

# Nudging people off of the couch

A nudge experiment on physical exercise in collaboration with SATS

Anton Goffe      Staffan Sundsmyr

Supervised by Paul Muller, PhD

A thesis presented for the degree of  
MSc in Economics



UNIVERSITY OF GOTHENBURG  
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Final version submitted to the Graduate School June 15, 2017

**Acknowledgements:** We would like to explicitly thank our supervisor, Paul Muller, for his valuable feedback and time. We also want to thank SATS for their cooperation and collection of the data. A special thanks to our liaisons at SATS, Meygol Tarahomi and Anna Backman.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Literature review</b>	<b>4</b>
2.1	Theory . . . . .	4
2.2	Empirical evidence . . . . .	8
<b>3</b>	<b>Hypotheses</b>	<b>13</b>
<b>4</b>	<b>Experiment overview</b>	<b>14</b>
4.1	Experimental design . . . . .	14
4.2	The treatment and mechanisms of effects . . . . .	15
4.2.1	Treatment 1 . . . . .	15
4.2.2	Treatment 2 . . . . .	16
4.2.3	Treatment 3 . . . . .	16
4.3	Data and baseline characteristics . . . . .	17
4.3.1	Randomization . . . . .	19
<b>5</b>	<b>Method &amp; Results</b>	<b>21</b>
5.1	Intention-to-treat (ITT) . . . . .	21
5.1.1	Time-Fixed Effects (FE) & Difference-in-differences (DiD) . . . . .	24
5.2	Local average treatment effects (LATE) . . . . .	27
5.3	Quartile treatment effects (QTE) . . . . .	29
<b>6</b>	<b>Discussion</b>	<b>30</b>
<b>7</b>	<b>Conclusion</b>	<b>34</b>
	<b>References</b>	<b>37</b>
	<b>Appendices</b>	<b>40</b>

### Abstract

In this paper we demonstrate that it is possible to increase average weekly exercise frequency by up to 9.45%, by sending the study participants e-mails. Several of the leading causes of death globally are closely linked to physical inactivity. In fact, more people die from coronary heart disease than from starvation every year. Exercise can mitigate the risks of these diseases and increase general health. We consider exercise as an intertemporal choice in the context of the Hyperbolic discounting model. To increase exercise frequency we perform a natural field experiment with the Nordic gym chain SATS, where we send three different types of e-mails to three randomized treatment groups over nine weeks. We then compare the treatments to a randomized control group with a difference-in-differences approach. The results show a difference in magnitudes in coefficients for the three different treatments, indicating that the textual contents of the e-mails matter for treatment outcome. This thesis contributes to the body of literature by including very large sample sizes compared to previous studies within the field, and furthermore, by having a completely randomized sample and not having any self selection into the study.

## 1 Introduction

“I’m going to be a healthier person this year!” or “This year I’m going to exercise two times a week!”. These are two examples of what people’s New Year’s resolutions could look like. People every year make these promises with great enthusiasm and they have every intention on keeping them, however, what really happens is a somewhat different story. The fact is that a lot of people sign up for expensive gym memberships, which they rarely use ([DellaVigna and Malmendier, 2006](#)), they fail to transform their good intentions into a sustained habit even though they know it is really good for them and it would improve their health. This behavior is problematic from a public health perspective as physical inactivity is closely tied to a variety of diseases. Ischaemic heart disease (also known as coronary heart disease) alone caused the death of 7.4 million people in 2012, which translated to 13% of all deaths that year. 7.4 million deaths are more than people dying from starvation, in fact ischaemic heart disease is the most common cause of death worldwide and

has been for the past decade. Stroke caused 6.7 million deaths in 2012 and was the second most common cause of death. Diabetes is another disease that claims a place in this, not so attractive, top ten list. Diabetes caused 1.5 million deaths in 2012 (WHO, 2014). WHO estimates that 3.2 million deaths each year are attributable to the lack of exercise (WHO, 2012a) and that the probability of dying from one of these diseases can be decreased greatly by physical exercise (WHO, 2012b). Arem et al. (2015) showed with 14-years of data, that individuals that exercised 75 high intensive minutes or 150 minutes of moderate physical exercise a week, reduced their probability of dying by 20%. The WHO- recommended amount is at least 150 minutes of exercise per week. This brings us to our main research question. We want to see if there is a positive effect of e-mail reminders on physical exercise and whether or not differentiating the text in said e-mail reminders has an impact on the magnitude of the effect of exercise frequency. If people clearly benefit from physical exercise, why will they not stick to their New Year's resolution and become new and healthier versions of themselves? A large factor is probably that people have time inconsistent preferences. Thaler (1981) performs a controlled experiment where individuals are faced with questions testing their intertemporal time preferences. A simple example of a question would be: An individual is faced with two scenarios, A and B. A: Choose between: A.1 One apple today or A.2 Two apples tomorrow. B: Choose between: B.1 One apple in a year or B.2 Two apples in a year and a day.

In Thaler's study most people choose alternative A.1 in question A but alternative B.2 in question B. The payoffs are exactly the same, it is only the time dimension that changed but still people have these time inconsistent preferences. People are willing to put in the extra self-control and prolong their reward and invest over time when the choice is set in the future. However, when faced with the same choice in present time they choose the direct reward, as waiting for a reward requires mental effort. With exercise the results are not immediate but will show after a few months and it is quite easy to make the connection between gym membership, exercise and time inconsistency. People sign for a binding gym membership, which they have every intent to use, and making the wise choice as in B.2. However, in reality people's behavior is probably more similar to the choice in A.1. When faced with the direct cost of going to the gym, the individual could either choose the A.2.

alternative (receiving the payoff months later), or instead choose the easier option A.1. (which is not going to the gym), thus avoiding the direct cost of exercising. A meta study conducted by [Davies et al. \(2012\)](#) on internet interventions on physical activity shows that web based intervention programs have a small but statistically significant positive effect on physical activity for participants. Most of the studies, however, are aging, and since 2005 internet activity has increased by more than 300% ([Consortium et al., 2015](#)). With ischaemic heart disease and stroke being the two most common causes of death, ([WHO, 2014](#)) and with more people online now than ever before ([Consortium et al., 2015](#)) we believe a new study in this field would be highly relevant.

In cooperation with the Nordic gym SATS, we conduct a randomized natural field experiment over nine weeks with tailored e-mails sent to individuals. The design of the e-mails is partly based upon the frame used in the study by [Allcott \(2011\)](#) on energy conservation. [Allcott \(2011\)](#) uses a descriptive norm and an injunctive norm to influence individual behavior regarding electricity consumption compared to the average consumption of the neighbors to the individual. [Allcott \(2011\)](#) showed that this comparison to neighbors had a negative causal effect on consumption. We implement a similar comparison, but instead compare workout frequency with the WHO recommended amount of 150 minutes of exercise per week, which we expect to increase exercise. Furthermore, in order to determine whether an observed effect is due to the specific content of the message or to the message itself we extend our analysis to include two additional treatment groups. One of the treatment groups receives an e-mail with a link to a schedule and encouragement to plan and book their next work out today. The third treatment group are sent e-mails including general health benefits derived from exercise. We have panel structured data, where we follow 5408 randomized SATS members over time. We observe a significant increase between 5.49 - 9.45% in average weekly exercise, depending on treatment. Furthermore, we investigate heterogeneous effects through a quartile analysis by comparing the groups by quartiles. We find significant differences between the treatments and the control group.

The remainder of the thesis is structured as follows, section 2 includes a review of

the literature, section 3 states our main hypotheses, section 4 gives an overview of our experiment design, in section 5 we explain how we estimate our data and present our results, section 6 is the discussion of the results, and section 7 concludes the thesis.

## 2 Literature review

### 2.1 Theory

For a lot of people the choice of going to the gym and exercising can be a difficult one. This is true even when the intention exists, such as for those who pay for gym memberships (DellaVigna and Malmendier, 2006). People paint grand pictures of their future selves, by e.g. planning to exercise twice a week. However, the picture often fades when it comes to committing to their plans. The problem seems to be that individuals are faced with an immediate cost in terms of exerted effort, only to receive the reward of exercise in the future, e.g. increased health and physical appearance. Considering this commitment issue in an intertemporal choice setting displays a clear dynamic inconsistency. There are several models dealing with intertemporal choice and utility. For a long period of time individual intertemporal choices were explained with an exponential discounting model. This model does not account for the commitment issue described above. While individuals discount their future utility, the model does not allow them to opt out of their own plans, rather they are assumed to exhibit absolute discipline. Thus, there is an inability of the exponential discounting model to explain dynamic inconsistency. The exponential discounting model has been shown to have a quite poor performance empirically (e.g. Frederick et al. (2002)), indicating that people are not acting as rationally as some economists would like to think. Phelps and Pollak (1968) theorized that individuals' discount rates differ depending on time, framing and size of the amount, and developed the hyperbolic discounting model, further developed by Laibson (1997), and O'Donoghue and Rabin (1999). The hyperbolic discounting model by O'Donoghue et al. (2006), presented below, is an intertemporal model in discrete time, where an

individual makes a choice of when to perform a one time activity.

$$U_t = u_t + \beta \sum_{\theta=t+1}^T \delta^{\theta-1} u_{\theta} \text{ for } t = 1 \dots T \quad (1)$$

$U_t$  is the intertemporal preference regarding the utility states  $u_t, u_{t+1}, \dots, u_T$ .  $\delta \in (0, 1)$  is the time discount factor, and explains how much an individual discounts the future for the present, e.g. an individual with  $\delta$  close to 0 is present biased and cares more about the present, i.e.  $t$ , than an individual with  $\delta = 1$ . The exponent of  $\delta$  is  $\theta - 1$ , where  $\theta = t + 1 \rightarrow t + 1 - 1 = t$ , and implies that there is a greater difference in the value of the discount function between the utility states  $u_{t+1}$  and  $u_{t+2}$  than between  $u_{t+10}$  and  $u_{t+11}$ , from the perspective of  $U_t$ . In other words,  $\delta$  would give  $U_t$  a negative exponential shape were we to ignore  $\beta$ . Indeed, that is what happens if  $\beta = 1$ , where the model reverts back to an exponential discounting model.  $\beta \in (0, 1)$  determines the individual time-inconsistent preference for contemporaneous utility. Based on this  $\beta$  can be said to represent self-control and determines how good the individual is at sticking to the plan, i.e. actually performing the activity.  $\beta$  is imposed on all future time periods and serves to amplify the discount rate given by  $\delta$ . This interaction gives the discount function a hyperbolic shape in the context of continuous time, and a quasi-hyperbolic shape in discrete time. The difference between the two shapes is that quasi-hyperbolic curves are not smooth but jagged. In spite of this, the interpretation of the two are very similar and distinctly different from the exponential case.

The dynamic inconsistency arises because preferences at time  $t$  are inconsistent with preferences at time  $t + 1$ . As indicated by Laibson (1997), this can be seen by looking at the difference in marginal rates of substitution between  $u_{t+1}$  and  $u_{t+2}$  from the perspectives of  $U_t$  and  $U_{t+1}$ . The aforementioned marginal rate of substitution from the perspective of  $t$  is  $u'_{t+1}/u'_{\delta t+2}$ . As we move to time  $t + 1$  it becomes  $u'_{t+1}/\beta\delta u'_{t+2}$ . In the first case both  $u_{t+1}$  and  $u_{t+2}$  are set in the future and therefore share the discount factor  $\beta$ , which can then be canceled out. However, in the second case only  $u_{t+2}$  is set in the future and thus  $\beta$  does not cancel out. Hence, preferences are inconsistent across time periods and this allows for reversal of preferences<sup>1</sup>.

---

<sup>1</sup>See appendix A for the derivation of the marginal rates of substitution.

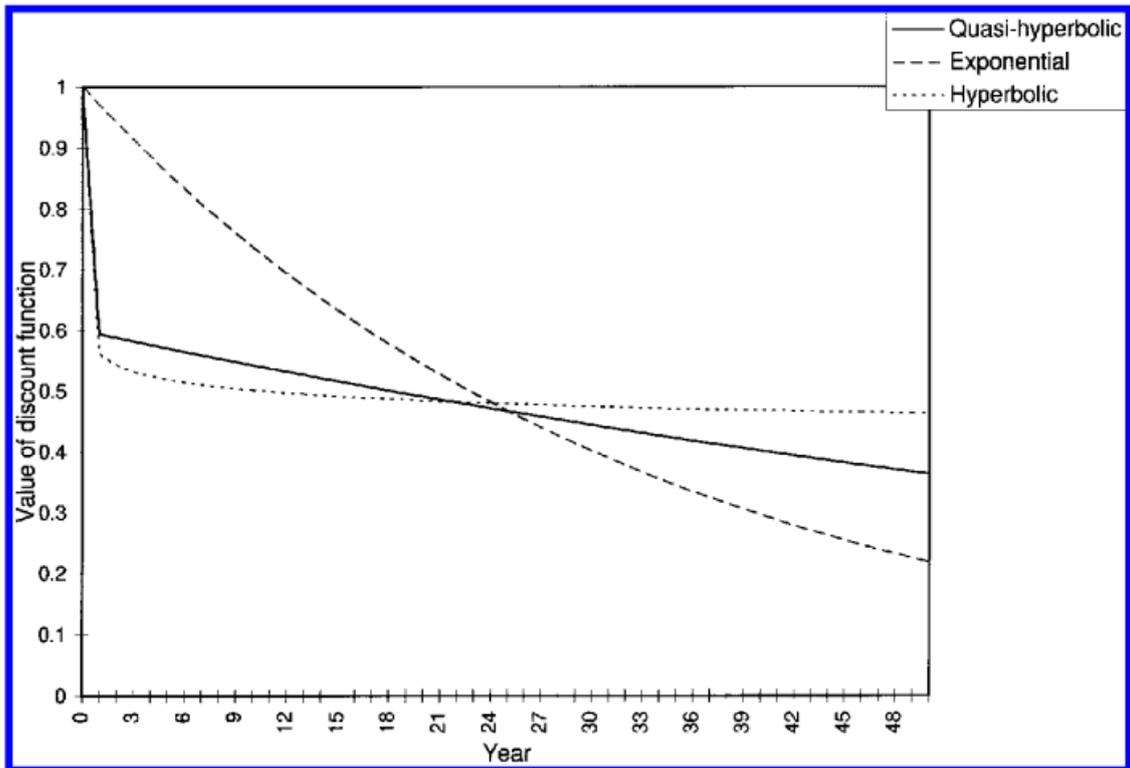


Figure 1: Hyperbolic v. exponential discounting (Laibson, 1997)

Furthermore, O'Donoghue and Rabin (1999, 2001) extend the hyperbolic discounting model by including individuals' sophisticated or naïve notions of their own  $\beta$ -value,  $\hat{\beta}$ . Both individuals with naïve and sophisticated perceptions can exhibit self-control issues ( $\beta < 1$ ), i.e. they have difficulties to perform the one time activity and gain the utility in future time periods. How the individual handles this is decided by the individual's perceptions. A naïve person has the intentions to follow through and actually go to the gym to exercise, but they lack awareness about their lack of self control and mistakenly behave like they have a  $\hat{\beta} = 1$ , effectively removing the  $\beta$ -parameter from equation 1 when making plans for the future. This implies that the naïve individual assumes that their time discounting function is exponential, but in reality, since their  $\beta$  is below one, the naïve individual is likely to default and not complete the task. A sophisticated individual, on the other hand, is aware of their lack of self-control ( $\hat{\beta} = \beta$ ) and can plan accordingly, taking active measures to complete their goals, such as scheduling exercise sessions in advance.

Another way of looking at intertemporal choice is through the model by [Thaler and Shefrin \(1981\)](#), which considers an individual as an organization of a planner and a doer. The planner is concerned with lifetime utility, whereas the doer is myopic and concerned primarily with contemporaneous utility. The utility of the doer is denoted by  $Z_t(\theta_t)$ , where  $\theta$  is a parameter indicating the doer's preferences. The utility of the planner is a function of the doer's utility in various time periods,  $V[Z_t(\theta_t), \dots, Z_T(\theta_T)]$ . The model is set in discrete time and individuals have a constant stream of income  $(y_t, \dots, y_T)$  and a non-negative level of consumption in each period,  $c_t \geq 0$ . The individual is subject to the budget constraint  $\sum_t^T c_t \leq \sum_t^T y_t$ , i.e. lifetime consumption cannot exceed lifetime income. The maximization problem that the planner is faced with is the following:  $\max V[Z_t(\theta_t), \dots, Z_T(\theta_T)]$ , subject to  $\sum_t^T c_t \leq \sum_t^T y_t$ . However, the myopic doer gets in the way of the planner's intentions as the doer attempts to maximize the contemporaneous utility of each period he finds himself in:  $\max Z_t(\theta_t)$ , subject to  $\sum_t^T c_t \leq \sum_t^T y_t$ . According to [Thaler and Shefrin \(1981\)](#) there are two different ways in which the planner can mitigate this problem. First, the preferences or incentives of the doer can be altered through  $\theta_t$ , changing the utility function of the doer,  $Z_t(\theta_t)$ . In our case we can think of our reminders as targeting  $\theta_t$ , aiming to change the preferences and incentives. Second, the planner can set up rules, or constraints, limiting the doer's choices, and possibly the optimized behavioral outcome. Thaler & Shefrin give examples of rules in various contexts. One of their examples is that dieters limit themselves by not keeping tempting sweets at home. Inversely, the reminder with information about booking exercise sessions online can be thought of as putting a healthy snack in the fridge, i.e. making the healthy option more easily available. Another example they give of a rule is that smokers could buy cigarettes by the pack instead of buying an entire carton, thus making each pack they consume more expensive in order to align their economic incentives with their intent to stop smoking. This is analogous to buying an annual or monthly gym contract instead of paying for each visit separately as it also aligns economic incentives with the intention to increase exercise frequency.

The concept of intertemporal choice has also been discussed in the psychological literature and according to [Berns et al. \(2007\)](#) there are three categories of mecha-

nisms underlying intertemporal choice. These are representation, anticipation and self-control. The concept of representation is considered the initial step in an intertemporal decision and concerns the mental image of future actions and its manifestation in the brain. This is akin to the concept of framing future events in the light of a gain or a loss, discussed by [Tversky and Kahneman \(1985\)](#). The way that a future event is represented or framed in the brain seems to influence contemporaneous utility through the effect of anticipation. Second, anticipation refers to the expected outcome of a future event and the utility, or disutility, drawn from that expectation in the period leading up to the event, and is predicated upon representation. Intuitively this would imply that people who are naïve about their self-control should have higher utility gains from anticipation than those who are sophisticated, as they are less aware of the extent of their self-control. Finally, self-control refers to the intertemporal struggle between the selves and is a central parameter in the models we employ. In the discounted utility models following [Phelps and Pollak \(1968\)](#) self-control is expressed in the parameter  $\beta$ . In the model by [Thaler and Shefrin \(1981\)](#) this represents the conflict between the planner and the doer, the preference modification parameter  $\theta$ .

## 2.2 Empirical evidence

[Thaler and Sunstein \(2008\)](#) coined the term nudge, which is a type of non-price intervention. Unlike the agents in classical economic models, often referred to as *Homo economicus*, real humans systematically fail to act rationally and consistently act on their biases. The idea of a nudge is to steer individuals into making the “correct” choice. This can be done by e.g. setting the default to opt-out rather than opt-in, changing the order of menu items, and so forth. *Homo economicus* would not respond to these nudges as they are non-price interventions, they do not alter any relative prices, and *Homo economicus* is governed by economic incentives. However, these non-price interventions have been seen to have considerable impacts on the behavior of real people in several contexts.

[Calzolari and Nardotto \(2016\)](#) perform a similar study to ours. They test the effectiveness of e-mail reminders on gym attendance. Calzolari and Nardotto send out simple weekly e-mails, reminding the individuals to visit the gym. Their treat-

ment period stretches for a little less than 6 months, with 247 individuals of which 89 were in the treatment group, and the rest assigned to the control group. They find the reminders to have a strong effect on gym attendance, especially on the low-attendance users. The low-attendance users increase their monthly visits by approximately 27%. [Calzolari and Nardotto \(2016\)](#) also demonstrate the decaying effect of reminders. During the first 24 hours after the reminder is sent the probability of going to the gym increases by 5.2- 7.9% (depending on model and sample). After that the effect diminishes greatly.

The effects of e-mail reminders have been investigated in several other contexts as well. [Apestequia et al. \(2013\)](#) study 50 000 individual users at the public libraries in Barcelona to investigate if the effect of e-mail reminders on compliance rate is sensitive to the information content of the reminders. They send out five e-mails with different information, where one e-mail mentions the penalties related to returning a book past its due date, while another one adds social pressure, and so on. [Altmann and Traxler \(2014\)](#) investigate if reminders can increase the frequency of check up at the dentist. To do this they cooperate with a German dentist to send out reminders via post, with either a neutral, positive, or negative framing. [Apestequia et al. \(2013\)](#) and [Altmann and Traxler \(2014\)](#) report no difference in compliance rates regarding different information in e-mail reminders. [Calzolari and Nardotto \(2016\)](#), [Apestequia et al. \(2013\)](#) and [Altmann and Traxler \(2014\)](#) all bring viable and important information to the table. However, the sample in [Calzolari and Nardotto \(2016\)](#) is relatively small and consists of only students. Both [Apestequia et al. \(2013\)](#) and [Altmann and Traxler \(2014\)](#) papers are in a field far from exercise, individual behavior in libraries and dentists might not be the same for exercise.

[Allcott \(2011\)](#) looks at the effectiveness of nudges on reducing domestic energy consumption in the United States. He uses data from 17 randomized natural field experiments, from projects carried out by the company OPOWER. From the 17 experiments combined, nearly 600 000 households are randomly assigned into a treatment or a control group. The treatment group is sent letters through the mail with information about their energy consumption, information on how they relate to their neighbors' average consumption (descriptive norm), a subjective rating on how well they are doing (injunctive norm), as well as energy conservation tips spe-

cific to the user’s historical consumption. The descriptive- and injunctive norms are nudges. The rating in the injunctive normative message; “Great”, “Good” and “Below Average”, corresponds to the relative energy usage levels lower than the the 20th percentile, between the 20th percentile and the mean, and above the mean consumption, respectively. The information in the letter sent to the treatment group is found below in figure 2.

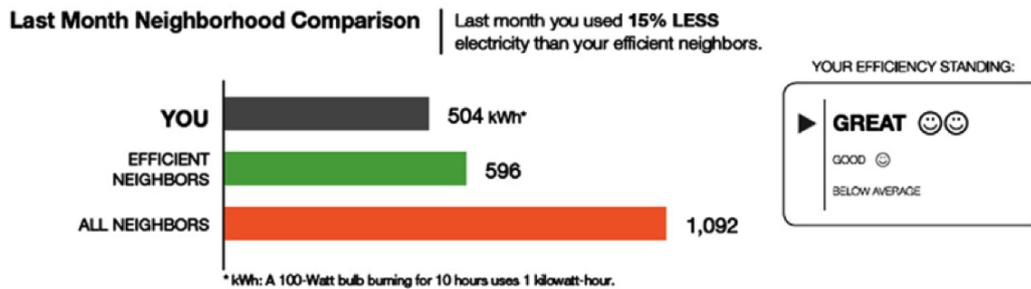


Figure 2: (Allcott, 2011)

The control group contains neighbors of the treatment group who are not sent this information. Allcott estimates the average treatment effect (ATE) for the experiments and finds that the mean ATE is -2.03%, meaning that the treatment caused an energy conservation of 2.03% on average. The treatment is more effective on those with an energy consumption above the mean. This could be due to the descriptive norms included in the letters sent to the treatment group, i.e. the statistical comparison with their neighbors’ average consumption. A regression to the mean is expected with use of descriptive norms because of conditional cooperation and social learning. However, a regression to the mean is not observed. Allcott finds that at most, 15% of the difference in the effects between “Great” and “Good” and 28% between “Good” and “Below Average” are due to the injunctive norm itself. Rather, Allcott argues, the absence of a regression to the mean, is likely due to the energy conservation tips enclosed in the letters.

In Allcott’s context, the energy consumption is directly related to the costs of electricity facing the consumers. Thus, even though relative prices remain unchanged, the consumers do have monetary incentives to reduce their consumption. In our case the individuals face no marginal costs when attending the gym. The

consumers face a prepaid fixed monthly or annual fee, which is essentially a sunk cost. However, it is possible that the individuals perform some mental accounting where they keep track of their attendance in relation to the price they paid for their membership.

[DellaVigna and Malmendier \(2006\)](#) look at how individuals choose between types of contracts differing in price and duration. At enrollment the members choose between monthly or annual contracts. The monthly contracts are automatically extended every month unless actively cancelled, whereas the annual contracts expire after one year unless the member actively renews it. The monthly contracts are more expensive than the annual contracts because the member can cancel their membership any time of the year. The authors follow 7752 individuals attending one of three health clubs in Boston, USA, over three years. They find that on average individuals predicted their exercise frequency to 9.50 visits per month, whereas the average actual frequency is 4.17 visits per month. On average they paid \$17 per gym visit, because they opt for the more expensive monthly or annual contracts. Had they instead paid per visit using a prepaid contract covering 10 visits to the gym they would have paid \$10 per visit. The authors explore possible reasons for this inconsistent/irrational behavior of categorical deviation from the optimal contract, in terms of prices given the number of visits. Their main explanation is that people overestimate their future efficiency and/or self-control, implying that the members initially intended to exercise more than they did.

The effect of internet- and e-mail based interventions on physical activity is a quite common in the medical literature. However, the medical literature focuses on this issue mainly in the context of improving the subjects' health, rather than investigating their behavioral incentives. [Davies et al. \(2012\)](#) and [Van den Berg et al. \(2007\)](#) each conduct a meta-study, investigating the literature published in the medical field more closely.

[Davies et al. \(2012\)](#) investigate the efficacy of internet based interventions to increase physical exercise in a meta-analysis. The authors review 34 articles published between 2001 and 2011, which investigated this issue in an experimental setting. The total sample size for the 34 studies was 11885 individuals and the average rate

of attrition was 20%. They find that the standardized mean difference<sup>2</sup> of the internet interventions as compared to the different comparison groups (some were control groups and others were alternative interventions) was 0.14 and statistically significant at the 1% level. Furthermore, they find that interventions that included some educational feature had a larger effect on average (again in terms of standardized mean difference) than those that did not, 0.20 as opposed to 0.08. However, only 74% of the studies had physical exercise as their primary target. Only 12% of the reviewed articles investigated e-mail interventions specifically and 62% of the studies looked at combinations of internet and e-mail based interventions. The remaining studies (26%) focused exclusively on internet based interventions.

Van den Berg et al. (2007) conduct a meta-analysis of 10 studies looking at internet based interventions on different types of physical activity. There is some overlap with Davies et al. (2012) in articles reviewed because of a partial overlap in the databases and journals searched as well as timespan. The total sample size of the 10 studies combined was 4133 before treatment and 3208 after, implying an average rate of attrition of 22%. There is a variation of the target outcomes used in the different articles reviewed (some target physical activity and some use weight loss) and because of this the results are hard to compare. Furthermore, there is variation in the use of control groups between the studies. Some do not include a control group at all but instead compare the effects of different intervention groups with each other. The authors conclude that there is tentative evidence of the efficacy of internet based interventions on physical activity, but that further research is needed and that this research should clearly define control groups and use objective measurements of physical activity.

Most of the samples in the articles reviewed by Van den Berg et al. (2007) are not representative of the general population, as they are drawn from specific patient groups, e.g. overweight people in risk of type 2 diabetes or patients with rheumatoid arthritis, or exclusively physically inactive adults. This is also the case for 16 out of the 34 articles reviewed by Davies et al. (2012) and limits the external validity of

---

<sup>2</sup>Because the articles reviewed in the meta-analysis use a range of different measurements the raw difference in means is not applicable. The authors therefore use a standardized difference in means; they divide the difference in means of each study with the standard deviation of that same study. This allows for comparability between measurements.

the articles considered in both of these meta-analyses. The samples we use in our study are drawn from all SATS members in Sweden. While this might not make our results representative of the entire population, as we only include people who at least had an intent to exercise when they bought the membership, the external validity is broader than most articles considered here. [Van den Berg et al. \(2007\)](#) and [Davies et al. \(2012\)](#) have a combined sample size of approximately 16 018 individuals spread out over 43 studies. Our study has a total sample size of 5 408 individuals which translates into 34% of the combined studies [Van den Berg et al. \(2007\)](#) and [Davies et al. \(2012\)](#). The volume of our sample is more than ten times larger than the average sample size of the articles reviewed here. Taking this into account, our study will have clear place in the literature and shed further light on this issue.

### 3 Hypotheses

Below we state our hypotheses. In section [4.2](#) we describe in detail the mechanisms, through which we expect the hypothesized effects to be operating. In section [5](#) we describe our methodology and test these hypotheses.

*H1<sub>a</sub>*: E-mail reminders sent to individuals, containing information regarding their personal workout frequency, cause an increase in exercise frequency compared to a control group.

*H2<sub>a</sub>*: E-mail reminders sent to individuals, containing a link to a schedule, cause an increase in a exercise frequency compared to a control group.

*H3<sub>a</sub>*: E-mail reminders sent to individuals, containing information regarding various health benefits from exercise, cause an increase in exercise frequency compared to a control group.

## 4 Experiment overview

### 4.1 Experimental design

We investigate if nudges can increase exercise frequency in an experimental setting. We perform our study in collaboration with the Nordic gym chain SATS. SATS is one of the largest gym chains in the Nordic region, with gyms in Norway, Finland, and, Sweden. Both gym and group sessions are available for the members of SATS. Our total sample size consists of 5408 individuals who are randomly drawn<sup>3</sup> from a subset of the total population of 418 000 Swedish SATS members, who are between 18-64 years old. Out of these 5408, 4021 individuals are assigned through randomization to three treatment groups, and 1384 members are assigned to a control group for observation. The three treatment groups are sent e-mails with different information for each group<sup>4</sup>. The e-mails are sent from SATS's e-mail platform, CARMA, and SATS has also been kind enough to measure exercise frequency and provide us with the data. The first e-mails were sent 1st of February, 2017, and the last were sent 30th of March, 2017. With this natural field experiment we are intervening in people's lives and thus is important to consider potential ethical concerns. A recent study found that 51% of the Swedish population is considered to be overweight ([Folkhälsomyndigheten, 2017](#)), which shows how important it is to exercise. Furthermore, [Arem et al. \(2015\)](#) showed that there are no negative health effects for exercise up to ten times the amount that WHO recommends, although the health benefits are diminishing for exercise at levels of three to four times the recommended amount. Since our study aims to nudge people in the right direction and that the members of SATS all have signed up for membership voluntarily, most likely with the intent to exercise and become healthier, we see no ethical concerns with doing an experiment such as this. Especially, since it is very easy to opt-out of treatment. Every e-mail sent includes a link that says "Unsubscribe from these e-mails from SATS".

An advantage of targeting only gym members is that they have no monetary marginal costs of exercise, as they have already paid for their membership. There-

---

<sup>3</sup>The individuals are randomly drawn and assigned through randomization to the different groups by SATS using their platform, CARMA.

<sup>4</sup>See appendix B for e-mail samples

fore, the decision of going to the gym is clearly a decision affected by behavioral preferences, rather than financial ones. We expect the e-mail itself will have a general reminder effect, which will influence the individual to visit the gym and exercise (see e.g. [Calzolari and Nardotto \(2016\)](#), and [Hurling et al. \(2007\)](#)). To further investigate if the content of the e-mail matters for exercise frequency the three treatment groups receive weekly e-mails over nine weeks, containing different information for each group. The data is collected continuously from the gym SATS over nine weeks. The structure of the dataset is a panel following members over time. The dependent variable is individual exercise frequency. We also have a number of control variables further examined in the data and method sections.

## 4.2 The treatment and mechanisms of effects

As mentioned earlier, we suspect a general reminder effect on exercise. This implies that subjects have to open the e-mail for the textual content of it to have an effect. For example, to schedule a group session through the provided link in the e-mail sent to group 2 the members of group 2 have to open the e-mail. Unopened e-mails, therefore, would only have the general reminder effect.

### 4.2.1 Treatment 1

Treatment group 1 are sent e-mails with a descriptive norm where we compare their exercise frequency to what is considered a healthy level of exercise (At least 150 minutes per week for individuals aged 18-64). Based on their exercise frequency compared to the WHO recommendation, the subjects in treatment group 1 will also receive an injunctive norm relating to their performance. However, since we do not have any data from SATS on for how long members usually stay at SATS and we do not observe any exercise outside of SATS, we have chosen a threshold level of two visits per week to SATS, to correspond with the 150 minutes recommended by the WHO. If the frequency is above the WHO recommendation, the e-mail will contain positive feedback, whereas if the frequency is below, the e-mail will encourage the subject to try harder next week.

We base the e-mail sent to group 1 on the same frame constructed by [Allcott \(2011\)](#) where an injunctive and a descriptive norm showed to have a causal effect on

participants' electricity consumption. We believe that our descriptive and injunctive norms would influence the individuals via a lowering of social costs when the subjects follow a norm. The descriptive and injunctive norms can possibly also nudge naïve individuals into sophisticated ones, when becoming aware of their exercise frequency. Furthermore, the descriptive norm could result in a boomerang effect where the individuals performing above the recommended amount decrease their efforts. As discussed by [Allcott \(2011\)](#), the inclusion of an injunctive norm acts as a safeguard mitigating this potential problem.

#### **4.2.2 Treatment 2**

Treatment group 2 receives an e-mail containing a link to a schedule with suggested work out sessions in their area. The reminder also includes a positively framed suggestion to the individual to make a reservation for a group exercise today.

When faced with an immediate cost a time-inconsistent individual is likely to “default”, thinking they would exercise tomorrow instead, as described in the theory section. Making the schedule more available and actually telling the the individuals to make an reservation, may function as a commitment device and help these individuals to plan ahead and bypassing their present bias, and thus increase their exercise frequency. This can also be related to the rule the planner uses to influence the doer in the model by [Thaler and Shefrin \(1981\)](#) described in the theory section.

Since it is only possible to schedule a group session workout through the provided link, this e-mail would only affect the individual's' habits at the gym via the reminder effect for those who do not open the e-mail. This will give us a possibility to estimate the effect of the general reminder as well as the relative importance of content.

#### **4.2.3 Treatment 3**

Treatment group 3 also receives an e-mail, but with a short informative message regarding the health benefits of regular physical exercise. The information in the e-mails varies between 4 different health benefits, e.g. one of the e-mails contains information that exercise can prevent depression, whereas one e-mail contains information about exercise and colon cancer. All of the recommendations are based on scientific research and the information of the health benefits are framed in a positive

way.

We hypothesize that when reminding individuals of the positive impact exercise can have on health in general, lifetime utility increases as a result of an increase in the reward of future health benefits, in relation to the cost or effort exerted in the present. The increase in future health stems from the fact that individuals are reminded of their possible future health benefits when exercising.

### 4.3 Data and baseline characteristics

Table 1: Summary statistics

VARIABLES	N	mean	sd	min	max
VisitsLast7d	45,076	1.197	1.589	0	16
VisitsGymLast7d	45,076	0.757	1.272	0	11
VisitsGXLast7d	45,076	0.441	1.067	0	15
Weekly moving average	45,076	1.245	1.362	0	13.50
age	5,408	36.84	10.47	18	64
Opened e-mails per individual	3,561	3.026	3.189	0	9

Note: The variables below the horizontal line are time invariant and are therefore reported only in time period 1.

Table 1 displays summary statistics of some of our variables over all of the 9 time periods in our sample. Our main dependent variable of interest is the variable "VisitsLast7d". It represents the total number of visits to a SATS centre within the last seven days and is compiled of the two variables "VisitsGymLast7d" and "VisitsGXLast7d", which are the number of gym visits and the number of group session visits, respectively, in the last seven days. The variable "Weekly moving average" is the weekly average number of gym visits calculated on the basis of the last month. Because these variables are all retrospective, the first observation, at Time=1, is a pre-treatment measurement and can be used as a baseline. The average age of our sampled members is 36.84. The minimum and maximum values are 18 and 64, respectively. We have chosen to sample members from the underlying population of SATS members who fall within this age range since the recommendations on physical activity from WHO apply specifically to this range. The variable "Opened

e-mails per individual” is an aggregated measurement of the number of opened e-mails over the entire time period. It is measured by downloaded content of the e-mails, such as images. The number of observations for this variable is naturally lower than that of the other variables, as no e-mails were sent to the control group. We can see that an individual in the treatment groups open on average 3.026 e-mails of the 9 sent, during the experiment.

Table 2 describes the distribution of membership types and distribution of gender across our treatment groups and control group. As we can see, the shares of the different membership types are fairly similar across the different groups. This is also true for the distribution of genders, as can be seen below the horizontal line. These categorical variables along with age make up the set of control variables later referred to in our regression analyses.

Table 2: Membership type and gender by group

	Treatment 1	Treatment 2	Treatment 3	Control
	Freq	Freq	Freq	Freq
Type	(Percent)	(Percent)	(Percent)	(Percent)
Corporate	495 (35.95)	497 (37.77)	485 (36.52)	535 (38.66)
Friend	43 (3.12)	43 (3.27)	37 (2.79)	33 (2.38)
Private	638 (46.33)	564 (42.86)	580 (43.68)	606 (43.79)
Senior	8 (0.58)	6 (0.46)	4 (0.30)	8 (0.58)
Student	193 (14.02)	206 (15.65)	222 (16.72)	202 (14.60)
Total	1377	1316	1328	1384
Female	763 (55.41)	719 (54.64)	759 (57.15)	763 (55.13)
Male	614 (44.59)	597 (45.36)	569 (42.85)	621 (44.87)
Total	1377	1316	1328	1384

Note: The data presented here is measured at Time=1.

### 4.3.1 Randomization

SATS was kind enough to collect and provide with the data and they also performed the randomization. Since we did not perform the randomization ourselves, we investigate if the groups are balanced, as an indication that the randomization was conducted properly.

At the baseline, we perform a Mann-Whitney U (Wilcoxon rank sum) test on our variable, "AverageLast7d", containing data on average number of visits to SATS over the last four weeks between the control groups and the three treatment groups. This is a non-parametric test that does not rely on any assumptions regarding the distribution of the data. It estimates the probability of a randomly drawn value from the control group is greater than a randomly drawn variable from the treatment groups. The resulting probability is 0.485 and is significantly different ( $p = 0.0872$ ) from the null hypothesis of  $P=0.5$ . This indicates that there may be some imbalances at the baseline between the groups.

There are cases where the non-parametric test is more reliable than a t-test, but it is also more likely that one commits a type II error with this test as it has less statistical power than the parametric t-test, (Lumley et al., 2002). The t-test relies on the assumption of normally distributed means. This is a problem mainly for small samples. It does not, however, require on normally distributed *data*, due to the central limit theorem.

To further investigate this baseline difference, we regress our independent variables on our treatment dummies at the baseline, i.e. pre-treatment, and perform an F-test for each regression. The null hypothesis of the test is that the coefficients of our group dummies are not jointly significantly different from zero. The results of the F-tests in table 3 show that we can not reject the null hypotheses of joint insignificance. However, there is a significant difference (on the 5%-level) in the number of gym visits in the last seven days for group 3 compared to the control group. This difference translates into differences in the compiled variable VisitsLast7d (significant on the 5%-level) and in the average variable (significant on the 10%-level). To address these baseline differences we perform a difference-in-differences (DiD) estimation in section 5.

When comparing the distribution of gender and age between the treatment and

Table 3: Baseline differences

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	AverageLast7d	VisitsLast7d	VisitsGymLast7d	VisitsGXLast7d	age	gender
Treatment 1	-0.0507 (0.0539)	-0.0414 (0.0615)	-0.0372 (0.0496)	-0.00422 (0.0414)	-0.788* (0.403)	0.00280 (0.0189)
Treatment 2	-0.0458 (0.0528)	-0.0299 (0.0627)	-0.0367 (0.0499)	0.00684 (0.0421)	-0.552 (0.404)	-0.00495 (0.0192)
Treatment 3	-0.0992* (0.0540)	-0.128** (0.0617)	-0.105** (0.0488)	-0.0229 (0.0418)	-0.288 (0.405)	0.0202 (0.0191)
Constant	1.348*** (0.0370)	1.290*** (0.0430)	0.834*** (0.0347)	0.456*** (0.0284)	37.25*** (0.287)	0.551*** (0.0134)
Observations	5,405	5,405	5,405	5,405	5,405	5,405
R-squared	0.001	0.001	0.001	0.000	0.001	0.000
F-test	1.127	1.559	1.614	0.173	1.423	0.648
Prob > F	0.337	0.197	0.184	0.915	0.234	0.584

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Note: This table presents six regressions at the baseline, i.e. pre-treatment. The variables in the top row are the dependent variables of these regressions, and the treatments are the independent variables. The control group has been omitted and is used as the benchmark to which the treatments are compared.

control group we get no significant results for gender, and for age there are no significant differences for group 2 and 3. However, we do get a significant difference in age distribution between group 1 and the control group (significant at the 10%-level). Table 3 also displays the average age for each of our groups. An individual in the control group is on average 37.25 years old, whereas in treatment group 1 the individual is on average 36.46 years. We believe that the significance we get is largely due to our large samples and the statistical power we have because of this. Since there should be no difference on exercise frequency whether you are 37.25 or 36.46 years old, i.e. the economic significance is negligible, and we can still argue that the sample is balanced and that there are no indications that the randomization has been conducted improperly.

## 5 Method & Results

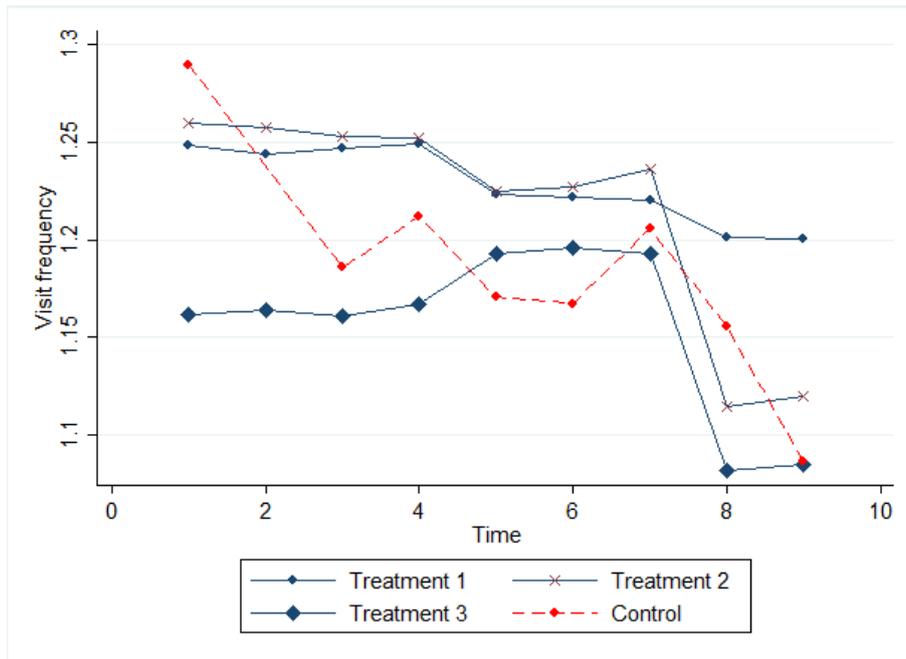


Figure 3: Average number of visits the last week per group

Figure 3 shows the trend of the mean values of our main outcome variable, `VisitsLast7d`, by group over time. There is a clear downward sloping trend, which is not surprising considering that the time period we are observing starts at the end of January (for this variable) and ends at the end of March. This could be a seasonal effect following New year’s resolutions or that people are choosing to exercise outdoors as temperatures are rising. Just by looking at figure 3 it looks like if the downward slope is steeper in the control group, suggesting that we will find an effect of our treatment.

### 5.1 Intention-to-treat (ITT)

Given that the individuals are randomly assigned to the treatment groups and the control group, as we showed earlier in section 4.3.1 (Randomization) there are no systematic differences between the groups in terms of descriptive statistics prior to the treatment. Therefore, since we have a randomized sample, we instead make the comparison, for the treated individuals, to an untreated control group (row 3 in equation 2)

Since we have an experimental design we would ideally want to analyze the average treatment effects (ATE) for the treatment. The mathematical definition of which is as follows:

$$ATE = E[Y_{1i} - Y_{0i}] \quad (2)$$

$$= E(y_1) - E(y_0)$$

$$= E(y | w = 1) - E(y | w = 0)$$

The first row of equation 2 takes the outcome of treatment  $w$  for individual  $i$  minus the untreated outcome for individual  $i$ . This is, however, a counterfactual since we cannot observe both outcomes at the same time point for one individual. Therefore, since we have a randomized sample we instead make the comparison for the treated individuals to an untreated control group (row 3 in equation 2).

However, ATE relies on the assumption that, given randomization, the group assignment is orthogonal to both the dependent variable and to any control variables, pre-treatment (Angrist and Pischke, 2008). Because of non-compliance, the treatment received in our case might be non-random since opening e-mails could be driven by certain unobserved behavioral traits. As we can see in the pre-treatment estimation in appendix C <sup>5</sup>, this seems to be the case. Therefore, we are not able to estimate the ATE, but we can estimate the ITT.

The ITT measures the effect of being assigned to the treatment group, whether treatment is actually received or not. Should the treatment groups have high rates of attrition or low rates of compliance the results are diminished. This is especially interesting from a policy perspective, since one cannot always control the rate of attrition or compliance. A treatment could have a high ATE but a low ITT, rendering a potential policy ineffective. This implies that ITT is a more conservative estimate, since the probability of committing a type I error is reduced but the probability of committing a type II error is increased (Gupta et al., 2011).

---

<sup>5</sup>Furthermore, male participants open more e-mails on average than female participants. The SATS members that belong to membership type "Private" open more e-mails than the "Corporate" members.

There are three possible channels for attrition in our experiment. The participants can either choose to opt out of the treatment, an opt out option is included in every e-mail. Another channel is that the participants can choose to end their membership with SATS, a third variant being actively selecting the option to not receive any e-mails at all from SATS what so ever. These channels allow attrition from the control group as well, even though they do not receive any of our e-mails. However, we are not able to distinguish between these channels of attrition as the individuals who somehow opted out simply disappear from our dataset, preventing us from following up their exercise frequencies after opt-out. Table 4 shows that attrition rates from opt-out lies between 8.5-9.7% for the treatment group and 5.9% for the control group, which is fairly low (see e.g. 22% for [Davies et al. \(2012\)](#), and 20% for [Van den Berg et al. \(2007\)](#)). The reason that the attrition rate is lower for the control group is that there are only two out of three possible channels available for attrition, since they do not receive any of our e-mails they can not choose to opt out of the treatment.

Table 4: Attrition

Visitslast7d	Control	Treatment 1	Treatment 2	Treatment 3
No. of obs. at time 1	1,384	1,377	1,316	1,328
No. of obs. at time 9	1,302	1,244	1,196	1,215
Attrition (%)	5.92	9.66	9.12	8.51

We first compare our three different treatment groups to the control group with t-tests. As we can see in table 5 there are no significant differences in exercise frequency between the treatment groups and the control group, according to our t-test. Furthermore, accounting for attrition makes no difference in terms of significance.

Table 5: ITT

Visitslast7d	Control	Treatment 1	Treatment 2	Treatment 3
Accounting for attrition				
Mean	1.1651	1.2156	1.2240	1.1377
Mean difference	0	0.0505	0.0589	-0.0273
p-value	.	0.3079	0.2393	0.5737
No. of obs.	1,384	1,377	1,316	1,328
Not accounting for attrition				
Mean	1.2081	1.2446	1.2224	1.1780
Mean difference	0	0.0365	0.0142	-0.0301
p-value	.	0.4744	0.7822	0.5520
No. of obs.	1,302	1,244	1,196	1,215

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Note: Table 5 shows the results from six t-tests performed between the control and the treatments, comparing average exercise frequency per person for all time periods. Above the horizontal line we include all individuals according to their group assignment at baseline. Below the horizontal line the individuals who opted out are omitted.

As noted in the beginning of the section, figure 3 depicts a decline in average exercise frequency over the experiment period, for all groups. Since the t-test does not account for this declining time trend and the fact that our data has a panel structure we extend the analysis by looking at a time-fixed effects model and a DiD estimation.

### 5.1.1 Time-Fixed Effects (FE) & Difference-in-differences (DiD)

We employ a fixed effects model rather than a pooled OLS because we are following the same individuals over time, implying that our observations are not independent across time. Moreover, we investigate the differences between our time-fixed effects model and a random effects model<sup>6</sup>. We perform both a Hausman specification test and a Sargan-Hansen test to determine whether the time-fixed effects model is more suitable than a random effects model or not. Both tests indicate that the time-fixed effects model is more appropriate.

<sup>6</sup> See appendix D for table 10 with the fixed- and random effect specifications, the results of the Hausman specification test, and the Sargan-Hansen test result.

The specification in equation 3 is the time-fixed effects model reported in the first two columns of table 6.

$$y_{kti} = \beta x_{kti} + \eta_t + \epsilon_{ti}, \quad (3)$$

where  $y_{kti}$  is the variable "VisitsLast7d",  $x$  is treatment  $k$  for individual  $i$  at time  $t$ , and  $\beta$  is the coefficient of these dummy variables.  $\eta$  is the time-fixed effect, i.e. the endogenous part of the error term, whereas  $\epsilon$  is the exogenous part of the error term.

Since we follow individuals over time our observations are not independent. This means that we have autocorrelated standard errors and that the robust standard errors are wrong. To account for this we use time-fixed effects and cluster our standard errors on Time. Technically, this is a case of cluster-specific fixed effects, as discussed by [Cameron and Miller \(2015\)](#).

The second column in table 6 includes a vector of controls, whereas the first column does not. As we can see, there is only a very small difference in coefficients and standard errors between the two columns. This is because the groups are fairly balanced in terms of control variables, as we showed in section on randomization (4.3.1). The coefficients for these two specifications both show that treatment 1 is positive and statistically significant at the 1%-level. The coefficient for treatment 2 is positive, but smaller than that of group 1, and statistically significant at the 5%-level, whereas the coefficient of the third group is statistically insignificant.

Table 6: Time-fixed effects and Difference-in-differences

	(1)	(2)	(3)	(4)
VisitsLast7d	Time-Fixed Effects	Time-Fixed Effects	Difference-in-differences	Difference-in-differences
Treatment 1	0.0531*** (0.0105)	0.0537*** (0.0107)	0.0954*** (0.0111)	0.0948*** (0.0110)
Treatment 2	0.0387** (0.0131)	0.0310** (0.0130)	0.0651*** (0.0133)	0.0636*** (0.0133)
Treatment 3	-0.0179 (0.0131)	-0.0241 (0.0134)	0.112*** (0.0134)	0.111*** (0.0135)
Constant	1.172*** (0.00746)	1.221*** (0.0225)	1.185*** (0.0061)	◇
Observations	39,671	39,669	45,076	45,074
R-squared	0.000	0.006	0.000	0.003
No. of time periods	8	8	9	9
Time-Fixed Effects	YES	YES	YES	YES
Controls	NO	YES	NO	YES

Clustered standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note: The standard errors are clustered on time for all models. The coefficients reported in the DiD models are the interaction terms i.e. DiD estimators. The vector of controls includes membership type, age and gender. ◇: Column 4 reports results from three different regressions, one for each treatment group, and when including control variables these regressions report three different constants. These are (clustered s.e. in parenthesis) 1.2254\*\*\* (0.0424), 1.332\*\*\* (0.0180), 1.236\*\*\* (0.0322) for treatment 1, 2 and 3 respectively. One of the control variables included is membership type and the type that is omitted in the regressions is "Corporate". Since this group contains all members with the membership type "Corporate" and not just the members of this type in the control group, the constants differ.

However, as we have shown in section 4.3.1, there are some baseline differences in the dependent variable between the treatment groups and the control group. To account for these we perform a DiD analysis. This estimation method relies on the assumption of parallel trends, which states that had the treatment groups not been treated they would exhibit trends parallel to that of the control group. Since the treatment groups were indeed treated this is a counterfactual, which we cannot observe. However, the members of the groups are randomly assigned to the different groups and so each member has an equal probability of being assigned to each group. This means that we can consider the parallel trend assumption satisfied.

$$y_{ikt} = \gamma_k + \lambda_t + \delta D_{kt} + \eta_t + \epsilon_{it} \quad (4)$$

$y_{ikt}$  is the dependent variable "VisitsLast7d" given individual  $i$ , treatment  $k$ , and time  $t$ .  $\gamma$  and  $\lambda$  are the intercepts for treatment  $k$  and time  $t$ , respectively. The coefficient  $\delta$  is the DiD estimator and  $D$  is the interaction term between  $\gamma$  and

$\lambda$ .  $\eta$  represents time-fixed effects, and  $\epsilon$  is an error term. Essentially, this is the same as the time-fixed effect specification in equation 3, except that we include the pre-treatment data of time period 1.

Columns 3 and 4 of table 6 display the results of our DiD model. The difference between the two is that column 4 includes a vector of control variables, whereas column 3 does not. The DiD estimator is an interaction term between the treatment dummy and the time dummy, and it is the coefficient of the interaction term that is reported in rows one through three. As we can see all three DiD estimators are significant at the 1%-level, whether we include control variables or not. The coefficients of column 3 and 4 are very similar in magnitude. The control variables include membership type, which is a categorical variable, and so one of the categories is omitted and influences the baseline. Since the coefficients are so similar and the interpretation of the results in column 4 is unnecessarily confusing we consider the results of column 3.

These results differ in significance from both those of the t-tests and those of the time-fixed effects analysis. While the coefficient of treatment group 1, 8.05% increase in the average weekly gym visits, is still greater than that of treatment group 2, 5.49%, the coefficient of treatment group 3, 9.45% is the largest one by this estimation. In the time-fixed effects specification we exclude the first time period as it contains only pre-treatment data. However, this time period is not excluded from the DiD analysis and if we look at figure 3 we can see that the largest mean difference at baseline is between treatment group 3 and the control group. It is in fact the only baseline difference that is significant. The DiD model accounts for this difference, whereas the time-fixed effects model does not.

The results of the DiD, which we consider the most appropriate approach, tells us to reject  $H1_0$ ,  $H2_0$  and,  $H3_0$ , implying that e-mail reminders do in fact increase exercise frequency.

## 5.2 Local average treatment effects (LATE)

The LATE is a treatment effect that measures the average treatment effect on those who actually receive the treatment (Angrist and Pischke, 2008) (for example, in an experiment for a radio ad, the LATE would be the effect for the people that actually

hear the ad and not everyone who are in range of the radio signal). In our experiment the LATE is the average treatment effect for the individuals who actually open an e-mail. The LATE is calculated by dividing the estimated coefficients for the ITT, obtained from our DiD model in table 6, with the average opening rate for each treatment. Figure 4 shows the average opening rate of e-mails per treatment. We compare these opening rates to the opening rate of a general newsletter sent by SATS to all Swedish members in the end of January, i.e. pre-treatment. As we can see, treatment 2 is significantly different and on average individuals open more e-mails than treatment 1 and 3. Furthermore, all of our treatments have significantly higher opening rate than the newsletter, yet they are within the same range.

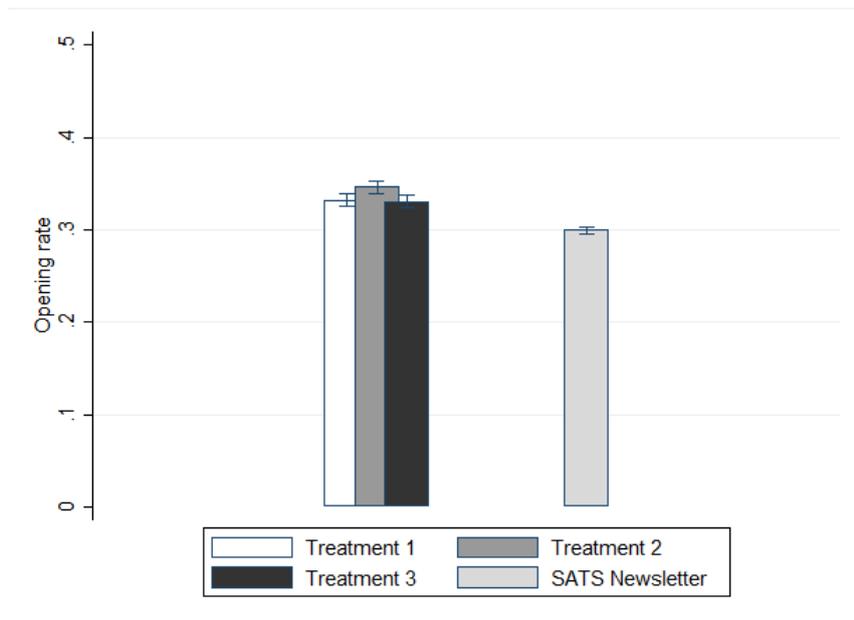


Figure 4: Average opening rate by e-mail type, 95% CI

Table 7 displays the results from the calculations of LATE. The LATE coefficients are as follows:

VisitsLast7d	Treatment 1	Treatment 2	Treatment 3
DiD coefficient	0.0954***	0.0651***	0.112***
Average opening rate	0.3320	0.3465	0.3305
LATE	0.2874	0.1879	0.3389
No. of obs.	22,453	21,940	22,091

To appreciate the magnitude of the LATE we translate these results into percentage increases from the mean. For treatments 1, 2 and 3, this implies an increase in the mean by 24.24%, 15.84% and 28.59%, respectively.

### 5.3 Quartile treatment effects (QTE)

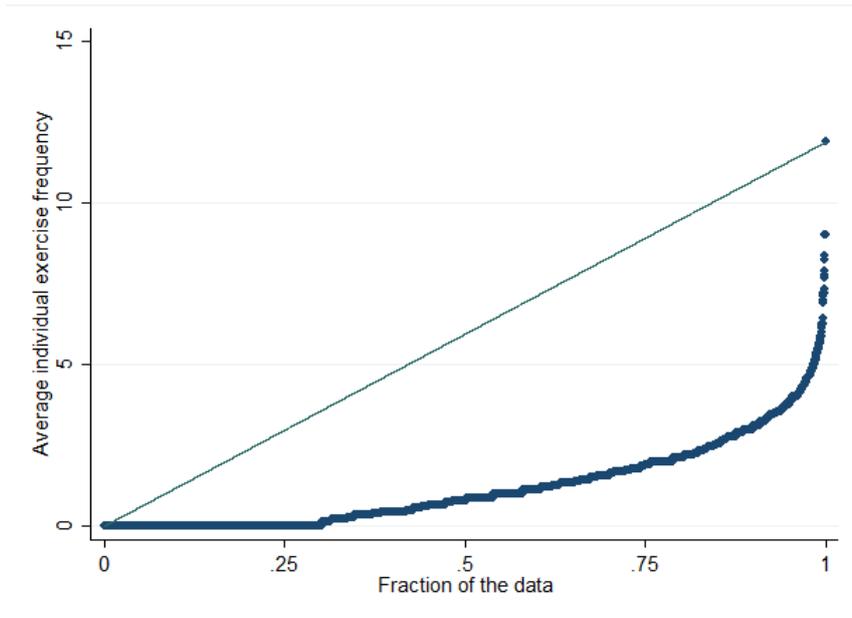


Figure 5: Quantiles of VisitsLast7d

Figure 5 displays the distribution of the individual average exercise frequency over all time periods and all groups, plotted against the uniform distribution (the diagonal line). If we look closer at the graph we can see that the lowest quartile, Q1, contains only zeros. This means that there is no variation in Q1, and that we can not make any statistical comparisons using this quartile. However, there is variation in the upper three quartiles and so we can compare these quartiles of the treatment groups to the corresponding quartiles of the control group. The results from table 8 show positive mean differences that are statistically significant at the 1%-level for the individuals in Q4 for all treatments (13.59% for treatment 1 , 13.40% for treatment 2 , and 10.15% for treatment 3). The mean difference for treatment 1 in both Q3 (5.94%) and and Q2 (13.77%) are statistically significant at the 5%-level. The mean differences for treatment 2 are statistically significant at the 1%-level for individuals both in Q3 (9.42%) and in Q2 (20.10%), and those of treatment 3 are insignificant for both Q3 and Q2. The intuition behind these results is that there

is a quite consistent effect of our reminders over the quartiles, regardless of textual content. We observe the strongest effect in Q2 for treatment 2, which shows an increase in exercise frequency by 20.10% compared to the control group

Table 8: Quartiles of VisitsLast7d

VisitsLast7d	Control	No. of obs	Treatment 1	No. of obs	Treatment 2	No. of obs	Treatment 3	No. of obs
Q4	2.8272	332	3.2114	333	3.2061	308	3.1142	304
	(0)		(0.3842***)		(0.3788***)		(0.2869***)	
	[.]		[0.0000]		[0.0000]		[0.0007]	
Q3	1.3349	379	1.4142	315	1.4606	319	1.3460	322
	(0)		(0.0793**)		(0.1257***)		(0.0111)	
	[.]		[0.0441]		[0.0031]		[0.7504]	
Q2	0.4852	346	0.5520	288	0.5827	270	0.4972	263
	(0)		(0.0668**)		(0.0975***)		(0.0119)	
	[.]		[0.0388]		[0.0048]		[0.6463]	
Q1	0	327	0	441	0	419	0	439

Mean values

(Mean differences in parentheses)

[p-values in squared brackets]

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ 

Note: Table 8 is compiled by six different t-tests. where we compare each treatment to the control, by quartile.

## 6 Discussion

There is a strong indication that the effect on the individuals in treatment 1 and 2 are the most robust, as they are significant in both the FE and the DiD specifications. As discussed earlier, treatment 3 is not significant in our fixed effects specifications, however, it becomes highly significant when including time period 1, controlling for baseline differences (see table 6). Considering the coefficient sizes, which are 8.05%, 5.49%, and 9.45% for treatments 1, 2, and 3, respectively, and the levels of significance in the DiD estimator, all significant at a 1%, the mechanism most susceptible our intervention seems to be the normative message to treatment group 1 and the health benefits in treatment 3. The effect is enhanced when taking into account that treatment group 2 had a higher average of opened e-mails than both treatment group 1 and 3. [Altmann and Traxler \(2014\)](#) and [Apesteguia et al. \(2013\)](#) find no difference in compliance rates when differentiating the contents in their reminders, [Altmann and Traxler \(2014\)](#) for an experiment at a German dentist,

and [Apesteguia et al. \(2013\)](#) for an experiment measuring returning rates at public libraries in Barcelona. The coefficients from our experiment indicate that we see a difference. For instance the coefficient for treatment 3 is almost twice the size of the coefficients for treatment 2. Indicating that the economic significance for group 1 and 3 is larger than that of group 2 (See figure 4).

This can be seen even more clearly by looking at table 7, showing the results of the LATE analysis. The increase in average weekly exercise for the individuals in treatment 1 and 3 who opened the e-mails is estimated to 24.25% and 28.59%, respectively. The corresponding effect for treatment 2 is estimated to 15.84%.

All of our treatments aim to influence the individual's  $\beta$ - parameter, which determines an individual's self control, through different mechanisms. For instance, treatment 1 and 2 aim to change the individual perception of  $\beta$ , i.e.  $\hat{\beta}$ , making a person aware of their own behavioral discrepancy. If we apply this to the theory about hyperbolic discounting from [O'Donoghue et al. \(2006\)](#) what happens in the treatment groups, where treatment was successful, is that  $\hat{\beta}$  converges to  $\beta$  in reference to the text following equation 1, increasing the value of the discount function in the time periods following treatment. This makes the shape of the curve slightly more similar to the rational behavior found in the exponential function, however, we have not removed the  $\beta$  from the function but the kink it produces should not be as jagged anymore. A convergence in  $\hat{\beta}$  to  $\beta$  would also imply that a naïve individual approaches a sophisticated one in this behavior. This would indicate that there is a learning effect and that one is not doomed to naïveté forever, as briefly mentioned by [O'Donoghue and Rabin \(2001\)](#) <sup>7</sup>.

For treatment 1, specifically, the convergence from  $\hat{\beta}$  to  $\beta$  is achieved by sending the members personalized statistics, shedding light on their efforts and thus revealing their  $\beta$ , allowing them to compare their  $\hat{\beta}$  and  $\beta$ . Furthermore, in the context of the Planner-doer model ([Thaler and Shefrin, 1981](#)) the descriptive and injunctive norms could be influencing the  $\theta$ , changing the preferences of their inner doers.

In the context of the same model, we encourage the treatment 2 members to make rules regarding their behavior by scheduling a group session, as a way for the planner to get the doer to commit to a plan against his own immediate interest.

<sup>7</sup>See figure 1 in the Theory section for a visual approach to what this entails.

As indicated by our results this worked, to an extent. We assume that there are sophisticated and naïve individuals in all of our treatment groups and in the control group. Furthermore, we assume that the distribution of these are fairly similar between the groups because of randomization. As the sophisticated individuals are aware of their shortcomings ( $\hat{\beta} = \beta$ ) they should be more likely to sign up for group sessions than their naïve counterparts who might think that they will attend the gym even if they do not schedule a session. It is possible that the naïve individuals are susceptible to this type of information as well, but we cannot distinguish between the two types in our data. This might explain the relatively low effect, compared to treatments 1 and 3, which target both sophisticated and naïve individuals.

The message sent to treatment 3 highlights the importance of future time periods (targeting  $\theta$  in the [Thaler and Shefrin \(1981\)](#) framework, and  $\beta$  in the hyperbolic discounting framework). Perhaps the subjects are already aware of the health benefits of exercise to such an extent that there is only a low marginal increase by our reminder e-mail - i.e. a weak effect. People who signed up for gym memberships might be more aware of these health benefits than the average population - i.e. control group subjects are also aware of this - small relative change in means between the two groups.

The result from the quartile treatment effect is really interesting. As we can see from figure 5 around 30% of our sample work out zero times during our nine week experiment period. Since we do not really know people’s incentives and life situations we should not be too judgmental on the lack of exercise, but we still feel quite comfortable in confirming the title by [DellaVigna and Malmendier \(2006\)](#), that a large share of people indeed are paying not to go the gym.

Unlike [Allcott \(2011\)](#), who observes a boomerang effect on those individuals with the highest electricity consumption from the descriptive norm (quintile 5), we see no such effect in our quartiles. In contrast, we see a quite evenly distributed increase in exercise frequency for Q4, Q3, and Q2. A reason why we don’t see this effect could be that the injunctive norm, included in the e-mail sent to treatment 1, mitigates and suppresses the boomerang effect of the descriptive norm for Q4. However, we see the exactly same effect for treatment 2 and 3 as well, and these e-mails do not include a mitigating injunctive norm.

Our results are also slightly different from the ones presented by [Calzolari and Nardotto \(2016\)](#). They perform a RCT, looking at exercise frequency at a gym. The results they get shows that their simple reminder has a higher positive effect on the, in their case, lower quintiles, i.e those who have a lower exercise frequency. Compared to our own experiment, we can see a very strong positive significant effect in the Q4 (13.59% for treatment 1 , 13.40% for treatment 2 , and 10.15% for treatment 3) and a still positive significant effect in Q3 (5.94% for treatment 1 and 9.42% for treatment 2) and Q2 (13.77% for treatment 1 and 20.10% for treatment 2). The study participants in [Calzolari and Nardotto \(2016\)](#) got to volunteer for an online survey about sports and of students' lives that was advertised via posters at the local. Roughly 55% of the gym population signed up for the survey and the randomized assignment into a treatment and control was later performed on these individuals. This is where our experiments differ from each other. Firstly, [Calzolari and Nardotto \(2016\)](#) only look at a student population and student behavior may differ from the rest of the population, as concluded by [Cunningham et al. \(1974\)](#), implying that the external validity is compromised. Secondly, the individuals that choose to participate in the study could, pretreatment, already have a systematically different behavior than those who do not choose to partake in the survey e.g. it could be that [Calzolari and Nardotto \(2016\)](#) filter out the low frequency observations in their selection process, i.e. individuals in Q1 in our sample. If we speculate, it may be that these individuals have a more positive mindset to exercise than those who do not partake.

This is one of the greatest strengths of our experiment, that individuals could in no possible way be aware that an experiment is performed and there is no or little selection bias whatsoever since the randomization is performed on almost the whole population of SATS (a subset containing all individuals between 18-64 years old). The sample might not be representative of the population at large but should be representative of the population attending gyms as the sample is randomly drawn from the population of SATS members in Sweden. This has consequences for the external validity of the inferences we are able to make based on this sample. Should the population of gym attendants differ systematically and significantly from the rest of the general population our ability to draw correct causal inference concerning the

general population will be limited. Furthermore, SATS is one of the more expensive gyms on the Swedish market, this may imply that there are inherent differences between individuals who are members at SATS and individual who work out at other gyms, further limiting our external validity.

When considering policy relevance, it is important to note that we cannot force people to open their e-mails, thus we have to consider the overall impact and account for both attrition and noncompliance. We would therefore need to consider both the ITT and the LATE in order to predict the efficacy of any potential policy recommendations. Alternatively, a policy could be designed in such a way that it is based on self-selection into the policy, in order to target the individuals who seem to benefit most from the intervention, As stated above, [Calzolari and Nardotto \(2016\)](#) have their participants self select into the study. Their average exercise frequency over a month is also notably higher than our average. Their average is over 2 visits per week, with some variation depending on group and time period, compared to our averages that range between 1.185-1.297 visits per week. Taking this into account and based on the results from our quartile analysis, making only people who self-select eligible for the policy intervention could maximize the effect of the intervention for the selected individuals. However, before implementing an intervention like this one must consider that sending an e-mail is costless, and by sending them to the whole gym population you would reach a larger share of individuals in absolute terms, given that around 30% - 35% of the members seem to open our e-mails.

## 7 Conclusion

It has been demonstrated numerous times that people are time inconsistent (See e.g. [Thaler \(1981\)](#) and [Frederick et al. \(2002\)](#)). In an exercise setting this expresses itself in that individuals procrastinate since they do not value their future well-being in the same sense as their present well-being. People are myopic and have a hard time visualizing future gains. As mentioned above, 51% of the Swedish population is considered to be overweight ([Folkhälsomyndigheten, 2017](#)) and [WHO \(2012b\)](#) estimates that 3.2 million deaths each year are preventable with regular exercise. By sending reminder to individuals we aimed to bypass the individuals

myopic view and highlight future gains by exercise. Over nine weeks we, with the help of SATS, sent three different e-mails every week to three randomized groups. We get significant results for all groups with our DiD model, for 2 out of 3 groups in the time-fixed-effects model. Our time-fixed-effects model and our DiD take into account the fact that we have a declining time trend. Therefore, we feel confident in rejecting  $H_{1_0}$ ,  $H_{2_0}$  and,  $H_{3_0}$ , indicating that e-mail reminders increase exercise frequency. The causal effects of the e-mails are as follows, a 8.05% increase in weekly exercise for treatment 1, the e-mail with an injunctive- and a descriptive norm, 5.49% for treatment 2, the e-mail with a link to a schedule, and 9.45% for treatment 3, the e-mail with information about health benefits. Unlike [Altmann and Traxler \(2014\)](#) and [Apesteguia et al. \(2013\)](#), we observe quite a large difference in our coefficients for the ITT, and therefore, we do feel comfortable in saying that textual content does matter for exercise frequency. In regards to [Calzolari and Nardotto \(2016\)](#) our effect is significantly smaller, their average effect is estimated to be 27%. This can be the cause of the self selection issues, discussed in section 6. The results are of importance in terms of economic significance as well. Suppose that a company is aiming to increase the exercise frequency of it's employees in order to decrease sick rates, increase productivity and general health. Our results indicate that they could increase exercise frequency by almost 10% just by sending e-mails. As a policy tool, sending e-mails is very cost efficient, since it can be automated and you can reach a lot of people with relative ease.

If we allow ourselves to be self-critical, we have a quite short time period. Our experiment only spans over 9 weeks, which is a bit short to estimate any long term effects of the treatment. A post-treatment analysis would also be preferable, but again, this is not within our time frame. Furthermore, as discussed in section 6, SATS is a quite expensive gym, compared to other gyms on the Swedish market, this may imply that SATS members are inherently different from the rest of the gym population, as well as the general population as. This might limit the external validity of our results. Another limitation is that the measurement of whether an individual opened an e-mail is imperfect. It is possible to disable the automated downloading of content in most e-mail clients. This would mean that we have a negative bias in terms of opening rate, which would in turn give us a positive bias of

the LATE. However, these biases affect our treatments in the same way and implies no relative differences between the groups.

Moreover, we have shown that it is quite easy to perform an experiment with a large sample. For example our sample is more than ten times the size than the average sample in [Davies et al. \(2012\)](#), [Van den Berg et al. \(2007\)](#), and [Calzolari and Nardotto \(2016\)](#). Most companies nowadays who interact with a large customer base have quite sophisticated CRM systems, which makes it easy and not too time consuming to collect data and perform experiments like this. Larger samples would give the research in the field a stronger external validity and policy makers would receive better material to base their policies on.

For further research it would be interesting to see what the effect would be by sending similar reminder through a different channel. If we hypothesize that the average opening rate for text messages on cell phones is larger than that of an e-mail it could be that the results from an experiment like this would produce a higher ITT, compared to the one we receive. Furthermore, we've chosen to frame all of e-mails in a positive manner due to simplicity and consistency, but people in general are loss averse ([Thaler, 1981](#)), meaning that they judge losses to be more significant than gains of equal value. Framing the same e-mails in a negative manner might enhance the effect of reminders. Another interesting aspect would be to estimate what the health increase due to increase in exercise would have on increased productivity in the workplace. There are numerous studies that confirm that an increase in health also gives an increase in productivity (See e.g. [Tomba \(2002\)](#)).

---

## References

- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9):1082–1095.
- Altmann, S. and Traxler, C. (2014). Nudges at the dentist. *European Economic Review*, 72:19–38.
- Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Apestequia, J., Funk, P., and Iriberry, N. (2013). Promoting rule compliance in daily-life: Evidence from a randomized field experiment in the public libraries of barcelona. *European Economic Review*, 64:266–284.
- Arem, H., Moore, S. C., Patel, A., Hartge, P., de Gonzalez, A. B., Visvanathan, K., Campbell, P. T., Freedman, M., Weiderpass, E., Adami, H. O., et al. (2015). Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship. *JAMA internal medicine*, 175(6):959–967.
- Berns, G. S., Laibson, D., and Loewenstein, G. (2007). Intertemporal choice—toward an integrative framework. *Trends in cognitive sciences*, 11(11):482–488.
- Calzolari, G. and Nardotto, M. (2016). Effective reminders. *Management Science*.
- Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.
- Consortium, W. W. W. et al. (2015). Internet users. *Internet Live Stats*.
- Cunningham, W. H., Anderson, W. T., and Murphy, J. H. (1974). Are students real people? *The Journal of Business*, 47(3):399–409.
- Davies, C. A., Spence, J. C., Vandelanotte, C., Caperchione, C. M., and Mummery, W. K. (2012). Meta-analysis of internet-delivered interventions to increase physical activity levels. *International Journal of Behavioral Nutrition and Physical Activity*, 9(1):52.

- 
- DellaVigna, S. and Malmendier, U. (2006). Paying not to go to the gym. *The American Economic Review*, 96(3):694–719.
- Folkhälsomyndigheten (2017). Övervikt och fetma - nationella resultat och tidsserier 2016.
- Frederick, S., Loewenstein, G., and O’donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, 40(2):351–401.
- Gupta, S. K. et al. (2011). Intention-to-treat concept: a review. *Perspectives in clinical research*, 2(3):109.
- Hurling, R., Catt, M., De Boni, M., Fairley, B., Hurst, T., Murray, P., Richardson, A., and Sodhi, J. (2007). Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. *Journal of medical Internet research*, 9(2):e7.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478.
- Lumley, T., Diehr, P., Emerson, S., and Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual review of public health*, 23(1):151–169.
- O’Donoghue, T., Rabin, M., et al. (2006). Incentives and self-control. *Econometric Society Monographs*, 42:215.
- O’Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, pages 103–124.
- O’Donoghue, T. and Rabin, M. (2001). Choice and procrastination. *The Quarterly Journal of Economics*, 116(1):121–160.
- Phelps, E. S. and Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *The Review of Economic Studies*, 35(2):185–199.
- Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics letters*, 8(3):201–207.

- 
- Thaler, R. and Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.
- Thaler, R. H. and Shefrin, H. M. (1981). An economic theory of self-control. *Journal of political Economy*, 89(2):392–406.
- Tompa, E. (2002). The impact of health on productivity: empirical evidence and policy implications. *The review of economic performance and social progress*, 2.
- Tversky, A. and Kahneman, D. (1985). The framing of decisions and the psychology of choice. In *Environmental Impact Assessment, Technology Assessment, and Risk Analysis*, pages 107–129. Springer.
- Van den Berg, M. H., Schoones, J. W., and Vlieland, T. P. V. (2007). Internet-based physical activity interventions: a systematic review of the literature. *J Med Internet Res*, 9(3):e26.
- WHO (2012a). Physical inactivity: a global public health problem. 2010. URL [www.who.int/dietphysicalactivity/factsheet\\_inactivity/en/](http://www.who.int/dietphysicalactivity/factsheet_inactivity/en/). Part I APPENDIX.
- WHO (2012b). Recommended levels of physical activity for adults aged 18–64 years.
- WHO (2014). The top 10 causes of death. 2014. URL [www.who.int/mediacentre/factsheets/fs310/en/](http://www.who.int/mediacentre/factsheets/fs310/en/).

## Appendices

### Appendix A

Comparing the marginal rates of substitution between  $u_{t+1}$  and  $u_{t+2}$  from the perspective of  $t$  and  $t + 1$ , in reference to the hyperbolic discounting model in section 2.1.

First, we take the partial differentials from the perspective of  $t$ :

$$\begin{aligned}\frac{\partial U_t}{\partial u_{t+1}} &= \beta \delta^t u'_{t+1} \\ \frac{\partial U_t}{\partial u_{t+2}} &= \beta \delta^{t+1} u'_{t+2}\end{aligned}$$

Then the total differential from the perspective of time  $t$ :

$$\begin{aligned}dU_t &= \left(\frac{\partial U_t}{\partial u_{t+1}}\right)du_{t+1} + \left(\frac{\partial U_t}{\partial u_{t+2}}\right)du_{t+2} \\ dU_t &= (\beta \delta^t u'_{t+1})du_{t+1} + (\beta \delta^{t+1} u'_{t+2})du_{t+2} = 0 \\ (\beta \delta^t u'_{t+1})du_{t+1} &= -(\beta \delta^{t+1} u'_{t+2})du_{t+2} \\ (\beta \delta^t u'_{t+1})/(\beta \delta^{t+1} u'_{t+2}) &= -du_{t+2}/du_{t+1} \\ du_{t+2}/du_{t+1} &= -(u'_{t+1})/(\delta u'_{t+2})\end{aligned}$$

Second, we take the partial differentials from the perspective of  $t + 1$ :

$$\begin{aligned}\frac{\partial U_{t+1}}{\partial u_{t+1}} &= u'_{t+1} \\ \frac{\partial U_{t+1}}{\partial u_{t+2}} &= \beta \delta u'_{t+2}\end{aligned}$$

Then the total differential from the perspective of time  $t + 1$ :

$$\begin{aligned}dU_{t+1} &= \left(\frac{\partial U_{t+1}}{\partial u_{t+1}}\right)du_{t+1} + \left(\frac{\partial U_{t+1}}{\partial u_{t+2}}\right)du_{t+2} \\ dU_{t+1} &= (u'_{t+1})du_{t+1} + (\beta \delta u'_{t+2})du_{t+2} = 0 \\ (u'_{t+1})du_{t+1} &= -(\beta \delta u'_{t+2})du_{t+2} \\ (u'_{t+1})/(\beta \delta u'_{t+2}) &= -du_{t+2}/du_{t+1} \\ du_{t+2}/du_{t+1} &= -(u'_{t+1})/(\beta \delta u'_{t+2})\end{aligned}$$

Finally, we compare the two different marginal rates of substitution:

$$du_{t+2}/du_{t+1} = -(u'_{t+1})/(\delta u'_{t+2})$$

and

$$du_{t+2}/du_{t+1} = -(u'_{t+1})/(\beta\delta u'_{t+2})$$

and conclude that they are different:

$$-(u'_{t+1})/(\delta u'_{t+2}) \neq -(u'_{t+1})/(\beta\delta u'_{t+2})$$

## Appendix B

## Håll koll på din träning



Genom att följa upp din träning får du bättre möjlighet att uppnå dina mål och skapa en sund rutin.

De senaste 4 veckorna har du i snitt besökt våra träningscenter **3** gånger per vecka. Världshälsoorganisationen rekommenderar att du som vuxen ska träna **2,5** timmar i veckan för en hälsosam livsstil!

Bra jobbat – du verkar ha hittat en skön rytm för din träning. Håll igång den här takten och säg till när du är redo för nya utmaningar.

Figure 6: Sample of the e-mails sent to treatment 1

The text in figure 6 translated into English reads: "Keep track of your exercise. By following up your exercise you get a better opportunity to reach your goals and establish a healthy routine.

The past 4 weeks you have on average visited our exercise facilities 3 times per week. The World Health Organization recommends that you as an adult should exercise 2.5 hours per week for a healthy lifestyle.

Well done - you seem to have found a good rhythm for your exercise. Keep up this pace and let us know when you are ready for new challenges."

In the sample e-mail above the individual had an exercise frequency above the threshold. Those who did not exercise above the threshold level received the following text: "Sometimes exercise is not the top priority for different reasons and we hope that you soon will return. As always we are available for encouragement and inspiration if it is only a matter of low spirits".

# SATS

## Boka in kommande träningspass



Få en energiboost genom **Crosstraining**, **Cycling** eller **MOVE!**. Oavsett om ditt mål är styrka, kondition eller kroppskontroll finns ett pass nära dig. Att träna tillsammans med andra kan ge dig det du behöver för att få ut det lilla extra av din träning.

[Boka nu](#)

Figure 7: Sample of the e-mails sent to treatment 2

The text in figure 7 translated into English reads: "Schedule upcoming group sessions. Get an energy boost through Crosstraining, Cycling or MOVE! Regardless if your goal is strength, stamina or coordination there is a session close to you. To

exercise with others can give you what you need to get that little extra from your workout. Schedule now."

The specific sessions recommended changed in each e-mail to appeal a wider variety of members.

# SATS

## Positiva effekter av träning



Visste du att regelbunden träning ökar den förväntade livslängden?

Med regelbunden träning kan du motverka hjärt- och kärlsjukdomar, högt blodtryck och stroke samtidigt som det hjälper dig hålla en jämn vikt och minska risken för metabolt syndrom.

[Läs mer om träning och hälsa](#)

Figure 8: Sample of the e-mails sent to treatment 3

The text in figure 8 translated into English reads: "Positive effects of exercise. Did you know that regular exercise increases life expectancy? With regular exercise you can prevent cardiovascular diseases, high blood pressure and stroke while also

helping you keep an even weight and reduce the risk of metabolic syndrome. Read more about exercise and health."

Treatment group 3 was sent five different text concerning health benefits associated with exercise, which were then cycled after five weeks. These texts translate into the following:

Text 2: "Studies have shown that the risks for type 2 diabetes, depression as well as breast cancer and colon cancer are reduced by regular exercise. Through exercise you also strengthen bones and muscles reducing the risk of hip- and spinal fractures."

Text 3: "Did you know that people who exercise regularly exhibit higher levels of oxygen absorption and a stronger muscle composition? Regular exercise also helps you keep an even weight, as well as getting a healthier body."

Text 4: "If you have a lot to do and are feeling stressed it can actually be a good idea to schedule a workout session. Regular exercise contributes to a higher stress threshold and to improved sleep for people with sleep difficulties."

Text 5: "Did you know that exercise increases general wellbeing and prevents anxiety and depression. Studies also show that regular exercise leads to an increased learning capability and cognitive function"

The button labeled "Read more about exercise and health" linked to the following websites:

1: <http://www.1177.se/Tema/Halsa/Motion-och-rorelse/Motion-och-traning/Rorelse-ar-livsviktigt/>

2: <https://www.folkhalsomyndigheten.se/far/inledning/fysisk-aktivitet-och-halsa/>

3: <http://www.1177.se/Vastra-Gotaland/Tema/Halsa/Motion-och-rorelse/Motion-och-traning/FaR—fysisk-aktivitet-pa-recept/>

4: <http://www.1177.se/Vastra-Gotaland/Fakta-och-rad/Mer-om/Motion-som-medicin-mot-psykisk-ohalsa/>

5: <http://www.1177.se/Fakta-och-rad/Sjukdomar/Stress/>

## Appendix C

In table 9 we regress our dependent variable, number of opened e-mails on "VisitsLast7d" and some control variables, at the baseline, i.e. at Time=1. The variable number of opened e-mails is an aggregated measurement taken at the end of our time period and thus each individual only has one unique observation of this variable. This regression gives us an indication of whether the aggregated number of e-mails an individual in the treatment groups opened is correlated with exercise frequency at the baseline, before the e-mails were sent. In other words, it is an indication of whether people who exercise a lot pre-treatment are more likely to open our e-mails. As we can see, this variable is highly significant indicating that this is indeed the case (significant at the 1% level), implying that we cannot compare the subset of the treatment groups who opened e-mails to the entire control group. However, this estimation should be interpreted with caution as it likely suffers from omitted variable bias, as we have very few control variables due to our experimental design. Furthermore, we can see that "gender" is positive and significant at a 1% level, indicating that men on average open more e-mails than women. The coefficient for "Private" is also positive and significant at the 10% level, meaning that members belonging to this membership type open more e-mails compared to the reference group, "Corporate".

Table 9: Baseline characteristics and no. of opened e-mails

(1)	
No. opened e-mails	OLS
gender	0.343*** (0.107)
age	-0.0316 (0.0380)
age2	0.000723 (0.000476)
Friend	-0.369 (0.299)
Private	0.199* (0.120)
Senior	-0.793 (0.650)
Student	0.0376 (0.175)
VisitsLast7d	0.150*** (0.0331)
Constant	2.671*** (0.746)
Observations	3,561
R-squared	0.017

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## Appendix D

Table 10: Time-fixed effects and random effects

	(1)	(2)	(3)	(4)
VisitsLast7d	Time-Fixed Effects	Time-Fixed Effects	GLS Random Effects	GLS Random Effects
Treatment 1	0.0531*** (0.0105)	0.0537*** (0.0107)	0.0547*** (0.0109)	0.0576*** (0.0122)
Treatment 2	0.0387** (0.0131)	0.0310** (0.0130)	0.0402*** (0.0130)	0.0348*** (0.0132)
Treatment 3	-0.0179 (0.0131)	-0.0241 (0.0134)	-0.0164 (0.0126)	-0.0203 (0.0126)
Constant	1.172*** (0.00746)	1.221*** (0.0225)	1.171*** (0.0148)	1.219*** (0.0323)
Observations	39,671	39,669	39,671	39,669
R-squared	0.000	0.006		
Number of Time	8	8	8	8
Time-Fixed Effects	YES	YES	NO	NO
Controls	NO	YES	NO	YES
Random Effects	NO	NO	YES	YES

Clustered standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 10 shows four different specifications, two with time-fixed effects (columns one and two) and two with random effects (columns three and four). The specifications in columns two and four include a set of control variables, whereas columns one and three do not. As we can see the results are very similar regardless of the exact specification used. It is not surprising that there is only a small change in coefficients when including control variables since the groups are fairly balanced at the baseline. The time-fixed effects model assumes that differences across time are correlated with our independent variables. The random effects model, on the other hand, assumes that this is not the case, i.e. that these time specific differences are orthogonal to our independent variables. We perform a Hausman specification test to determine if time-fixed effects or random effects are more appropriate. Under the null hypothesis of the test the random effects model is preferred as it is both consistent and efficient, whereas the time-fixed effects model is consistent but inef-

ficient. Under the alternative hypothesis the time-fixed effects model is consistent and the random effects model is inconsistent. The resulting chi2-statistic=162.20 with  $\text{Prob}>\chi^2=0.0000$ , implying that we can reject the null hypothesis and that the time-fixed effects model is more appropriate as it is consistent. Furthermore, we get the same results when we perform a Sargan-Hansen test (Sargan-Hansen statistic = 14.000, p-value=0.0156).