Summer 2019

The impacts of social comparison information on physical activity

Lianjun Li
University of Iowa

Recommended Citation
Li, Lianjun. "The impacts of social comparison information on physical activity." PhD (Doctor of Philosophy) thesis, University of Iowa, 2019.
https://doi.org/10.17077/etd.u86x-pkuo

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the Economics Commons

THE IMPACTS OF SOCIAL COMPARISON INFORMATION ON PHYSICAL ACTIVITY

by

Lianjun Li

A thesis submitted in partial fulfillment
of the requirements for the
Doctor of Philosophy
degree in Economics in the
Graduate College of
The University of Iowa

August 2019

Thesis Supervisor: Associate Professor David Frisvold

ACKNOWLEDGEMENTS

This work would not have been possible without the financial support 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.

I would like to express my deepest appreciation to my main advisor, David Frisvold, who has mentored me in formulating my research topic, pushing the process of my experiment, and providing invaluable advice on methodology. Without his guidance and persistent encouragement this dissertation would not have been possible. I would like to thank my committee members: Linnea Polgreen, Julia Garlick, Suyong Song, and Lucas Carr for their insightful comments and professional guidance. In addition, a thank you to Renea Jay, who has managed my academic situation and assisted and supported me throughout my time in the program. I would also like to thank the Department of Economics, DSP, ITS, HSO, and CLAS server group for their help in offering me the resources to run my experiment.

I must thank Professor Jiuchang Wei and other professors and students from the University of Science and Technology of China, and my friends for their continual supports and friendship. I am also grateful to members of my family for their unconditional love, encouragement, and supports. I wish to thank my loving fiancé, Roy Crihfield, for his patience and constant assistance throughout my graduate studies. Special thanks to my father, my grandmother, and other family members, who have been and will always be the source of my strength. Most importantly, special thanks to my mother, who passed away ten years ago. I am indebted to her everlasting love and encouragement for my achievements.

ABSTRACT

My dissertation focuses on changes in health-related behavior in react to information-based intervention. The first chapter analyzes the results of a field experiment to investigate the effects of comparative information on the daily number of steps taken of adults. The second chapter further explores the effects using qualitative analysis. The third chapter intends to offer explanations in the mechanisms of the results from the first chapter.

My first chapter uncovers how patterns in the daily number of steps of adults are affected by information that compares one with unknown peers. I conducted a field experiment that used fitness trackers to collect daily and minute-by-minute 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. I examined whether individuals in the two groups behaved differently during and after the intervention period. I find no clear evidence of an aggregate impact of social norms on the daily number of steps taken. However, I find individuals who are not overweight or nor married or cohabiting are more likely to be influenced by social norms. Greater treatment effects are found among individuals whose number of steps that are at the tails of the distribution curve.

My second chapter reports the results of the textual data from the survey in the field experiment. I present dominant themes that emerged from answers to the open-ended essay questions in the survey. The results support that health concern, body image, appearance, psychological factors, peers and friends are major motives for being physically active. For participation in the study specifically, text messages that contain comparative information produced some improvement of the exercise level. However, participants also requested more interactions with peers, additional information provision, rewards for reaching goals. The results imply external incentives play a smaller role in promoting daily number of steps.

In the third chapter, I conduct a survey experimentation to test the effectiveness of informing descriptive social norms and types of text messages in predictions about health-related behaviors. First, I investigate if errors in beliefs about activity levels exist and I find no evidence of over- or under-confidence in one's own activity levels. Further analysis provide preliminary evidence of negative effects of informedness in predictions about one's own behavior. However, the intention-to-treat effects of comparative information are unclear. The data provide evidence in favor of the correlation between first-order personal beliefs, not higher-order normative beliefs, in predicting an increase number of steps taken in response to intervention with text messages.

PUBLIC ABSTRACT

My dissertation focuses on changes in health-related behavior in react to information-based intervention. The first chapter analyzes the results of a field experiment to investigate the effects of comparative information on the daily number of steps taken of adults. The second chapter further explores the effects using qualitative analysis. The third chapter intends to offer explanations in the mechanisms of the results from the first chapter.

My first chapter uncovers how patterns in the daily number of steps of adults are affected by information that compares one with unknown peers. I conducted a field experiment that used fitness trackers to collect daily and minute-by-minute 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. I examined whether individuals in the two groups behaved differently during and after the intervention period. I find no clear evidence of an aggregate impact of social norms on the daily number of steps taken. However, I find individuals who are not overweight or nor married or cohabiting are more likely to be influenced by social norms. Greater treatment effects are found among individuals whose number of steps that are at the tails of the distribution curve.

My second chapter reports the results of the textual data from the survey in the field experiment. I present dominant themes that emerged from answers to the open-ended essay questions in the survey. The results support that health concern, body image, appearance, psychological factors, peers and friends are major motives for being physically active. For participation in the study specifically, text messages that contain comparative information produced some improvement of the exercise level. However, participants also requested more interactions with peers, additional information provision, rewards for reaching goals. The results imply external incentives play a smaller role in promoting daily number of steps.

In the third chapter, I conduct a survey experimentation to test the effectiveness of informing descriptive social norms and types of text messages in predictions about health-related behaviors. First, I investigate if errors in beliefs about activity levels exist and I find no evidence of over- or under-confidence in one's own activity levels. Further analysis provide preliminary evidence of negative effects of informedness in predictions about one's own behavior. However, the intention-to-treat effects of comparative information are unclear. The data provide evidence in favor of the correlation between first-order personal beliefs, not higher-order normative beliefs, in predicting an increase number of steps taken in response to intervention with text messages.

TABLE OF CONTENTS

LIST OF TABLES . . . . **ix**

LIST OF FIGURES . . . . **x**

CHAPTER I THE IMPACTS OF SOCIAL COMPARISON INFORMATION ON PHYSICAL ACTIVITY: EVIDENCE FROM A FIELD EXPERIMENTB **1**

Introduction 1

Literature And Background 4

The Experiment 7

Results 11

Conclusion And Discussion 22

CHAPTER II THE EFFECTS OF SOCIAL COMPARATIVE INFORMATION ON PHYSICAL ACTIVITY: A QUALITATIVE STUDY **26**

Introduction 26

Results 28

Descriptive Statistics: Qualitative And Quantitative Results 36

Conclusion And Discussion 39

CHAPTER III HETEROGENEOUS EFFECTS OF INFORMEDNESS AND TEXT MESSAGES IN PREDICTING EXPERIMENTAL RESULTS **41**

Introduction 41

Survey And Data 46

Results 52

Conclusion And Discussion 61

REFERENCES **63**

APPENDIX A: APPENDIX TO CHAPTER I **71**

APPENDIX B: APPENDIX TO CHAPTER II **88**

APPENDIX C: APPENDIX TO CHAPTER III **102**

LIST OF TABLES

Table 1. Descriptive statistics 9

Table 2. Average treatment effects in different periods 12

Table 3. Weekly treatment effects 14

Table 4. Treatment effects by different groups 17

Table 5. Average daily steps per week for each decile by group 21

Table 6. Descriptive statistics: demographics and other characteristics 29

Table 7. Descriptive statistics 51

Table 8. Joint distribution of self-reported daily steps and prior beliefs 53

Table 9. The empirical allocation function for beliefs 53

Table 10. Joint distribution of self-reported daily steps and updated beliefs 54

Table 11. The empirical allocation function for updated beliefs 54

Table 12. Predictions about oneself: main results 1 58

Table 13. Predictions about oneself: main results 2 60

LIST OF FIGURES

Figure 1. Average daily steps by group 11

Figure 2. Weekly treatment effects 15

Figure 3. Steps/day by group 19

Figure 4. Quantile treatment effects by different periods 23

Figure 5. Daily number of steps (respondents and non-respondents) 30

Figure 6. Average daily steps 37

Figure 7. Average daily time of wearing of the fitness trackers 38

Figure 8. Survey design: groups and sample size 46

Figure 9. Distribution for outcomes of interests by group 49

CHAPTER I THE IMPACTS OF SOCIAL COMPARISON INFORMATION ON PHYSICAL ACTIVITY: EVIDENCE FROM A FIELD EXPERIMENTB

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, Liebman, & Katz, 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, Hoekstra, & West, 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 & 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, I studied one specific behavior – daily physical activity – from the results of a field experiment conducted in a university setting. I 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

[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.

with such information. I also documented the persistence, or lack thereof, of changes in physical activity after individuals no longer receive such information.

I focused my study on physical activity because of its inherent correlation with health outcomes. In 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, Karvinen, Peavler, Smith, & Cangelosi, 2013). High blood pressure and high BMI accounted for over 600,000 U.S. deaths in 2005 (Danaei, Ding, Mozaffarian, 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, Ding, Mozaffarian, 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) for the citizens of almost two thirds of countries in 2004.[2]

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, Fulton, Pratt, Yang, & Adams, 2015).[3]

Accordingly, I conducted a randomized field experiment to test how social comparison information influences behavior. I 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 I 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).

[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).

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] I randomly selected participants for a control and treatment group. The treatment period spanned four weeks, during which each 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.

My 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, Rottensteiner, Kaprio, & Kujala, 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, I 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

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

[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). (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 I would not be able to get the data tracked by the device.

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 person to sustain or improve her current status. Finally, because forgetfulness is a common barrier to individuals engaging in regular behaviors (McKenzie-Mohr & Schultz, 2014), a consistent provision of information may also function as a prompt to remind people to stand up and move. Therefore, I 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.

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 ((Charness & Gneezy, 2009; John, Loewenstein, Troxel, et al., 2011; Volpp, John, Troxel, et al., 2008)) and commitment devices (Royer, Stehr, & Sydnor, 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 & Price, 2013). The take-up rate of commitment devices is low,[8] possibly because people are not precisely aware of their time preferences. Understanding how

[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, John, Troxel, 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, Stehr, and Sydnor (2015) find positive and lasting effects in gym attendance, but with a low take-up rate, a little over 20%.

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" 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 & Rogers, 2014). Similar desirable results are also found in household water use ((Brent, Cook, & Olsen, 2015; Ferraro & Price, 2013)). 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). My goal is to investigate whether individuals are affected by only receiving information about their peers, even if no financial incentive factors are involved.

Among other studies that make use of interventional messages, Beshears, Choi, Laibson, Madrian, and Milkman (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.

I was 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, Wang, Yang, et al., 2011), to remind people to comply with pharmacological procedures, and to support smoking cessation and weight loss maintenance (Gerber, Stolley, Thompson, Sharp, & Fitzgibbon, 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, Wilbur, Ingram, & Fogg, 2013). My study combines the practice of text messaging and comparison-based information and examines the impact of information provision on number of daily steps taken. I also test for the existence of heterogeneity.

The participants of my experiment were not restricted to undergraduate students. In fact, about half of the participants were university staff. Using college students as experimental subjects should not cause problems for the external validity of a study (Druckman & 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. In addition, the outcomes of interest in my study use daily activity, rather than gym attendance as in most previous literature, allowing me to analyze behavior at a finer level of detail. I use the standard difference-in-difference model as my primary analysis method. To investigate heterogeneity, I apply dummy variables indicating different characteristics, like gender, marital status, BMI status, ex-ante level of daily number of steps etc. I use the unconditional exogenous quantile treatment effect method (QTE) to examine distributional effects of treatment.

To preview the results, I 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.[9] 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 and below the average. For a habitually active individual, the marginal benefits are little, because intrinsic motivation is more important for maintaining such activity (Aaltonen, Rottensteiner, Kaprio, & Kujala, 2014). Marginal costs are high for those who are not active due to the discomfort of starting an exercise routine. However, I find significantly negative treatment effects among over-weight and married or cohabiting individuals: their activity decreased compared to their counterparts in the control group. The effects persisted for about three weeks after the intervention was removed.

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.[10] I 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

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

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

the study began.[11][12] Participants were not required to return the fitness tracker at the end of the experiment.

I 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.[13]

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. Therefore, the sample size for the experiment is 144.[14] In addition, a follow-up 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 analyzed is 143.[15] The initial survey collected information about participants' demographics, physical activity levels,[16] height and weight, time discount factors, hyperbolic discount factors,[17] level of procrastination, and resistance to peer influences.

The summary statistics for selected variables are in Table 1. Resistance to Peer Influence (RPI) is measured using a series of questions constructed in Steinberg and Monahan (2007) with

[11] 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. I 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.

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

[13] The screening survey collected information on participants' level of physical activity (Marcus, Forsyth, et al., 2009) and beliefs about the physical activity level of their peers. However, I was prevented from comparing the interest levels of those participating in the study with those who and scheduled a meeting but did not show up.

[14] Using the average active minutes of American adults in C (2005) and an expected 10 minutes difference in active mintues between the control and the treatment groups, the effective sample size is about 126.

[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. I used the formula in List, Sadoff, and Wagner (2011) and calculated the minimum statistically significant sample size for one group be about 60. Therefore, a sample size of 143 is more than adequate.

[16] I used IPAQ (2002) and GPAQ (2002) to get estimates of minutes of self-reported physical activity.

[17] I intended to get individual levels of time discount factors and hyperbolic discount factors by selecting questions from Hardisty, Thompson, Krantz, and Weber (2013). However, many participants appeared not to fully understand the questions, invalidating the values.

Table 1: Descriptive statistics

Variable name	(1) Mean value of the control group	(2) Obs.	(3) Treatment group - Control group	(4) Obs.
Age	35.514	70	-4.709**	142
			(1.983)	
Female	0.686	70	0.013	143
			(0.078)	
Weight (*kg*)	73.994	70	-0.312	143
			(2.654)	
Overweight	0.471	70	0.022	143
			(0.084)	
College and above	0.718	71	-0.033	144
			(0.077)	
Married	0.493	71	-0.02	143
			(0.084)	
Baseline statistics				
Ave. daily steps (first week)	9588.649	71	-939.848	142
			(784.885)	
Ave. days of wearing (first week)	6.761	71	-0.391	144
			(0.238)	
Ave. daily wearing hours (first week)	15.877	71	-1.147	144
			(0.783)	
Ave. daily steps(first two weeks)	9666.386	70	-733.383	144
			(725.091)	
Ave. days of wearing(first two weeks)	13.423	71	-0.573	144
			(0.423)	
Ave. daily wearing hours(first two weeks)	15.684	71	-0.834	144
			(0.755)	

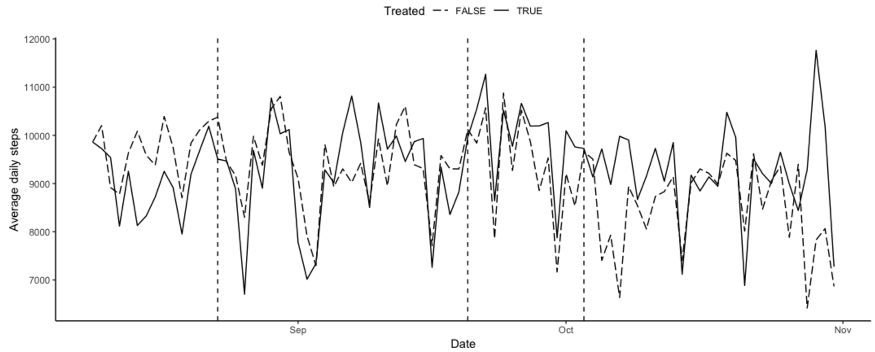
Notes: Standard errors are in parentheses; $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$. The table shows the average in individual characteristics for control group and how the treatment average differs from the control group. Column (3) is derived from regressing the variable of interest on the binary variable indicating the treatment group; IPS: Irrational procrastination scale (Steel, 2010); RPI: Resistance to peer influence (Steinberg & Monahan, 2007).

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. The bottom six rows of Table 1 summarize the average number of daily steps and the compliance behaviors of participants (via information collected from the fitness trackers). The final two columns show the calculated difference between the two groups, which was derived using regression analysis. Summary statistics of other variables collected through the survey are shown in the appendix table A.2.

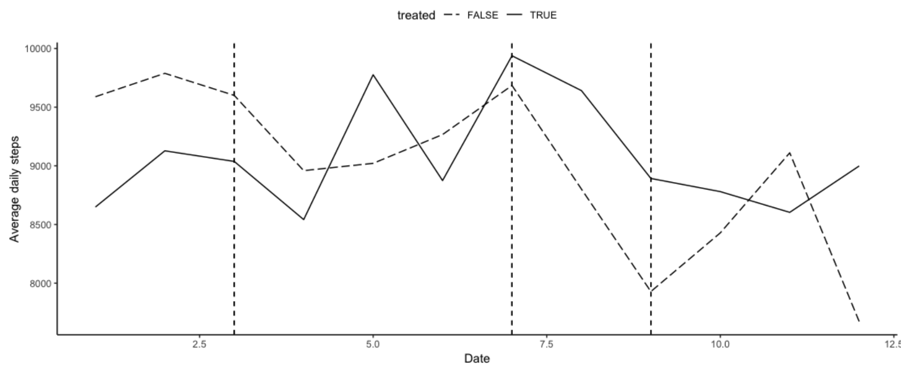
The groups were largely similar, although some differences were observed. Participants in the experimental group were on average five years younger and more physically active. They incurred 113 more minutes of moderate activity and 12 more minutes of vigorous activity each week. However, these participants also walked 733 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, I failed to reject the null hypothesis that the groups were the same.

Appendix table A.1 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 follow-up survey was emailed to participants two weeks after the text message were no longer sent. Participants who completed the survey were mailed $10.

Figure 1: Average daily steps by group



(a) Average daily steps



(b) Average daily steps (smoothed on a weekly basis)

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.

Results

Baseline Results

The key outcome of interest is the number of steps taken per day. Figure 1 shows the trend of total number of steps taken using the raw daily data and a smoothed version when the number of steps is averaged on a weekly basis.

Table 2 shows the primary regression results for average treatment effects for total number of steps taken daily, number of steps taken due to brisk walking, total active minutes due to walking or running, and total daily wearing hours of the tracker.[18] I use a standard difference-in-difference

[18] In the minute-by-minute data, the column "mode" indicates different status for each minute of a user, like "shallow

Table 2: Average treatment effects in different periods

Variables	(1) After-treated Week 3 - 12	(2) Treatment period Week 3 - 6	(3) Post-treatment period Week 7 - 12	(4) Post-treatment 1 Week 7 - 8	(5) Post-treatment 2 Week 10 - 12
Total number of steps/day	429.143	467.433	368.950	758.231	-3.964
	(318.985)	(293.240)	(448.272)	(407.879)	(557.958)
R^2	0.00	0.00	0.00	0.01	0.01
N	10,447	10,447	10,447	10,447	10,447
Steps/day (brisk walk)	314.415	331.781	287.010	659.649	-59.681
	(292.147)	(262.854)	(411.948)	(380.381)	(500.789)
R^2	0.00	0.00	0.00	0.01	0.01
N	10,363	10,363	10,363	10,363	10,363
Total active minutes/day	4.852	7.030	2.103	8.332	-3.669
(walk or run)	(4.974)	(4.745)	(6.612)	(6.270)	(8.296)
R^2	0.00	0.00	0.00	0.00	0.00
N	10,379	10,379	10,379	10,379	10,379
Total wearing time/day (hrs)	0.239	0.795	-0.133	0.153	-0.275
	(1.094)	(0.912)	(1.401)	(1.398)	(1.612)
R^2	0.01	0.10	0.10	0.15	0.15
N	15,158	15,158	15,158	15,158	15,158

Notes: Robust standard errors (clustered at the individual level) in parentheses. Fixed effects are included in all specifications. The table shows the estimates of the average treatment effects for total daily number of steps taken, number of daily steps due to brisk walking, total active minutes (minutes due to walking or running). The first column combines the treatment period and the post-treatment period, showing an aggregate effects once the text messages were started. The second column specifies the average treatment effects during the treatment period. The follow-up survey was sent on the first day of week 9.

model to estimate the treatment effects. I first show the aggregate average effects during the entire period once the text messages began (first column), then show the results for treatment and post-treatment periods (I also divide the post-treatment period into two sub-periods, one before the follow-up survey was sent and the other after). On average, the difference between the treatment group and the control group is an insignificant 400 steps total and 300 steps of brisk walking per day

sleep," "walking," "running," etc. I use this information to obtain the number of steps/minutes due to walking or running. The column "activeness" 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 I use the minutes when the active level is greater than 85 to get steps due to brisk walking.

after the first text message was sent. Similar differences can be seen during the treatment period. The differences become greater for the first two weeks of the post-treatment period, yet negative for the the four weeks of the post-treatment period. Overall, the effects decay after the intervention is removed. The magnitudes are small: a 300-400 steps of brisk walking per day represents about 3 minutes of moderate intensity activity per day or about 20 minutes per week.[19] Using the minute-by-minute data I am able to distinguish active minutes during which participants are walking or running. An insignificant difference of 7 minutes of total active minutes (minutes due to walking or running) as a dependent variable (the third panel) is observed between the control and the treatment groups during the treatment period. During the first two weeks of the post-treatment period, this difference becomes only slightly larger, then turns negative afterwards. All the magnitudes are comparatively small and insignificant. Therefore, I do not find clear evidence of any treatment effects during any time period.

Table 3 shows the average treatment effects if I divide the study period into weeks using the same dependent variables. The weekly effects are consistently small and insignificant for the four-week treatment period (week 3 through 6) with an exception of the third week during this period. The difference in daily average steps between the two groups for each week is around 100 steps per day, while it was more than 1000 steps in week 5. After the intervention, the effects do not fade away right after the incentive is removed. The values of the effects increase substantially but then become negative. A similar trend can be found if the dependent variable is the number of steps of brisk walking. Little change is detected during the treatment period. For the first three weeks following the treatment period, the difference becomes larger but insignificant. For the last three weeks in the post-treatment period, the difference turns negative. 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. Figure 2 displays graphical demonstration of the weekly treatment effects.

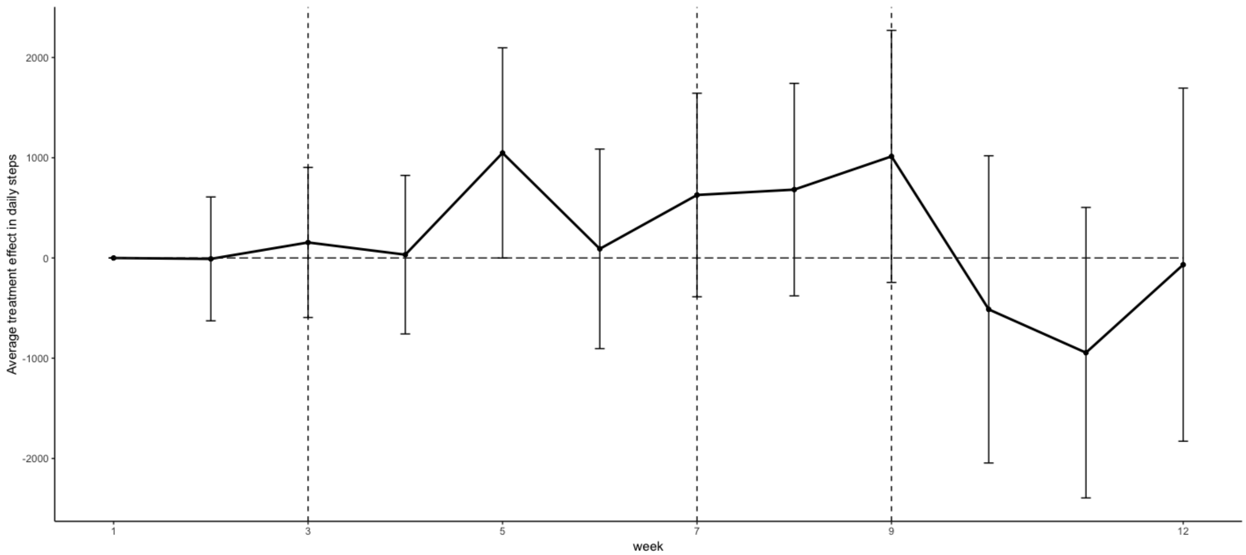
[19] I get very similar results if I use number of steps due to walking or running as dependent variable.

Table 3: Weekly treatment effects

VARIABLES	(1) Average daily steps	(2) Steps due to walking or running	(3) Total wearing time (hrs)
Week 2	-9.132	-121.1	0.556
	(312.6)	(287.9)	(0.398)
Week 3	154.9	4.181	-0.478
	(378.9)	(341.5)	(0.443)
Week 4	32.90	-79.24	-0.262
	(400.1)	(358.3)	(0.500)
Week 5	1,048*	884.9*	0.271
	(530.5)	(506.7)	(0.593)
Week 6	90.99	117.9	-0.710
	(503.5)	(453.3)	(0.635)
Week 7	628.3	487.2	-0.00434
	(513.5)	(483.9)	(0.647)
Week 8	682.1	692.5	-0.524
	(536.3)	(506.5)	(0.743)
Week 9	1,014	861.5	0.0640
	(636.1)	(570.5)	(0.799)
Week 10	-512.9	-479.6	-1.766*
	(775.3)	(729.9)	(0.921)
Week 11	-945.0	-618.4	-1.581*
	(733.2)	(646.4)	(0.872)
Week 12	-66.98	-537.3	-1.524
	(891.2)	(773.4)	(1.069)
Observations	9,059	8,999	9,286
R-squared	0.016	0.013	0.016
Number of individuals	144	144	144

Notes: Robust standard errors (clustered at individual level) in parentheses. Individual fixed effects are included in all models. Week 1 is the base week, week 1 & 2 are the pre-treatment period, week 3 through 6 are the treatment period, week 7 through week 10 are the post-treatment period. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$;

Figure 2: Weekly treatment effects



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.

Heterogeneity In The Effects

Next, I investigate the treatment effects for different groups. I divided the sample into groups based on gender, average daily steps taken during the pre-treatment period (above or below the sample average), overweight,[20] marital status (married or cohabiting), education (college education or above), and stages of motivational readiness for change for physical activity. In the health psychology model, the stages of change model (or the transtheoretical model) in Prochaska and DiClemente (1982) and Prochaska, DiClemente, and Norcross (1992), predicts a persons success or failure in achieving a behavioral change. Individuals move through stages of precontemplation (PC), contemplation (C), preparation (PR), action (A), and maintenance (M) (Verbeek & Nijman, 1992). In the context of physical activity, the first four stages represent an inadequate, or adequate but not habitual activity level. The last stage implies that being physically active is a habit. I use the five-item survey question to determine the stages of change for participants. I then divide

[20]The BMI is calculated mostly based on self-reported height and weight from the initial survey; since the height and weight were asked again in the follow-up survey, missing values from the initial survey was replaced by the values in the follow-up survey.

them into two groups: those with physically active habits and those without active habits. The stages of change survey questions were used twice, once in the survey when recruiting participants (about three months before study began), then again two weeks after the incentive was removed (the follow-up survey).

The regression specification I use is the standard triple difference-in-difference model as the following:

$$\begin{aligned} y_i \;=\; & \alpha_0 + \beta_1 Group_j + \beta_2 Treated + \beta_3 Group_j \cdot Treated + \\ & \sum_k \delta_{0k} Period_k + \sum_k \delta_{1k} Period_k \cdot Group_j + \sum_k \delta_{2k} Period_k \cdot Treated + \\ & \sum_k \delta_{3k} Period_k \cdot Group_j \cdot Treated + \varepsilon_i \end{aligned}$$

where "Group" means the individual's group (female, for instance), "Period" refers to treatment period, post-treatment period, or the two-week post-treatment period before the follow-up survey was sent. The triple interaction term coefficient is the triple difference-in-difference model estimate and detects the time change in mean for individuals specified by different characteristics in the treatment group less the change in mean for their counterparts in the control group and that for the individuals with no such characteristics in the treatment group. I run the regression separately for each characteristic specification.

Table 4 shows the results of the regressions. First, I do not find significant treatment effects for groups of female, baseline daily steps above the average, college degree or above, regularly active (claimed before or after the intervention). Qualitatively, a female, an overweight individual, an individual who is married or is cohabiting, or an individual who after the intervention claims to be physically active for the past six months, 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, an individual who has a college or above degree, or an individual who claimed to be regularly physically active before study, walks more steps if she is treated.

Table 4: Treatment effects by different groups

VARIABLES	(1) After-treated (week3-10)	(2) Treatment period (week3-6)	(3) Post-treatment period (week7-10)	(4) Post-treatment 1 (week7-8)	(5) Post-treatment 2 (week 9-12)
Baseline above average	642.740	637.844	687.602	437.511	1136.274
	(617.411)	(522.374)	(913.332)	(878.694)	(1085.549)
Female	-1003.999	-754.661	-1296.504	-927.596	-1431.889
	(769.588)	(634.644)	(1237.97)	(1150.853)	(1452.927)
Overweight	-1449.447 **	-948.261*	-2026.508**	-1428.638*	-2436.067
	(660.579)	(566.946)	(940.139)	(857.444)	(1152.97)
Married (or cohabiting)	-1304.677*	-980.795*	-1776.557 **	-1878.263 ***	-1986.524
	(677.903)	(566.219)	(989.102)	(883.232)	(1210.888)
College and above	354.192	89.466	685.966	807.511	538.665
	(811.661)	(678.938)	(1190.268)	(1148.632)	(1376.33)
Regularly physically active	724.749	824.361	600.497	990.269	462.864
(Before study began)	(653.197)	(545.267)	(950.839)	(848.043)	(1180.133)
Regularly physically active	-462.0013	-151.029	-763.630	-259.383	-1014.301
(Follow-up survey)	(856.408)	(702.147)	(1201.733)	(1023.451)	(1574.014)

Notes: Robust standard errors (clustered at the individual level) in parentheses. Individual fixed effects are included in all models. Column (1) includes the entire period starting from the first day the text messages were sent. Column (2) compares the treatment period only to the pre-treatment period observations; column (3) includes the entire period starting from the day the text messages were no longer sent; column (4) includes the post-treatment period before the follow-up survey was sent. Individual stages of change of physical activity were surveyed twice, once about a few months before study began(screening survey), the other after the incentive was removed. *p<0.1; **p<0.05; ***p<0.01.

Particularly, the magnitudes of the effects for the overweight group and the married (or cohabiting) group are statistically significant. An overweight individual would walk around 1000 steps less during the intervention period. After the intervention is removed, the effects for overweight individuals are even greater (about 1500 steps per day less). Similarly, married or cohabiting individuals walk less during the intervention period compared to their counterparts in the control group (about 980 steps per day) and drop further after the incentive is removed (about 1300 steps per day). Appendix figure A.1 displays the distributions of daily steps for these groups by period. Figures 3 indicates that the married individuals and the overweight individuals did not necessarily walk fewer steps. The treatment effects are driven by increases in the number of daily steps walked by non-married (or non-cohabiting) or non-overweight individuals. However, for non-married (or non-cohabiting) individuals, the increase might be driven by an increase in total amount wearing time of the fitness trackers (see Figure A.2).

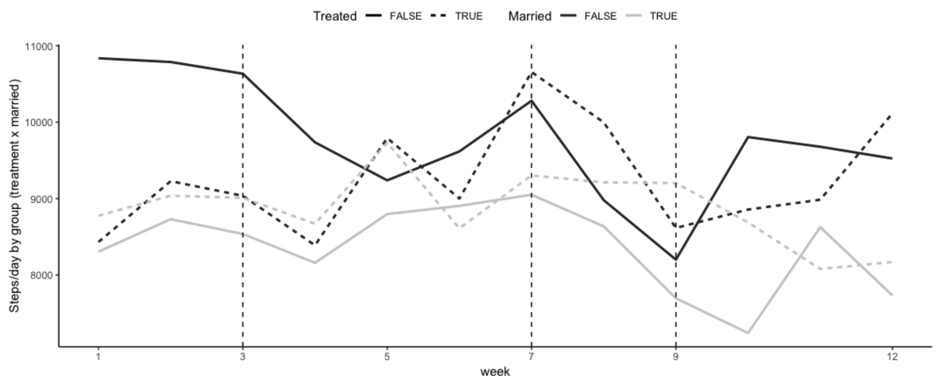
For individuals who claim to be regularly active during the study period, their treatment effects are negative qualitatively. This might be driven by an overestimation by the activity about their peers. They might have realized their peers are not as active as they think, therefore reduce their activity level after treatment.

Quantile Treatment Effects

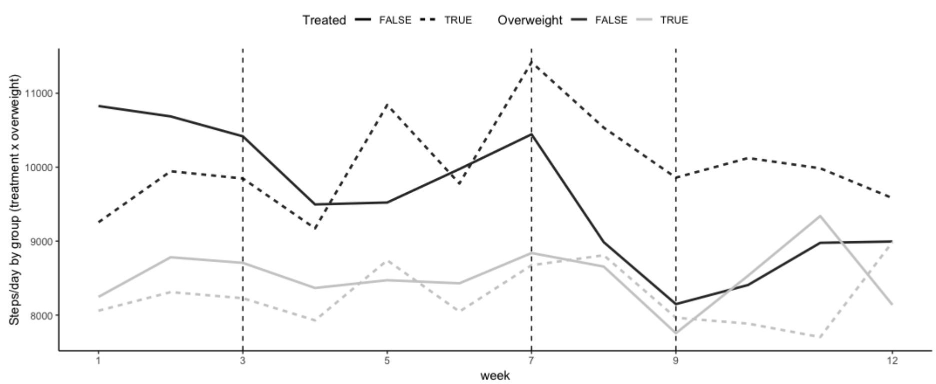
I estimate the quantile treatment effects (QTE) of receiving peer-comparison information on number of daily steps walked. The estimation of QTE yields insights of the impacts of independent variables on the entire distribution of the dependent variable. Using quantile regressions also allows us to deal with sensitivity to outliers (Frölich & 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.[21]

[21] 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, Gelbach, and Hoynes (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, Gelbach, & Hoynes, 2006). Andrews, Li, and Lovenheim (2012) estimate QTE of college

Figure 3: Steps/day by group



(a) By married x treatment



(b) By overweight x treatment

Notes: The black lines in (a) and (c) are for the non-married (or non-cohabiting) individuals and those in (b) and (d) are for non-overweight individuals. The grey lines in (a) and (c) are for married or cohabiting individuals and those in (b) and (d) are for overweight individuals. Dashed lines in all panels represent the treatment group and solid lines represent the control group.

The QTE allows me 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.

I 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.[22] 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.[23] 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.

I explain the weighting method I use for QTEs in the appendix. Table 5 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. I first divide the entire study period into a pre-treatment period and any time after that. I 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

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.

[22] On the other hand, the conditional QTE recovers features of the conditional outcome distribution. The conditional quantiles are more difficult to interpret (Andrews, Li, & Lovenheim, 2012).

[23] 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 & Melly, 2008). Another advantage is including covariates that are independent from the treatment can change limit of the estimated conditional QTEs (Frölich & Melly, 2008).

Table 5: Average daily steps per week for each decile by group

		Percentiles								
Study period	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.

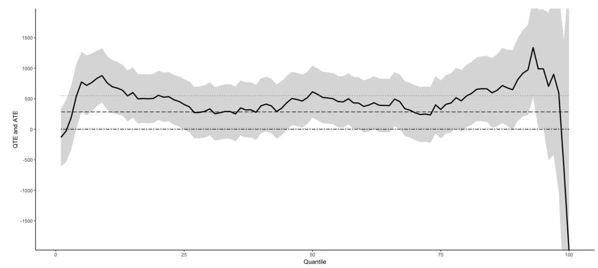
are mostly significantly positive for each decile. The magnitudes of the effects do not differ largely across the distribution, with a few hundred steps for each decile. The treatment effects reach a minimum between the second and the third deciles. During the treatment period, slightly greater treatment 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 4 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.3 and figure A.3 reports the deciles of daily average steps per week for each group and the weekly QTE.

Conclusion And Discussion

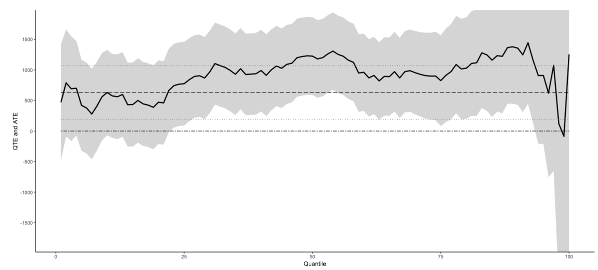
In this paper, I investigated the pattern of daily physical activity level measured by number of daily steps in response to social-comparison information. I conducted a field experiment, distributed fitness trackers to collect data, and used text messages for disseminating information. The purpose of 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, I 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. 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

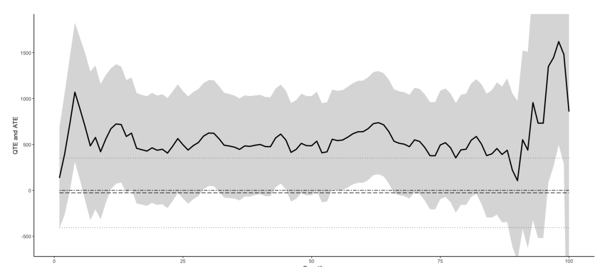
Figure 4: Quantile treatment effects by different periods



(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.

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, Murphy, and Boone-Heinonen (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.[24] The minute-by-minute data of my study shows an increase of 7 minutes walking time (not statistically significant) falls in the range of the results of prior intervention studies. In terms of number of steps, Gilson, Faulkner, Murphy, et al. (2013) found a maximum of 1800 steps and a minimum of 900 steps increase among employees. My results for heterogeneity, however, show a smaller range of changes in the daily number of steps taken.

In addition, I test heterogeneity using a triple difference-in-difference model and find that married (or cohabiting) and overweight individuals are more likely to be influenced negatively by being exposed to their peers' information. But gender, education, claimed exercise level, and baseline level do not affect the treatment effects.

I also find attrition among the participants. Appendix Figure A.5 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, Anthony, Carr, et al., 2018). The regression results are reported in the appendix Table A.4. In addition, the "study begin" effects seem to be larger than the effects of daily text messages containing comparative information. Appendix Figure A.4 shows the line graph of average daily

[24] The study lacks control group.

steps by group including the enrollment period. Note that the enrollment took 3 weeks in total. On the day before the study began, I sent a text message to everyone to let them know that study would begin the next day. The line graphs show a clear sharp increase at the first dashed line (the begin date of the study). This indicates an interesting point that participating a field experiment itself has some effects.[25]

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 my study are compared with someone they do not know. Finally, compared to other physical activity studies, my 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.

[25] This is called the Hawthorne effect (Mayo, 1949).

CHAPTER II THE EFFECTS OF SOCIAL COMPARATIVE INFORMATION ON PHYSICAL ACTIVITY: A QUALITATIVE STUDY

Introduction

In a randomized controlled trial, we used fitness trackers and text messages to explore how individual daily physical activity level changes in response to comparative information. Participants are randomized into a control and a treatment group. Each participant in the treatment group received text messages that contained comparative information, whereas each participant in the control group received text messages that contained no such information.[26] We tracked their activity levels by using a fitness band that was distributed to every participant. We find about 3% increase in the daily number of steps walked during the intervention period and a short-term habit formation after the information was no longer provided to the participants. At the second week since the information was stopped to be sent to the participants, we sent a follow-up survey. After that, we observe the effects waned quickly and a great amount of participants stopped using the fitness band.

The motivation of using information-based strategies to improve individual daily exercise level stems from its established role in directing people's behavior in a broad range of areas. Numerous field-based studies have found evidence that household utility bills that contain descriptive norms, comparing their energy or water use with their neighbors, can improve residential energy or water conservation (e.g. Allcott, 2011; Ferraro & Price, 2013). Empirical evidence has also confirmed the role of normative information in predicting or intervenning behaviors like college student gambling (Larimer & Neighbors, 2003), adolescent druge use (Donaldson, Graham, & Hansen, 1994), and littering (Reno, Cialdini, & Kallgren, 1993), etc. In addition, positive normative triggers produce improved performance on muscular endurance task (Randazzo, 2016). Descriptive norms has a greater influence in taking stairs instead of elevators than proclaming

[26]The comparative type of text message reads as: "Yesterday, you walked 7015 steps, you were ahead of 48% of your peers. If you walked 1600 more steps, you would be ahead of 58% of your peers." And the non-comparative type of text message reads as "Yesterday, you walked 7015 steps. Please remember to sync your [fitness tracker]." The numbers in the text messages would be replaced with real values for each participant each day.

benefits of using stairs (Burger & Shelton, 2011).

Research on the impacts of social norms has relied primarily on mere data analysis. Although previous studies demonstrate convincingly social influence-based strategies have strong directory effects in socially significant behaviors and suggest viability of applications of such intervention, little focus on an comprehensive and in-depth understanding in how participants thought or felt in the study. The quality of the implementation of field experiments of this type is rarely reported.

The follow-up survey in our field experiment included a number of multiple-choice questions, as well as two open-ended questions that are designed to record what "behind the numbers" (Schutt, 2018, p. 321). Our primary interest is on text – on qualitative data rather than on data or a few discrete variables – and we set out to investigate what participants did or thought about the study and about exercise in general. Thus we contribute to this strand of literature by adding on richness of social experience.

By analyzing the texts, it allows us to better understand why certain things work or not. It is also important to learn how interventions could be enhanced or modified for potential future study. In contrast to quantitative data analysis, our study reports the textual data with no predetermined directions or categories. Therefore we examine the factors influencing the performances of the information-based strategies in the area of physical activity. By categorizing and counting, we also transpose the texts into quantitative data and report some descriptive statistics based on these categories.

The two open-ended essay questions help to determine participants' experience of their participation in the study and their general motivations to be physically active. We follow Toda, Njeru, Zurovac, et al. (2017) and summarize the qualitative data into thematic areas and analyze the answers to these two questions separately in the next section.

The survey was administered to the 151 participants, 119 completed responses were collected. 88 answered the question about the experiment and 96 answered the questions on the general motivations of being physically active. A full list of responses to those questions are shown in the appendix. We excluded individuals who requested to quit the study and also those whose data were largely missing in the descriptive analysis.[27] Table 6 displays demographics information and other individual characteristics for open-ended questions' respondents and non-respondents. The proportion of participants who are in the treatment group is the same. Respondents are about 6 years older than non-respondents. Respondents reported their moderate and vigorous active minutes more conservatively. Figure 5 displays daily number of steps for each group graphically throughout the study period. The daily number of steps taken for respondents was larger than that for non-respondents once they started to receive text messages. I used triple difference-in-difference analysis and found no statistically significant difference in the number of steps between the two groups. Appendix Table B.1 shows daily number of steps and daily time participants wear the fitness trackers for each week during the study period, for respondents and non-respondents of the open-ended questions. The difference in the number of steps between the two groups might be due to the better compliance behavior among respondents, that is, they wore the fitness trackers for longer time compared to non-respondents. In addition, participants were allowed to keep the fitness trackers and they were told to wear the trackers for as long as they wish in the individual meetings at the beginning of the study. Accordingly we are able to collect their activity data after the follow-up survey was taken. Although no clear difference in the number of steps taken (or the time of wearing) is detected, it appears a much faster attrition among non-respondents. In week 12, about half of the respondents stayed in the study, whereas only 37% of the non-respondents stayed.

Results

We now concentrate on each of the response to the essay questions. The survey was sent out two weeks after the text messages were no longer provided. Therefore, the answers might reflect

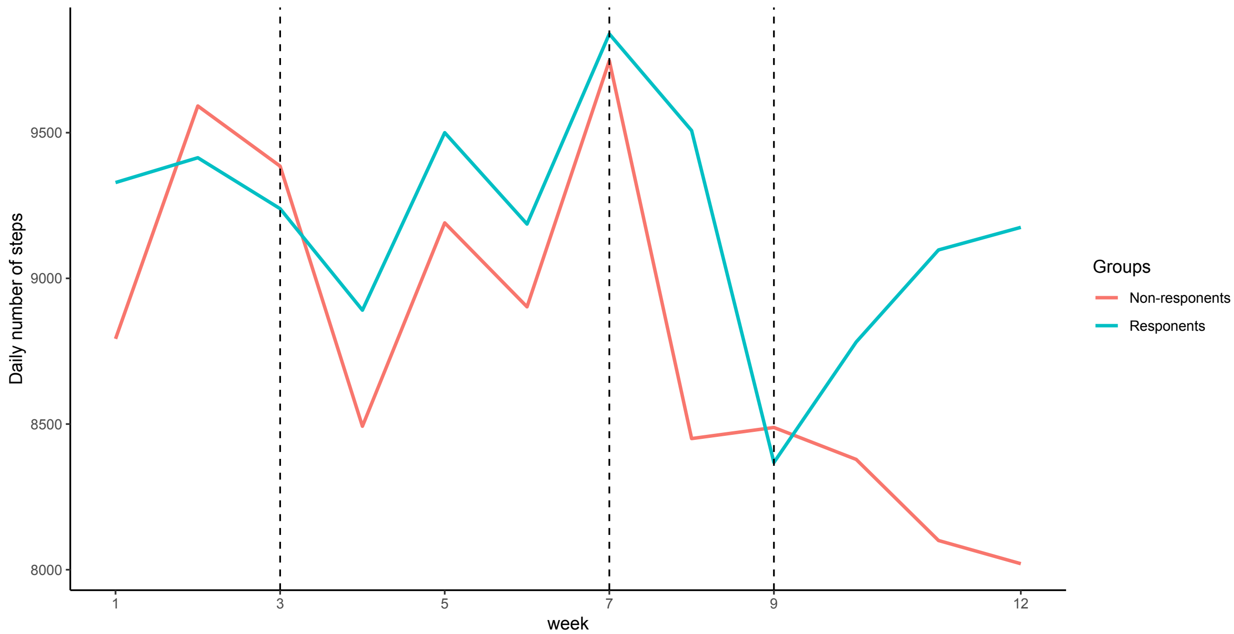
[27] However, a few participants, although did not show any sign of participation (for instance, wearing the fitness tracker for only one day), finished the follow-up survey. Their data is not analyzed, but their answers are included in the appendix.

Table 6: Descriptive statistics: demographics and other characteristics

	(1) *Respondents*			(2) *Non-respondents*		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Groups1						
Age	35.39	12.16	92	29.39	10.88	46
Female	0.70	0.46	93	0.65	0.48	46
Married	0.59	0.49	93	0.30	0.47	46
College or above	0.79	0.41	94	0.52	0.51	46
Treatment group	0.51	0.50	94	0.52	0.51	46
Group2						
Weight (*kg*)	73.22	15.01	93	76.31	17.27	46
Height (*m*)	1.69	0.09	94	1.70	0.09	46
Overweight	0.46	0.50	93	0.54	0.50	46
RPI	2.99	0.37	94	2.94	0.38	46
IPS	23.16	6.00	94	24.70	5.70	46
Vigorous/week	161.57	238.79	89	250.54	446.04	46
Moderate/week	544.17	600.30	87	589.78	739.93	45
N	94			46		

Notes: IPS: Irrational procrastination scale; RPI: Resistance to peer influence. Statistical results show that differences in age, married, and college or above are significant between the two groups. Specifically, the respondents are older, more likely to be married, and more likely to have college or higher education.

Figure 5: Daily number of steps (respondents and non-respondents)



Notes: This line graph depicts the daily number of steps by respondents and non-respondents of the open-ended questions.

not only participants' thought and experience about receiving the text messages, but those about information withdraw as well. We show the main themes of the answers in separate subsections

Which Elements Of The Study, If Changed, Do You Feel Would Motivate You To Be More Physically Active?

About six dominant themes emerge from these answers. The themes include: (1) the text messages or participation in the study worked or did not work, because of peer effects, or the feedback and the reminding features; (2) functionality of the fitness trackers we use for the study; (3) additional information would be helpful; (4) goals related; (5) nothing is motivating; (6) others.

Working Through Peers, Feedback Or Reminding Feature

About one out of four of the respondents approve that participating in the study was motivating through the peers or the text messages, either by feeling competitive, by knowing others are also participating, or being encouraged by the text messages.

"Knowing how much I am walking in comparison to theirs."

"Knowing others were participating."

"I was pretty motivated to wear the device every day, walk to my car when the weather was good, etc."

Some respondents have positive experience because they received feedback or reminders, or the quantification.

"I enjoyed having the feedback from the day before, it was super motivational to be receiving..."

"step count motivate me to walk more"

Participants who received the control group information also mentioned they wish to receive the other version of the information: *"If I received the other type of message."* Some also mentioned a *"longer duration of text reminders"* would be more motivating.

In addition, a few also think it would be more motivating to be able to interact with others: *"... if there was an anonymous chat group that let us talk to each other about how we're doing," "a community of people that not only said what they were doing to exercise but also invited others along with them to exercise."*

However, comparing with peers may also work in a negative way: *"After the 6 week I was a little demoralized since I never won."* Similarly, some respondents also suggested more *"encouraging"* text messages would be more motivating.

"Encouragement in the texts instead of just a report of how many steps I took with a reminder to sync," "... just a text telling me I was doing well - something encouraging would have been nice to get," "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""

Therefore, text messages that compare individuals with their peers work for some people, but according to the responses, interactions with others, longer duration of the intervention period would work better. It also risks discouraging those who did not do well comparatively. The information based strategies should be implemented based on individual characteristics. For those who do badly compared to their peers, the text messages could concentrate less on their low ranking but more on cheering them up. Alternatively, we could inform individuals information about their comparative peers, like peers at the similar age or that have similar level of physical activities. The design of the information delivered to participants should be tailored to enhance engagement in physical activity.

Functionality of The Fitness Trackers

Most of the responses that are about the fitness trackers are negative. Mainly participants wish to have a fitness tracker that could sync with the corresponding phone app on their smart phones automatically, or that could record activities other than just steps walked and sleep time, or that has information on the tracker instead of viewing their activity information on the app in their smart phones.

"I was disappointed because the tracker didn't record the time I spent rowing, either on the river or at the gym. Having that data included would have helped understand my overall activity level & made me more motivated." "My band didn't keep track of anything except for steps & sleep at night. didn't track any exercise or naps during the day."

"I wish the band would auto-sync. Sometimes I did not sync the band in time, so the number of steps was incorrect in the text message. I wish it was an option to automatically sync at midnight."

"The fitness band needs to be more interactive, this one was way too basic."

The usage of the fitness trackers in this study is meant to be a tool for data collection, however, the choice of an inadequate device may deter participation. One takeaway from these responses is to use better fitness trackers in the future study. It will not only benefit by promoting incentives to participating, also will allow researchers to collect data in broader dimensions and therefore to obtain richer data analysis. We should also be aware that previous studies on effects of fitness trackers show limited effects of fitness trackers. Ledger and McCaffrey (2014) find one-third of the U.S. consumers who have owned an activity monitor stopped using the device within six months of receiving it: they often forget to wear or lose interest in wearing the monitor over time. Also, personal quantification can decrease continued engagement in physical activity and subjective well-being (Etkin, 2016). According to Patel, Asch, and Volpp (2015), "the notion is that by recording and reporting information about behaviors such as physical activity or sleep patterns, these devices can educate and motivate individuals toward better habits and better health. The gap between recording information and changing behavior is substantial, however, and while these devices are increasing in popularity, little evidence suggests that they are bridging that gap." Ideally, if technology allows, some better device - convenient to carry, sync automatically, and measure accurately - would benefit both the researchers and study participants.

Additional Information Would Be Helpful

Many respondents wish the text messages to include additional information.

"Measure of heart rate exertion and not just counting steps," "Also providing updates about sleeping patterns," "If information from my other exercise (lap swimming) were included so I had a holistic picture of my exercise for the day," "Continual heart rate monitor, ability to do behavior tagging, expansion of activity and behavior tagging to include yoga."

Others want to be informed with *"tips for how to get more steps"* or *"activities happening in the area"*.

Setting Goals

A lot of responses mention the values of setting goals.

"Texts with a countdown to goal, like one at noon that you are X% from your goal for the day," "Updates on if I'm close to my daily goal," "I think it could have been fun to have a "goal" to work towards in monthly steps with facts about exercise or something that were unlocked with each milestone met."

People believe goal-setting would work well in promoting preferable behaviors. In fact, goal-setting has limited effects in different areas (Carrera, Royer, Stehr, Sydnor, & Taubinsky, 2018, e.g). A large amount of psychological literature study the correlation between goal-setting and physical activity, together with other features like perceived behavioral control, self-satisfaction, etc. In a field-based study, Polgreen, Anthony, Carr, et al. (2018) conclude that eliciting personal step goals did not lead to increases in daily steps. Specifically, although participants reached their goals, the result can be confounded by endogeneity effects of reverse causality issues. Future studies that intend to explore the effects of goal setting may have to set goals for participants exogeneously instead of letting them set their own goals.

Nothing is Motivating

About ten percent of the question respondents mention that *"nothing"* would make them more active, partially due to the reason that they are already active enough.

"I don't think anything would make me more active," "Probably nothing. The work I do and the workout class I take is enough activity for me right now."

Therefore, for future research study, selecting potential participants that have sedentary lifestyle might be less likely to get a non-results. From the perspective of research in the effects of information on behavioral change, it is safe to conclude a non-linear reaction curve. It is worth to explore the effects across the distribution of the outcome variables.

Design of The Experiment

Some of the responses provide insights in other details of experimental design of this type.

"I would replace the texts with a graphic or even a gif that displayed [the] my step count and other info," "If I could wear the exercise band when I shower then I would not take it off and forget to put it back on."

In fact, in the social norms literature on energy or water conservation behaviors, bar charts are used in households' utility or water bills (Allcott, 2011; Brent, Cook, & Olsen, 2015). However, the effects of different forms of information displayed in outcomes are a different research question. Indeed, in the context of cafeterias, the orders of the food displayed on menus could affect the consumption of each food item by as large as 25% (Thaler & Sunstein, 2009). We expect to see greater treatment effects when the statistics are more visualized. Future research might have to design randomized controlled trails to quantitatively measure differences in the effects.

In addition, in the experiment, we informed each participant that they should be wearing the fitness trackers at least 2 hours per day. A better way to implement is to ask all the participants to wear the band all the time, so that they would not take it off and then forget to put it back on.

In General, What Motivates You To Be More Physically Active?

Now we turn to the second question. 96 participants responded to this question. Most of the answers to this question are shorter and more succinct and there are clearer themes emerged for this question. We categorize the answers into five main themes and they include: (1) Health concern; (2) Mental health related, feeling better, stress relief, etc.; (3) Friends/peers matter; (4) Appearances, body image, losing weight/gain weight/keep fit; (5) External incentives, knowledge, work/personal experience.

First, about one-third of the responses of this question mention that health concerns are general motivation of being physically active. Examples are: *"Concerns about my health," "... I get sick less often when I am active," "Desire to be healthier."*

This also explains why the magnitude of the treatment effects of the field experiment is small. A lot of participants are intrinsically motivated to be physically active. It might be less likely to affect them by using external incentives. Those who concern health may have already been active, therefore, there is not so much room for them to improve.

The second theme is associated with mental health or stress relief, etc. About one-fourth of the respondents think being physically active is good for their mental health and for other jobs:

"I feel healthier and happier and I am more productive during the day," "... feeling better physically and mentally," "... Consistent exercise helps some against anxiety and depression as well," "I work out as a stress reliever..."

Third, we also see responses that emphasize the importance of friends and peers. Specifically, friends or people they know can be supportive for them to be physically active and hold them accountable.

"I am more motivated when my friends want to work out with me..." "Having someone to exercise with," "A teem/peers. It holds me accountable," "... having an exercise partner."

Fourth, one third of answers also highlight how their appearance, body image, and weight control motivate them to be physically active. For instance: *"Wanting to look and feel good," "Seeing myself lose muscle definition and gaining more fat ... " "A general dislike for how my body looks."* [28]

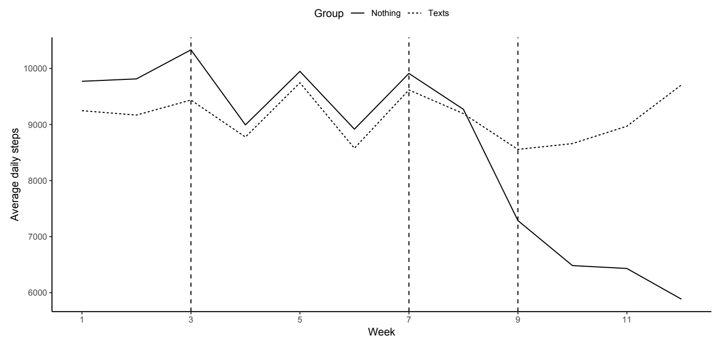
Lastly, more than one third responses also mention the importance of various of external incentives in motivating them to be physically active. One of them is about knowledge, for instance: *"... I am also more motivated when I know more about how food works in the human body. I have learned a lot about how food breaks down chemically and helps fuel the body, which pushes me to work out." "I work with patients that are injured and ill and see how a lifetime of inactivity and poor health maintenance can negatively impact a person in the future."* Other factors include *"good weather," "having a goal," "free time," "having a goal," and being healthy for "family"* are most often mentioned.

Descriptive Statistics: Qualitative And Quantitative Results

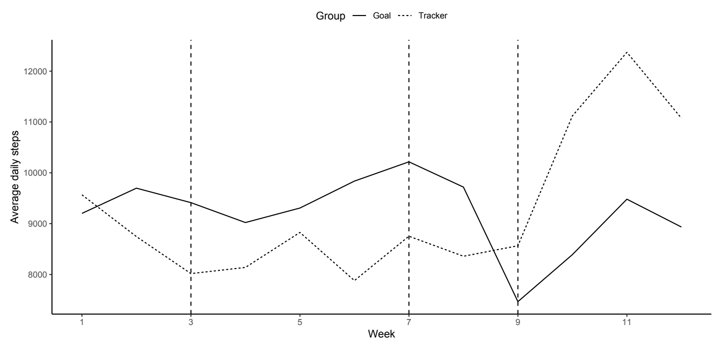
In this section we visualize how participants performed in the study based on the categories of their answers to the first question. We use the answers to the first question since this question is directly related to their experience in participation in the study. We plot a series of graphs showing trend in different outcomes by the six groups we categorized above. Figure 6 depicts the average daily steps for each of the six groups throughout the 12 weeks of the study. Most of the groups started with an initial level of between 9000 to 10000 daily steps, except that the group who

[28] Body dissatisfaction may be motivating, however, if the dissatisfaction comes from comparing with peers, it can affect exercise negatively. (Wasilenko, Kulik, & Wanic, 2007) find exposure to a fit peer has undermining effects on women's body satisfaction and exercise duration.

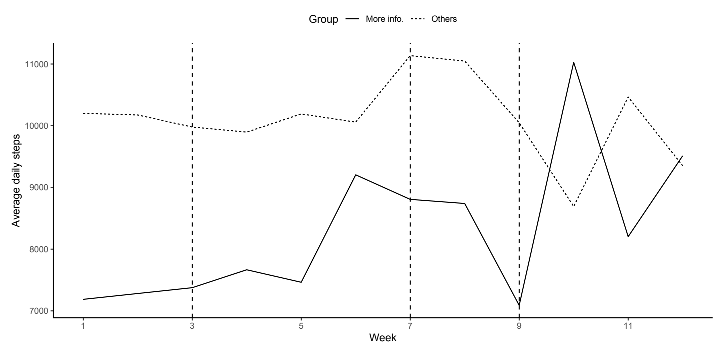
Figure 6: Average daily steps



(a) For groups “nothing” and “text”



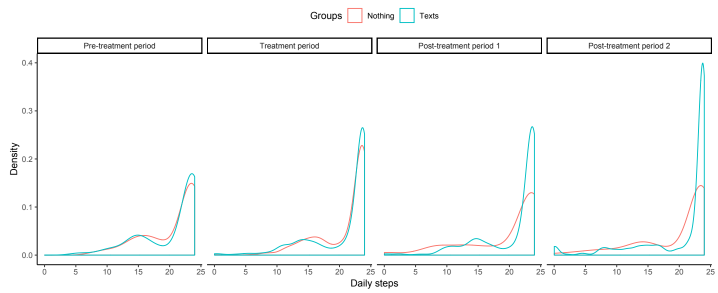
(b) For groups “goal” and “trackers”



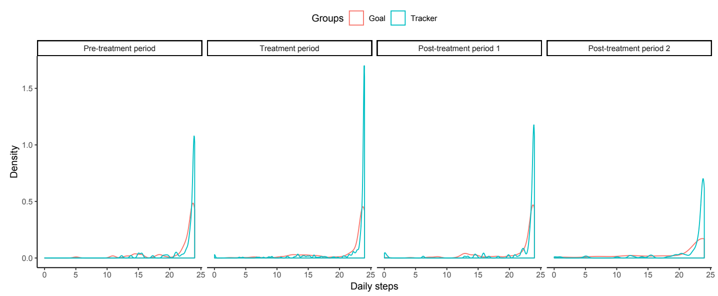
(c) For groups “additional info.” and “others”

Notes: The first vertical line divides the pretretment period and the treatment period. The second vertical line divides the treatment period and the post-treatment period. The third vertical line indicates the time when the follow-up survey was sent.

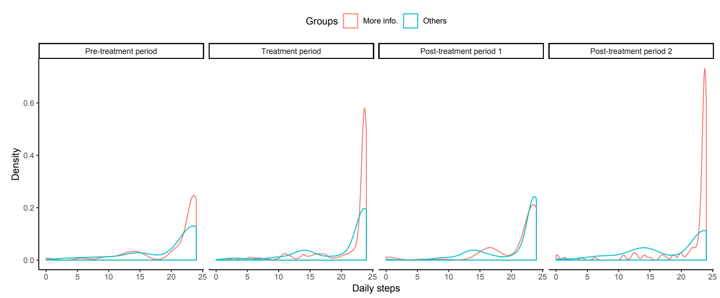
Figure 7: Average daily time of wearing of the fitness trackers



(a) For groups "nothing" and "text"



(b) For groups "goal" and "trackers"



(c) For groups "additional info." and "others"

Notes: The first vertical line divides the pretretment period and the treatment period. The second vertical line divides the treatment period and the post-treatment period. The third vertical line indicates the time when the follow-up survey was sent.

requested for additional information started off low (less than 8000 steps daily), and the group that belongs to "others" category (above 10000 steps per day). In panel (a), there is a sharp drop in the number of steps taken after the text messages were no longer sent for the group of participants that claimed nothing motivates. It would be interesting to test if the removal of the intervention cause backfire in the activity levels, like in Charness and Gneezy (2009), negative effects are found after the intervention (in this context, financial incentives) is removed. A sharp increase are seen for both the "more info." and the "tracker" groups after the follow-up survey was taken. Figure 7 shows the distributions of wearing hours for each group for each period. Surprisingly, more participants among those who complained about the functionality of the fitness trackers wear longer hours compared to other groups. Individuals who hoped more information would have been delivered and who commented about the direct effects of participation increased the time of wearing.

Conclusion And Discussion

Our results provide a valuable source in future research aiming to promote health behaviors. Our analysis reveals that being physically active is motivated mainly by health concern, psychological states, appearances/looking better, being fitter, etc. A lot of participants also report the importance of peers and quantification in holding them accountable. The results are consistent with previous study about motives for physical activity. For instance, Aaltonen, Rottensteiner, Kaprio, and Kujala (2014) conclude that intrinsic motives are associated with consistent leisure-time physical activity.

Although the results of the qualitative study provide additional evidence supporting the results from the quantitative statistics, it is important to note the limitations of this research. None of the themes is exclusive from another. Some responses include multiple themes. We simplified the descriptive analysis by assuming only one theme is reflected in each answer. In addition, the sample used in this study was not through a random selection process. Our dataset was limited to voluntarily recruited participants whose initial activity level was relatively high. Therefore, our sample is not representative of a larger campus population, and our findings are not generalizable beyond our sample. The present study also did not address any mechanism of behavioral differences between each group. We also did not incorporate relationship between individual beliefs and their attitudes reflected in the survey responses.

Despite the limitations, the present study has gained a richer insight in how program participants felt and thought. Some of them also offered insightful suggestions for improvement of future experimental design.

CHAPTER III HETEROGENEOUS EFFECTS OF INFORMEDNESS AND TEXT MESSAGES IN PREDICTING EXPERIMENTAL RESULTS

Introduction

Informational social influence plays a profound role in our daily decisions. In social psychology it is defined as "relying on other people as a source of information to guide our behavior" (Aronson, Wilson, Akert, & Sommers, 2016, p. 231). Do you choose to go to the polls and vote once you are eligible? What cutlery should you use when eating at a French restaurant? In situations like these when we are unsure what a good choice is, we internalize the behaviors of other people to direct our own actions. Informational social influence leads to conformity, which refers to a voluntary change in one's behavior due to the behaviors of others, when people exhibit compliance to their referencing group's social norms (Aronson, Wilson, Akert, & Sommers, 2016, p. 230).

One type of social norms is descriptive norms – what people commonly do (Cialdini, Reno, & Kallgren, 1990). A great amount of literature has established the power of descriptive norms in spurring or guiding people's behavior (Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007). For instance, households reduce their energy consumption when they receive utility bills that contain information comparing their energy use to that of their neighbors (e.g. Allcott, 2011; Allcott & Rogers, 2014; Ayres, Raseman, & Shih, 2013). Likewise, social comparison messages help to decrease residential water demand (Brent, Cook, & Olsen, 2015; Grossman, 1972). Other behaviors have been proved to be improved by descriptive norms include tax compliances (Hallsworth, List, Metcalfe, & Vlaev, 2017), charitable giving (Agerström, Carlsson, Nicklasson, & Guntell, 2016), etc. Other programs have been implemented to reduce undesirable behaviors like binge drinking (Peeler, Far, Miller, & Brigham, 2000), drug use (Donaldson, Graham, & Hansen, 1994), and gambling (Larimer & Neighbors, 2003).

In health related behaviors, field-based studies have provided evidence of positive effects of normative messages. Burger and Shelton (2011) found that the elevator use at the site where a descriptive norm was presented decreased by 46% from the first to the second week.[29] Priebe and Spink (2015) examined the effects of norm messages delivered via emails and found positive effects on light physical activity levels and negative effects on sedentary behavior in the workplace. When adults are told that 80% of their peers could hold their second plank position at least 20% longer than their first attempt, they can hold their plank longer on the second attempt than those who are not provided such information (Priebe & Spink, 2014).

Nevertheless, the magnitudes of the effects of descriptive norms vary across individuals and different behaviors. The reduction in energy usage is found to range from 0.81% to 2.55% in the randomized controlled trials using descriptive households' utility bills (Jachimowicz, Hauser, O'Brien, Sherman, & Galinsky, 2018). Little to no effect is found when information about their co-workers' retirement savings is revealed to employees (Beshears, Choi, Laibson, et al., 2015). In a field experiment in a large university setting, there is no clear evidence that disseminating descriptive norms affect gym use among college students (Beatty & Katare, 2018).

In a field experiment we conducted in a university setting, we attempted to uncover how patterns in the physical activity level of adults are affected by descriptive comparative information. We randomized our participants into a control and a treatment group. During the intervention period, each participant in the treatment group received a text message every day stating what their activity level of the previous day was in relation to that of other participants (comparative text messages), whereas each individual in the control group received a text message only describing their own activity levels of the previous day (plain text messages). We found a 3% increase in the number of steps taken during the intervention period, a greater effects after the information was no longer provided to the participants for about two weeks, and a rapid disappearance of the effects after that. Additionally, a greater treatment effect (as large as a 12% increase) was found among non-married and overweight individuals.

Accordingly, we have become interested in explaining these variations across individuals for a specific behavior. A number of scholars have argued that individuals' perceptions or beliefs about

[29]The descriptive norm says that 90% of individuals took the stairs instead of the elevator.

social norms play a critical role in predicting behaviors. People change their behavior or conform due to their misperceptions of social norms, or, imagined influence of other people. Therefore, this study offers an analysis to understand the mechanisms underlying the effects of normative triggers on health-related behavioral intentions.

The aim of this study is threefold. The first sets out to investigate if systematic errors – overestimation or underestimation – in individual beliefs about one's activity level exist. In addition, once they are exposed to or informed about the activity level of their peers, are there any changes in their self-assessment? The second examines the effectiveness of this informedness and comparative messages in potential health-related behavioral change. For instance, are people who are provided with information about their peers (informedness) more likely to change their behaviors? Are people more likely to expect a change in their behaviors in response to the type of information that compares their activity level with that of their peers? How do these two effects – informedness and comparative messages – interact with each other? For example, given that one is offered the facts about others, does a comparative message contribute to any difference in their predictions about their behaviors? The third seeks to shed some light on the role of first- and second-order normative beliefs in predicting activity level. We analyze results from cross-sectional survey data to answer our research questions. We explain each of these below.

First, we attempt to provide some evidence if individuals are mistaken in their beliefs about their activity levels. Previous research has found that people exhibit overconfidence or flawed self-assessment in abilities like driving skills (Horswill, Waylen, & Tofield, 2004) and committment to gym use (DellaVigna & Malmendier, 2006). Following Burks, Carpenter, Goette, and Rustichini (2013), we test if people show signs of overinflated self-views in their exercise levels. We ask our participants to rate how they believe the number of steps they take each day compares to that of an average university student and employee – above, at, or below the average. Next we attempt to alter their self-assessment by informing them of the actual average number of steps taken. We ask them to rate their own activity level again. We examine if self-assessment or beliefs change after this informedness. This first step allows us to compare our results about inflated self-assessment with those previous literature. It also motivates the results about intention-to-treat effects later. Intuitively, if people are mostly accurate about their own level, they are less likely to need to be nudged. On the other hand, if they are relatively underconfident (or overconfident), knowing others' true activity levels makes it less (or more) likely for them to increase their current level.

Next, we explore the link between informedness (the first steps) and the likelihood of changing one's behavior. We describe the field experiment to survey respondents and ask them to predict the change in their number of steps taken if they were the participants. Half of the participants are only shown the type of message that contains comparison information; the other half are shown the plain type that does not include comparison information. The details of the survey design will be explained in the next section. Although we are not able to measure the actual behavioral change for the same sample, according to Manski (2004, p. 1370), "persons respond informatively to questions eliciting probabilistic expectations for personally significant events." Our results elucidate the relative effectiveness of norms and comparative information in potential behavioral change.

Lastly, we collect participants' first- and second-order beliefs about the contribution of walking to health. The first-order personal beliefs are one's understanding of the events in the world, and second-order normative beliefs are one's understanding of someone else's beliefs. A recent line of research has stressed the importance of higher order beliefs in predicting people's behavioral change. Perkins and Berkowitz (1986) find that a college student drinks more alcohol if her belief about drinking differs from what she believes what others think about drinking. Second-order beliefs are also found to be a better way to predict culturally consistent behavior (e.g. Chiu, Gelfand, Yamagishi, Shteynberg, & Wan, 2010; Zou, Tam, Morris, et al., 2009). Jachimowicz, Hauser, O'Brien, et al. (2018) document that second-order normative beliefs have a causal effect on energy conservation behavior, over and above first-order personal beliefs, based on the results from a set of randomized controlled trials and an experimental study that manipulated second-order normative beliefs. Relatively little is understood about the role of personal beliefs in the effects or null effects of exposure to descriptive norms in health-related behaviors. We thus ask participants to report their beliefs and we control them in our analysis.

To preview our results, we find no clear evidence of overconfidence or underconfidence in individuals' activity level, which confirms the results in Byrne, Nauze, and Martin (2018) that no systematic overestimation or underestimation exists in evaluating one's energy usage.[30] In addition, we find informedness contributes to a smaller likelihood (by about 8 to 9 percentage points) that people expect to walk more steps if they are provided with either comparative or non-comparative text messages.[31] One possible explanation of the negative results is due to a larger number of people who are underconfident about their activity level. Once they update their self-assessment, they are less likely to walk more steps in response to the information intervention. Therefore, the result about the predictions is consistent with that of the self-evaluation. However, controling the variable indicating overestimation or underestimationi reveals that those who were correct about their activity levels prior to the informedness are associated negatively with the prediction results. Future study will have to either improve the experimental design or to explain the mechanism of this phenomenon. Finally, in opposition to the correlation between the second-order normative beliefs and the increased energy savings rate, we provide evidence that first-order personal beliefs, but not second-order normative beliefs, are associated with a larger likelihood of increasing the daily number of steps taken. This finding is less surprising if we consider the feature of exercise, which is less likely to be viewed as a socially significant behavior (like energy and water conservation, charity, recycling, etc.). As in the qualitative analysis of our field experiment, walking more steps or exercising is motivated by individual health or body image concerns. Social norms are unwritten rules about how people should behave, and it is less likely that social pressure would be exerted on those who do not engage in exercise. A possible reason for the positive results in field experiment might be that people feel accountable because of other people's participation.[32] These results are also consistent with those of Priebe and Spink (2011), that others' behavior is less influential in one's decision to be physically active.

[30] In our context, measured by the number of steps taken per day.

[31] Similar negative effects are found when the dependent variable is that people predict they are walking 1000 or more additional steps in response to the descriptive text messages.

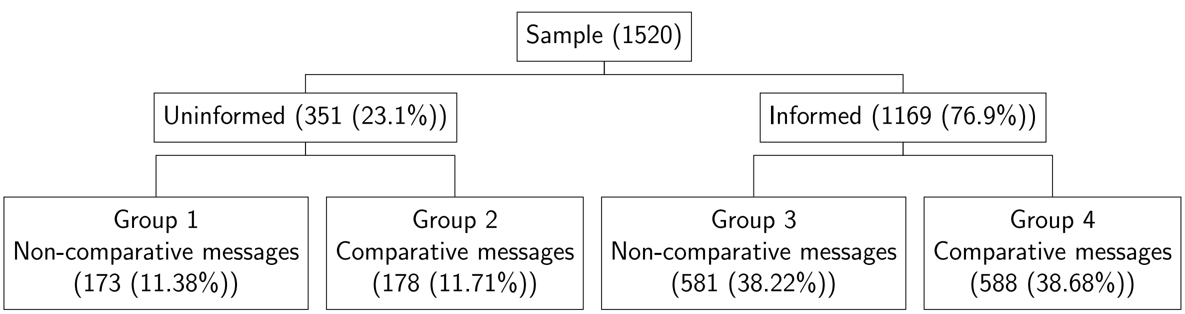
[32] Some survey responses support this guess.

Survey And Data

To provide evidence for the effectiveness of exposure to descriptive norm information (informedness) and the type of the text messages in change in behavioral intentions, we conducted a survey experiment. We sent it to the university employees and students through our mass email system in May 2019.[33] The first few questions asked the participants to approximate their daily activity level in terms of moderate-active minutes, vigorous-active minutes, and number of steps taken (or distance walked). Next, we asked respondents to report if they believe their daily numbers of steps walked was above, below, or at the average level, in relation to the daily number of steps of other university employees and students.

The survey uses two layers of randomization. Figure 8 demonstrates how participants are randomized into different subgroups. For the first layer, some respondents were presented with descriptive norm information, whereas others were not presented with such information:

Figure 8: Survey design: groups and sample size



Notes: This flow chart demonstrates how the randomization is achieved and the sample size of each subgroups. The "informedness" group (Groups 3 & 4) was presented to the question: *"On average, university staff and students walk approximately 7000 steps per day. For the average person, 7000 steps are approximately 3.2 miles (5.15 kilometers). Taking 7000 steps takes about 40 min (running) to an hour (brisk walking). Given this fact, please re-answer the question: Compared to the number of steps taken per day (including walking, jogging, or running) by staff and students at the University of Iowa, which of the following do you believe best describes your daily amount of walking?"*, while this question was not shown to the "no update beliefs" group (Groups 1 & 2).

[33] A copy of the survey is shown in the appendix. The survey was not sent to faculty members.

"On average, university staff and students walk approximately 7000 steps per day. For the average person, 7000 steps are approximately 3.2 miles (5.15 kilometers). Taking 7000 steps takes about 40 min (running) to an hour (brisk walking).

We call the group that was shown to the information above the "informed" group and the other "uninformed." After displaying to them the descriptive information, the "informed" group was asked to re-evaluate their activity level (in terms of the number of steps taken per day):

Given this fact, please re-answer the question:

Compared to the number of steps taken per day (including walking, jogging, or running) by staff and students at the University of Iowa, which of the following do you believe best describes your daily amount of walking?"

We use the descriptive text and the re-evaluation question for two purposes. We first test if self-assessment changes among those who are informed about the actual average activity level of others. We also use this sequence of questions to explore how informedness affects their predictions about their own behavior.

The second layer of randomization is achieved by showing participants different versions of the descriptive text messages in the field experiment. In that experiment, participants were randomly selected to receive a daily text message that contained comparative information or a plain text message with no comparative information. So in the survey, we described the field experiment and only one type of the text messages to each respondent. The description of the field experiment reads as follows:

We ran an experiment that was conducted at the University of Iowa in August 2018. The purpose was to determine if people walk more or less once they are provided information about how much they walked compared to other participants. 144 university staff and students participated in the study. On average, they initially walked about 8400 steps per day.

During the study, we provided daily text messages to the participants for a month. Half of the participants would receive a text message of the following format:

"Yesterday, you walked 7015 steps. "

(Or *Yesterday, you walked 7015 steps, you were ahead of 48% of your peers. If you walked 1600 more steps, you would be ahead of 58% of your peers.* for the other group.)

Everyone's text message reflected the specific number of steps they walked. Therefore, the number above would be replaced with real values.

All participants in the study received the same descriptive norm information by group.[34] We then asked them to predict if and how they would change their number of steps taken if they had received such text messages.[35]

The participants were asked to pick among nine choices: "> 1501 fewer steps," "1001 - 1500 fewer steps," "501 - 1000 fewer steps," "0 - 500 fewer steps," "No change," "0 - 500 more steps," "501 - 1000 more steps," "1001 - 1500 more steps," "> 1500 more steps."

The answers to this question serve as our dependent variables. The distributions of the responses by group is shown in Figure 9. Panel (a) shows the distribution of the answers when participants were asked to predict others' behaviors, and panel (b) shows that of their own predictions. The distributions suggest that most of the respondents believe they, as well as the participants of the field experiment, will increase their number of steps (73% predicted they would walk more steps, and 81% predicted others would walk more steps).

The last part of the survey contained a set of questions for demographics information, other individual characteristics, and questions about respondents' first-order and second-order beliefs.[36] Specifically, we asked them to report to what extent they believed the following statements:

"Walking contributes to multiple health benefits."

"Most people that I know think walking contributes to multiple health benefits."

"The majority of University of Iowa staff and students think walking contributes to multiple health benefits."

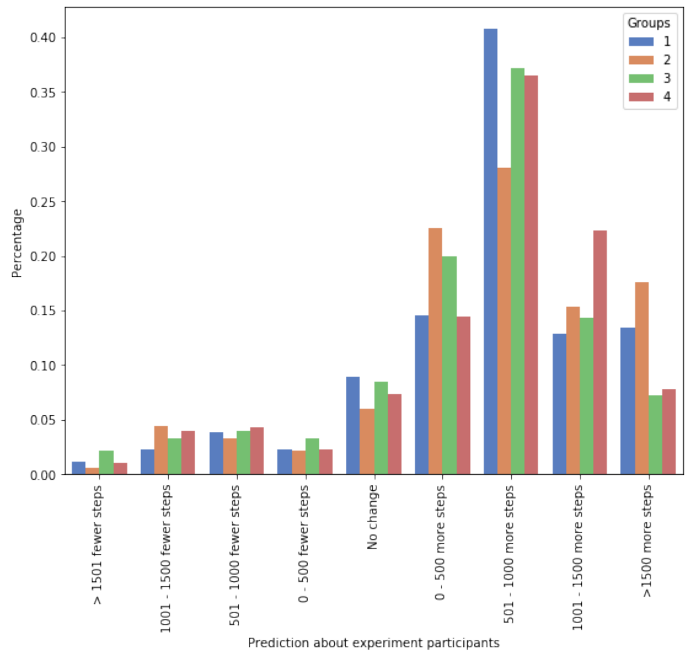
The response to these statements was given on a 7-point scale ranging from 1.00 (not at all)

[34] That they walked 7015 steps the previous day, and they are ahead of 48% of their peers and they would beat 58% of their peers if they had walked 1600 more steps.

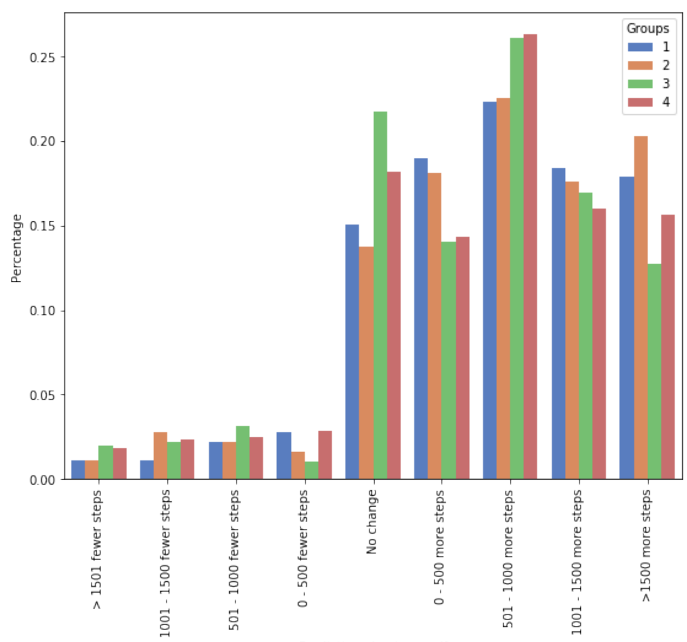
[35] We also asked them to predict how the actual participants of that field experiment performed.

[36] Variables of demographics information include age, gender, education, marital status, and employment status; variables of other characteristics include height, weight, ideal weight, ideal number of steps taken every day, if they consider themselves overweight, underweight, or about right, etc.

Figure 9: Distribution for outcomes of interests by group



(a) Predictions about experiment participants



(b) Predictions about oneself

Notes: Group 1: no update in beliefs and control group information; Group 2: no belief update and treatment group info; Group 3: with belief update question and control group info.; Group 4 refers: with belief update and treatment group info.

to 7.00 (very much) using a slide bar. Previous research categorized different levels of perceived norms depending on social distance between individuals and their peers. One level of social norms is proximate peers (e.g. friends and members of the same social groups), another is more distant peers (e.g. other students or a typical student at the same university) (Park, Klein, Smith, & Martell, 2009). Therefore, we use the second statement to control for second-order beliefs about proximal peers, and the third for second-order beliefs about distant peers. The first statement is to control for first-order personal beliefs.

To summarize our survey set-up, we randomized the informedness (descriptive norm information) and the types of the text messages delivered to participants, held constant the contents of the text messages by group and measured first- and second-order beliefs about individuals' activity levels. In this way we are able to do a pairwise estimation – for instance, given that people all receive the non-comparative type of text message, does informedness increase or decrease the likelihood of walking more steps?

The total number of responses add up to around 1800. We excluded suspicious responses, illogical responses, and those that were missing key information, for instance, responses that report the difference between their ideal weight and their actual weight being positive while they want to weigh more; daily distance of walking greater than 15 miles, etc. were invalidated. The sample of this study was taken from staff and students from the same university and it is comprised of 1520 responses, with about 73% female, 63% who obtained a college degree or higher, 44% married or living with a partner, and over 80% working part-time or full-time. They are on average 32 years old. About 43% of the respondents are overweight. They walk approximately 8500 steps per day.[37] Participants have a higher level of first-order personal beliefs (mean = 6.5, s.d. = 0.88) compared to their second-order normative beliefs (proximal norms: mean = 5.76, s.d. = 1.27; distal norms: mean = 5.61, s.d. = 1.26). The summary statistics about demographics and other characteristics by group are displayed in Table 7. Overall, the four groups are balanced in demographics, other characteristics, and self-reported activity levels.[38]

[37]This is comparable to the participants in our field experiment.

[38]To ensure that participants enter the best approximation for their daily number of steps, we allowed them to enter

Table 7: Descriptive statistics

	(1) Group 1		(2) Group 2		(3) Group 3		(4) Group 4	
	Mean	Obs	Mean	Obs	Mean	Obs	Mean	Obs
Panel A: Demographics								
Age	33.73	171.00	32.63	177.00	31.89	575.00	31.48	581.00
Female	0.68	171.00	0.78	178.00	0.75	580.00	0.71	585.00
Married	0.46	171.00	0.42	174.00	0.45	567.00	0.42	578.00
College	0.65	172.00	0.66	178.00	0.61	581.00	0.63	587.00
Employed	0.86	173.00	0.80	178.00	0.80	581.00	0.79	587.00
Panel B: Characteristics and self-reported activity levels								
Height (cm)	169.96	172.00	169.31	178.00	169.47	580.00	170.08	588.00
Weight (kg)	74.39	172.00	75.80	177.00	73.56	575.00	74.71	583.00
BMI	25.60	171.00	26.42	177.00	25.55	575.00	25.79	583.00
Overweight	0.41	171.00	0.50	177.00	0.42	575.00	0.43	583.00
Ideal weight (kg)	65.93	166.00	66.55	174.00	66.65	565.00	67.11	567.00
Moderate activity (min.)	101.27	173.00	110.51	177.00	102.50	576.00	96.18	587.00
Vigorous activity (min.)	40.23	173.00	33.71	178.00	38.59	580.00	34.29	588.00
Daily number of steps	8533.58	173.00	8573.80	178.00	8531.39	580.00	8301.61	588.00
Ideal daily num. of steps	11 340.20	170.00	11 076.47	178.00	10 874.94	578.00	10 597.03	586.00
Panel C: First- and Second-order beliefs								
First-order personal belief	6.54	173.00	6.55	178.00	6.49	581.00	6.50	588.00
Second-order normative belief (Proximal)	5.73	173.00	5.74	178.00	5.74	581.00	5.78	588.00
Second-order normative belief (Distal)	5.57	173.00	5.71	178.00	5.59	581.00	5.61	588.00
N	173		178		581		588	

Notes: College equals 1 if highest level of education completed is obtained an associate's or professional degree or higher; Married indicates married or living with partner. Employed include working part-time and full-time.

daily walking distance. We also offered a table that shows step equivalence by minute of different types of activities. We then applied Barreira, Rowe, and Kang (2010) and calculated stride length using their heights and then number of steps.

Results

Overestimation And Underestimation

Following Byrne, Nauze, and Martin (2018) and Burks, Carpenter, Goette, and Rustichini (2013), we provide preliminary evidence of overplacement or underplacement with respect to individuals' self-evaluation about their exercise level, measured in the number of daily steps walked. We assume that respondents' self-reported number of daily steps is close enough to their actual number of steps taken each day. We also assume they state the truth about their self-assessment (below, at, or above the average). Table 8 presents the joint distribution of individuals' daily number of steps taken and their beliefs. Since we are asking participants to evaluate based on their perceived average level of a typical staff and student of the same university, the distribution of the daily number of steps is not symmetric.[39] If all individuals held correct beliefs, the entries in parentheses along the left diagonal (the diagonal that starts at the extreme left of the top row) would have been 30.48%, 24.36%, and 45.16%, respectively. However, about half of the respondents (49.7%) believe they are "average." And 49% of the respondents are "correct."[40] There is an approximately equal split between those who "overestimate" (21.79%) and those who "underestimate" (29.03%).[41]

Table 10 displays the joint distribution of their number of steps and their "updated" beliefs.[42] It shows that 77% of the respondents are "correct" after they are informed about the true average number of steps of their comparative peers. Table C.2 presents the joint distribution of how "correct" respondents are about their self-evaluations prior to and after the informedness. 69% of

[39]Recall that in the survey, we stated the average number of daily steps taken by a typical university staff and student is about 7000 steps.

[40]Obtained from the summation of the percentages along the left diagonal: 12.9% + 14.68% + 21.59%

[41]Following Byrne, Nauze, and Martin (2018), we also looked if individuals whose number of daily steps that are at the tails of the distribution predict better than others. Among those whose number of daily steps falls in the top 20 percentile, 62% believe they are above average. Among those whose number of daily steps falls in the bottom 20 percentile, 49% of them believe they are below average. Among those whose number of steps in the middle 20 percentile, indeed, 55% of them predict they are "average." Therefore, participants are relatively correct how they are adhere to the norm.

[42]Because a proportion of participants were randomized to report their beliefs after the informedness, the sample size in this table differs from table 8. Table C.1 shows the joint distribution of the number of steps and their prior beliefs for this subgroup, the table suggests the descriptive statistics is similar to the whole sample.

Table 8: Joint distribution of self-reported daily steps and prior beliefs

	Beliefs about daily steps			
Daily steps (self-reported)	Below average (j_1)	Average (j_2)	Above average (j_3)	**Total**
Below the university average (k_1)	196	211	56	463
	(12.90)	(13.89)	(3.69)	(30.48)
Average of the university level (k_2)	83	223	64	370
	(5.46)	(14.68)	(4.21)	(24.36)
Above the university average (k_3)	37	321	328	686
	(2.44)	(21.13)	(21.59)	(45.16)
Total	316	775	448	1519
	(20.80)	(49.70)	(29.49)	(100.00)

Notes: The university average is about 7000 steps walked daily. Those who reported walking between 6000 to 8000 steps each day fall in the "average of the university level," those who reported walking more than 8000 steps fall in the "above average" category, and those who reported walking less than 6000 steps are therefore "below average."

Table 9: The empirical allocation function for beliefs

	Beliefs about daily steps		
Daily steps ("True" performance)	Below average (j_1)	Average (j_2)	Above average (j_3)
Below the university average (k_1)	42.33	45.57	12.10
Average of the university level (k_2)	22.43	60.27	17.30
Above the university average (k_3)	5.40	46.79	47.81

Notes: Assuming that each individual truthfully report their number of steps taken per day and their assessment about the level, the empirical allocation function indicates for each level of performance k, what fraction of individual put themselves in performance level j. Note that for every true level of performance k, the summation of the allocation to each believed performance level should be 100.

Table 10: Joint distribution of self-reported daily steps and updated beliefs

	Updated beliefs			
Daily steps	Below average	Average	Above average	**Total**
(Self-reported)	(l_1)	(l_2)	(l_3)	
Below the university average (k_1)	267	84	14	365
	(22.86)	(7.19)	(1.20)	(31.25)
Average of the university level (k_2)	38	163	66	267
	(3.25)	(13.96)	(5.65)	(22.86)
Above the university average (k_3)	10	54	472	536
	(0.86)	(4.62)	(40.41)	(45.89)
Total	315	301	552	1168
	(26.97)	(25.77)	(47.26)	(100.00)

Notes: The university average is about 7000 steps walked daily. Those who reported walking between 6000 to 8000 steps each day fall in the "average of the university level," those who reported walking more than 8000 steps fall in the "above average" category, and those who reported walking less than 6000 steps are therefore "below average."

Table 11: The empirical allocation function for updated beliefs

	Beliefs about daily steps		
Daily steps	Below average	Average	Above average
("True" performance)	(l_1)	(l_2)	(l_3)
Below the university average (k_1)	73.15	23.01	3.84
Average of the university level (k_2)	14.23	61.05	24.72
Above the university average (k_3)	1.87	10.07	88.06

Notes: Similar to table 9, the empirical allocation function indicates for each level of performance k, what fraction of individual put themselves in performance level l. Note that for every true level of performance k, the summation of the allocation to each believed performance level should be 100.

those who overestimate or underestimate updated their beliefs.

We now formally show if overestimation or underestimation exists based on the method in Burks, Carpenter, Goette, and Rustichini (2013) using Bayes' theorem. The Bayesian model is expressed as

$$Pr(Hypothesis|Data) = \frac{Pr(Data|Hypothesis)Pr(Hypothesis)}{Pr(Data)}.$$

The "Hypothesis" is individual's posterior beliefs about her number of steps taken per day. We use this model to test the relationship between the true distribution of the number of the steps and relative beliefs about the number of steps.[43] Given the pair of observations – their actual performance and the stated relative performance (below, at, or above the average), we are able to test how individuals' self-assessments given their number of steps taken per day. If individuals state the truth about their posterior beliefs in the survey, the largest group of individuals thinking they are in group k must belong to that group (Burks, Carpenter, Goette, & Rustichini, 2013). Table 9 demonstrates empirical results of an empirical allocation function, which is defined as the proportion of individuals in true group k assigning themselves to group j. Therefore, each cell of the matrix represents the share of individuals in group k who believe they are in group j. The Bayesian updating implies the diagonal condition: the entries with the largest values are on the diagonal Burks, Carpenter, Goette, and Rustichini (2013). Table 9 suggests that the diagonal condition is satisfied.

To test formally overestimation or underestimation, we numerically calculate values of a theoretical allocation function that best fit the data in Table 8 and satisfy the diagonal condition, by solving a constrained maximum likelihood (ML) problem proposed in Burks, Carpenter, Goette, and Rustichini (2013). If the empirical frequencies do not violate the diagonal condition, the values for the theoretical allocation function should be close to the empirical values. Indeed, the numerical calculation converges quickly and the resulting matrix is almost exactly the same as the empirical

[43] Relative belief, according to Burks, Carpenter, Goette, and Rustichini (2013), refers to how individuals believe about their performance compared to others. In contrast, absolute belief refers to how individuals believe about their performance compared to their actual level of performance.

results. Therefore, we fail to reject the null hypothesis that individuals beliefs are consistent with Bayesian updating.[44]

Table 9 demonstrates the empirical allocation function based on beliefs following informedness. Similarly, the diagonal condition satisfies and a larger group of people in group k actually think they are in that group. This implies the manipulation of the informedness is successful.

Therefore, we conclude that there is no over- or under-estimation with respect to self-assessment in one's activity level. This finding is consistent with the results in Byrne, Nauze, and Martin (2018), which find no evidence of systematic over- or under-confidence in households' pre-treatment beliefs about their energy use. [45]

Predictions

We now investigate how informedness and text messages with comparative information affect one's predictions about their future behaviors. We also looked at how these affect one's predictions about others' behaviors, the results are displayed in the appendix.

We study if individuals with "informedness" (they were provided with the information of true average number of steps by a typical university staff and student) predict differently from those with no informedness. In addition, to test the role of comparative information, we compare predictions by those who are presented with different types of text messages.[46]

We analyze the survey data using Logit and OLS regressions specified as following:

[44] We are also aware that simple "better-than-average" data alone can not be used to test for overconfidence (Benoit & Dubra, 2011). There might also be ambiguities when individuals are only asked to choose if they are above or below the mean (Burks, Carpenter, Goette, & Rustichini, 2013). In addition, our question asked the sample how their self-perceived level of exercise is compared to a different sample – the average of a university staff and student, not the average of all the participants. This also makes it difficult to interpret the results.

[45] Although exploring overconfidence or "illusionary superiority" in different contexts seems interesting, the current paper mainly focuses on a different research question.

[46] We would also compare how intention-to-treat effects differ from the actual effects estimated in the field experiment, but that is not the main research question of the current article.

$$
\begin{aligned}
Y_i &= \alpha + \beta_1 G_{-Informedness\,\&\,Comp.} + \beta_2 G_{Informedness\,\&\,-Comp} + \beta_3 G_{Informedness\,\&\,Comp.} \\
&\quad + \delta Norms_i + \gamma X_i + \varepsilon_i
\end{aligned}
$$

where Y_i is a binary variable. The original outcome variable – changes in the number of steps walked daily – consists of nine discrete values. A cursory glance at Figure C.1 reveals two ways the binary dependent variables are derived from the original prediction variable. First we leave the majority of the responses to the right, so $Y_i = 1$ if an individual believes she/he would walk more steps in response to the text message, and 0 if she expects no change or to walk fewer steps. We have 1106 (72.91%) responses that predict to increase a positive number of steps. The second way leaves the majority of the responses to the left, and thus $Y_i = 1$ if an individual believes she/he would increase the daily number of steps taken by 1000 or a larger magnitude, and 0 otherwise. The economic interpretation of this division is how likely it is people are walking a lot more steps. 490 (32.3%) responses predict that they will walk a lot more steps in response to the text messages.

The variables $G's$ are indicator variables representing each subgroup from the sample. The minus sign, $-$, is used to represent negation of the feature: for instance, $-Informedness$ indicates the group with no informedness. Using binary variables G allows pairwise comparisons between groups. For those who are not informed about the actual average level of activeness of others, the difference in the conditional probability that the outcome equals 1 between those who receive comparative messages and who do not receive comparative information is β_1. For those who received non-comparative information, the difference in the conditional probability that the outcome is 1 between those with and those without informedness is represented by β_2. For those who did receive comparative information, the difference in the conditional probability such that $Y = 1$ between those with and those without informedness is $(\beta_3 - \beta_1)$. The difference in the outcome between those who received and those who did not receive comparative messages, for those who were informed about the descriptive norm information, is $(\beta_3 - \beta_2)$.

Table 12: Predictions about oneself: main results 1

	Dependent variable: walk > 1000 more steps - oneself					
		Logit			OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Uninformed & Comparative info.	0.00840	0.0190	0.0156	0.00814	0.0194	0.0160
	(0.0513)	(0.0527)	(0.0523)	(0.0498)	(0.0515)	(0.0513)
Informed & Non-comparative info.	−0.0674	−0.0556	−0.0615	−0.0681	−0.0567	−0.0619
	(0.0411)	(0.0426)	(0.0424)	(0.0404)	(0.0422)	(0.0420)
Informed & Comparative info.	−0.0492	−0.0321	−0.0295	−0.0497	−0.0326	−0.0295
	(0.0412)	(0.0427)	(0.0426)	(0.0403)	(0.0421)	(0.0419)
First-order beliefs	0.0599***	0.0553**	0.0442*	0.0555***	0.0500**	0.0396*
	(0.0169)	(0.0177)	(0.0176)	(0.0158)	(0.0164)	(0.0165)
Second-order beliefs (Proximal peers)	0.00234	−0.00197	−0.000904	0.00279	−0.00102	0.000373
	(0.0132)	(0.0137)	(0.0135)	(0.0132)	(0.0137)	(0.0136)
Second-order beliefs (Distal peers)	−0.0163	−0.0147	−0.0114	−0.0168	−0.0152	−0.0118
	(0.0129)	(0.0134)	(0.0133)	(0.0131)	(0.0136)	(0.0136)
Correct self-eval. (prior)	−0.00761	−0.00161	−0.0112	−0.00710	−0.000569	−0.00949
	(0.0240)	(0.0249)	(0.0248)	(0.0240)	(0.0250)	(0.0249)
Female		0.149***	0.147***		0.154***	0.150***
		(0.0335)	(0.0335)		(0.0380)	(0.0380)
Age		0.00211	0.00225*		0.00216	0.00230*
		(0.00113)	(0.00113)		(0.00117)	(0.00116)
Observations	1516	1415	1403	1516	1415	1403
R^2/Pseudo R^2	0.0104	0.0217	0.0395	0.013	0.026	0.049
F statistic/Chi2	19.92	38.78	69.76	2.768	2.374	3.534
Prob > F/Chi2	0.0057	0.0012	0.000	0.007	0.002	0.000

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Columns (1) through (3) are Logit regression results, marginal effects are reported those columns. Columns (4) through (6) are OLS regression results. Other demographics includes education (indicatin college graduates or above), marital status (indicating married or living with a partner), employment (indicating full-time or part-time employment), height (in *cm*), weight (in *kg*), ideal weight (in *kg*); Activitie levels average daily moderate active minutes, vigorous active minutes, daily number of steps taken, ideal number of steps per day.

The marginal effects of the Logit and OLS regressions when the dependent variable indicates weather or not increase the number of steps by a positive number are shown in Table 12. First of all, given that participants receive non-comparative information, it is less likely that individuals who were informed about the descriptive norm information expect themselves to walk more steps in react to the text messages, by about 8 to 9 percentage points. Similarly, informedness contribute negatively to intention-to-treat effects for those who receive comparative text messages, by about 5 percentage points. Comparative text messages increases the likelihood that individuals predict to walk more steps, by about 2 percentage points if they are informed about the descriptive norms. However, for individuals with no informedness, those who were to receive the comparative messages are less likely to predict themselves to walk more steps, compared to those who were to receive the non-comparative information. In addition, unlike the previous studies that emphasize the importance of higher-order normative beliefs in predicting behaviors, our results yield no significant correlation between higher-order normative beliefs and predictions in number of steps taken. However, the results provide evidence that first-order personal beliefs are related to the predictions in one's future health-related behavior. An interesting side finding is that female respondents are more likely to predict an increase in daily number of steps taken in response to information provision, by as large as 20 percentage points.[47]

We use dependent variable that equals 1 when participants predict that they will walk a lot more (by at least 1000 steps) steps and re-do the Logit and OLS regressions. The results are shown in table 13. We obtain consistent effects in the conditional probability of outcome being 1 between the groups that received comparative text messages and that received non-comparative text messages, and between the groups that was informed about the descriptive norms and that was not informed with the norms. Specifically, receiving comparative text messages increases the likelihood of walking 1000 or more additional steps by about 1 percentage point. Informedness, likewise, contributes negatively to the likelihood of the intention-to-treat effects, by about 5 to 6

[47] A possible reason for this discrepancy between male and female respondents is that female respondents are more likely to be underconfident about their present level of exercise, therefore they are more willing to increase the number of steps given information of exercise levels about others. However, we do not find significant evidence that those who are underconfident are more likely to increase their number of steps taken.

Table 13: Predictions about oneself: main results 2

	Dependent variable: walk more steps - oneself					
		Logit			OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Uninformed & Comparative info.	-0.00823	-0.0243	-0.0253	-0.00793	-0.0248	-0.0263
	(0.0441)	(0.0444)	(0.0443)	(0.0470)	(0.0479)	(0.0482)
Informed & Non-comparative info.	-0.0835^{*}	-0.0912^{*}	-0.0974^{**}	-0.0827^{*}	-0.0911^{*}	-0.0955^{*}
	(0.0364)	(0.0365)	(0.0365)	(0.0382)	(0.0392)	(0.0395)
Informed & Comparative info.	-0.0554	-0.0692	-0.0716^{*}	-0.0547	-0.0698	-0.0705
	(0.0361)	(0.0361)	(0.0360)	(0.0381)	(0.0391)	(0.0393)
First-order belief	0.0572^{***}	0.0484^{***}	0.0471^{**}	0.0629^{***}	0.0531^{***}	0.0523^{***}
	(0.0139)	(0.0144)	(0.0145)	(0.0149)	(0.0152)	(0.0155)
Second-order belief	-0.0157	-0.0187	-0.0190	-0.0160	-0.0188	-0.0188
(Proximal peers)	(0.0125)	(0.0128)	(0.0128)	(0.0125)	(0.0128)	(0.0128)
Second-order belief	0.0134	0.0129	0.0119	0.0136	0.0131	0.0118
(Distal peers)	(0.0122)	(0.0124)	(0.0125)	(0.0124)	(0.0127)	(0.0127)
Correct self-eval. (prior)	-0.0470^{*}	-0.0428	-0.0366	-0.0474^{*}	-0.0437	-0.0380
	(0.0226)	(0.0231)	(0.0232)	(0.0227)	(0.0232)	(0.0234)
Female		0.229^{***}	0.229^{***}		0.225^{***}	0.226^{***}
		(0.0396)	(0.0399)		(0.0354)	(0.0356)
Age		0.000253	0.000375		0.000248	0.000283
		(0.00109)	(0.00110)		(0.00109)	(0.00109)
Observations	1516	1415	1403	1516	1415	1403
R^2/Pseudo R^2	0.019	0.051	0.0551	0.023	0.061	0.065
F statistic/Chi2	33.62	83.79	89.95	5.034	5.667	4.789
Prob > F/Chi2	0.000	0.000	0.000	0.000	0.000	0.000

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Columns (1) through (3) are Logit regression results, marginal effects are reported those columns. Columns (4) through (6) are OLS regression results. Other demographics includes education (indicatin college graduates or above), marital status (indicating married or living with a partner), employment (indicating full-time or part-time employment), height (in *cm*), weight (in *kg*), ideal weight (in *kg*); Activitie levels average daily moderate active minutes, vigorous active minutes, daily number of steps taken, ideal number of steps per day.

percentage points. Additionally, first-order personal beliefs again contributes to a greater chance of increasing a log of steps, not the second-order normative beliefs. We also find female respondents are more likely to walk more steps, by about 15 percentage points.

We also show results using OLS regressions without combining values of dependent variables in the appendix (see Table C.3). The data suggests further evidence of positive association between first-order personal beliefs and predictions of one's behaviors. It also confirms that female respondents are more likely to predict increase in the number of steps taken. As shown in Table C.4 in appendix, the results in predictions about others' behavior find consistent signs for effects of types of text messages and informedness.

Conclusion And Discussion

To conclude our findings, informedness contributes negatively to the likelihood that individuals predict to walk more steps in response to descriptive text messages. It might seem counterintuitive, but if we consider the positive correlation between first-order personal beliefs and the prediction results, the descriptive norm information might be playing a much smaller role in potential behavioral change. The important role that first-order beliefs play is consistent with the argument that physical activity is an intrinsically motivated behavior Aaltonen, Rottensteiner, Kaprio, and Kujala (2014). This is also consistent with the the survey responses from our field experiment.

Our study has limits. At present we are not in a position to determine the mechanism underlying the negative effects of informedness. Our findings suggest a need for examining causal effects of first- and higher-order normative beliefs in health-related behaviors. Additionally, future research may have to design randomized controlled trials where the change in behaviors due to informedness can actually be measured. And personal beliefs and other characteristics might be collected both before and after the treatment, or multiple times through out the study. Future research might also have to manipulate the contents of information delivered to participants based on their beliefs prior to treatment. For instance, given

We should also be cautious about the validity of the survey results. In literature where participants are asked to make predictions or self-evaluations, incentives or rewards are offered according to the accuracy of their answers so that values of the variables of interest can be measured more precisely. For instance, the payment in DellaVigna and Pope (2018) is calculated based on the formula: $1000 - (\textit{Mean Squared Error}/200)$. In Burks, Carpenter, Goette, and Rustichini (2013), experiment participants are paid additionally if their self-assessment is "correct."

Besides, as mentioned earlier, the self-evaluation uses below or above average. This could be problematic. More detailed information is needed to examine over- or under-estimation, like "percentage of people who believe they rank above each decile" and the "strengths of these beliefs" (Benoit & Dubra, 2011).

Other potential problems with regard to this study includes possible repetition of responses. People can use the link to take the survey multiple times. Also, a fixed number of respondents were randomly picked to be compensated for their participation and there was no other criteria for payment, this increased the probability of individuals taking the survey multiple times to increases their chances to be paid. If participants selectively chose different answers for outcome variables and held other personal characteristics the same, the results could be problematic. In addition, participants may predict the answers that researchers are looking for, they might also have picked answers that do not reflect their true potential reactions.

REFERENCES

Aaltonen, S., Rottensteiner, M., Kaprio, J., & 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.

Agerström, J., Carlsson, R., Nicklasson, L., & Guntell, L. (2016). Using descriptive social norms to increase charitable giving: The power of local norms. *Journal of Economic Psychology*, *52*, 147–153.

Allcott, H. (2011). Social norms and energy conservation. *Journal of public Economics*, *95*(9-10), 1082–1095.

Allcott, H., & 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., & Lovenheim, M. F. (2012). *Quantile treatment effects of college quality on earnings: Evidence from administrative data in texas*. National Bureau of Economic Research.

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

Aronson, E., Wilson, T. D., Akert, R. M., & Sommers, S. (2016). *Social psychology* (9th ed.). Pearson.

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.

Ayres, I., Raseman, S., & Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *The Journal of Law, Economics, and Organization*, *29*(5), 992–1022.

Barreira, T. V., Rowe, D., & Kang, M. (2010). Parameters of walking and jogging in young adults. *International Journal of Exercise Science*, *3*(1), 4–13.

Beatty, T. K., & Katare, B. (2018). Low-cost approaches to increasing gym attendance. *Journal of Health Economics*.

Benoit, J.-P., & Dubra, J. (2011). Apparent overconfidence. *Econometrica*, *79*(5), 1591–1625.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., & 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., & 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., & 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., & Fogg, L. (2013). Physical activity text messaging interventions in adults: A systematic review. *Worldviews on Evidence-Based Nursing*, *10*(3), 163–173.

Burger, J. M., & Shelton, M. (2011). Changing everyday health behaviors through descriptive norm manipulations. *Social Influence*, *6*(2), 69–77.

Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2013). Overconfidence and social signalling. *Review of Economic Studies*, *80*(3), 949–983.

Byrne, D. P., Nauze, A. L., & Martin, L. A. (2018). Tell me something i don't already know: Informedness and the impact of information programs. *Review of Economics and Statistics*, *100*(3), 510–527.

C, C. (2005). National health and nutrition examination survey. *NCFHS (NCHS), editor.(Hyattsville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention*.

Carlson, S. A., Fulton, J. E., Pratt, M., Yang, Z., & 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., & Cangelosi, K. (2013). Multicomponent intervention to reduce daily sedentary time: A randomised controlled trial. *BMJ open*, *3*(10), e003261.

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

Carrera, M., Royer, H., Stehr, M., Sydnor, J., & Taubinsky, D. (2018). The limits of simple implementation intentions: Evidence from a field experiment on making plans to exercise. *Journal of health economics*, *62*, 95–104.

Cawley, J., & 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. Retrieved August 29, 2017, from https://www.cdc.gov/nchs/fastats/exercise.htm

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

Chiu, C.-Y., Gelfand, M. J., Yamagishi, T., Shteynberg, G., & Wan, C. (2010). Intersubjective culture: The role of intersubjective perceptions in cross-cultural research. *Perspectives on Psychological Science*, *5*(4), 482–493.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, *58*(6), 1015.

Danaei, G., Ding, E. L., Mozaffarian, D., Taylor, B., Rehm, J., Murray, C. J., & 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.

DellaVigna, S., & Malmendier, U. (2006). Paying not to go to the gym. *american economic Review*, *96*(3), 694–719.

DellaVigna, S., & Pope, D. (2018). Predicting experimental results: Who knows what? *Journal of Political Economy*, *126*(6), 2410–2456.

Donaldson, S. I., Graham, J. W., & Hansen, W. B. (1994). Testing the generalizability of intervening mechanism theories: Understanding the effects of adolescent drug use prevention interventions. *Journal of behavioral medicine*, *17*(2), 195–216.

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

Etkin, J. (2016). The hidden cost of personal quantification. *Journal of Consumer Research*, *42*(6), 967–984.

Ferraro, P. J., & 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., & Melly, B. (2008). Estimation of quantile treatment effects with stata.

Gerber, B. S., Stolley, M. R., Thompson, A. L., Sharp, L. K., & 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.

Gilson, N. D., Faulkner, G., Murphy, M. H., Meyer, M. R. U., Washington, T., Ryde, G. C., ... Dillon, K. A. (2013). Walk work: An automated intervention to increase walking in university employees not achieving 10,000 daily steps. *Preventive medicine*, *56*(5), 283–287.

GPAQ. (2002). global physical activity surveillance. Retrieved August 29, 2017, from http://www.who.int/ncds/surveillance/steps/GPAQ

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political economy*, *80*(2), 223–255.

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, *148*, 14–31.

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

Horswill, M. S., Waylen, A. E., & Tofield, M. I. (2004). Drivers' ratings of different components of their own driving skill: A greater illusion of superiority for skills that relate to accident involvement 1. *Journal of Applied Social Psychology*, *34*(1), 177–195.

IPAQ. (2002). international physical activity questionnaire. Retrieved August 29, 2017, from http://www.ipaq.ki.se

Jachimowicz, J. M., Hauser, O. P., O'Brien, J. D., Sherman, E., & Galinsky, A. D. (2018). The critical role of second-order normative beliefs in predicting energy conservation. *Nature Human Behaviour*, *2*(10), 757.

John, L. K., Loewenstein, G., Troxel, A. B., Norton, L., Fassbender, J. E., & 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., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), 83–119.

Larimer, M. E., & Neighbors, C. (2003). Normative misperception and the impact of descriptive and injunctive norms on college student gambling. *Psychology of Addictive Behaviors*, *17*(3), 235.

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

Ledger, D., & McCaffrey, D. (2014). Inside wearables: How the science of human behavior change offers the secret to long-term engagement. *Endeavour Partners*, *200*(93), 1.

Liang, X., Wang, Q., Yang, X., Cao, J., Chen, J., Mo, X., … 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., & 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.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, *72*(5), 1329–1376.

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

Mayo, E. (1949). Hawthorne and the western electric company. *Public Administration: Concepts and Cases*, 149–158.

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

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

Park, H. S., Klein, K. A., Smith, S., & Martell, D. (2009). Separating subjective norms, university descriptive and injunctive norms, and us descriptive and injunctive norms for drinking behavior intentions. *Health communication*, *24*(8), 746–751.

Patel, M. S., Asch, D. A., & Volpp, K. G. (2015). Wearable devices as facilitators, not drivers, of health behavior change. *Jama*, *313*(5), 459–460.

Peeler, C. M., Far, J., Miller, J., & Brigham, T. A. (2000). An analysis of the effects of a program to reduce heavy drinking among college students. *Journal of Alcohol and Drug Education*, *45*(2), 39.

Perkins, H. W., & 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., … 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.

Priebe, C. S., & Spink, K. S. (2011). When in rome: Descriptive norms and physical activity. *Psychology of Sport and Exercise*, *12*(2), 93–98.

Priebe, C. S., & Spink, K. S. (2014). Blood, sweat, and the influence of others: The effect of descriptive norms on muscular endurance and task self-efficacy. *Psychology of Sport and Exercise*, *15*(5), 491–497.

Priebe, C. S., & Spink, K. S. (2015). Less sitting and more moving in the office: Using descriptive norm messages to decrease sedentary behavior and increase light physical activity at work. *Psychology of Sport and Exercise*, *19*, 76–84.

Prochaska, J. O., & DiClemente, C. C. (1982). Transtheoretical therapy: Toward a more integrative model of change. *Psychotherapy: theory, research & practice*, *19*(3), 276.

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American psychologist*, *47*(9), 1102.

Publications, H. H. (n.d.). walking: Your steps to health. Retrieved August 29, 2017, from https://www.health.harvard.edu/staying-healthy/walking-your-steps-to-health

Randazzo, K. D. (2016). How do social norms affect physical activity and performance on an endurance task?

Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The transsituational influence of social norms. *Journal of personality and social psychology*, *64*(1), 104.

Royer, H., Stehr, M., & 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, pp. 249–277). Elsevier.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, *18*(5), 429–434.

Schutt, R. K. (2018). *Investigating the social world: The process and practice of research*. Sage Publications.

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

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

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Toda, M., Njeru, I., Zurovac, D., Kareko, D., Shikanga, O., Mwau, M., Morita, K., et al. (2017). Understanding msos: A qualitative study examining the implementation of a text-messaging outbreak alert system in rural kenya. *PloS one*, *12*(6), e0179408.

Verbeek, M., & Nijman, T. (1992). Testing for selectivity bias in panel data models. *International Economic Review*, 681–703.

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

Wasilenko, K. A., Kulik, J. A., & Wanic, R. A. (2007). Effects of social comparisons with peers on women's body satisfaction and exercise behavior. *International Journal of Eating Disorders*, *40*(8), 740–745.

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

Yun, D., & 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.

Zou, X., Tam, K.-P., Morris, M. W., Lee, S.-l., Lau, I. Y.-M., & Chiu, C.-y. (2009). Culture as common sense: Perceived consensus versus personal beliefs as mechanisms of cultural influence. *Journal of personality and social psychology*, *97*(4), 579.

APPENDIX A: APPENDIX TO CHAPTER I

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, I 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 I 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 & Melly, 2008).

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

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

Sample selection

Since I 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 I use is selected on the basis of the response variable.

Appendix figure A.6 and Table A.7 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-committment is correlated with observable or unobservable individual time-unvarying characteristics. However, sample selection in a fixed effects context is a problem if selection is related to the errors.

I 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. I 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. I 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

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

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.[48] I use treatment condition and other individual characteristics (baseline observations, demographics)[49] 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), I 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:

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

A t test of $H_0 : \rho_1 = 0$ is a test of the null hypothesis of no sample selection. If the sample

[48] The choice of variables for x_{it} depend on availabe variables.

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

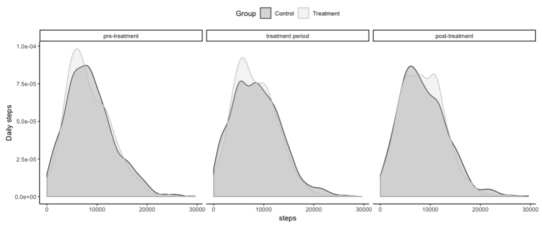
selection bias does exist, the estimate β_4 in the above model will be the interested estimate: the treatment effect during the period that I apply the above two steps.

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

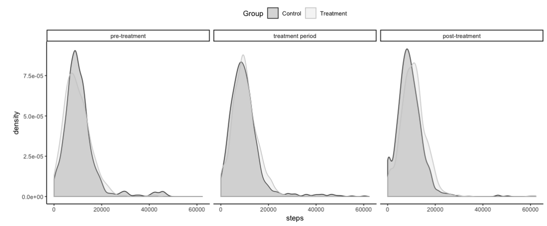
The stage two result indicate a statistically insignificant estimate for the Inverse Mills Ratio, therefore I fail to reject the null hypothesis that the sample selection bias is not an issue during the treatment period.[50]

[50] The results for stage 1 and stage 2 are shown in the appendix table A.8

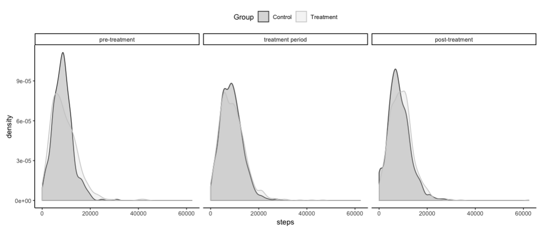
Figure A.1: Daily step distributions by period



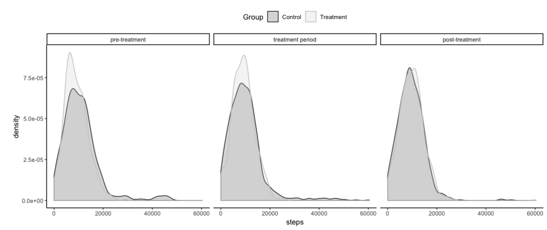
(a) Overweight individuals



(b) Non-overweight individuals



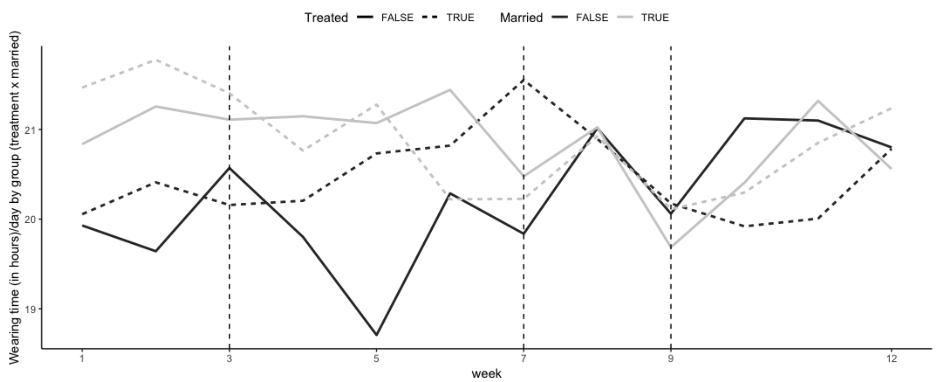
(c) Married or cohabiting individuals



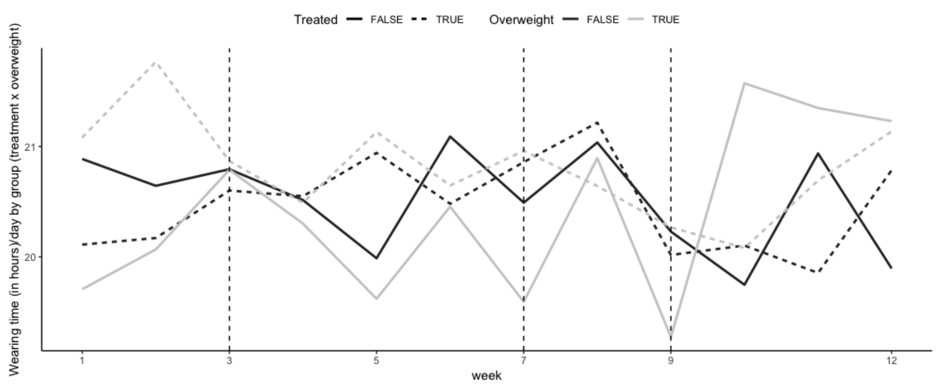
(d) Non-married or non-cohabiting individuals

Notes: This figure depicts the distributions of daily number of steps taken by group and by period. The post-treatment period refers to the six-week period after the intervention was removed.

Figure A.2: Wearing hours/day and Wearing time (in hours)/day by group



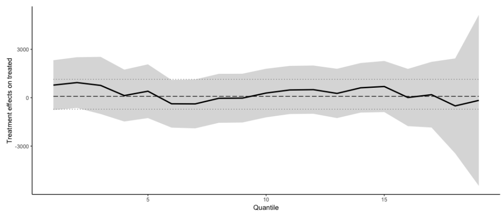
(a) By married x treatment



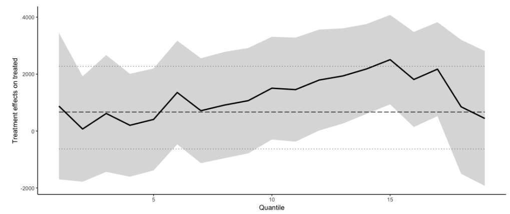
(b) By overweight x treatment

Notes: The black lines in (a) and (c) are for the non-married (or non-cohabiting) individuals and those in (b) and (d) are for non-overweight individuals. The grey lines in (a) and (c) are for married or cohabiting individuals and those in (b) and (d) are for overweight individuals. Dashed lines in all panels represent the treatment group and solid lines represent the control group.

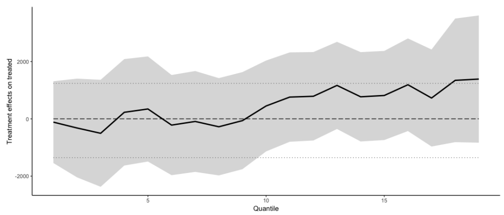
Figure A.3: Quantile treatment effects (weekly)



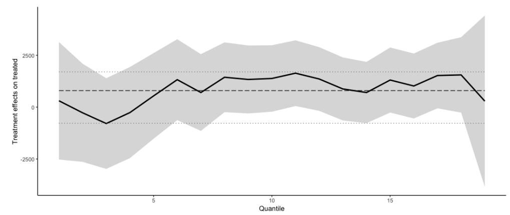
(a) week 3



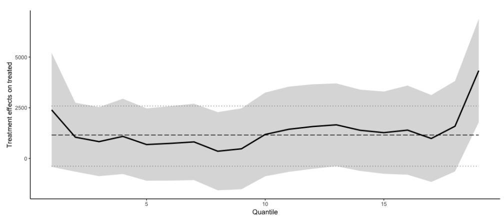
(b) week 8



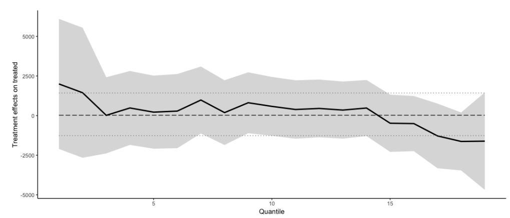
(c) week 4



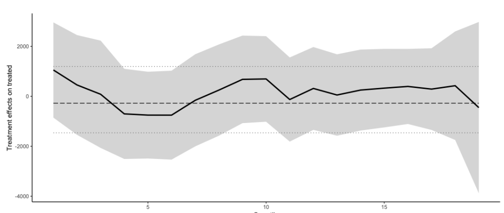
(d) week 9



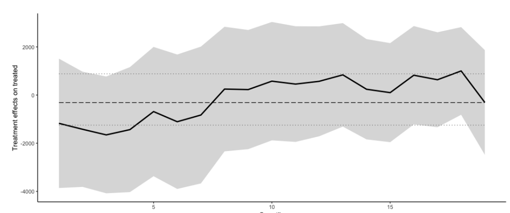
(e) week 5



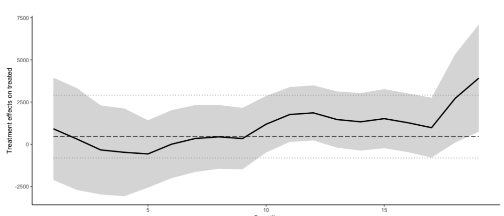
(f) week 10



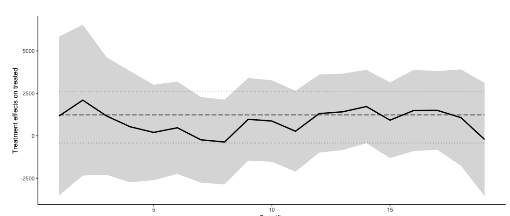
(g) week 6



(h) week 11



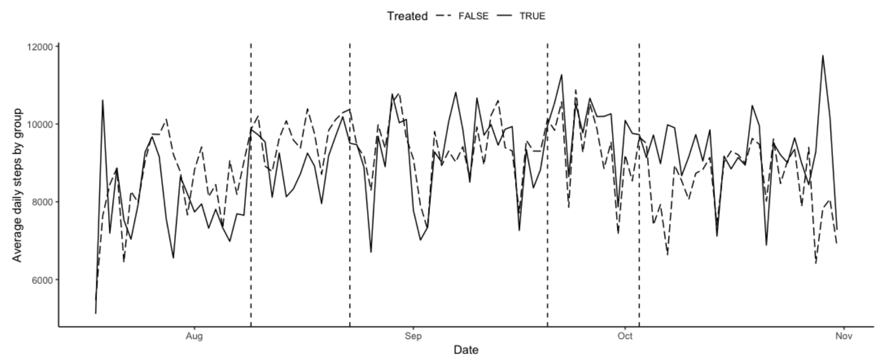
(i) week 7



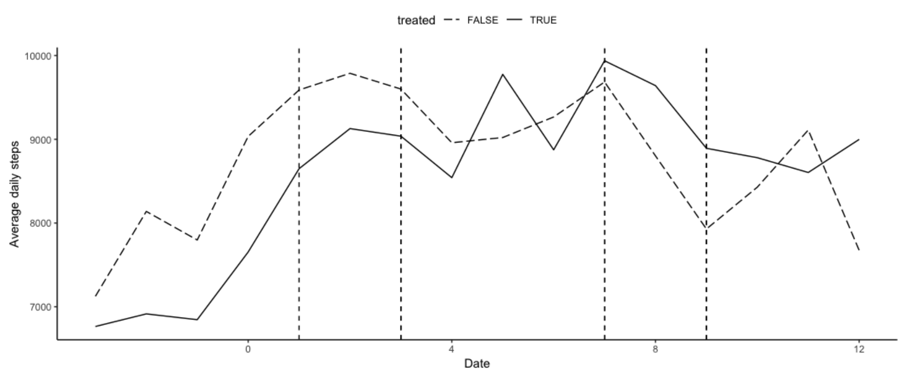
(j) week 12

Notes: This figure depicts the results for quantile treatment estimates for each week.

Figure A.4: Average daily steps by group, including the enrollment period



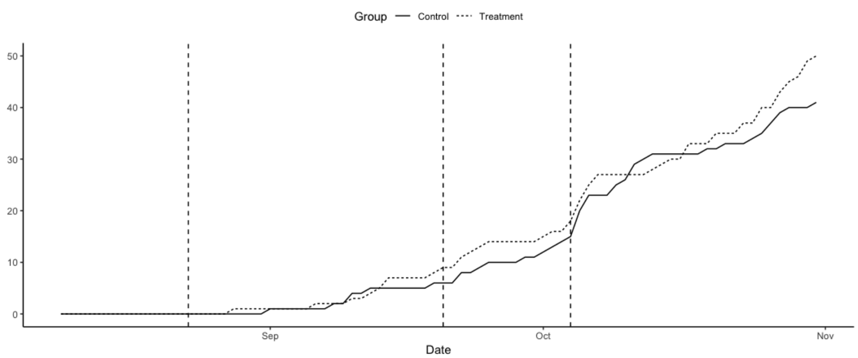
(a) Average daily steps (including the enrollment period)



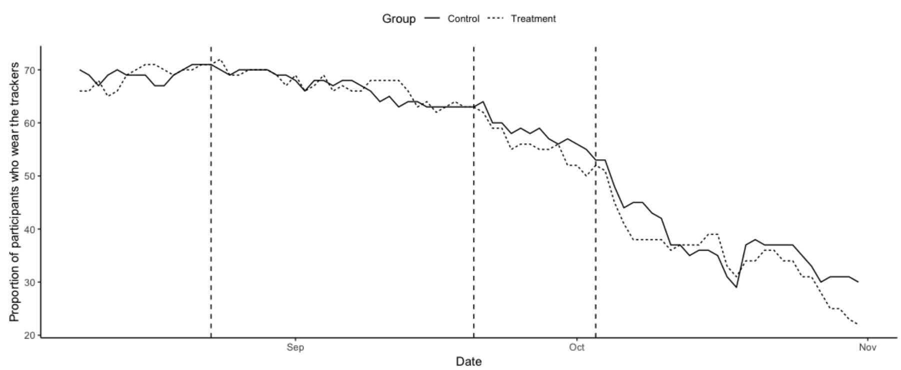
(b) Average daily steps (smoothed on a weekly basis; including the enrollment period)

Notes: The first vertical dashed line indicates the beginning of the study; the second indicates the beginning of the treatment period; the third one indicates the end of the treatment period; the last 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.5: Tracker adherence



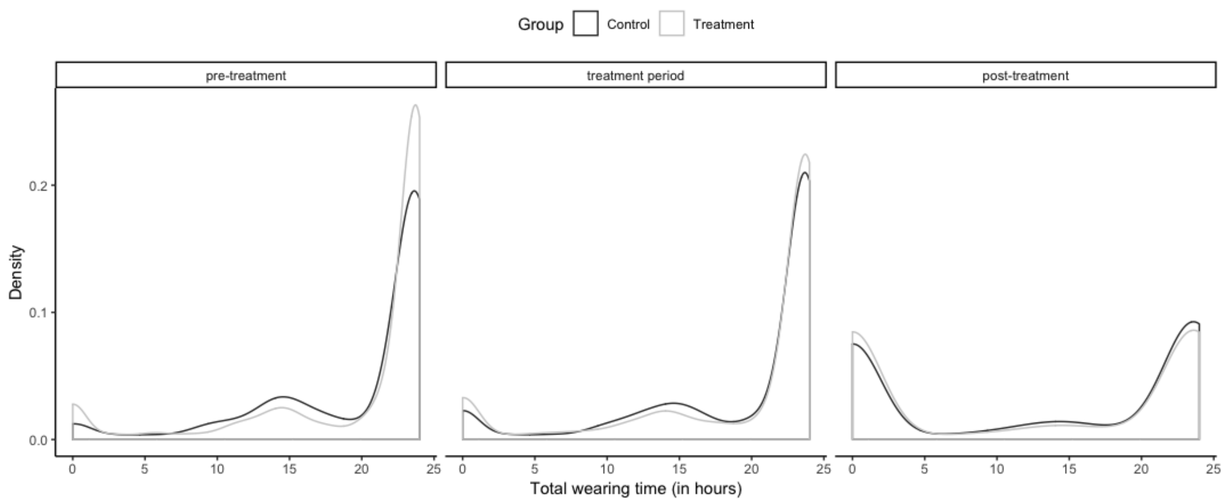
(a) Number of dropouts



(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.6: 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.

Table A.1: 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 and after	
Oct.4	A follow-up survey sent through emails.
Jan, 2019	End of fitness tracker's data collection.

Notes: The entire recruitment period took three weeks, during which some participants have started using the fitness tracker.

Table A.2: Descriptive statistics (self-reported)

Variable name	Mean value of the control group	Obs.	Difference between the treatment and the control groups	Obs.
BMI	25.860	70	-.278	143
			(0.795)	
Obese	0.171	70	0.020	143
			(0.065)	
IPS score	22.831	71	1.484	144
			(0.978)	
RPI score	2.946	71	0.056	144
			(0.064)	
Moderate active/week (minutes)	494.735	66	112.908	136
			(109.901)	
Vigorous active/week (minutes)	183.358	67	12.058	139
			(54.813)	
Sedentary time/day (hours)	6.186	64	0.189	132
			(0.466)	
Want to weigh less	0.732	71	0.007	144
			(0.074)	
Ideal weight (*kg*)	69.087	53	-1.209	111
			(2.257)	

Notes: Standard errors are in parentheses;Column of the difference: regression estimates. IPS (Steel (2010)) ranges from 9 to 35. A lower score indicates "first things first", while a higher score implies high level of procrastination; RPI: Resistance to peer influence (Steinberg & Monahan (2007)) *p<0.1; **p<0.05; ***p<0.01

Table A.3: 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.4: Dependent variable: wearing hours

VARIABLES	(1) After-treated	(2) Treatment period	(3) Post-treatment period	(4) Post-treatment period (Before the follow-up survey taken)
Treatment effects	0.529	0.704	0.549	0.348
	(0.600)	(0.490)	(0.911)	(0.913)
Observations	9936	6048	9936	8064
R-squared	0.044	0.069	0.178	0.000
Number of individuals	144	144	144	144
FE	YES	YES	YES	YES

Notes: Standard errors (clustered at the individual level) in parentheses; $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ This table shows the statistical results of treatment effects when the dependent variable is the hours of wearing of the fitness trackers from difference-in-difference regressions.

Table A.5: 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.6: 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.7: 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 *i*th week, i = 1, 2, $\cdots$, 12.

Table A.8: 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.

APPENDIX B: APPENDIX TO CHAPTER II

A full list of the open-ended question responses

I. Which elements of the study, if changed, do you feel would motivate you to be more physically active?

1. Participation in the study, mainly receiving text messages, worked or did not work, because of reasons like peer effects, either by competition or supportive activities, receiving feedbacks, or the reminding feature. (*labeled as "texts"*)

    (1) Knowing others were participating.

    (2) Text encouragement

    (3) Competition with peers.

    (4) I would like to see how my steps compared to the steps of other people.

    (5) Texts about your peers and where you are 100 percent of the experiment

    (6) Knowing how I was doing compared to other people

    (7) see how I compare to others during the day.

    (8) Knowing how much I am walking in comparison to others.

    (9) Knowing where I stand relative to other participants would have been a likely motivator for me. Additionally, having fun goals of objectives may have helped me to do more.[51]

    (10) Perhaps the statistics of where I was in relation to other people because there would be a competitive side to it.

    (11) If I received the other type of message

    (12) The daily text

    (13) The nagging text messages...about syncing the band

    (14) Longer duration of text reminders, and adding the percentiles into the texts

    (15) Knowing where I stand compared to my peers

    (16) Tell you are where you compare to others

[51] This response is categorized as "texts" in the quant analysis, although it is also in the "goal" category.

(17) I think if I saw the progress of the others in the study, or if there was an anonymous chat group that let us talk to each other about how we're doing, it would help motivate me more. The mild competition mixed with group support would help.

(18) Text messages with steps and how I compare to others

(19) I was pretty motivated to wear the device every day, walk to my car when the weather was good, etc

(20) Texts messages for longer than 6 weeks

(21) Encouragement in the texts instead of just a report of how many steps I took with a reminder to sync

(22) 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.

(23) I enjoyed having the feedback from the day before, it was super motivational to be receiving

(24) I think it is simply a good reminder to keep moving

(25) step count motivate me to walk more

(26) 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.

2. Functionality of fitness trackers (*labeled as tracker*)

(1) I wish the band would auto-sync. Sometimes I did not sync the band in time, so the number of steps was incorrect in the text message. I wish it was an option to automatically sync at midnight.

(2) If device worked better.

(3) If I could read the number of my steps on the band vs having to check my phone during the day. Access to my phone is limited by my job, so I didn't know how many steps I'd taken until the end of the work day.

(4) Tracker with information displayed on band

(5) Visually tracking the physical activity or steps[52]

[52] Couldn't match this response to steps (fitness tracker) data, because the participant did not enter the assigned study name.

(6) Better tracking device
(7) If the tracker showed information on it.
(8) Easier to get feedback directly from device and not just the app
(9) The fitness band needs to be more interactive, this one was way too basic
(10) I was disappointed because the tracker didn't record the time I spent rowing, either on the river or at the gym. Having that data included would have helped understand my overall activity level & made me more motivated
(11) Band that shows steps on screen
(12) If the tracker was able to automatically detect exercising.
(13) my band didn't keep track of anything except for steps & sleep at night. didn't track any exercise or naps during day. I didn't get it, it didn't do anything for me & I took it off one day to do something & never put it back on again-kinda forgot abt it. had no info to motivate me. I can bring it back to you if you want.
(14) Interactive wearable device to monitor goals without logging in and syncing with app. A drawing that is opened up only to the top performers for one greater sum.

3. Additional information would be helpful (*labeled as "add info"*)

(1) More information to compare my status of exercise during the week to other users.
(2) Measure of heart rate exertion and not just counting steps. Non-impact aerobic activity doesn't necessarily show up in the step count.
(3) Also providing updates about sleeping patterns.
(4) If information from my other exercise (lap swimming) were included so I had a holistic picture of my exercise for the day. There were days I swam over a mile (for 50 minutes) but didn't meet the step goal. It was frustrating not to have the complete picture of my activity represented in the app data.
(5) Continual heart rate monitor, ability to do behavior tagging, expansion of activity and behavior tagging to include yoga
(6) If the study measured more than just steps. During the study period I have also done strength training and cycled over 300 miles.

(7) reporting actual steps compared to goal steps, incorporating measurements of exercise other than only steps (e.g., minutes of heart rate elevation), more accurate step counter[53]

(8) Tips for how to get more steps. For instance - walk around the block would add 400 steps to your daily total steps

(9) List of activities happening in the area

(10) A buzz feature after long periods of being sedentary.[54]

(11) Send updates, how well i walked and slept the previous day in the morning. Updates on if i'm close to my daily goal. Reminder to update weight.[55]

4. Goals related (*labeled as "goal"*)

(1) If there were goals that I had to attain.

(2) Help in setting goals rather than texts comparing me to others

(3) incentives for reaching your goal

(4) goals

(5) Study pre-sets goal and offers monetary incentive to meet those goals

(6) seeing my goal not reach was possibly encouraging

(7) I think it could have been fun to have a "goal" to work towards in monthly steps with facts about exercise or something that were unlocked with each milestone met.

(8) Daily goals.

(9) Putting monthly/weekly incentives (monetary) on reaching certain exercise goals within a specific time

(10) Texts with a countdown to goal, like one at noon that you are X% from your goal for the day

(11) Small goals might help like: take one more flight of stairs today. Part a little further from your destination today.

[53] This response is catagorized as "goal" in the quantitative analysis.
[54] The respondent quit; not included in the quantitative analysis.
[55] The respondent quit; not included in the quantitative analysis.

(12) reporting actual steps compared to goal steps, incorporating measurements of exercise other than only steps (e.g., minutes of heart rate elevation), more accurate step counter[56]

(13) reached goal

(14) I think maybe a reminder of personal goals, instead of comparing me to peers (who I will never know) would have been more motivating. For example, something like "Yesterday you were 200 steps from completing your goal! If you walk 10 more minutes today, you'll pass your goal!"

(15) Send updates, how well i walked and slept the previous day in the morning. Updates on if I'm close to my daily goal. Reminder to update weight.[57]

(16) Maybe pointers on being more active/ simple ways to become more active. That, or putting in current goals as well as future goals.

(17) Knowing where I stand relative to other participants would have been a likely motivator for me. Additionally, having fun goals of objectives may have helped me to do more.

5. Nothing motivates me to walk more steps (*labeled as nothing*)

(1) I don't think anything would make me more active

(2) Nothing, I was already very physically active.

(3) Probably nothing. The work I do and the workout class I take is enough activity for me right now.

(4) I was not challenged by text messages

(5) None come to mind - I typically exercise before work

(6) Needs to be a personal decision.

(7) Nothing, I don't think it changed my motivation level at all.

(8) health concern[58]

(9) None

6. Others

[56]This response is categorized as "goal" in the quant analysis.
[57]The respondent quit; not included in the quantitative analysis.
[58]The respondent quit; not included in the quantitative analysis.

(1) Having an exercise routine to complete during the week.
(2) More accountability.[59]
(3) Interactive wearable device to monitor goals without logging in and syncing with app. A drawing that is opened up only to the top performers for one greater sum. [60]
(4) If fitness was a requirement
(5) Letting me know how much more I should walk
(6) Working toward a reward
(7) Calories burned
(8) More money offered. More awareness if the weekly drawing even happened.
(9) Allowing more flexibility in band sync times, i work full time nights so it was easier to sync in the morning when i went to bed.
(10) Incentives for being active
(11) If I could wear the exercise band when I shower then I would not take it off and forget to put it back on.
(12) I would replace the texts a graphic or even a gif that displayed the my step count and other info
(13) More regular text messages and different time (after work).
(14) Telling you how far behind/ahead you are mid-day instead of day after.
(15) 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 peerr"
(16) a community of people that not only said what they were doing to exercise but also invited others along with them to exercise.

II. In general, what motives you to be more physically active?

1. Healthy ("health")

(1) Health and appearance
(2) weight loss and health

[59]The respondent quit; not included in the quantitative analysis.
[60]This response is more of a "tracker" category, which is how it is categorized in the quantitative analysis.

(3) Personal health
(4) Desire to be healthier
(5) I feel healthier and happier and I am more productive during the day...
(6) ... and staying healthy
(7) My health.. health benefits of being physically active.
(8) To be strong and healthy.
(9) It is important for me to be active and healthy.
(10) In general, being healthy and feeling good, but this is mainly for light exercises...
(11) Health and stress management.
(12) Health concerns
(13) Concerns about my health
(14) being more healthy
(15) My job: I work with patients that are injured and ill and see how a lifetime of inactivity and poor health maintenance can negatively impact a person in the future. Also, my parents are young (under 60) and have collectively gone through breast cancer and several heart attacks. I don't want this for myself or my future family. Finally, I simply feel better when I'm regularly active.
(16) Wanting to be in better health.
(17) ... , desire to be healthier
(18) I want to be healthy and maintain a good weight.
(19) Health and appearance
(20) Want to be healthy
(21) ... and because I get sick less often when I am active
(22) improve health
(23) The health benefits
(24) To stay fit and healthy.
(25) physical health
(26) ... for my own health and productivity
(27) enjoyment, healthly [healthy] living
(28) Health concerns

(29) Feel healthier, less aches, more energy...

(30) Health

(31) Overall health

(32) being healthy & feel better because not sitting around all time.

(33) It makes me feel better when I do it consistently. Helps keep my joints more limber, helps with my arthritis symptoms.

(34) health concern

(35) Feel healthier, less aches, more energy. Consistent exercise helps some against anxiety and depression as well.

(36) Knowing my genetic risks for heart disease and diabetes... Having new things offered to me, like new classes at the CRWC or new events that involve physical activity.

(37) Being healthy for my family.

2. Mental related, feeling better, stress relief ("mental")

(1) Feeling better

(2) how I feel mentally and ...

(3) Feel good/relaxed...

(4) I feel healthier and happier and I am more productive during the day. I also look better and feel more confident.

(5) In general, being healthy and feeling good, but this is mainly for light exercises. What drives me to do high impact or consistent exercise is building skills, competition/games/fun, stress/anger relief, and looking a little fitter.

(6) Health and stress management.

(7) ... emotional well-being

(8) weight, stress

(9) ... Finally, I simply feel better when I'm regularly active.

(10) ... feeling better physically and mentally.

(11) Makes me feel good and helps with stress.

(12) happiness, ...

(13) Stress management and how I feel afterward. This is especially important as I work 40 hours a week and commute to take graduate classes.
(14) stress, i am active to think through my issues
(15) Weight fluctuations and the high I get from exercising.
(16) I work out as a stress reliever ...
(17) More or less just wanting to feel confident in physical abilities.
(18) ... and feeling good afterward exerciing [exercising]...
(19) To feel better, look better and for my own health and productivity. exercise improves my mental function and makes me more focused, less distracted and more motivated to work.
(20) ... I like the way exercise makes me feel like I accomplished something.
(21) enjoyment, healthly [healthy] living
(22) Feel healthier, less aches, more energy. Consistent exercise helps some against anxiety and depression as well.
(23) being healthy & feel better because not sitting around all time.

3. Friends/peers matter ("peers")

(1) Not having any injuries currently and being surrounded by other fit people who go workout.
(2) I am more motivated when my friends want to work out with me...
(3) ... being able to play sports recreationally with others
(4) ...seeing peers exercise...
(5) Having someone to exercise with
(6) A friend to exercise with and hold me accountable.
(7) ...What drives me to do high impact or consistent exercise is building skills, competition/games/fun, stress/anger relief, and looking a little fitter.
(8) someone to work out with...
(9) ..., having an exercise partner
(10) Friends' support mostly...
(11) Peers, support from friends

(12) ... making physical activity social
(13) ... Having friends to exercise with...
(14) Someone to workout with
(15) Getting up in the morning and doing more that my peers.
(16) A team/peers. It holds me accountable.
(17) working out with friends, especially for some friends I only get to see when I go to the gym
(18) friends and family, good weather, keeping in shape

4. Appearance, body image, losing weight/gain weight ("appearance")

(1) Health and appearance
(2) ... and how my clothes fit
(3) weight loss and ...
(4) Physical appearance, or weight loss (seeing results)
(5) ... Lose weight
(6) Seeing physical results
(7) Physical appearance...
(8) body image...
(9) I feel healthier and happier and I am more productive during the day. I also look better and feel more confident.
(10) Losing weight and ...
(11) ... , or if I want to lose weight
(12) Gaining weight
(13) ...What drives me to do high impact or consistent exercise is building skills, competition/games/fun, stress/anger relief, and looking a little fitter.
(14) wanting to look and feel good
(15) Seeing myself lose muscle definition and gaining more fat in my abdomen
(16) weight, stress
(17) ..., feeling like I'm out of shape
(18) I want to be healthy and maintain a good weight.

(19) Losing weight, feeling better physically and mentally.

(20) ..., physical fitness, and physical appearance.

(21) A goal to lose weight, look better, etc

(22) Health and appearance

(23) Wanting to stay in shape.

(24) Weight fluctuations and the high I get from exercising.

(25) My goal for weight loss

(26) wanting to look better

(27) Having people find me attractive

(28) To ... look better and ...

(29) Looking in a mirror, stepping on a scale, or pants not fitting

(30) A general dislike for how my body looks. Occasionally, I'll get a mental push to just be healthier, but more often than not my motivation lies in just wanting to feel better about my body.

(31) friends and family, good weather, keeping in shape

(32) Higher quality of life, physical fitness, and physical appearance.

5. Other external incentives, knowledge, work/personal experience ("external")

(1) When I can be outside

(2) Not having any injuries currently and being surrounded by other fit people who go workout.

(3) ... 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 down chemically and helps fuel the body, which pushes me to work out

(4) ... career aspirations

(5) incentives

(6) I try to make my step goal everyday...

(7) ..., seeing the hard work pay off

(8) observing the benefits of exercise - strength, flexibility and emotional well-being.

(9) Feeling tired.

(10) The desire to accomplish my personal goals regarding physical fitness.

(11) a goal, such as a race like a marathon

(12) Certain times of the year, certain moods, etc.

(13) my own well being

(14) My job: I work with patients that are injured and ill and see how a lifetime of inactivity and poor health maintenance can negatively impact a person in the future. Also, my parents are young (under 60) and have collectively gone through breast cancer and several heart attacks. I don't want this for myself or my future family...

(15) Having free time...

(16) Good weather

(17) Free time. My exercise is inversely proportional to my work-related commitments

(18) Higher quality of life, physical fitness, and physical appearance.

(19) how I physically feel, mobility, flexibility

(20) ..., free time

(21) A plan and accountability

(22) good weather...

(23) ... , and sometimes the knowledge that my body needs me to exercise and stay fit.

(24) Desire to be stronger and more flexible

(25) I would be more physically active if I wasn't so fatigued all the time...

(26) Having a goal...

(27) So that I can not feel so bad to eat treats.

(28) Life events/group events, for example school reunion, family gathering

(29) an all-inclusive gym membership; I used to have one when I lived in Cedar Rapids and when I moved away, I lost my motivation

(30) Financial incentives work.

(31) I am motivated to be more physically active when I have free time.

(32) Physical surroundings

(33) Wanting to live a long and full life surrounded by others.

(34) It makes me feel better when I do it consistently. Helps keep my joints more limber, helps with my arthritis symptoms.[61]

[61] This is catogrized in the "health" group for the quantitative analysis.

(35) goal

(36) friends and family, good weather, keeping in shape

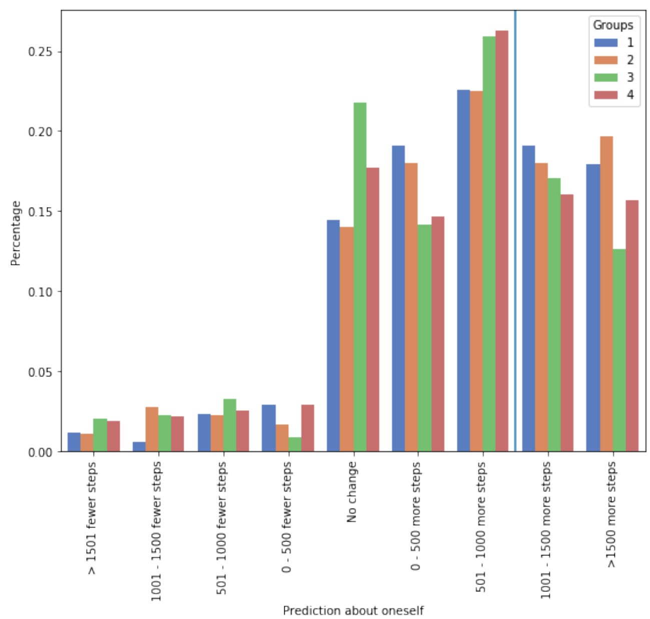
Table B.1: Descriptive statistics: steps and wearing time

	Respondents			***Non-respondents***		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Average daily steps						
Week 1	9328.89	4113.34	93	8792.50	5817.05	45
Week 2	9414.03	3633.77	94	9591.71	5457.86	45
Week 3	9239.17	3645.06	94	9385.07	5672.26	46
Week 4	8890.78	3538.14	93	8492.38	5061.10	45
Week 5	9500.04	3907.95	93	9190.47	4102.08	43
Week 6	9186.35	3555.48	90	8902.28	5478.13	40
Week 7	9838.87	3703.29	88	9746.67	5939.09	36
Week 8	9507.11	3231.26	80	8450.15	3420.98	29
Week 9	8368.08	3337.50	73	8488.24	3214.61	27
Week 10	8781.64	3449.75	56	8378.57	2959.79	22
Week 11	9097.55	3303.93	54	8100.29	3675.44	20
Week 12	9175.09	3795.46	48	8020.70	3073.33	17
Average wearing hours each day						
Week 1	20.91	4.08	93	19.09	5.39	45
Week 2	20.98	3.99	94	18.91	5.42	46
Week 3	21.09	3.35	94	19.42	4.60	46
Week 4	20.37	5.02	94	18.97	4.89	45
Week 5	20.47	4.81	93	18.56	6.20	44
Week 6	21.03	4.10	90	19.36	4.95	40
Week 7	20.71	4.36	88	19.45	5.84	37
Week 8	20.66	4.25	81	19.00	6.24	31
Week 9	19.36	5.14	75	19.84	5.01	29
Week 10	19.31	6.35	58	18.78	6.67	24
Week 11	19.91	4.72	55	20.94	3.85	21
Week 12	19.83	5.81	50	20.09	4.97	18
N	94			46		

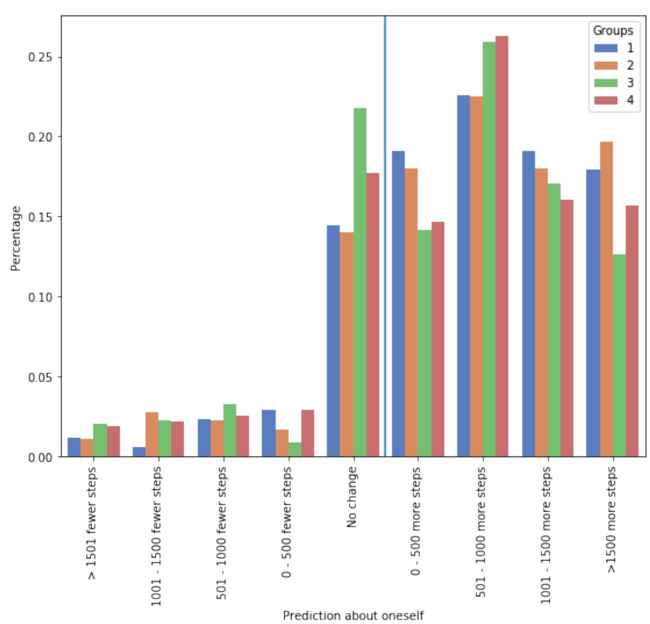
Notes: This table report average daily steps and average daily time when participants wear the fitness trackers for each week, for respondents of the open-ended questions and non-respondents.

APPENDIX C: APPENDIX TO CHAPTER III

Figure C.1: Catogrize values of the predictions into two categories as dependent variables



(a) Equals 1 if walk > 1000 more steps per day



(b) Equals 1 if walk more per day

Notes: The vertical line shows where the the distribution of the predictions is cut to derive the binary dependent variables.

Table C.1: Joint distribution of self-reported daily steps and prior beliefs - subsample

Daily steps (self-reported)	Beliefs about daily steps: Below average (j_1)	Average (j_2)	Above average (j_3)	**Total**
Below the university average (k_1)	161	160	44	365
	(13.78)	(13.70)	(3.77)	(31.25)
Average of the university level (k_2)	58	166	43	267
	(4.97)	(14.21)	(3.68)	(22.86)
Above the university average (k_3)	29	251	256	536
	(2.48)	(21.49)	(21.92)	(45.89)
Total	248	577	343	1168
	(21.23)	(49.40)	(29.37)	(100.00)

Notes: This table displays the joint distribution of daily number of steps and prior beliefs (before informedness) among the group of participants who were presented with the true average number of steps taken by their comparative peers.

Table C.2: Joint distribution of “correctness” of self-assessment

	Self-assessment (posterior)			
Self-assessment (prior)	Underestimate	Correct	Overestimate	**Total**
Underestimate	72	262	4	338
	(6.16)	(22.43)	(0.34)	(28.94)
Correct	28	499	56	583
	(2.40)	(42.72)	(4.79)	(49.91)
Overestimate	2	141	104	247
	(0.17)	(12.07)	(8.90)	(21.15)
Total	102	902	164	1168
	(8.73)	(77.23)	(14.04)	(100.00)

Notes: This table displays the joint distribution of “correctness” of their self-evaluations prior to and following the informedness.

Table C.3: OLS results - dependent variable: predictions about oneself

	(1)	(2)	(3)
No informedness & Comparative info.	−0.0283	−0.0160	−0.0204
	(0.191)	(0.195)	(0.196)
Informedness & Non-comparative info.	-0.322^{*}	−0.291	−0.295
	(0.155)	(0.160)	(0.160)
Informedness & Comparative info.	−0.220	−0.203	−0.194
	(0.155)	(0.160)	(0.160)
First-order beliefs	0.298^{***}	0.255^{***}	0.237^{***}
	(0.0604)	(0.0620)	(0.0627)
Second-order norm	0.00705	−0.00968	−0.00397
(Proximal peers)	(0.0508)	(0.0521)	(0.0523)
Second-order norm	−0.0149	−0.0203	−0.0145
(Distal peers)	(0.0504)	(0.0517)	(0.0520)
Age		0.00566	0.00544
		(0.00442)	(0.00444)
Female		0.686^{***}	0.675^{***}
		(0.144)	(0.145)
Characteristics	*NO*	*YES*	*YES*
Acitivity level	*NO*	*NO*	*YES*
Observations	1517	1415	1408
R^2	0.025	0.043	0.044
F statistic	6.413	4.160	3.562
Prob > F	0.000	0.000	0.000

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
The outcome variable takes on 9 values: 1 if "> 1501 fewer steps," 2 if "1001 - 1500 fewer steps," 3 if "501 - 1000 fewer steps," 4 if "0 - 500 fewer steps", 5 if "no change," 6 if "0 - 500 more steps," 7 if "501 - 1000 more steps," 8 if "1001 - 1500 more steps," and 9 if "> 1500 more steps."

Table C.4: Results in predictions about experiment participants.

	Dependent variable: walk > 1000 more steps - experiment participants					
	Logit			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
No informedness & Comparative info.	0.0706	0.0628	0.0604	0.0706	0.0642	0.0617
	(0.0483)	(0.0498)	(0.0501)	(0.0470)	(0.0487)	(0.0490)
WInformedness & Non-comparative info.	−0.0431	−0.0511	−0.0552	−0.0434	−0.0510	−0.0551
	(0.0373)	(0.0391)	(0.0393)	(0.0381)	(0.0399)	(0.0401)
Informedness & Comparative info.	0.0422	0.0443	0.0425	0.0421	0.0445	0.0425
	(0.0382)	(0.0400)	(0.0403)	(0.0381)	(0.0398)	(0.0400)
First-order norm	0.0345*	0.0276	0.0270	0.0315*	0.0250	0.0245
	(0.0160)	(0.0165)	(0.0167)	(0.0148)	(0.0155)	(0.0157)
Second-order norm (Proximal peers)	−0.0007	−0.0022	−0.0021	−0.000628	−0.00205	−0.00192
	(0.0125)	(0.0129)	(0.0129)	(0.0124)	(0.0129)	(0.0130)
Second-order norm (Distal peers)	0.0037	0.0002	0.0003	0.00363	0.000413	0.000547
	(0.0124)	(0.0127)	(0.0128)	(0.0124)	(0.0129)	(0.0129)
Age		0.0012	0.0014		0.00124	0.00140
		(0.0010)	(0.0011)		(0.00110)	(0.00111)
Female		0.0822*	0.0790*		0.0810*	0.0756*
		(0.0361)	(0.0363)		(0.0359)	(0.0362)
Characteristics	*NO*	*YES*	*YES*	*NO*	*YES*	*YES*
Activity level	*NO*	*NO*	*YES*	*NO*	*NO*	*YES*
Observations	1520	1417	1405	1520	1417	1405
R^2/Pseudo R^2	0.013	0.023	0.032	0.014	0.027	0.037
F statistic/Chi2 statistic	22.62	38.49	52.24	3.710	2.754	2.650
Prob > F/Chi2	0.001	0.000	0.000	0.001	0.000	0.000

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Marginal effects are reported in the columns of Logit models. Other demographics includes education (indicatin college graduates or above), marital status (indicating married or living with a partner), employment (indicating full-time or part-time employment), height (in *cm*), weight (in *kg*), ideal weight (in *kg*); Activitie levels average daily moderate active minutes, vigorous active minutes, daily number of steps taken, ideal number of steps per day.