 days for the period after the second survey was taken, close to the level at the treatment period. The weekly treatment effects, shown in Figure 4, show a consistent positive pattern, unlike that for the daily number of steps taken.

In addition to the quantitative results, the appendix tables A.20 — A.23 report the full results of the follow-up survey. Survey results show that generally, exercise is more of an intrinsically rather than an extrinsically motivated behavior. Individuals value the benefits of being physically active such as being healthy and having a better body image. Participants also suggested that they would be more motivated by financial incentives, goal-setting, and support of known peers.

[21]Using the minute-by-minute data we are able to distinguish active minutes during which participants are walking or running. An insignificant difference of 5 minutes of total active minutes (minutes due to walking or running) as a dependent variable (the last column of Table A.8) is observed between the control and the treatment groups during the treatment period. During the post-treatment period, this difference decreased, then turned to negative afterwards. All the magnitudes are comparatively small and insignificant. Therefore, we are not able to conclude any treatment effects during any time period.

4.3 Heterogeneity in the effects

Next, we investigate the treatment effects for different groups, using daily number of steps taken as the outcome variable. First, when we consider how the text messages are associated with the daily number of steps taken among the participants in the treatment group, we use the elements "your steps" and "your rank" that were contained in the text messages as the main independent variables, and only the participants in the treatment group during the treatment period to run a simple linear regression models. Results are shown in Table A.10. We see a positive correlation between a person's rank of the previous day and the number of steps taken. Therefore we divided the sample into groups based on their baseline (pre-treatment period) level of number of steps taken. We also divided the sample into groups based on gender, procrastination level (using the IPS scale), and marital status (married or cohabiting).

Table 5 shows the stratified sample regression results. We unfortunately did not find significant treatment effects for any of the groups. Qualitatively, a female, an individual whose number of daily steps taken was at bottom 20 or bottom 50 percentile within her group, an individual who is married or is living with partner walks fewer steps per day if she is treated. In contrast, an individual who walks more than the sample average during the pre-treatment period walks more steps if she is treated.

We do see a similar pattern in the change of the treatment effects for all groups (and the baseline results from the previous subsection) that the treatment effects, although statistically insignificant, increased even after the incentive was removed, which seems to be consistent with the results in Charness and Gneezy (2009).

4.4 Sample selection

Since we use fitness trackers to collect individual daily number of steps taken, if a participant does not wear the tracker for a day, her number of steps for that day would be missing. The regression equation should represent all adults regardless a person is wearing the tracker or not. The data on the primary dependent variable – daily number of steps taken – is observable only for participants who choose to wear the fitness tracker. Individuals chose to use the fitness trackers and that directly determines whether or not the dependent variable is available. Therefore, the sub-sample we use is selected on the basis of the response variable.

Appendix figure A.3 and table A.9 show that the incidental truncation (people do not disappear from the panel but variables are observable for some periods) problem appear. When the sample selection is not systematically related to the outcome of interest, a standard fixed effects analysis is consistent, even if attrition or non-commitment is correlated with observable or unobserved individual time-unvarying characteristics. However, sample selection in a fixed effects context is a problem if selection is related to the errors.

We therefore test for sample selection bias. This can be done by extending Heckman's test to the unobserved effects panel data context.

The equation of interest is

$$y_{it1} = x_{it1}\beta_1 + c_{i1} + u_{it1}, \qquad t = 1, \cdots, T \tag{1}$$

where y_{it1} is the daily number of steps taken and is not observable for each day. We use x_{it} to denote a set of exogenous variables at t. All elements of x_{it} are time varying and are assumed to be observable in every time period in theory. Here the time period is restricted to the treatment period and the unit of t is a day. We use c_i to represent a set of variables that are individual specific and time-invarying.

Let s_{it} as an indicator that equals 1 if (x_{it}, y_{it}) is observed. For each t,

$$s_{it2} = \mathbb{1}[x_{it}\delta_2 + c_{i2} + a_{it2} > 0], \qquad a_{it2}|(x_i, c_i1, c_i2) \sim Normal(0, 1) \tag{2}$$

where

$$y_{it}^{*} = max(0, x_{it}\delta_2 + c_{i2} + v_{it2}), \qquad v_{it2}|x_i \sim Normal(0, \sigma_{t2}^2),$$

where y_{it1} is observed if $y_{it}^{*} > 0$, and x_{it1} is assumed to be a subset of x_{it}.

Specifically, the two-step procedure to test sample selection bias is as following:

1. Estimate a pooled probit model (2). The variable y^*_{it} is wearing hours/day. Since in theory it assumes x_{it} to be individually specific, time-varying, and also exogenous. I use the following variables for the set of x_{it}: wearing hours and the lagged term of wearing hours, percentage of peers an individual is ahead of within her group the previous day and the next day.[22] I use treatment condition and other individual characteristics (baseline observations, demographics)[23] to represent c_i (binary variable). In addition, individual dummy variables and dummy variables for weeks are controlled. Therefore, the specification of equation (2) is:

$$\begin{aligned} s_{it2} = \mathbb{1}[&\delta_1 WearHrs_{it} + \delta_2 Percentage_{i,t-1} + \\ &\delta_3 WearHrs_{i,t+1} + \delta_4 Percentage_{i,t+1} + \\ &\delta_5 Treatment_i + \delta_6 Char_i + \delta_{7,i} + \delta_{8,week} + a_{it2} > 0]; \\ &a_{it2}|(x_i, c_i1, c_i2) \sim Normal(0,1). \end{aligned} \tag{3}$$

Using (3), We obtain the inverse Mills ratio $\hat{\lambda}$.

2. Regress y_{it2} on x_{it2} and the inverse Mills ratio from the first steps ($\hat{\lambda}$) on selected sample (that is, if $s_{it2} = 1$) to obtain estimates in the equation (2). I use a subset of x_{it1} for x_{it}.

   The specification of equation (2) is:

$$\begin{aligned} Steps_{it} = &\beta_1 WearHrs_{it} + \beta_2 Percentage_{i,t-1} + \\ &\beta_3 Treatment_i + \beta_4 Char_i + \beta_{5,i} + \beta_{6,week} + \\ &\boldsymbol{\rho_1 \hat{\lambda}} + u_{it1}; \end{aligned} \tag{4}$$

A t test of $H_0 : \rho_1 = 0$ is a test of the null hypothesis of no sample selection. If the sample selection bias does exist, the estimate β_4 in the above model will be the interested estimate: the treatment effect during the period that we apply the above two steps.

[22]The choice of variables for x_{it} depend on available variables.

[23]I use age, gender, weight, average wearing time of the tracker per day and average daily steps during the pre-treatment period as the covariates.

We use the data as the pooled cross-sectional data. Using the above procedure we could test if there is sample selection and if the decision of selection is affected by the treatment assignment.

The stage two result indicate a statistically insignificant estimate for the Inverse Mills Ratio, therefore we fail to reject the null hypothesis that the sample selection bias is not an issue during the treatment period.[24]

4.5 Quantile treatment effects

We estimate the quantile treatment effects (QTE) of receiving peer-comparison information on number of daily steps walked. The estimation of QTE yields insights of the impacts of independent variables on the entire distribution of the dependent variable. Using quantile regressions also allows us to deal with sensitivity to outliers (Frölich and Melly, 2008) and relax the assumptions of the normal distribution. On the other hand, focusing on mean effects may average positive and negative responses and miss important heterogeneous treatment effects.[25]

The QTE allows us to examine if the peer-comparison information with the same format affects the entire distribution of number of daily steps walked equally. The reported mean effects may differ conditionally on whether the whole activity-level-distribution shifts or whether the changes are concentrated at the bottom, top, or middle of the distribution. For an individual who is already active, her response to comparison information is unclear from the mean-effects model. She might be willing to walk as much the next day because of the positive effects made by the information, or be less likely to walk as much because of the misperception of her peers' activity level.

[24]The results for stage 1 and stage 2 are shown in the appendix table A.12

[25]Recent studies have shown advantages of applying QTE compared to studies relying on mean effects under some contexts. To evaluate the economic effects of a welfare program, Bitler et al. (2006) estimate QTE of welfare reform across the distributions of wealth and find no effects at the bottom of the distribution, positive in the middle, and negative effects at the top. Although existing studies show mixed or little mean effects, the results generated by quantile regressions, proving that the Connecticut's JObs First program have increased income for a large group of women (Bitler et al., 2006). Andrews et al. (2012) estimate QTE of college quality on earnings and demonstrate a negative effect on earnings overall but a positive effect at the top of the earnings distribution. Their findings suggest that mean effects miss significant uncertainty of the returns for any given student and public subsidies to higher education do not benefit students equally.

We use specifically unconditional QTE with random treatment assignment reviewed in Frölich and Melly (2008). The unconditional QTE focuses on features of the outcome distribution, not conditional on explanatory variables.[26] The unconditional effects are not a function of covariates. If we are interested in the bottom of the distribution, unconditional QTE allows us to summarize the effects with a low absolute quantile.[27] Estimating the treatment effects (on the treated) at a 25th quantile is to estimate the difference between the 25th quantile in the treatment and the control group not conditional on certain individual characteristics.

We explain the weighting method we use for QTEs in the appendix. Table 7 shows the magnitudes of daily steps for the control and the treatment group and their differences. The periods are defined in the same manner as the baseline results present. We first divide the entire study period into a pre-treatment period and any time after that. We then investigate the differences between the two groups for other sub-periods. Overall, the distributions for both the control and the treatment groups are stable with a little outward shift across time. Within each period, the treatment effects are mostly significantly positive for each decile. The magnitudes of the effects do not differ largely across the distribution, with a few hundred steps for each decile. The treatment effects reach a minimum between the second and the third deciles. During the treatment period, slightly greater treatment effects are shown at the two tails of the distribution. The effects show a sign of fading away after the incentive is removed (week 9-12). Figure 5 plots the steps QTE for the corresponding periods. Those whose number of steps are at the top or the bottom are slightly more affected, resulting in greater treatment effects. All the effects are small in magnitude but significant at the 10% level. The overall treatment effects range from 500 steps to 1000 steps, implying a 5 - 10 minutes more walking per day. The mean effects shown in the graph are consistent with the mean effects from the baseline model, except that these averages are significant at the 10% level during the treatment period, week 3 through 12, and week 7 to 8. The appendix table A.11 and figure A.4 reports

[26]On the other hand, the conditional QTE recovers features of the conditional outcome distribution. The conditional quantiles are more difficult to interpret (Andrews et al., 2012).

[27]The conditional QTE is different from the unconditional QTE in that a low absolute quantile might be at a comparative high quantile conditional on X. An advantage of unconditional QTE estimators compared to conditional QTE estimators is they can be estimated consistently without any parametric restrictions (Frölich and Melly, 2008). Another advantage is including covariates that are independent from the treatment can change limit of the estimated conditional QTEs (Frölich and Melly, 2008).

the deciles of daily average steps per week for each group and the weekly QTE.

5 Conclusion and discussion

In this paper, we investigated the pattern of daily physical activity level measured by number of daily steps in response to social-comparison information. We conducted a field experiment, distributed fitness trackers to collect data, and used text messages for disseminating information. The purpose of the study is first to rule out any financial incentives and second to evaluate if information about unknown peers effectively influence one's own daily physical activity.

Overall, we find no clear evidence of average treatment effects neither during the treatment period nor the post-treatment period using the difference-in-difference specification. The weekly average treatment effects only show incidental positive effects for a week or two, and the effects do not show a consistent pattern for different time periods. The results add more evidence to the results from a prior study by Beatty and Katare (2018), where social norms have little effects in promoting gym attendance among college first-year students. However, the results from the QTE show statistically significant (at 10%) mean effects of about 300 steps per day during the intervention period, and greater mean effects (600) two weeks after the incentive is removed. The QTEs do not show great variation across the distribution of the dependent variable. There is a slightly greater effect at the top and the bottom of the distribution. The middle of the distribution is more likely to be affected after the incentive is removed.

Several other intervention studies designed to improve physical activity level have measured walking as an outcome. Murtagh et al. (2010) summarize that pedometer-, mobile phone- and computer-based programs that use walking requirements as a direct intervention method are effective for increasing walking levels (but the magnitudes are not specified). University employees, tasked with increasing workday brisk walking, add 6-10 minutes of walking per day compared to their baseline level as a result of an automated, web-based walking intervention.[28] The minute-by-minute data of our study shows an increase of 5 minutes walking time (not statistically significant) which is close to the range of the results of prior intervention

[28]The study lacks control group.

studies. In addition, we test heterogeneity using a stratified analysis and also find no clear evidence of different effects across groups of people with different characteristics.

We also find attrition among the participants. Appendix figure A.2 panel (a) shows the number of dropouts in each group and panel (b) shows the proportion of participants who wear the trackers in each group. Compliance behavior is not the main concentration of the study, but this phenomenon implies a downside of using fitness trackers as a tool for data collection. Individuals may be more likely to wear the fitness tracker when they know they are going to exercise (Polgreen et al., 2018).

To compare the effects of social comparison with other behaviors that are significantly affected by such incentive, for instance, energy conservation in Allcott and Rogers (2014). Firstly, physical activities do not have an immediate return or reward. The return to perseverance exercise can only be shown later in life, but remembering to turn off lights or AC can save money for current consumption. Also, behaviors like remembering to turn off AC or lights takes little time or exert little physical costs, while physical activities yield significant time cost and physical discomfort. Social comparison is mostly found effective in the peer effects literature because subjects are usually compared with people they know, or a fixed group of people will have direct social interactions, it is more likely that the social pressure generated under such contexts have more effects. On the other hand, the participants of our study are compared with someone they do not know. Finally, compared to other physical activity or health-related behavior studies, our group does not have much room for improvement. Selection bias is a problem. The survey results support the above hypothesis since most of people responded that they are already physically active. While being asked what would motivate them more, many mentioned goal setting, more tips or guides offered, supports from known peers, or financial incentives are more preferred as intervention strategies.

References

Aaltonen, S., Rottensteiner, M., Kaprio, J. and Kujala, U. M. (2014), 'Motives for physical activity among active and inactive persons in their mid-30s', *Scandinavian journal of medicine*

& science in sports **24**(4), 727–735.

Allcott, H. and Rogers, T. (2014), 'The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation', *American Economic Review* **104**(10), 3003–37.

Andrews, R. J., Li, J. and Lovenheim, M. F. (2012), Quantile treatment effects of college quality on earnings: Evidence from administrative data in texas, Technical report, National Bureau of Economic Research.

Angrist, J. D. (2014), 'The perils of peer effects', *Labour Economics* **30**, 98–108.

Austin, P. C. (2011), 'An introduction to propensity score methods for reducing the effects of confounding in observational studies', *Multivariate behavioral research* **46**(3), 399–424.

Bandura, A. (2006), 'Guide for constructing self-efficacy scales', *Self-efficacy beliefs of adolescents* **5**(1), 307–337.

Beatty, T. K. and Katare, B. (2018), 'Low-cost approaches to increasing gym attendance', *Journal of Health Economics* .

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C. and Milkman, K. L. (2015), 'The effect of providing peer information on retirement savings decisions', *The Journal of finance* **70**(3), 1161–1201.

Bitler, M. P., Gelbach, J. B. and Hoynes, H. W. (2006), 'What mean impacts miss: Distributional effects of welfare reform experiments', *American Economic Review* **96**(4), 988–1012.

Brent, D. A., Cook, J. H. and Olsen, S. (2015), 'Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials', *Journal of the Association of Environmental and Resource Economists* **2**(4), 597–627.

Buchholz, S. W., Wilbur, J., Ingram, D. and Fogg, L. (2013), 'Physical activity text messaging interventions in adults: a systematic review', *Worldviews on Evidence-Based Nursing* **10**(3), 163–173.

Carlson, S. A., Fulton, J. E., Pratt, M., Yang, Z. and Adams, E. K. (2015), 'Inadequate physical activity and health care expenditures in the united states', *Progress in cardiovascular diseases* **57**(4), 315–323.

Carr, L. J., Karvinen, K., Peavler, M., Smith, R. and Cangelosi, K. (2013), 'Multicomponent intervention to reduce daily sedentary time: a randomised controlled trial', *BMJ open* **3**(10), e003261.

Carrell, S. E., Hoekstra, M. and West, J. E. (2011), 'Is poor fitness contagious?: Evidence from randomly assigned friends', *Journal of Public Economics* **95**(7-8), 657–663.

Cawley, J. and Price, J. A. (2013), 'A case study of a workplace wellness program that offers financial incentives for weight loss', *Journal of health economics* **32**(5), 794–803.

CDC (n.d.), 'Exercise or physical activity, national center for health statistics'. https://www.cdc.gov/nchs/fastats/exercise.htm[Accessed:2017-08-29].

Charness, G. and Gneezy, U. (2009), 'Incentives to exercise', *Econometrica* **77**(3), 909–931.

Courtemanche, C., Heutel, G. and McAlvanah, P. (2015), 'Impatience, incentives and obesity', *The Economic Journal* **125**(582), 1–31.

Danaei, G., Ding, E. L., Mozaffarian, D., Taylor, B., Rehm, J., Murray, C. J. and Ezzati, M. (2009), 'The preventable causes of death in the united states: comparative risk assessment of dietary, lifestyle, and metabolic risk factors', *PLoS medicine* **6**(4), e1000058.

Druckman, J. N. and Kam, C. D. (2009), 'Students as experimental participants: A defense of the'narrow data base'. ssrn elibrary'.

Ferraro, P. J. and Price, M. K. (2013), 'Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment', *Review of Economics and Statistics* **95**(1), 64–73.

Frölich, M. and Melly, B. (2008), 'Estimation of quantile treatment effects with stata'.

Gerber, B. S., Stolley, M. R., Thompson, A. L., Sharp, L. K. and Fitzgibbon, M. L. (2009), 'Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study', *Health informatics journal* **15**(1), 17–25.

GPAQ (2002), 'global physical activity surveillance'.
URL: *http://www.who.int/ncds/surveillance/steps/GPAQ*

Hardisty, D. J., Thompson, K. F., Krantz, D. H. and Weber, E. U. (2013), 'How to measure time preferences: An experimental comparison of three methods', *Judgment & Decision Making* **8**(3), 236–250.

Harvard Health Publications (n.d.), 'Walking: your steps to health'. https://www.health.harvard.edu/staying-healthy/walking-your-steps-to-health[Accessed:2017-08-29].

IPAQ (2002), ' international physical activity questionnaire'.
URL: *http://www.ipaq.ki.se*

John, L. K., Loewenstein, G., Troxel, A. B., Norton, L., Fassbender, J. E. and Volpp, K. G. (2011), 'Financial incentives for extended weight loss: a randomized, controlled trial', *Journal of general internal medicine* **26**(6), 621–626.

Kamenica, E. (2012), 'Behavioral economics and psychology of incentives', *Annu. Rev. Econ.* **4**(1), 427–452.

Kling, J. R., Liebman, J. B. and Katz, L. F. (2007), 'Experimental analysis of neighborhood effects', *Econometrica* **75**(1), 83–119.

Lazear, E. P. (2001), 'Educational production', *The Quarterly Journal of Economics* **116**(3), 777–803.

Liang, X., Wang, Q., Yang, X., Cao, J., Chen, J., Mo, X., Huang, J., Wang, L. and Gu, D. (2011), 'Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis', *Diabetic medicine* **28**(4), 455–463.

List, J. A., Sadoff, S. and Wagner, M. (2011), 'So you want to run an experiment, now what? some simple rules of thumb for optimal experimental design', *Experimental Economics* **14**(4), 439.

Marcus, B. H., Forsyth, L. H. et al. (2009), *Motivating people to be physically active.*, number Ed. 2, Human Kinetics.

McKenzie-Mohr, D. and Schultz, P. W. (2014), 'Choosing effective behavior change tools', *Social Marketing Quarterly* **20**(1), 35–46.

Murtagh, E. M., Murphy, M. H. and Boone-Heinonen, J. (2010), 'Walking–the first steps in cardiovascular disease prevention', *Current opinion in cardiology* **25**(5), 490.

Perkins, H. W. and Berkowitz, A. D. (1986), 'Perceiving the community norms of alcohol use among students: Some research implications for campus alcohol education programming', *International journal of the Addictions* **21**(9-10), 961–976.

Polgreen, L. A., Anthony, C., Carr, L., Simmering, J. E., Evans, N. J., Foster, E. D., Segre, A. M., Cremer, J. F. and Polgreen, P. M. (2018), 'The effect of automated text messaging and goal setting on pedometer adherence and physical activity in patients with diabetes: A randomized controlled trial', *PloS one* **13**(5), e0195797.

Royer, H., Stehr, M. and Sydnor, J. (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.

Sacerdote, B. (2011), Peer effects in education: How might they work, how big are they and how much do we know thus far?, *in* 'Handbook of the Economics of Education', Vol. 3, Elsevier, pp. 249–277.

Steel, P. (2010), 'Arousal, avoidant and decisional procrastinators: Do they exist?', *Personality and Individual Differences* **48**(8), 926–934.

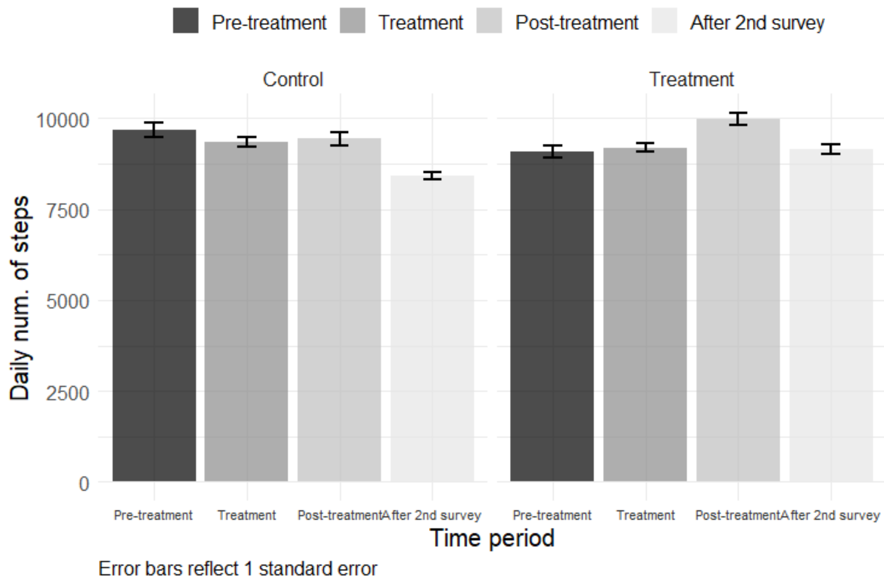
Steinberg, L. and Monahan, K. C. (2007), 'Age differences in resistance to peer influence.', *Developmental psychology* **43**(6), 1531.

Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J. and Loewenstein, G. (2008), 'Financial incentive–based approaches for weight loss: a randomized trial', *Jama* **300**(22), 2631–2637.

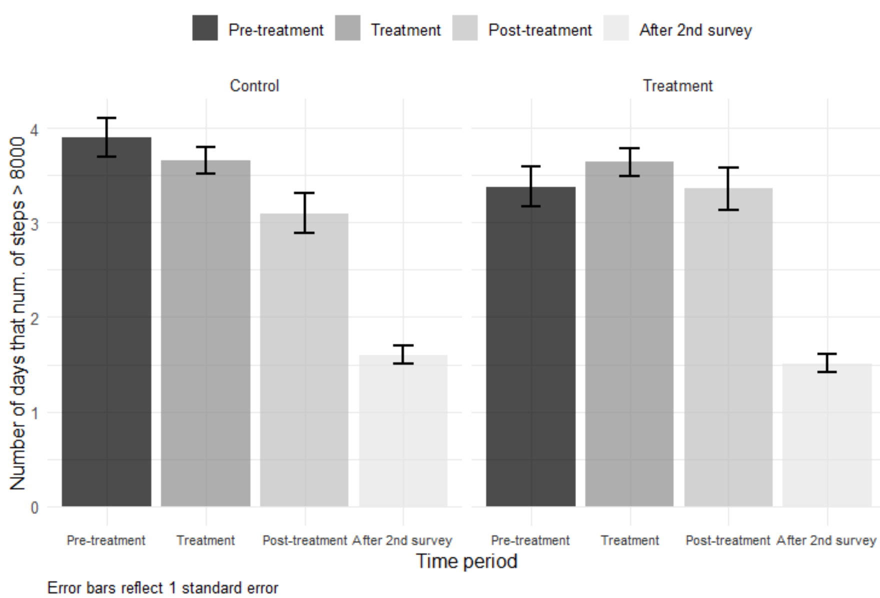
World Health Organization (2009), 'Global health risks: Mortality and burden of disease attributable to selected major risks.'.

Yun, D. and Silk, K. J. (2011), 'Social norms, self-identity, and attention to social comparison information in the context of exercise and healthy diet behavior', *Health communication* **26**(3), 275–285.

Figure 1: Outcomes of interest.

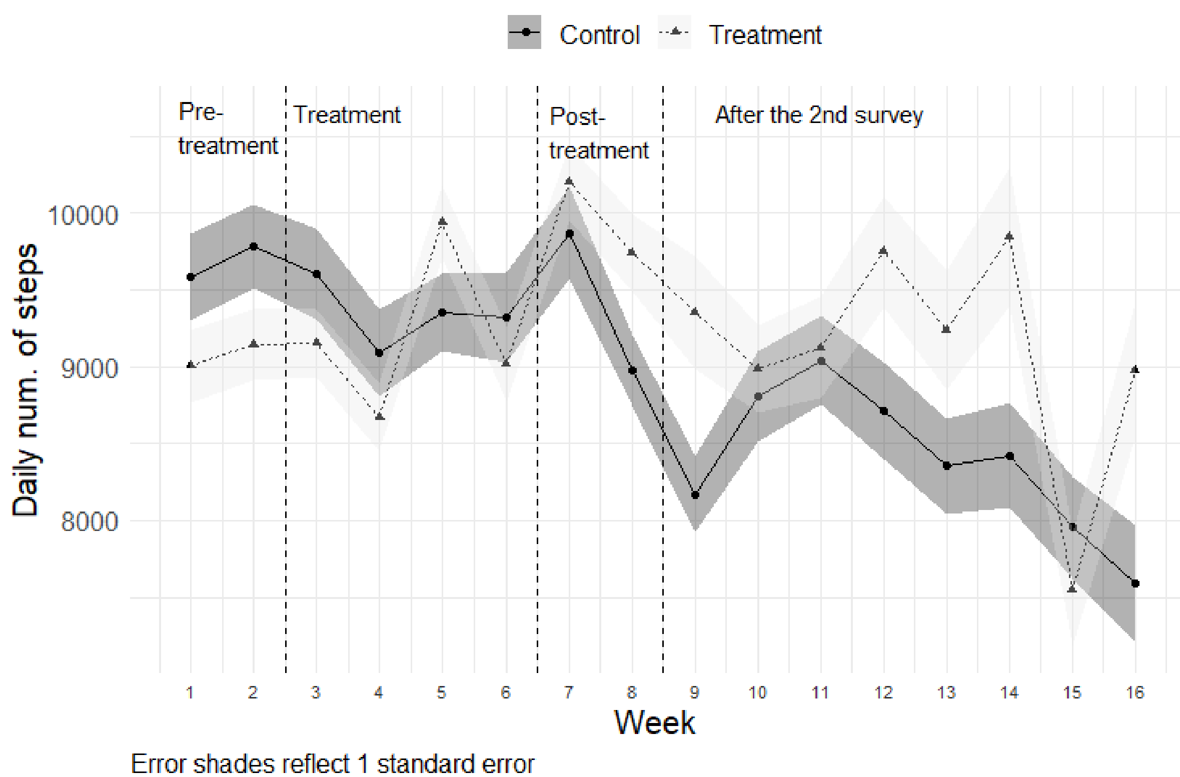


(a) Average daily steps



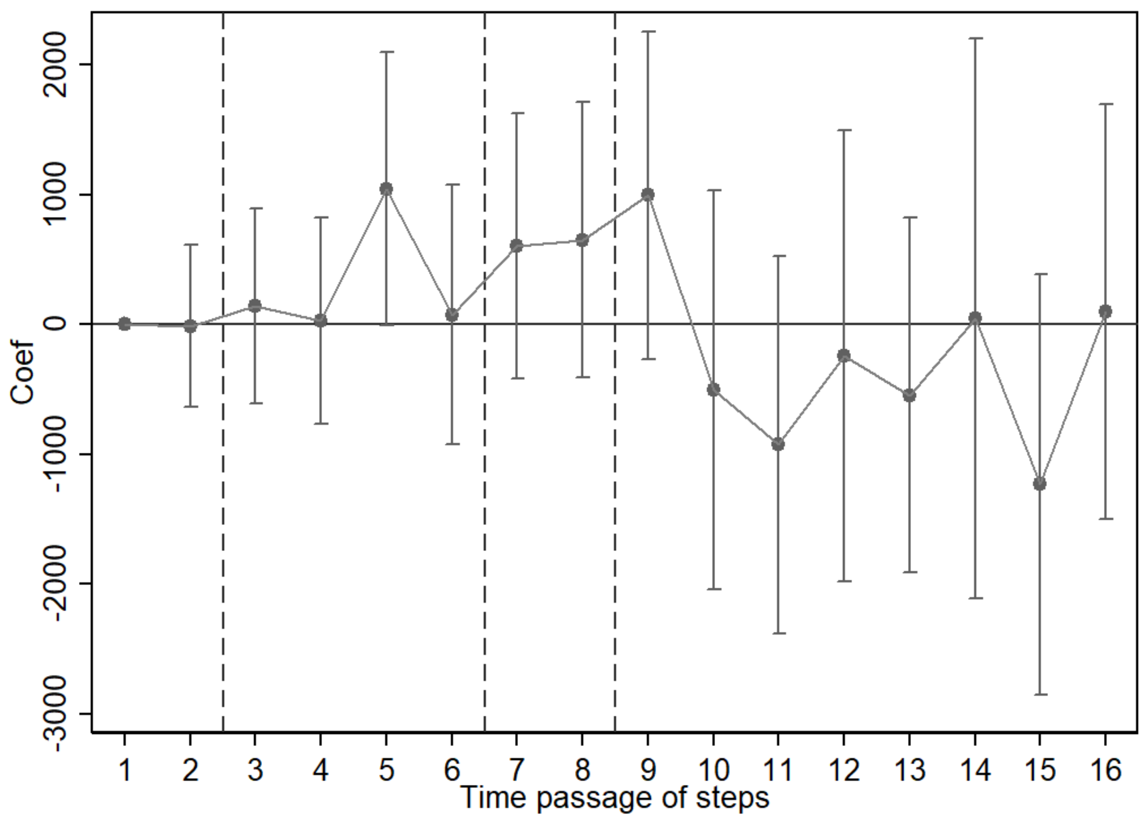
(b) Num. of days per week such that daily steps ≥ 8000

Figure 2: Daily num. of steps for each group by week



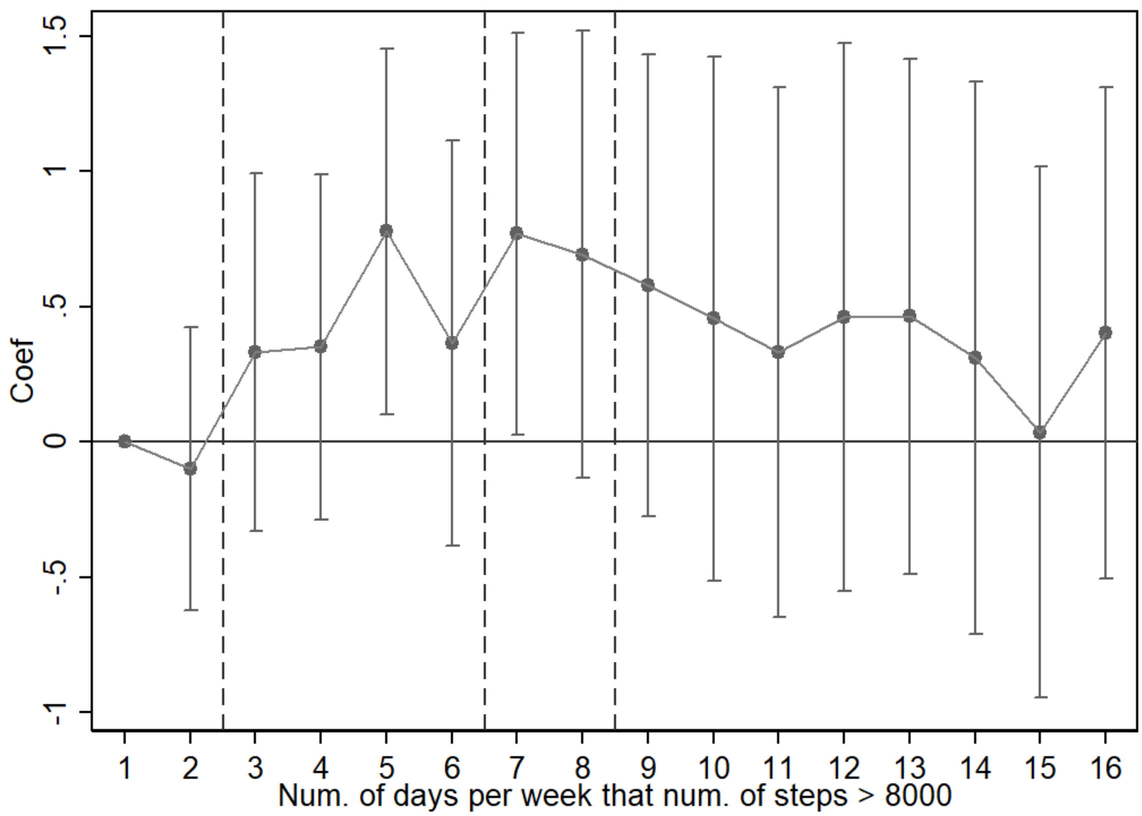
Notes: The graph shows the weekly average of daily steps for 16 weeks.

Figure 3: Weekly treatment effects for daily number of steps taken



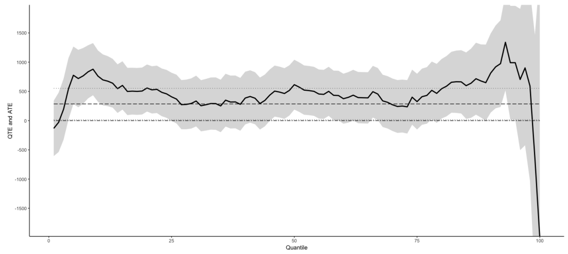
Notes: Vertical ranges represent 95% confidence intervals.

Figure 4: Weekly treatment effects for number of days per week such that the daily number of steps taken $\geq$ 8000.

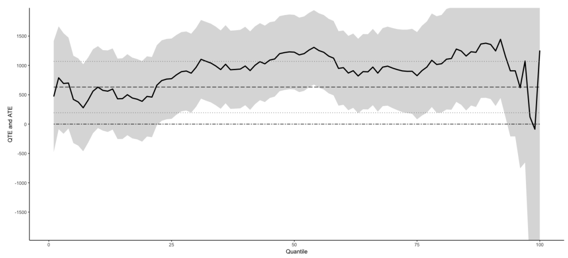


Notes: Vertical ranges represent 95% confidence intervals.

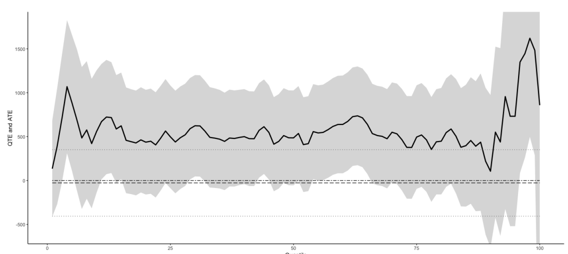
Figure 5: Quantile treatment effects by different periods



(a) Treatment period (week 3 - 6)



(b) Post-treatment period *before* follow-up survey (week 7-8)



(c) Post-treatment period *after* follow-up survey (week 9-12)

Notes: The solid line is the QTE; the shaded area provides bootstrapped 90% confidence intervals; the dashed line is the ATE using inverse probability weighting, the dotted lines are 95% confidence intervals for the ATE.

Table 1: Descriptive statistics

	Treatment		Control		Control - Treatment
	Mean	Obs	Mean	Obs	
A. Survey results					
Age	31.13	70.00	35.71	68.00	4.58**
Female	0.69	71.00	0.68	68.00	-0.01
College and above	0.69	72.00	0.69	71.00	-0.00
Married (or live with partner)	0.47	72.00	0.48	71.00	0.01
Height (meters)	1.70	71.00	1.69	69.00	-0.00
Weight (kg)	74.03	71.00	74.27	68.00	0.24
BMI	25.69	71.00	25.79	68.00	0.11
Want to weigh less	0.74	72.00	0.70	71.00	-0.03
Overweight (Self perceived)	0.53	72.00	0.46	71.00	-0.06
Resistance to Peer Influence	2.99	71.00	2.95	69.00	-0.04
Irrational procrastination scale score	24.21	71.00	23.03	69.00	-1.18
Ideal steps	9916.18	68.00	9405.76	66.00	-510.42
Ideal weight (kg)	68.28	56.00	69.19	52.00	0.92
Total (mod. minutes)	808.73	71.00	668.88	69.00	-139.86
Total (vig. minutes)	238.31	71.00	217.39	69.00	-20.92
Total PA MET	5141.41	71.00	4414.64	69.00	-726.77
Weekday *in hours*	6.67	71.00	6.35	67.00	-0.32
Weekend *in hours*	5.89	71.00	5.59	68.00	-0.31
Total sedentary/day	6.12	71.00	5.73	67.00	-0.38
B. Baseline stats					
Pre-treatment period					
Avg. of daily num. of steps	8971.56	72.00	9666.39	71.00	694.83
Num. of days that num. of steps $\geq$ 8000	3.38	72.00	3.90	71.00	0.52
First week					
Avg. of daily num. of steps	8726.11	70.00	9588.65	71.00	862.54
Num. of days that num. of steps $\geq$ 8000	3.36	72.00	3.83	71.00	0.47
Second week					
Avg. of daily num. of steps	9127.82	72.00	9788.24	70.00	660.42
Num. of days that num. of steps $\geq$ 8000	3.40	72.00	3.97	71.00	0.57

Notes: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. The table shows the mean values in individual characteristics for the two groups and the difference between them. IPS: Irrational procrastination scale (Steel, 2010); RPI: Resistance to peer influence (Steinberg and Monahan, 2007).

Table 2: Timeline of the experiment

Recruitment	
July 18-Aug.8	Recruitment: individual meetings with potential participants: enrollment, initial survey
Aug. 4th	A reminder text message that the study would begin the next day was sent to enrolled participants.
Aug. 8th	Another text message that the study begin day sent to all participants.
Pre-treatment period	
Aug. 9 — Aug.22	Pre-treatment period (baseline observation.)
Treatment period	
Aug.23 — Sep.19	Four-week intervention period: daily text messages to both control and the treatment group.
Post-treatment period	
Sep. 20 — Oct. 3	Two-week post-treatment period
Oct.4 — Dec. 25	A follow-up survey sent through emails on Oct. 4th.
Dec. 25	End of fitness tracker's data collection.

Notes: During the recruitment period participants who signed up early started using the fitness tracker.

Table 3: Descriptive statistics of wearing time for two groups at different time periods

	Treatment		**Control**		**Diff.**
	Mean	Obs	Mean	Obs	
Num. of days that wearing hours $\geq$ 2 hours					
Pre-treatment	13.17	72.00	13.45	71.00	0.28
Treatment	25.69	72.00	26.07	71.00	0.38
Post-treatment	10.61	72.00	11.32	71.00	0.71
After 2nd survey	21.57	72.00	25.86	71.00	4.29
Average wearing hours per day					
Pre-treatment	20.43	72.00	20.08	71.00	-0.35
Treatment	20.29	72.00	20.25	71.00	-0.04
Post-treatment	20.32	63.00	20.04	65.00	-0.28
After 2nd survey	18.93	54.00	19.49	56.00	0.56

Notes: *p<0.1; **p<0.05; ***p<0.01. The table shows the mean values in variable indicating wearing behaviors for the two groups and the difference between them. Missing values were excluded from the calculation.

Table 4: Difference-in-difference regression results for different time periods

	Daily num. of steps taken	Days per week that daily num. of steps $\geq$ 8000
Treatment	324.554	0.538*
	(289.082)	(0.232)
Post-treatment	629.704	0.866*
	(443.969)	(0.333)
After 2nd survey	-204.037	0.480
	(603.724)	(0.391)
Cons	9116.403***	2.305***
	(149.035)	(0.131)
R^2	0.433	0.657
N	10425	2288

Notes: Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001.

Table 5: Stratified DID regression results I

	Female	Male
Treatment	77.525	845.259
	(336.607)	(554.645)
Post-treatment	474.690	1247.529
	(396.079)	(1096.551)
After 2nd survey	-572.072	535.200
	(589.631)	(1285.159)
Cons	8849.832***	9727.651***
	(159.538)	(304.376)
R^2	0.406	0.458
N	6828	3359
	Top 20 percentile baseline	Bottom 20 percentile baseline
Treatment	590.257	-33.620
	(771.929)	(449.098)
Post-treatment	926.991	-273.648
	(1457.114)	(1052.314)
After 2nd survey	31.806	-567.271
	(1402.992)	(1518.769)
Cons	13088.117***	5971.911***
	(461.209)	(221.293)
R^2	0.394	0.344
N	2253	1754
	Top 50 percentile baseline	Bottom 50 percentile baseline
Treatment	477.377	225.573
	(441.623)	(334.108)
Post-treatment	839.486	550.465
	(669.989)	(564.680)
After 2nd survey	567.161	-289.505
	(873.946)	(627.128)
Cons	10927.546***	7014.280***
	(238.139)	(148.998)
R^2	0.372	0.293
N	5355	4999

Notes: Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001.

Table 6: Stratified DID regression results II

	Proactive	Procrastinator
Treatment	297.509	284.904
	(419.343)	(388.593)
Post-treatment	996.789	221.884
	(611.800)	(606.756)
After 2nd survey	-263.464	-227.153
	(715.707)	(926.850)
Cons	9713.652***	8622.199***
	(185.387)	(224.346)
R^2	0.522	0.306
N	4878	5547
	Never married = 1	Never married = 0
Treatment	963.593	-194.544
	(488.648)	(339.036)
Post-treatment	1987.188*	-442.105
	(839.707)	(424.058)
After 2nd survey	1008.759	-1251.433
	(1071.582)	(705.813)
Cons	9337.531***	8922.888***
	(292.650)	(147.012)
R^2	0.426	0.445
N	4645	5780

Notes: Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001.

Table 7: Average daily steps per week for each decile by group

		Percentiles								
Week	Group	10	20	30	40	50	60	70	80	90
After treated	Control	3097	4837	6253	7597	8695	9902	11298	12780	15134
(Week 3-12)	Treatment	3754	5305	6762	8136	9351	10608	11757	13390	16013
	Diff.	657	468	509	539	656	706	459	610	879
Treatment period	Control	2819	4706	6094	7557	8623	9941	11258	12816	15184
(Week 3-6)	Treatment	3673	5264	6499	7978	9325	10483	11690	13390	15862
	Diff.	854	558	405	421	702	542	432	574	678
Post-treatment period	Control	3197	4990	6334	7633	8738	9801	11234	12628	15048
(Week 7-12)	Treatment	3844	5396	7018	8362	9521	10781	11841	13392	16099
	Diff.	647	406	684	729	783	980	607	764	1051
Post-treatment 1	Control	3633	5446	6593	7825	9036	10243	11568	13192	15534
(Week 7-8)	Treatment	4263	5918	7561	8815	10262	11205	12523	14296	16891
	Diff.	630	472	968	990	1226	962	955	1104	1357
Post-treatment 2	Control	2917	4758	6163	7408	8441	9601	10901	12297	14638
(Week 9-12)	Treatment	3475	5195	6787	7908	8927	10240	11451	12745	14744
	Diff.	558	437	624	500	486	639	550	448	106

Notes: The percentiles are calculated using the inverse probability of treatment weighting using the propensity score (IPTW). The Appendix explains how the propensity score and the weights are derived.

Appendix

Unconditional QTE

The propensity score is defined as the probability of treatment assignment conditional on observed baseline covariates: $p_i = Pr(Z_i = 1|X_i)$. Propensity scores are used to balance the distribution of baseline covariates between treated and untreated subjects (Austin, 2011).

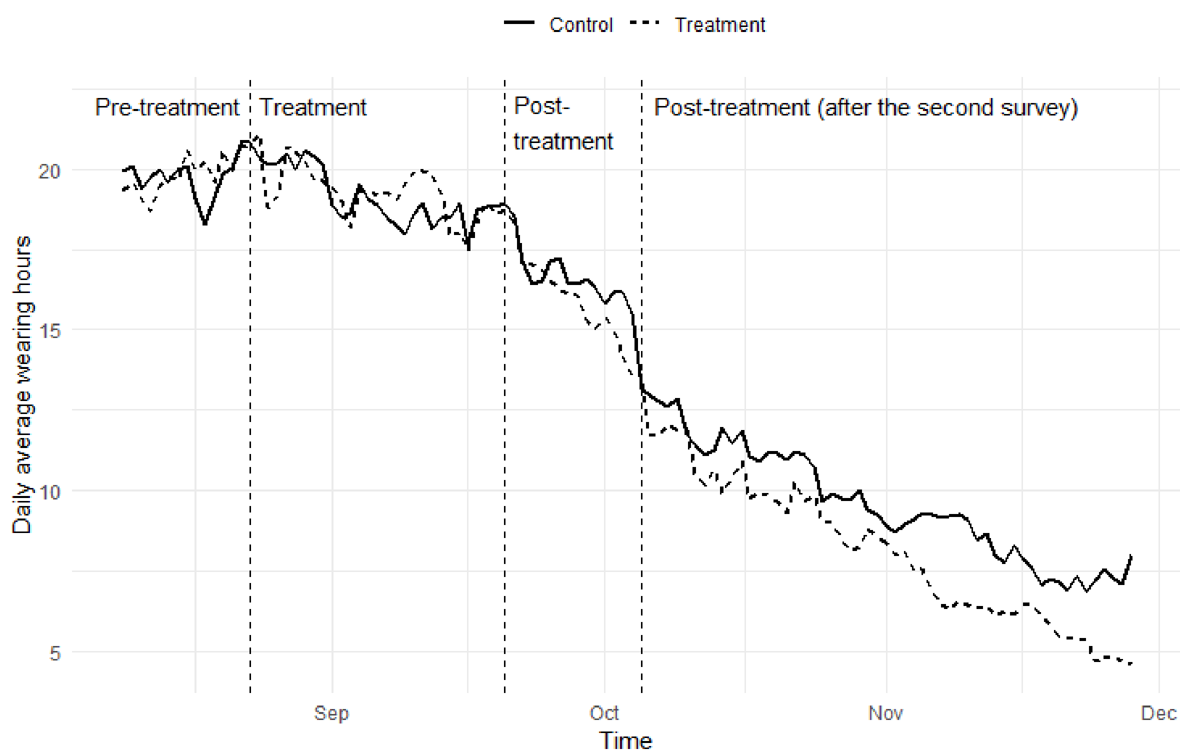
To proceed to estimate the QTE, we use the inverse probability of treatment weighting (IPTW) of the propensity score to get percentiles of the dependent variable. IPTW is the weights derived from the propensity score to create a synthetic sample in which the assignment in the treatment and the control group are independent.

The propensity score is predicted from a logit regression where the dependent variable is the dummy variable indicating treatment or control assignment. The covariates I use for this logit regression include: Age, Age squared, female, weight(in kg), married, college degree or above, RPI score, IPS score, baseline(first 2 weeks) average daily steps, baseline average days per week a participant wears the tracker, physical activity stages of change, binary variable indicating if a participant wants to weigh less, a binary variable indicating if the survey is incomplete. Note that even though the assignment is random and we use unconditional QTE, using covariates can improve the efficiency of the estimator. The covariates are used in the first stage and are integrated out (Frölich and Melly, 2008).

The inverse probability weights are defined as $w_i \equiv \frac{Z_i}{\hat{p}_i} + \frac{(1-Z_i)}{1-\hat{p}_i}$.

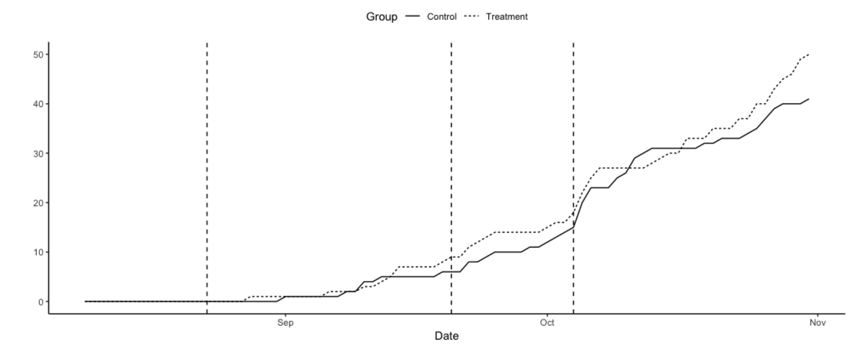
We estimate the QTE assuming the treatment is exogenous and assuming selection on observables.

Figure A.1: Tracker adherence: wearing hours trend

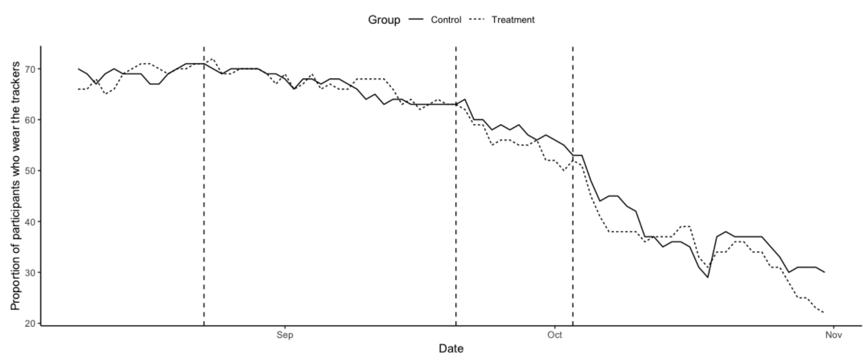


<u>Notes:</u> The graph shows the trend of wearing hours for 16 weeks. After the second survey was taken, more participants quit wearing the fitness trackers.

Figure A.2: Tracker adherence



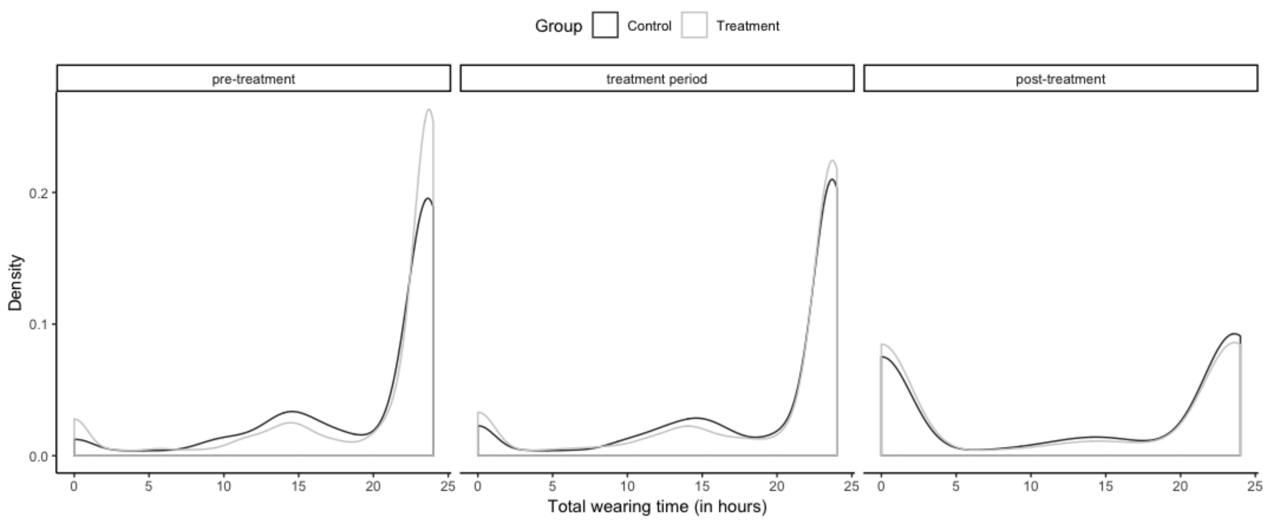
(a) Number of dropouts



(b) Proportion of participants who wear the trackers

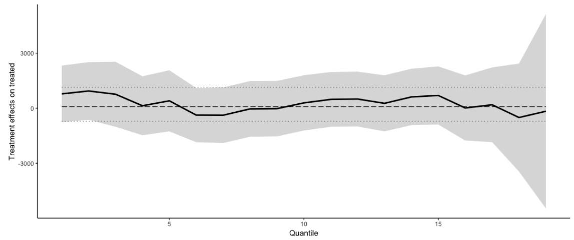
Notes: The first vertical dashed line indicates the beginning of the treatment period; the second one indicates the end of the treatment period; the third dashed line indicates the day when the follow-up survey was sent. Participants knew that after the follow-up survey, they would not be contacted.

Figure A.3: Tracker adherence: wearing hours distributions

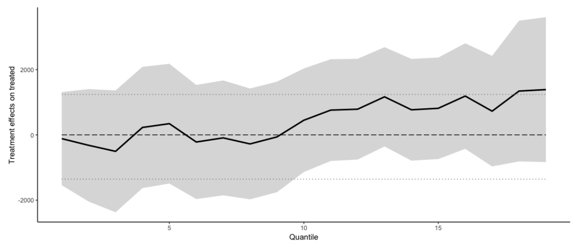


Notes: The graph shows the distribution of wearing hours for the three periods by group. As the incentive removed, more participants quit wearing the fitness trackers.

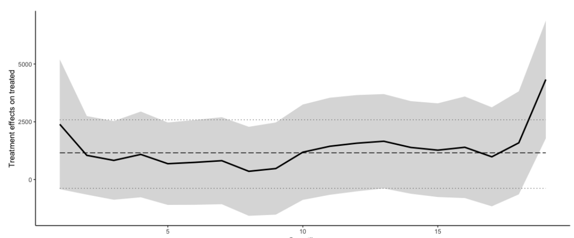
Figure A.4:



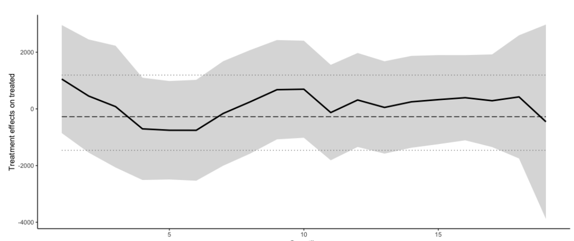
(a) week 3



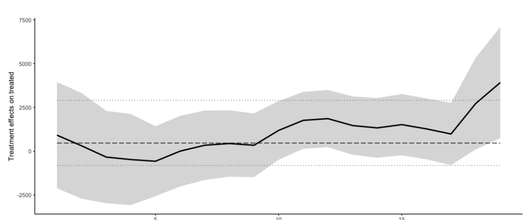
(b) week 4



(c) week 5



(d) week 6



(e) week 7

Table A.8: Regression results for number of steps due to walking, running, or brisk walking

	steps_walking	steps_running	steps_brisk	act_min
Treatment	185.808	116.232	277.156	5.115
	(258.684)	(246.318)	(262.917)	(6.710)
Post-treatment	406.124	382.212	612.887	1.758
	(378.527)	(512.293)	(419.303)	(10.833)
After 2nd surveyr	-397.873	255.481	-135.589	-2.097
	(507.706)	(558.420)	(540.529)	(9.888)
Cons	7972.738***	1419.031***	7148.438***	195.739***
	(129.494)	(168.895)	(137.194)	(3.205)
R^2	0.406	0.455	0.455	0.532
N	10378	3073	10350	3073

Notes: *p<0.05; **p<0.01; ***p<0.001.

Table A.9: Wearing adherence statistics

Weeks committed	Num. of individuals
1	1
3	1
4	4
5	5
6	5
7	12
8	14
9	22
10	5
11	17
12	58
Total	144

Notes: This table shows how many participants (out of 144 participants) appear in the sample for the ith week, i = 1, 2, $\cdots$, 12.

Table A.10: OLS results for daily num. of steps for treatment group during treatment period

	Dependent var.: daily num. of steps			
	(1)	(2)	(3)	(4)
Your rank	83.029***	75.399***		
	(5.199)	(5.313)		
Your steps			0.339***	0.304***
			(0.024)	(0.024)
Age		77.869***		77.276***
		(14.170)		(13.981)
Female		-345.494		-398.900
		(291.332)		(288.884)
College and above		-457.136		-390.523
		(265.437)		(262.257)
Married (or live with partner)		-743.501**		-832.499**
		(261.943)		(258.247)
Height (meter)		20100.831**		20715.415**
		(7690.757)		(7595.941)
Weight (kg)		-228.302**		-234.288**
		(85.865)		(84.574)
BMI		425.038		453.753
		(242.644)		(238.685)
Cons.	5863.482***	-23393.442	6557.858***	-24020.668
	(231.580)	(12939.088)	(205.494)	(12779.362)
R^2	0.134	0.181	0.143	0.186
N	1846	1810	1846	1810

Notes: Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001.

Table A.11: Average daily steps per week for each decile by group

		Percentiles								
Week	Group	10	20	30	40	50	60	70	80	90
week 3	Control	4636	6178	7689	8156	8880	9545	10364	11667	14008
	Treatment	5575	6311	7313	8119	9170	10046	10979	11680	13500
week 4	Control	4792	5344	7244	8137	8581	9459	10130	10831	12660
	Treatment	4473	5573	7026	7860	9031	10248	10900	12022	14004
week 5	Control	4540	5427	6656	7987	8969	9777	10734	12581	13671
	Treatment	5588	6516	7400	8345	10153	11351	12124	13979	15260
week 6	Control	4448	6354	7470	7908	8599	9971	10781	11403	13311
	Treatment	4901	5648	6714	8154	9292	10294	11030	11797	13734
Week 7	Control	4657	6756	8319	8957	9437	10080	11060	12060	13338
	Treatment	4966	6281	8325	9398	10627	11942	12393	13337	16059
Week 8	Control	5448	6398	6883	7944	8656	9463	9929	11202	13082
	Treatment	5518	6600	8237	8857	10161	11254	12110	13012	13933
Week 9	Control	4750	5984	6319	7177	7943	8735	9568	10381	11071
	Treatment	4477	5722	7643	8617	9325	10091	10272	11400	12624
Week 10	Control	3441	5920	6848	8173	8807	9650	10099	11504	12998
	Treatment	4885	6403	7132	8362	9390	10102	10578	11001	11367
Week 11	Control	5619	6400	7579	8093	9163	9765	11068	11685	12512
	Treatment	4198	4965	6472	8342	9741	10335	11310	12509	13513
Week 12	Control	2268	5216	6488	7776	8145	9011	9289	10556	12261
	Treatment	4367	5749	6960	7408	9014	10316	11012	12045	13331

Notes: Average daily steps per week means for each week, the number of The deciles are calculated using the inverse probability of treatment weighting using the propensity score (IPTW). The Appendix explains how the propensity score and the weights are derived.

Table A.12: Sample selection results table

	Stage 1	Stage 2
Wearing time (hours)	0.119**	228.722**
	(0.011)	(41.779)
Wearing time (hours) at (t+1)	-0.001	
	(0.012)	
Percentile at (t+1)	0.011**	
	(0.003)	
Percentile at (t-1)	0.016**	18.553**
	(0.003)	(6.642)
Treatment	-4.616	6,149.482**
	(3.243)	(1,944.581)
Inverse Mills Ratio		-1,649.036
		(1,025.859)
N	988	716
R^2		0.38

Notes: This table shows the results of the two stages in the sample selection session in the appendix. The insignificant of the coefficient for the Inverse Mills Ratio indicate that there is likely no sample selection bias.

Table A.13: Initial survey full list

(1)	(2) Treatment (Std. Dev.)	(3) Control (Std. Dev.)	(4) Treat - Control (Std. Err.)
I. Basic information			
1. What is the highest grade or level of school you have completed or the highest degree you have received?			
Did not graduate high school			
Current college student OR Attended college but did not obtain a degree			
Obtain an associate's or professional (not 4-year college) degree			
College graduate			
Attended graduate school but did not obtain a degree			
Obtained a graduate degree			
Refused			
Don't know			
Education: College or above	0.6944	0.6901	0.0043
	(0.4657)	(0.4639)	(0.0777)
2. Are you now married, widowed, divorced, separated, never married or living with a partner?			
Married			
Widowed			
Divorced			
Separated			
Never married			
Living with partner			
Refused			
Never married	0.4722	0.4366	0.0356
	(0.5027)	(0.4995)	(0.0418)
Married or living with partner	0.4722	0.4788	-0.0067
	(0.5031)	(0.5027)	(0.0841)
3. Your gender			
Female			
Male			
4.Your date of birth			
Month			
Year			
Age			

Notes: Column (1) presents the questions and choices in the follow-up survey; columns (2,) (3) and (4) present mean and std. dev. for the treatment group, control group, the difference in mean and the std. err.

Table A.14: Initial survey full list

(1)	(2) Treatment (Std. Dev.)	(3) Control (Std. Dev.)	(4) Treat - Control (Std. Err.)
II. Time discounting and present bias factors (Hardisty et al., 2013, Courtemanche et al., 2015)			
5. Imagine you could choose between **receiving** \$300 immediately, or another amount **6 months** from now. How much would the future amount need to be to make it as attractive as receiving \$300 immediately? (answer $\leq$ 1000)	544.46	516.23	28.23
	(217.13)	(206.62)	(36.54)
6. Imagine you could choose between **receiving** \$300 immediately, or another amount **1 year** from now. How much would the future amount need to be to make it as attractive as receiving \$300 immediately? (answer $\leq$ 10800)	1044.59	872.08	172.51
	(1326.49)	(636.57)	(181.30)
7. Imagine you could choose between **receiving** \$300 immediately, or another amount **10 years** from now. How much would the future amount need to be to make it as attractive as receiving \$300 immediately? (answer $\leq$ 10800)	4618	4107.69	510.31
	(3925.10)	(3736.27)	(660.65)
8. Imagine you could choose between **paying** \$300 immediately, or another amount **6 months** from now. How much would the future amount need to be to make it as unattractive as paying \$300 immediately? (answer $\leq$ 1000)	404.48	421.45	-16.97
	(153.13)	(181.08)	(28.91)
9. Imagine you could choose between **paying** \$300 immediately, or another amount **1 year** from now. How much would the future amount need to be to make it as unattractive as paying \$300 immediately? (answer $\leq$ 10800)	551.75	516.97	34.78
	(337.12)	(380.19)	(62.24)
10. Imagine you could choose between **paying** \$300 immediately, or another amount **10 years** from now. How much would the future amount need to be to make it as unattractive as paying \$300 immediately? (answer $\leq$ 10800)	1882.41	1113.70	768.71
	(3022.72)	(1792.47)	(435.95)
III. IPS - Irrational Procrastination Scale (Steel, 2010)			
11.Please rate how true the following statements are			
(1). I put things off so long that my well-being or efficiency unnecessarily suffers.			
(2). If there is something I should do, I get to it before attending to lesser tasks.			
(3) My life would be better if I did some activities or tasks earlier.			
(4) When I should be doing one thing, I will do another.			
(5) At the end of the day, I know I could have spent the time better.			
(6) I spend my time wisely			
(7) I delay tasks beyond what is reasonable.			
(8) I procrastinate.			
(9) I do everything when I believe it needs to be done.			
IPS score (higher, procrastinate more)	24.21	23.03	1.18
	(6.14)	(5.50)	(0.99)

Notes: Column (1) presents the questions and choices in the follow-up survey; columns (2,) (3) and (4) present mean and std. dev. for the treatment group, control group, the difference in mean and the std. err.

Table A.15: Initial survey full list

(1)	(2)	(3)	(4)
	Treatment	Control	Treat - Control
	(Std. Dev.)	(Std. Dev.)	(Std. Err.)
IV. RPI - Resistance to Peer Influence (Steinberg and Monahan, 2007)			
For each of the following ten questions, describe which sort of person you are most like - the one described on statement (A) or the one described on statement (B). Then decide if that is ”sort of true” or ”really true” for you, and mark that choice. For each question mark only ONE of the four choices.			
12(A) Some people go along with their friends just to keep their friends happy.			
BUT			
12(B) Other people refuse to go along with what their friends want to do, even though they know it will make their friends unhappy.			
13(A) Some people think it’s more important to be an individual than to fit in with the crowd.			
BUT			
13(B)Other people think it is more important to fit in with the crowd than to stand out as an individual.			
14(A) For some people, it’s pretty easy for their friends to get them to change their mind.			
BUT			
14(B) For other people, it’s pretty hard for their friends to get them to change their mind.			
15(A) Some people would do something that they knew was wrong just to stay on their friends’ good side.			
BUT			
15(B) Other people would not do something they knew was wrong just to stay on their friends’ good side.			
16(A) Some people hide their true opinion from their friends if they think their friends will make fun of them because of it.			
BUT			
16(B) Other people will say their true opinion in front of their friends, even if they know their friends will make fun of them because of it.			
17(A) Some people will not break the law just because their friends say that they would			
BUT			
17(B) Other people would break the law if their friends said that they would break it.			
18(A)Some people change the way they act so much when they are with their friends that they wonder who they “really are”.			
BUT			
18(B) Other people act the same way when they are alone as they do when they are with their friends.			
19(A) Some people take more risks when they are with their friends than they do when they are alone.			
BUT			
19(B) Other people act just as risky when they are alone as when they are with their friends.			
20(A) Some people say things they don’t really believe because they think it will make their friends respect them more.			
BUT			
20(B) Other people would not say things they didn’t really believe just to get their friends to respect them more.			
21(A) Some people think it’s better to be an individual even if people will be angry at you for going against the crowd.			
BUT			
21(B) Other people think it’s better to go along with the crowd than to make people angry at you.			
RPI score	2.99	2.95	0.04
	(0.36)	(0.40)	(0.06)

Notes: Column (1) presents the questions and choices in the follow-up survey; columns (2,) (3) and (4) present mean and std. dev. for the treatment group, control group, the difference in mean and the std. err.

Table A.16: Initial survey full list

V. Physical activity (IPAQ, 2002, GPAQ, 2002)

We are interested in finding out physical activities people do as their everyday life. The questions will ask you about the time you spent being physically active in the last month.

Work or classes related

Think about the time you spend doing work in the last month. Think of work as the things that you have to do such as paid or unpaid work, study/ training, household chores, harvesting food/crops, fishing or hunting for food, seeking employment.

Think about all the “vigorous-intensity” and “moderate-intensity” activities that you did in the last month. In answering the following questions “vigorous-intensity activities” are activities that require hard physical effort and cause large increases in breathing or heart rate, “moderate-intensity activities” are activities that require moderate physical effort and cause small increases in breathing or heart rate.

22. Do you currently have a job or do any unpaid work outside your home OR do you currently take classes?
23. Does your work or classes involve vigorous-intensity activity that causes large increases in breathing or heart rate like carrying or lifting heavy loads, digging or construction work for at least 10 minutes continuously?

    If answer no, skip the next two questions.
24. In a typical week, on how many days do you do vigorous intensity activities as part of your work or course work?
25. How much time do you spend doing vigorous-intensity activities at work or for school on a typical day?

26. Does your work or classes involve moderate-intensity activity that causes small increases in breathing or heart rate such as brisk walking or carrying light loads for at least 10 minutes continuously?
27. In a typical week, on how many days do you do moderate intensity activities as part of your work or course work?
28. How much time do you spend doing moderate-intensity activities at work or for school on a typical day?

Travel to and from places

The next questions exclude the physical activities at work or for the school that you have already mentioned. Now I would like to ask you about the usual way you travel to and from places. For example to work, for shopping, to market, to a place of worship.

29. Do you walk or use a bicycle (pedal cycle) for at least 10 minutes continuously to get to and from places?

    If answer no, skip the next two questions.
30. In a typical week, on how many days do you walk or bicycle for at least 10 minutes continuously to get to and from places?
31. How much time do you spend walking or bicycling for travel on a typical day?

Physical activities around home

The next questions are about the physical activities you might have done in the last month in and around your home, like housework, gardening, yard work, general maintenance work, and caring for your family.

32. Do you do any vigorous-intensity activities that cause large increases in breathing or heart rate like heavy lifting, chopping wood, shoveling snow, digging in the garden or yard for at least 10 minutes continuously?

    If answer no, skip the next two questions.
33. In a typical week, on how many days do you do vigorous-intensity housework or house maintenance?
34. How much time do you spend doing vigorous-intensity housework or house maintenance on a typical day?

35. Do you do any moderate-intensity activities that cause a small increase in breathing or heart rate such as carrying light oads, sweeping, washing windows, and raking in the garden or yard OR carrying light loads, washing windows, scrubbing floors and sweeping inside your home?

    If answer no, skip the next two questions.
36. In a typical week, on how many days do you do moderate-intensity housework, house maintenance, or caring for family?
37. How much time do you spend doing moderate-intensity housework, house maintenance, or caring for family on a typical day?

Table A.17: Initial survey full list

(1)	(2) Treatment (Std. Dev.)	(3) Control (Std. Dev.)	(4) Treat - Control (Std. Err.)
V. Physical activity (IPAQ, 2002, GPAQ, 2002)			
Leisure			
The next questions exclude the work and transport activities that you have already mentioned. Now I would like to ask you about sports, fitness and recreational activities (leisure).			
38. Do you do any vigorous-intensity sports, fitness or recreational (leisure) activities that cause large increases in breathing or heart rate like running, football, or tennis for at least 10 minutes continuously?			
If answer no, skip the next two questions.			
39. In a typical week, on how many days do you do vigorous intensity sports, fitness or recreational (leisure) activities?			
40. How much time do you spend doing vigorous-intensity sports, fitness or recreational activities on a typical day?			
41. Do you do any moderate-intensity sports, fitness or recreational (leisure) activities that cause a small increase in breathing or heart rate such as brisk walking, cycling, swimming, volleyball for at least 10 minutes continuously?			
If answer no, skip the next two questions.			
42. In a typical week, on how many days do you do moderate-intensity sports, fitness or recreational (leisure) activities?			
43. How much time do you spend doing moderate-intensity sports, fitness or recreational (leisure) activities on a typical day?			
Sedentary			
The following question is about sitting or reclining at work, at home, getting to and from places, or with friends including time spent sitting at a desk, sitting with friends, travelling in car, bus, train, reading, playing cards or watching television, but do not include time spent sleeping.			
44. How much time do you usually spend sitting or reclining on a typical weekday?			
45. How much time do you usually spend sitting or reclining on a typical weekend day?			
Vigorous activity time per week (in minute)	238.31	217.39	20.92
	(331.83)	(385.38)	(60.72)
Moderate activity time per week (in minute)	808.73	668.88	139.86
	(694.38)	(681.76)	(116.34)
Total physical activity MET (Metabolic Equivalents)	5141.41	4414.64	726.77
	(4702.17)	(5217.92)	(838.99)
Sedentary hours (weekday)	6.67	6.35	0.32
	(3.77)	(3.70)	(0.64)
Sedentary hours (weekend)	5.89	5.59	0.31
	(4.06)	(3.11)	(0.62)

Notes: Column (1) presents the questions and choices in the follow-up survey; columns (2,) (3) and (4) present mean and std. dev. for the treatment group, control group, the difference in mean and the std. err.

Table A.18: Initial survey full list

VI. Self-efficacy and decision balance (Marcus et al., 2009)

46.Please rate how important each of these statements are in your decision of whether to be physically active.

Scale: Not at all important - slightly important - moderately important - very important - extremely important

(1) I would have more energy for my family and friends if I were regularly physically active.

(2) Regular physical activity would help me relieve tension.

(3) I would feel more confident if I were regularly physically active.

(4) I would sleep more soundly if I were regularly physically active.

(5) I would feel less stressed if I were regularly physically active.

(6) I would feel good about myself if I kept my commitment to be regularly physically active.

(7) I would feel more comfortable with my body if I were regularly physically active.

(8) Regular physical activity would help me have a more positive outlook on life.

(9) It would be easier for me to perform routine physical tasks if I were regularly physically active.

(10) I think I would be too tired to do my daily work after being physically active.

(11) I would find it difficult to find a physical activity that I enjoy and that is not affected by bad weather.

(12) I feel uncomfortable when I am physically active because I get out of breath and my heart beats very fast.

(13) Regular physical activity would take too much of my time.

(14) I would have less time for my family and friends if I were regularly physically active.

(15) At the end of the day, I am too exhausted to be physically active.

(16) Equipment or gym membership is too expensive for me to engage in exercise.

(17) I would like to exercise with a companion or in a group.

Table A.19: Initial survey full list

(1)	(2) Treatment (Std. Dev.)	(3) Control (Std. Dev.)	(4) Treat - Control (Std. Err.)
VII. Health information			
47. How tall are you without shoes? (in meters)	1.70	1.69	0.0020
	(0.09)	(0.10)	(0.0159)
48. How much do you weigh? (in kg)	68.23	69.19	-0.92
	(10.54)	(13.23)	(2.2940)
49. Do you consider yourself now to be ...			
Overweight			
Underweight			
About the right weight			
Refuse			
Consider overweight	0.53	0.46	0.06
	(0.50)	(0.50)	(0.08)
50. Would you like to weigh ...			
More			
Less			
Stay about the same			
Refuse			
Want to weigh less	0.74	0.70	0.04
	(0.44)	(0.46)	(0.08)
51. Ideally, how much would you like to weigh?	68.28	69.19	-0.92
	(10.54)	(13.23)	(2.29)
52. Have you ever been advised by a physician to lose weight?			
53. During the past 12 months, have you tried to lose weight?			
54. Ideally, how many steps should you walk per day?	9916	9405	511
	(4576.24)	(2462.71)	(637.61)
55. Please enter the nickname that we assigned to you. It is in the form you entered with your name, phone number, email address, address, and Mi Fit user ID.			
56. Please enter your XX app user ID(that is, your XX user ID). Please copy and paste the user ID shown on our study web app on the computer screen.			

Notes: Column (1) presents the questions and choices in the follow-up survey; columns (2,) (3) and (4) present mean and std. dev. for the treatment group, control group, the difference in mean and the std. err. We use "XX" to represent the name of the phone app due to the confidential agreement with the company that offers the data access.

Table A.20: Survey response: the follow-up survey 1

(1)	(2)	(3)	(4)
1. Which type of messages did you receive?	Treatment	Control	None
Obs.	57	56	4
2. Height (meter)			
Mean	1.69	1.69	1.76
Std. Dev.	(0.087)	(0.104)	(0.091)
3. Weight (kg)			
Mean	72.79	70.96	84.94
Std. Dev.	(17.145)	(14.189)	(13.983)
4. Employment status			
Employed full time (40 or more hours per week)	31	32	3
Employed part time	4	6	0
Not working	2	2	0
Student while working	15	10	1
Student	5	6	0
5. Which type of messages did you wish to receive?			
Treatment text message	41	44	1
Control text message	10	10	2
None	6	2	1
6. Did the information feedback you received motivate you to exercise more?			
Yes	19	11	3
No	6	29	0
Maybe	22	16	1
7. Did the information feedback you received motive you to exercise less?			
Yes	1	0	0
No	52	55	4
Maybe	4	1	0
8. Which effort(s) did you make to walk more steps?			
I did not make any effort to walk more steps.	8	14	0
Park farther from destination.	14	12	1
Walk to bus stop.	0	4	1
Use stairs instead of elevators.	26	26	3
Take an extra lap at the grocery store.	3	2	1
Walk and talk on the phone.	14	10	0
Walk to run an errand.	7	14	1
Wear the fitness tracker more.	25	27	1
Exercise more/go to the gym.	20	20	2
Others, such as	9	10	0

Notes: Column (1) presents the questions and choices in the follow-up survey; columns (2,) (3) and (4) present number of observations for each choice of each question by group. The groups are divided based on the answer of Q1.

Table A.21: Survey response: the follow-up survey 2

(1)	(2)	(3)	(4)
1. Which type of messages did you receive?	Treatment	Control	None
Obs.	57	56	4
9. Which of the following gave you more incentive to exercise or walk more steps?			
Using the wearable device	38	39	3
Functions like setting goals in the smartphone app	26	27	1
Text messages you received	29	11	2
None	6	7	0
Others, please specify	3	3	0
10. Please rank the following information you received in order of most to least motivating:			
"You walked X steps." ranked the first	8	-	-
"You were ahead of Y (Z%) of your peers." ranked the first	11	-	-
"You would be ahead of more of your peers if you walked Z more steps." ranked the first	8	-	-
11. Did any of the following information in the text messages cause you to be discouraged?			
No	39	-	-
Yes, "you walked X steps."	0	-	-
Yes, "you were ahead of Y (Z%) of your peers."	6	-	-
Yes, "you would be ahead of more of your peers if you walked Z more steps."	12	-	-
12. Experience of owning another wearable device as a fitness tracker:			
Did not own another wearable device	38	36	2
1 week	1	1	0
1 month	1	0	1
6 months	3	2	0
one year	1	3	0
more than a year	12	14	1
13. Stage of change for physical activity			
Precontemplation stage	2	2	0
Contemplation stage	10	7	0
Preparation stage	7	3	0
Action stage	6	9	0
Maintenance stage	32	35	4
14. Beliefs:			
Most of my friends engage in regular physical activity.	37	31	2
Most people who are important to me engage in regular physical activity.	30	29	2
Most people whose opinion I value engage in regular physical activity.	36	29	3
The majority of University of X students engage in regular physical activity.	24	24	4
People like me engage in regular physical activity.	35	33	4

Notes: Column (1) presents the questions and choices in the survey questions; columns (2), (3) and (4) present number of observations for each choice of each question by group. The groups are divided based on the answer of Q1. For Q14, participants were asked to rank the extent to which they agree (strongly disagree to strongly agree) with each statement. This table reports only the number of observations that shows agree or strongly agree with the statement.

Table A.22: Survey response: the follow-up survey 3

(1)	(2)	(3)	(4)
1. Which type of messages did you receive?	Treatment	Control	None
Obs.	57	56	4
15. Sefl-efficacy to regulate exercise (perform exercise routine regularly - three or more times a week)			
Degree of confidence rated by a number from 0 to 100:			
0 is Cannot do at all; 50 is Moderately can do; and 100 is Highly certain can do.			
When I am feeling tired.	48.42	45.82	
When I am feeling under pressure from work.	59.45	59.77	
During bad weather.	60.21	46.91	
After recovering from an injury that caused me to stop exercising.	40.05	42.09	
During or after experiencing personal problems.	62.47	65.02	
When I am feeling depressed.	48.93	52.67	
When I am feeling anxious.	67.37	66.34	
After recovering from an illness that caused me to stop exercising.	52.63	45.50	
When I feel physical discomfort when I exercise.	44.30	44.79	
After a vacation.	76.41	69.48	
When I have too much work to do at home.	44.37	43.84	
When visitors are present.	38.85	34.81	
When there are other interesting things to do.	52.03	53.18	
If I don't reach my exercise goals.	69.93	66.49	
Without support from my family or friends.	64.13	65.93	
During a vacation.	51.53	51.20	
When I have other time commitments.	40.36	37.69	
After experiencing family problems.	63.82	59.55	
Average value	57.01	52.59	

Notes: Column (1) presents the questions and choices in the survey questions; column (2) and (3) present mean value of level of confidence (self-efficacy) (Bandura, 2006) for each item in Q15 for the treatment and the control groups.

Table A.23: Survey response: examples of answers to the open-ending questions in the follow-up survey

16. Which elements of the study, if changed, do you feel would motivate you to be more physically active?
(Number of responses for this question: 93/144 = 64.58% response rate)
Tips for how to get more steps. For instance - walk around the block would add 400 steps to your daily total steps.
Encouragement in the texts instead of just a report of how many steps I took with a reminder to sync.
I enjoyed having the feedback from the day before, it was super motivational to be receiving; Step count motivate me to walk more.
If device worked better; The fitness band needs to be more interactive, this one was way too basic.
To continue to get text messages telling me how many steps I took yesterday. Also, just a text telling me I was doing well – something encouraging would have been nice to get.
Telling you how far behind/ahead you are mid-day instead of day after.
If there were goals that I had to attain.
Showing how ahead of others I am as well as stats on who completed wearing the band every day that week. After the 6th week I was a little demoralized since I never won.
Help in setting goals rather than texts comparing me to others; Incentives for reaching your goal; Study pre-sets goal and offers monetary incentive to meet those goals.
Longer duration of text reminders, and adding the percentiles into the texts.
I would replace the texts with a graphic or even a gif that displayed the my step count and other info.
Continual heart rate monitor, ability to do behavior tagging, expansion of activity and behavior tagging to include yoga.
Working toward a reward.
More money offered. More awareness if the weekly drawing even happened.
Make sure the messages are actually attainable "if you would have taken 40,000 more steps you'd be ahead of 75 of 75 of your peers."
17. In general, what motives you to be more physically active?
Number of responses for this question: 100/144 = 69.44%)
I am more motivated when my friends want to work out with me. I am also more motivated when I know more about how food works in the human body. I have learned a lot about how food breaks.
Personal health; Desire to be healthier; To be strong and healthy; My own well being; Wanting to be in better health.
Losing weight and staying healthy; Physical appearance, or weight loss (Seeing results).
I feel healthier and happier and I am more productive during the day. I also look better and feel more confident; Health and stress management.
Having someone to exercise with; A friend to exercise with and hold me accountable.
Good weather, desire to be healthier.
Losing weight, feeling better physically and mentally; Makes me feel good and helps with stress; Stress management and how I feel afterward.
Desire to be stronger and more flexible.
Peers, support from friends.
Having a goal; making physical activity social.
Financial incentives work.
Being healthy for my family.
Goals.
A general dislike for how my body looks $\cdots$

Notes: Different answers with similar meanings are listed in the same row separated by a semicolon.