The B.E. Journal of Economic Analysis & Policy

Topics

Volume 10, *Issue* 1	2010	*Article* 91

The Effect of Competition on Physical Activity: A Randomized Trial

Magnus Johannesson* Robert Östling† Eva Ranehill‡

*The Stockholm School of Economics, magnus.johannesson@hhs.se
†Institute for International Economic Studies, robert.ostling@iies.su.se
‡The Stockholm School of Economics, eva.ranehill@hhs.se

Recommended Citation
Magnus Johannesson, Robert Östling, and Eva Ranehill (2010) "The Effect of Competition on Physical Activity: A Randomized Trial," *The B.E. Journal of Economic Analysis & Policy*: Vol. 10: Iss. 1 (Topics), Article 91.
Available at: http://www.bepress.com/bejeap/vol10/iss1/art91

The Effect of Competition on Physical Activity: A Randomized Trial*

Magnus Johannesson, Robert Östling, and Eva Ranehill

Abstract

Recent literature in economics has highlighted that competition and symbolic awards can provide non-monetary incentives. In this paper, we report on a step contest that we carried out at a large Swedish workplace in order to test whether competition for symbolic awards can be used to promote physical exercise. Each individual was equipped with a pedometer and registered the number of steps daily during a four week period. Participants competed both in teams and individually and the winning team and individual received symbolic prizes. To evaluate the effect of the competition, we randomized teams into a control group and two treatment groups. We found that the step contest significantly increased both the fraction of subjects that completed the step contest and the number of steps. The number of steps was about 1,000 steps higher in the main treatment group than in the control group (an increase by about 10 percent). This is a conservative estimate as the dropouts on average walked fewer steps than individuals completing the study. In an additional treatment, we also included a daily step goal in the contest. The step goal had no additional significant effect on the number of steps, which may be due to the relatively low step goal used (7,000 steps per day).

KEYWORDS: competition, incentives, exercise, obesity, pedometer

*We thank Mia Hultberg, Anders Raustorp and three anonymous referees for helpful comments and Hanna Christiansson for editorial assistance. Financial support from the Jan Wallander and Tom Hedelius Foundation, the Swedish Research Council, and the Swedish Council for Working Life and Social Research is gratefully acknowledged.

1. Introduction

Given the increasing trend in obesity in Western countries (Cutler et al. 2003), it is of central importance to increase physical exercise. Both physical exercise and obesity are important determinants of health (Brown 1992; Miller et al. 1997; Lee et al. 1999; Speck and Looney 2001; Alevizos 2005; Guh et al. 2009; Whitlock et al. 2009). Physical exercise has both a direct effect on health, as well as an indirect effect through affecting important risk factors of diseases like obesity, blood pressure, and cholesterol levels (Brown 1992; Miller et al. 1997; Lee et al. 1999; Speck and Looney 2001; Alevizos 2005). The economic costs to society of obesity and inactivity are substantial (Keeler et al. 1999; Cox et al. 1998; Colditz 1999; Fiebelkorn et al. 2003; Branca et al. 2007). It is therefore important to test new and innovative approaches to promote exercise. A recent study by Charness and Gneezy (2009) showed that physical exercise can be increased by paying individuals to go to the gym, and another study showed some success in using cash incentives to reduce smoking (Volpp et al. 2009). The evidence is more mixed regarding financial rewards as a means of reducing obesity (Finkelstein et al. 2007; Volpp et al. 2008; Cawley and Price 2007). In this study we tested another incentive to increase exercise and health: a contest to win a symbolic award. We carried out a step contest at a major Swedish hospital in which employees, equipped with pedometers, competed (both individually and in teams) to win a symbolic prize.

Recent research in economics has shown that both competition per se and symbolic awards can act as non-monetary incentives. The introduction of a competitive element (for a given level of monetary incentives) has been shown to increase performance in a number of economic experiments performed on students and children (Gneezy, Niederle and Rustichini 2003; Gneezy and Rustichini 2004; Niederle and Vesterlund 2007; Booth and Nolen 2009; Niederle, Segal and Vesterlund 2010; Dreber, von Essen and Ranehill 2010; Sutter and Rützler 2010). These studies measure the impact of competition in a one-shot experimental setting, mainly using tasks such as solving mazes, math exercises, or running short distances (less than 100 meters). We contribute to this literature by testing if a competitive setting can also promote exercise. Compared to the above studies we also study behavior during a longer period and in a more representative sample.

In our experiment participants compete for a symbolic prize. Using the terminology of Frey (2006, 2007), the symbolic prize is a form of award, and Frey refers to such incentives as "non-material, extrinsic compensation".[1] One argument for why awards may improve performance is that it signals relative

[1] As examples of awards Frey (2006) for instance includes Academy Awards (Oscars), prizes in sports competitions, the Pulitzer Prize, and awards in academia such as Nobel Prizes.

position, which is valuable if individuals care about their rank compared to other individuals (Besley and Ghatak 2008). There is a sizeable literature in economics on status and positional goods, see for instance Hirsch (1978) and Frank (1985). Another related argument is that awards are associated with increased social esteem that is directly valued by individuals (Frey 2006, 2007). For recent models of social esteem in economics, see Bénabou and Tirole (2006), Ellingsen and Johannesson (2008) and Andreoni and Bernheim (2009).[2] The widespread use of awards suggests that they are effective in enhancing performance. Although there is only limited empirical work testing this proposition, three recent studies suggest that awards can increase contributions to public goods (Neckerman and Frey 2007), increase the supply of blood donations (Lacetera and Macis 2008) and increase worker performance (Kosfeld and Neckermann 2010).[3] We contribute to this line of empirical research by testing whether competition for a symbolic prize can be used to increase physical exercise.

Pedometer use is a well-established way to increase physical activity.[4] A recent systematic review of studies testing the effects of using a pedometer showed that according to the randomized clinical trials of pedometer use, the use of a pedometer on average increased the number of steps per day by about 2,500; an increase by about 30 percent over the baseline activity level (Bravata et al. 2007). The corresponding effect based on observational studies was an increase by 2,183 steps per day. The effect of pedometer use was especially pronounced if a step goal was used, e.g. a step goal of at least 10,000 steps per day. According to the study by Bravata et al. (2007), the use of a pedometer also led to a significant reduction in body mass index (BMI) and systolic blood pressure.

The step contest studied in this paper was not invented by us, but is a common practice at many workplaces in Sweden. Out of a labor force of about 4.5 million, about 150,000 individuals participate in step contests every year.[5] To what extent these step contests increase physical activity is thus an important public health issue in Sweden. In a typical step contest the participants form teams and compete both individually and in teams. The participants in a step contest are equipped with a pedometer and then register the number of steps on a homepage created for the contest; on this homepage they can also see the number of steps of all other participants as well as their position in the contest. The winner is the individual and the team with the highest number of steps during the period of the

[2] Ariely et al. (2009) provides a recent empirical test of the social esteem model.

[3] See also the related work in psychology and management science about the effects of social recognition on performance (Stajkovic and Luthans 2003).

[4] The review article by Tudor-Locke et al. (2002) compares pedometers with other measures of physical activity and find strong correlations verifying that steps measured with a pedometer is a valid measure of physical activity.

[5] This is an estimate based on talking to some of the providers of step contests in Sweden (as the companies are unwilling to reveal their exact market size it is difficult to get the exact number).

contest (normally four weeks).[6] The reward is typically a symbolic award like a cup.

Step contests are a way of trying to promote increased exercise among the employees, but we know little about the actual effects on exercise and health. The studies evaluating the effects of the use of pedometers suggest that using a pedometer per se can have substantial positive effects (Bravata et al. 2007). In the present study we tested, in a randomized field experiment, if the competitive aspect of step contests has an additional positive effect on exercise.[7] Compared to the previous clinical trials included in the Bravata et al. (2007) review our study is the first to test the importance of symbolic rewards and competition. The study also investigates a "real world" intervention in a realistic setting.

The study was carried out at a major hospital in Sweden, which is similar to other workplaces that organize steps contests (a previous step contest had for instance been carried out at this very same hospital). Three experimental treatments were included. To be able to compare the effects of the contest per se we included a control group in the study in which individuals were equipped with a pedometer and registered their steps in the same way as in a step contest, but without being able to see the number of steps by other participants or their position in the contest. The second experimental group was a standard step contest, with the only difference compared to the control group being that individuals could see the number of steps of other participants and that the winning individual and team received a cup at the end of the four week period. The treatment we are looking at is the combination of an individual contest and a team contest to win a symbolic award. Individuals form teams which have a common goal and receive information about the number of steps both within and between teams to generate a competitive environment. Individuals in the control group share no information within or across teams. We also included a third experimental group that added a step goal to the experimental treatment. In this treatment the individuals had to walk at least 7,000 steps per day to participate in a random draw of prizes during the study (everyone in the other two groups were also randomly allocated prizes irrespective of their number of steps). With this third experimental group we tested the hypothesis that adding a step goal to a step contest will further increase physical activity. We chose a relatively modest step goal in order to provide the least physically active participants with an additional incentive to walk more.

Our design isolates the effect of the step contest per se, i.e. the effect over and above using a pedometer to monitor physical activity. A limitation of our design is that we cannot separate the effect of the symbolic reward (the award)

[6] The winner of the team competition is the team with the highest number of steps per team member.

[7] In the terminology used by Harrison and List (2004) the study is a framed field experiment.

from other aspects of the competition. Our estimate does also not separate the effect of the individual and the team contest, but accounts for the joint effect of these. With more experimental treatments it would have been possible to disentangle the effects of individual competition from team competition, but that was not within the scope of the present study. In principle a team contest could be either more or less effective than an individual contest. The team contest introduces possibilities to free ride which could reduce performance, but the effect of peer pressure would go in the other direction.

The design of the study is presented below, followed by a section about the hypotheses and statistical tests. Thereafter we present the results. The paper ends with a discussion of our findings.

2. Design of study

The study was carried out during the spring of 2009 at a major hospital in Sweden (Södersjukhuset with about 4,200 employees); we refer to Södersjukhuset as "the hospital" below. The step contest was handled by Select Wellness AB who handles a large number of step contests in Sweden (they have carried out step contests for more than 300,000 individuals).[8] In previous step contests organized by Select Wellnes AB the average number of steps has been 11,357 for hospitals and 10,674 for other employers. Although the number of steps among hospital employees has been slightly higher than for average participants, these numbers still suggest that hospitals are relatively representative for the workplaces that typically participate in step contests.

The hospital offered all their employees to participate in the study and all employees received information about the study. The employees were told that the step contest was part of a scientific study conducted to evaluate the effects of step contests. They were also told that teams would be randomized to one of three groups and the three groups were briefly described. The information also stated that the study was voluntary and that individuals could discontinue the study at any point in time.[9]

[8] Select Wellness AB charge the employers SEK 365 (approximately $50 based on the exchange rate at the time of the study) per participant in a step contest (including the cost of the pedometers). It is thus a relatively low cost intervention to increase physical exercise.

[9] The study was approved by the Regional Ethics Committee in Stockholm.

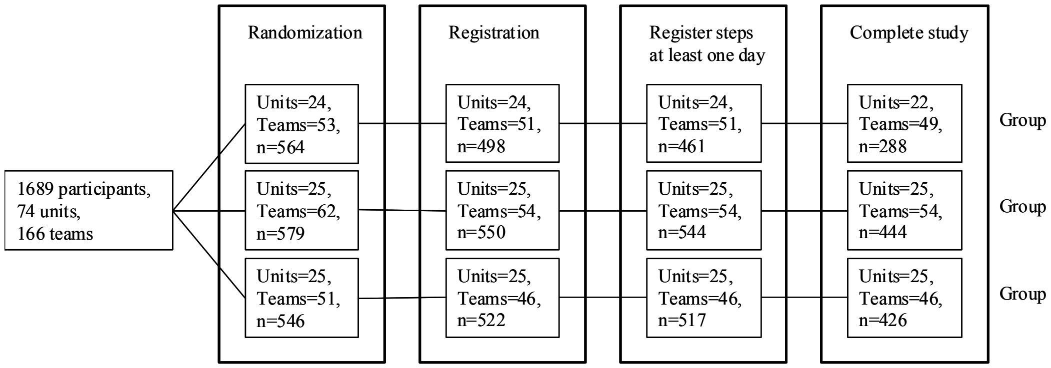

Figure 1. Design of the study

The project setup is illustrated in Figure 1. Teams consisting of 5–15 individuals could sign up for the study (individuals self select into teams). At the end of the entry period 166 teams with in total 1689 individuals had signed up for the study. This means that about 40 percent of the employees at the hospital signed up for the study. The population participating in the study is a much more unselected population than the patient populations typically taking part in the clinical trials on pedometer use included in the overview by Bravata et al. (2007). However, this does not imply that the participants are representative of the general population in Sweden, nor of the employees at the hospital (the 40 percent selecting into the study can differ from the 60 percent not choosing to participate).

The teams were divided into 74 randomization units by the hospital so that teams in the same work unit were in the same randomization unit, i.e. a randomization unit consisted of all the teams in the same work unit (as defined by the hospital). This was done to make sure that everyone in the same work unit would face the same incentives and it was also a prerequisite by the hospital to carry out the study. The randomization was stratified with respect to the number of individuals in each randomization unit and the 74 randomization units were thereafter randomly allocated into the following three experimental groups:[10]

Group 1 (control group): Each participant received a pedometer and registered their number of steps daily during the four weeks of the study. The participants could not see the number of steps by other team members or other participants in

[10] There were 44 randomization units with 22 or fewer individuals in each. Out of these, 14 were randomly selected to group 1 and 15 to group 2 and 3. Of the remaining 30 units with 23 or more individuals, 10 were randomly selected into each group.

the study.[11] The participants could win the prizes that were randomly allocated to all participants in the study.

Group 2 (the contest treatment group): Each participant received a pedometer and daily registered the number of steps during the four weeks of the study. The participant could observe the number of steps of all other participants, as well as see his/her position in the contest and the position of his/her team. Prizes were randomly allocated to all participants in the study. A symbolic prize (a cup) was given to the individual with the most steps and the team with most steps per team member.

Group 3 (the contest with a step goal): This treatment was the same as in group 2, with the difference that only participants with an average number of steps over 7,000 per day could win the prizes that were randomly allocated.[12]

After the randomization all team leaders (each team had to have a team leader) received entry packages for all participants in the team and allocated one package to each team member. The package consisted of brief information about the study and the group to which the individual had been randomized, a pedometer, as well as information about how to register on the home page for the contest.[13] At the registration individuals were asked to fill out a questionnaire with some background information. They were also asked to register the number of steps they took on a daily basis.[14] After the four weeks the individuals were asked to fill out another questionnaire.[15]

---

[11] The steps of the control group were never revealed to any subject in any of the groups during the study.

[12] 7,000 steps corresponds to a distance of about five kilometers if each step is counted as 75 centimeter (which is a standard conversion used).

[13] Participants received a step diary where they could write down the number of daily steps and they also registered the number of steps at a homepage. They were encouraged to register their steps every day on the home page, but it was also for instance possible to enter the homepage once a week and then register the steps from the step diary for each day in that week. There were no explicit instructions for what to do if a participant failed to use the pedometer a day. The pedometer is for measuring the activity of walking. For other physical activities like running individuals were told to convert 30 minutes of the activity to 2,700 steps.

[14] The number of daily steps was not verified by anyone; it was completely based on self-reported number of steps (which was also the standard in the clinical trials included in the Bravata et al. 2007 overview).

[15] It would have been ideal to continue to follow-up the participants after the end of the four week intervention period. Unfortunately this was not possible, and we have no data on the number of steps or other measures of physical activity or health after the end of the four week intervention period.

Some prizes were randomly allocated among all participants (in group 3 only participants with over 7,000 steps per day could win prizes). Three weekly draws were conducted with 75 “winners” each time; each individual in these draws won a lottery ticket worth SEK 25.[16] After the end of the study, three SEK 1,000 gift certificates were also randomly allocated among all participants (but in group 3 only participants with over 7,000 steps per day could win).

The entry questionnaire consisted of a number of background questions (age, sex, education, marital status, if they had children, if they worked full time or not and if they had a desk job or not), questions on exercise habits (the hours of exercise per week in a typical week, the hours of exercise the previous week, and their self-assessed physical activity level compared to other individuals (below average, average, or above average), as well as questions on health status (length and weight to measure BMI; self-assessed health status: poor, fair, good, very good, or excellent).[17] The exit questionnaire only included questions about exercise habits and health status (but not height).

3. Hypotheses and tests

We tested two main hypotheses with this design. The first hypothesis was that the competition aspect of the step contest increases the levels of physical activity, i.e. that the number of steps would be higher in group 2 than in group 1. The second hypothesis was that the inclusion of a step goal would further increase the activity level, i.e. that the number of steps would be higher in group 3 than in group 2. As a secondary hypothesis we tested if the step contest led to a reduction in self-assessed BMI as an indicator of improved health.[18] It should be noted that self-assessed BMI is a noisy measure of overweight. It would have been preferable to measure BMI by measuring weight and height in the study. BMI as such is also an imperfect measure of overweight as muscle mass and bone structure can differ between individuals, and exercise can also lead to an increase in muscle mass. It is also unlikely to be able to detect a change in weight during just a four week intervention period. These limitations should be borne in mind with respect to our secondary hypothesis concerning BMI, but it is still of interest to report these results.

---

[16] The exchange rate at the time of the study was approximately $1 = SEK 7.5.

[17] All employees at the hospital, including both nurses, physicians and other workers, were invited to participate in the study. Unfortunately, we did not collect data about occupation in the study apart from the variables about desk job or not and if the individuals worked full time or not, so we cannot directly control for occupation in the analyses (though occupation is likely to be correlated with years of education that we do control for).

[18] The BMI change was based only on changes in the reported weight during the study; the height was only measured in the beginning of the study.

As some individuals dropped out of the study we evaluated our hypotheses both based on the individuals that completed the study (referred to as "completers" below) and on all the possible information about dropouts (i.e. participants were included with the number of steps they walked before they dropped out; referred to as "all subjects" below).[19]

For "completers" the statistical tests were based on an OLS regression with dummy variables for the three experimental groups and the daily average number of steps as the dependent variable. To account for that the observations within teams and units may be correlated with each other, we estimated the OLS regression with clustered standard errors at the level of the randomization units (Wooldridge 2003). We carried out these tests with and without controlling for the background characteristics collected in the entry questionnaire (and to test if the background variables differed significantly between the groups we also used OLS regressions using clustered standard errors). The same statistical methodology was also used to test our secondary hypothesis about the effects on BMI as measured in the questionnaire responses (based on all individuals that filled out both the entry and the exit questionnaire).

In the estimation for "all subjects" we estimated the number of steps per individual during each day of the competition including all subjects that were still in the study; e.g. an individual that dropped out on day 11 would be included in the estimate of the number of steps per individual in days 1-10, but not in days 11-28. We did this estimation with and without controlling for background characteristics (when we controlled for background characteristics we ran a separate OLS regression for every day of the contest with the group dummy variables and the background characteristics as explanatory variables). As the "all subjects" sample (due to the dropouts) differed on different days in the study, it is not straightforward to carry out standard statistical tests of the difference in mean steps between the groups. To carry out the statistical tests we therefore used bootstrap methods (Efron and Tibshirani 1993).[20]

---

[19] An individual with zero steps in a day was defined as having dropped out of the study if the individual had not entered steps at any subsequent day. The dropout day is the day after the last day that an individual reported steps. If an individual for instance reported steps for days 1–10, but not for the subsequent days, the dropout day is day 11. Some individuals did not register steps for every day in the period, and any blank days were counted as zero steps during that day (for some individuals it was evident that they summed the steps for several days, e.g. a week, and entered the sum for a single day; but it was impossible to separate these observations from true missing values).

[20] The p-values were based on generating 1,099 bootstrap replications, which according to Davidson and MacKinnon (2000) should yield a relatively high precision in the estimated p-values. The reason for choosing numbers of replications ending with 99 is that Monte Carlo tests are exact only if $p(N + 1)$ is an integer, where N is the number of replications and p is the significance level (Davidson and MacKinnon 2000). In each bootstrap replication we randomly drew 74 randomization units with replacement and based on each draw we calculated the

Comparing the results for completers with the results for "all subjects" indicates in what direction dropouts biased the results (i.e. if dropouts prior to dropping out walked more or less steps than non-dropouts). As a further check of the impact of dropouts on our results we therefore compared the number of steps between dropouts and non-dropouts. We constructed two measures for each individual that dropped out. In the first, we estimated the difference between the average number of steps per day prior to dropping out and the average number of steps per day in the same period for the individuals that did not drop out at that day in that group (e.g. if an individual in group 3 that dropped out by day 11 had walked 11,500 steps per day in days 1-10 and the average number of steps in days 1-10 in group 3 for individuals that did not drop out on day 11 was 12,000 steps, this measure was equal to –500 (11,500 – 12,000)). This measure shows if individuals that dropped out on a specific day had walked fewer steps up until that day compared to individuals in their group that had not yet dropped out of the study. The second measure is similar, but with the difference that the steps per day prior to dropping out was now compared with the steps per day for completers in the same period (e.g. if an individual in group 3 that dropped out on day 11 had walked 11,500 steps per day in days 1-10 and the average number of steps in days 1-10 in group 3 for individuals that completed the study was 12,500 steps, this measure was equal to –1,000 (11,500 – 12,500)). This measure shows if individuals that dropped out on a specific day had walked fewer steps up until that day compared to individuals in their group that completed the study.

These two measures give an indication of if individuals dropping out of the study were walking fewer or more steps than those that continued in the study. To statistically test if dropouts and non-dropouts differed in the number of steps based on the two above measures, we ran OLS regressions with dummy variables for the three experimental groups using clustered standard errors at the level of randomization units.

4. Results

4.1 Subjects and dropouts

As can be seen in Figure 1, 1,689 individuals were initially randomized in the study. Out of these individuals 1,570 (93 percent) registered for the study and filled out the entry questionnaire. The dropout between randomization and registration was somewhat higher in group 1 (11.7 percent), than in groups 2 and

---

differences in mean number of steps between the three groups. The reported *p*-values were calculated under the assumption that the bootstrapped estimates were approximately normal (using the standard bootstrap procedure in STATA 10), but the *p*-values calculated directly from the 1,099 replications were highly similar.

3 (5.0 percent in group 2 and 4.4 percent in group 3). Unfortunately we have no information about these individuals as the questionnaire was filled out at registration. The background characteristics of those individuals who filled out the questionnaire are shown in Table 1. As expected with the randomized design, the groups were similar in terms of background characteristics. In only four of the 33 pairwise comparisons were the differences statistically significant at the five percent level (there is a higher rate of married/cohabiting individuals in group 3 than in group 1 or 2; the hours of exercise are lower in group 3 than in group 2; and the self-assessed physical activity level is higher in group 1 than in group 3).

For the individuals that registered for the study we have at least some information about the number of steps for 97 percent of the individuals; i.e. they filled in the number of steps for at least one day. Also at this stage the dropout rate was larger in group 1 (7.4 percent of those who registered never reported any information about steps) than in groups 2 and 3 (1.1 percent in group 2 and 1.0 percent in group 3).[21]

The dropout rate during the study is shown in Figure 2 for the individuals that registered steps during the first day. It is evident that the dropout rate during the study was much higher in the control group than in the two other groups. In group 1, 37.5 percent of the individuals that started the study dropped out during the study, whereas these fractions were 18.4 percent in group 2 and 17.6 percent in group 3. The difference between group 1 and the other two groups was highly significant (the *p*-values are below 0.01 in both the comparison between group 1 and group 2 and in the comparison between group 1 and 3).[22] This means that in terms of the initially randomized individuals only 51.1 percent completed the study in the control group whereas 76.7 percent completed the study in group 2 and 78.0 percent completed the study in group 3. From these results it is evident that one important effect of the step contest was that it substantially increased the likelihood of completing the study; i.e. to continue using the pedometer and registering steps. Given that the use of a pedometer per se has been shown to increase physical activity and health in other studies (Bravata et al. 2007), this effect is important in itself.

[21] A possible explanation for the higher dropout rate in the control group between randomization and registration and between registration and registering any steps is that subjects became disappointed when they found out that they would not actually be part of a "real" contest. If the pattern in terms of the number of steps follows the pattern for subjects dropping out later in the study shown below, this effect would go towards underestimating the treatment effects.

[22] This statistical test as most of the other statistical comparisons across groups in the paper were based on an OLS regression with dummy variables for the experimental groups using clustered standard errors at the level of randomization units.

Table 1. Background characteristics.*

	Group 1 (n=498)	Group 2 (n=550)	Group 3 (n=522)	*p*-value of difference**		
				Group 1 vs 2	Group 1 vs 3	Group 2 vs 3
Age (years)	46.29 (10.41)	44.31 (10.89)	43.97 (11.68)	0.130	0.133	0.836
Women (%)	86	89	90	0.434	0.330	0.783
Married/Cohabitant (%)	72	68	77	0.125	0.030	<0.001
Children (%)	73	70	72	0.487	0.888	0.627
Education (years)	14.27 (2.04)	14.31 (1.95)	14.57 (1.88)	0.827	0.062	0.157
Fulltime work (%)	71	77	68	0.279	0.630	0.101
Desk job (%)	33	31	30	0.856	0.768	0.887
BMI	24.36 (4.07)	24.17 (4.44)	24.46 (6.60)	0.547	0.795	0.468
Exercise (hours per week)	2.03 (1.47)	2.13 (1.47)	1.87 (1.41)	0.328	0.145	0.021
Self-assessed physical activity:				0.654	0.042	0.082
Less active than others (%)	26	24	30			
As active as others (%)	41	46	43			
More active than others (%)	33	30	27			
Self-assessed health status:				0.865	0.243	0.355
Poor or fair (%)	16	16	16			
Good (%)	43	43	46			
Very good (%)	31	29	28			
Excellent (%)	11	12	9			

* Standard deviations of continuous variables in parentheses.
** The *p*-values were based on OLS regressions with dummy variables for the three experimental groups using clustered standard errors at the level of randomization units.

We also tested if the probability of dropping out of the study was related to the observed characteristics in the study by running a linear probability model (with clustering on the randomization units as in the other regression models). In the regression we included all the individuals that registered in the study and filled out the entry questionnaire with background information. Only two of the observed characteristics were significantly related to the probability of dropping out at the five percent level; higher age and having a desk job significantly decreased the probability of dropping out of the study.[23]

4.2 Steps

Figure 3 shows the number of steps during the study in each of the three randomized groups using data for all individuals that were still in the study on each day (i.e. the "all subjects" sample). Figure 3 indicates that there are some day-of-the-week effects, with a drop in the number of steps during the weekend.[24] This indicates higher physical activity during weekdays, which is probably due to working per se involving physical activity (both at the workplace and when getting to and from work).[25] This pattern could be expected to vary between occupations.

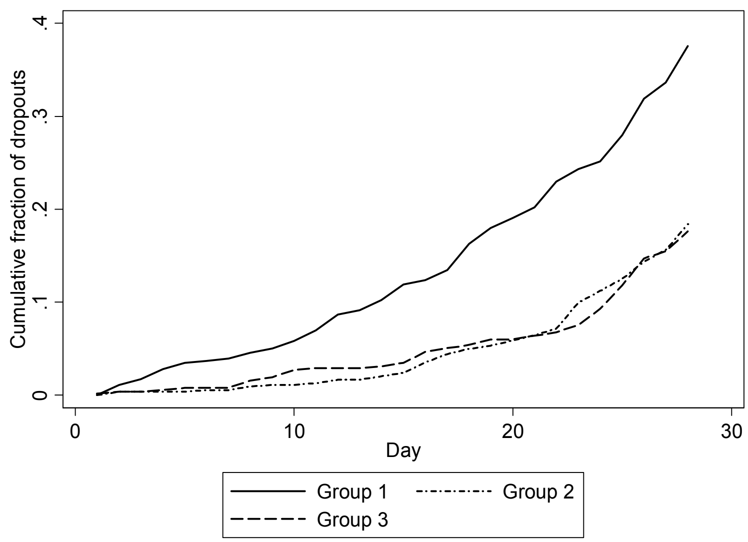

Figure 2. Dropouts during the study in each treatment group (estimated as a fraction of the subjects that registered steps on the first day).

[23] The coefficient of age was –0.0065 and the coefficient of desk job was –0.074.

[24] The day-of-the-week effects introduces some additional noise in the daily step measures, but we have refrained from controlling for these effects in our figures to show the actual pattern of steps across the four weeks of the study. Our estimate of the number of steps per day is based on the average number per day and will average out any day-of-the-week effects.

[25] One alternative explanation is that participants do not report steps during weekends. However, the day-of-the-week effects are highly similar also for those participants that reported positive number of steps during all days of the study.

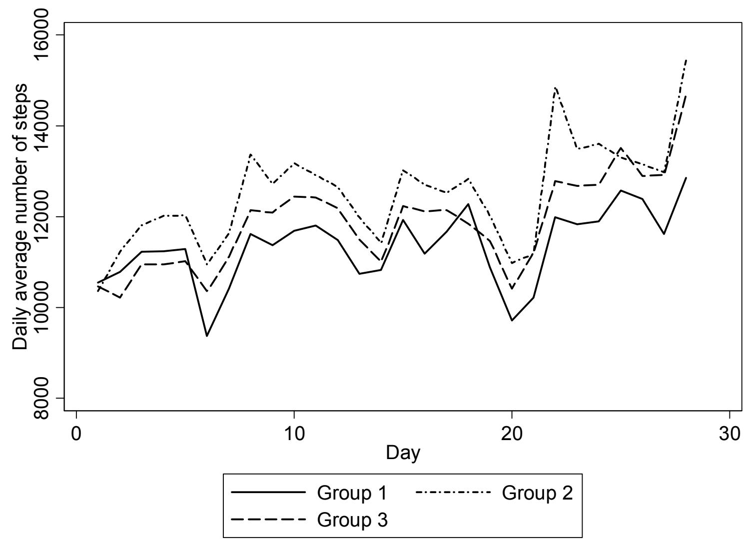

Figure 3. The number of steps in each treatment group during the study. The data includes all subjects that were still in the study each day (i.e. including those that subsequently dropped out).

Table 2. Step per day in the three experimental groups.

	Group 1	Group 2	Group 3	Difference in steps (*p*-value)**		
				Group 2 vs 1	Group 3 vs 1	Group 2 vs 3
All subjects*:						
No control for background variables	11337	12512	11872	1175 (0.008)	536 (0.131)	640 (0.199)
Control for background variables	11337	12244	11980	908 (0.003)	643 (0.027)	265 (0.482)
Subjects that completed the study:						
No control for background characteristics	11649	12701	12062	1052 (0.017)	413 (0.266)	639 (0.196)
Control for background characteristics	11649	12381	12150	731 (0.024)	501 (0.087)	230 (0.543)

* "All subjects" means that the estimate each day was based on all subjects that were still in the study that day.
** The *p*-values were estimated using bootstrap methods for the first two rows, whereas the *p*-values in the last two rows were based on OLS regressions with dummy variables for the three groups using clustered standard errors (at the level of randomization units).

Consistent with the first hypothesis, the number of steps was higher in group 2 than in the control group during the study. The average number of steps per day was 11,337 in the control group and 12,512 in group 2. The increase of 1,175 steps per day was statistically significant at the one percent level; see the results in Table 2. However, we failed to find support for our second hypothesis that a step goal would further increase physical activity. The number of steps in group 3, 11,872, was in between the two other groups and did not differ significantly from either of the two other groups.

In order to control for any differences in background characteristics between the groups we estimated the results controlling for the background characteristics collected in the study (as shown in Table 2 and in Figure 4). This reduced the difference in steps between group 1 and group 2 somewhat to 908 steps, but this difference was still significant at the one percent level. The difference between group 1 and group 3 increased somewhat to 643 steps, whereas the difference between groups 2 and 3 decreased to 265 steps.

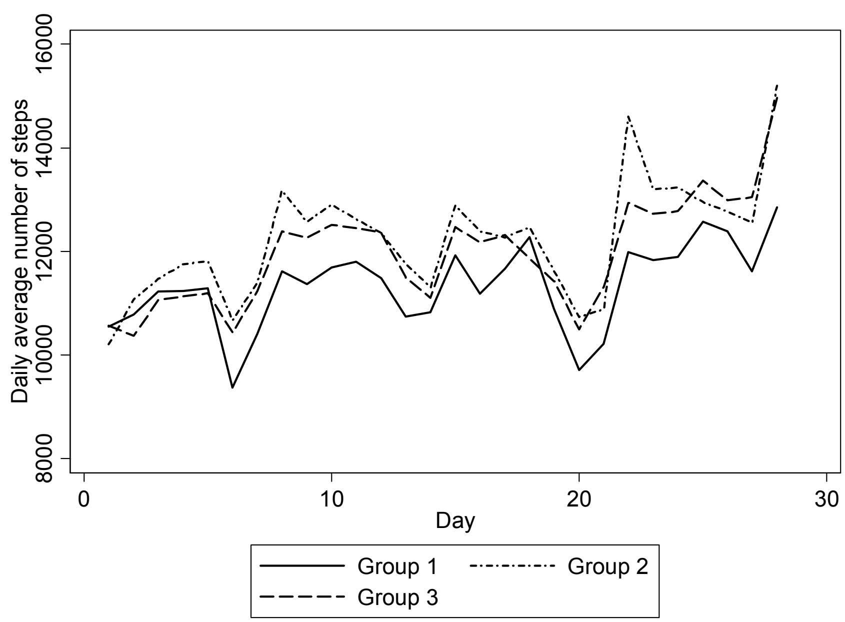

Figure 4. The number of steps in each treatment group during the study controlling for background characteristics. The data includes all subjects that were still in the study each day (i.e. including those that subsequently dropped out).

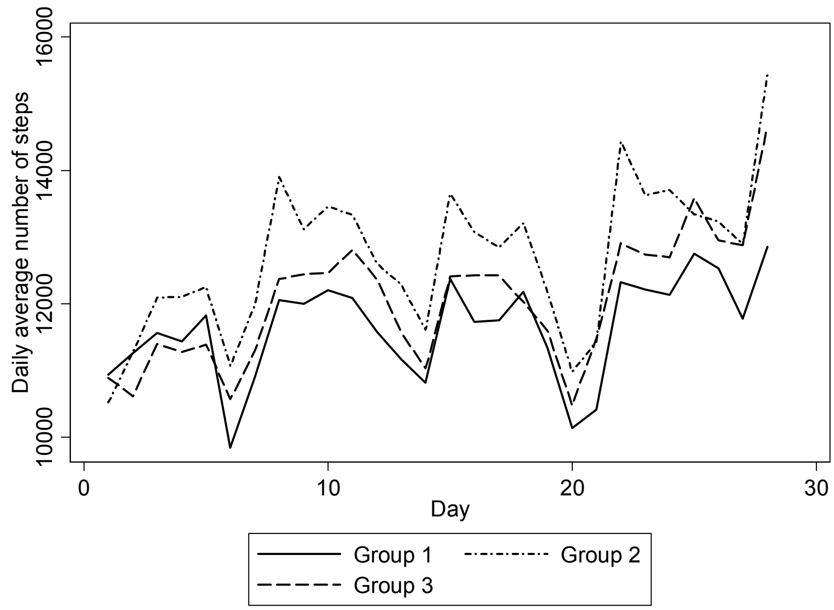

Figure 5. The number of steps in each treatment group during the study. The data includes only subjects that completed the study.

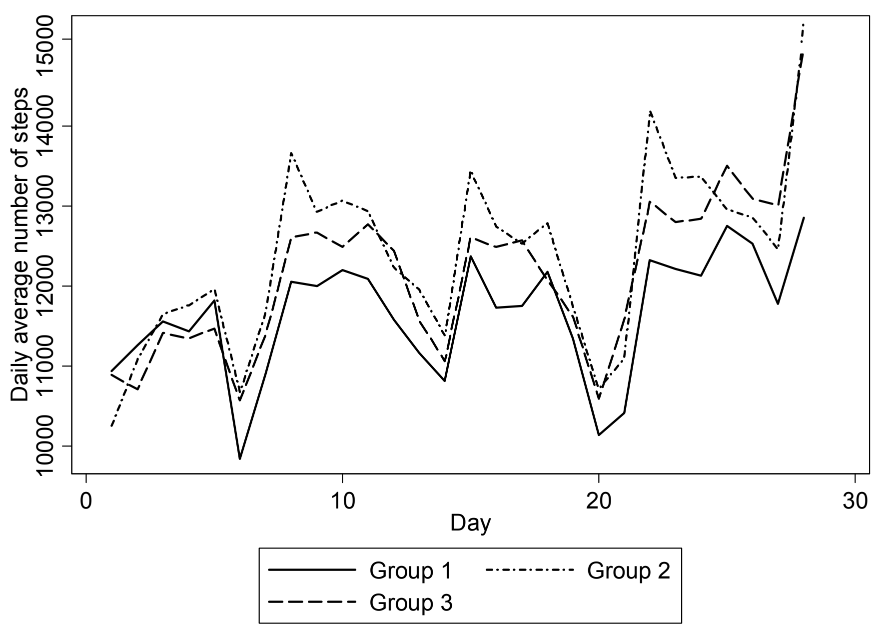

Figure 6. The number of steps in each treatment group during the study controlling for background characteristics. The data includes only subjects that completed the study.

An alternative way of analyzing the data is to only compare the number of steps between the subjects that completed the study. This is done in Figure 5 (without controlling for background factors) and in Figure 6 (controlling for background factors). This led to similar results as in the "all subjects" sample, although the difference between groups 1 and 2 decreased somewhat to 1,052 steps (without controlling for background characteristics) and 731 steps (with a control for background factors). As can be seen from Table 2, the difference was significant at the five percent level in both cases. The differences between the other groups were not significant at the five percent level.

Figure 5 and 6 indicate that there is an upward trend in the number of steps over time. We estimated a linear time trend in all groups. The time trend was significant (at the one percent level) in all groups with a point estimate of 46 steps per day in group 1, 82 steps per day in group 2, and 90 steps in group 3. The time trend was also significantly greater (at the five percent level) in groups 2 and 3 than in group 1, but did not differ significantly between groups 2 and 3. This implies that the difference in the number of daily steps between the two step contest groups and the control group increased over time in the study. A possible explanation for this is that the competition aspect spurs an increasing trend over time in groups 2 and 3 as individuals can observe the number of steps of other participants and teams. That there is a positive time trend also in the control group is consistent with the meta-analysis by Bravata et al. (2007); using a pedometer per se has a sizeable effect on the number of steps. It is likely that this increase occurs gradually over time after starting to use the pedometer.

Table 3 shows the full regression results for completers with all the background characteristics (corresponding to the last row of Table 2). Several background characteristics were significantly related to the number of steps reported during the study. Older and more educated participants took significantly fewer steps, as did participants with desk jobs and a high BMI. Full time workers, healthier and more physically active subjects as well as participants that exercise more tended to take more steps.

Table 3. OLS regression of the effects of background characteristics on the number of steps per day for subjects that completed the study.

Variable	Coefficient	Standard error*	*p*-value
Group 1 (baseline)			
Group 2	731.23	317.30	0.024
Group 3	500.84	288.15	0.087
Age (years)	-37.02	12.78	0.005
Woman	219.12	414.43	0.599
Married/Cohabitant	-441.62	268.80	0.105
Children	198.83	245.63	0.421
Education (years)	-182.26	60.11	0.003
Full time work	510.95	204.04	0.015
Desk job	-1011.08	237.37	0.000
BMI	-34.74	17.71	0.054
Exercise (hours per week)	807.22	89.26	0.000
Self-assessed physical activity:			
Less active than others (baseline)			
As active as others	597.62	290.32	0.043
More active than others	1023.10	343.24	0.004
Self-assessed health status:			
Poor or fair (baseline)			
Good	149.54	275.62	0.589
Very good	663.12	313.21	0.038
Excellent	263.15	377.57	0.488

*The standard errors were estimated with clustering at the level of randomization units.

We also carried out a sub-group analysis to investigate the effect of step contests in different sub-groups. These results are shown in Table 4. In one analysis we divided the subjects into desk workers and non-desk workers. In a second analysis we divided the sample into two BMI groups: above 25 (the definition of being overweight) and below 25. Finally, we divided the groups into different exercise levels: less than two hours per week and more than two hours per week. The effect of the step contest is weaker for desk workers, presumably because it is more difficult for desk workers to influence how many steps they take at work. The step contest seems to have a stronger effect for overweight people, but a weaker effect for people that exercise less than two hours per week. However, none of these differences between sub-groups are close to being statistically significant.[26]

Table 4. Step per day in different sub-groups (estimated for all subjects with controls for background variables).*

	Group 1	Group 2	Group 3	Difference in steps (*p*-value)**		
				Group 2 vs 1	Group 3 vs 1	Group 2 vs 3
Worker type:						
Desk workers	10787	11315	11411	528 (0.110)	623 (0.163)	-96 (0.814)
Non desk workers	11638	12744	12322	1106 (0.008)	684 (0.049)	423 (0.410)
BMI:						
BMI > 25	10356	11729	11123	1373 (0.005)	767 (0.046)	607 (0.263)
BMI ≤ 25	11957	12565	12481	608 (0.051)	523 (0.135)	84 (0.840)
Exercise:						
≤ 2 hour per week	10502	11113	10932	611 (0.114)	429 (0.155)	181 (0.685)
> 2 hour per week	12417	13638	13362	1221 (0.003)	945 (0.017)	276 (0.585)

* ”All subjects” means that the estimate each day was based on all subjects that were still in the study that day.
** The *p*-values were estimated using bootstrap methods.

[26] For each of the three sub-group analyses we compared the three estimates for the treatment effects in the two sub-samples using a standard Z test based on the bootstrapped estimates of the standard deviations. The lowest among the nine different p-values was 0.18.

Table 5. Comparison of steps between dropouts and non-dropouts.*

	Group 1	Group 2	Group 3	*p*-value of difference****		
				Group 1 vs 2	Group 1 vs 3	Group 2 vs 3
					vs 3	
Dropouts vs non-drop-outs**	-738 (0.013)	-1192 (0.039)	-1697 (<0.001)	0.477	0.053	0.464
Dropouts vs completers***	-1020 (0.001)	-1327 (0.022)	-1815 (<0.001)	0.629	0.109	0.482

* *p*-values within parentheses.
** Based on a comparison of the number of steps per day before the dropout day for subjects that dropped out a specific day versus subjects that did not drop out that day (but may have dropped out later in the study).
*** Based on a comparison of the number of steps per day before the dropout day for subjects that dropped out a specific day versus subjects that completed the study.
**** The *p*-values were based on OLS regressions with dummy variables for the three experimental groups using clustered standard errors at the level of randomization units.

The fact that the effect of the step contest decreased somewhat when we used "completers" instead of "all subjects", suggests that individuals that dropped out of the study were less physically active than other individuals. To analyze this issue further we compared the steps between "dropouts" and "non-dropouts". The data is summarized in Table 5. The dropouts on average walked fewer steps than the individuals that continued the study. The dropouts in group 1 walked 738 steps less per day compared to the non-dropouts. For groups 2 and 3, the dropouts walked 1,192 and 1,697 steps less, respectively. If non-dropouts were defined as completers these differences increased somewhat. Dropouts now walked 1,020 steps less per day in group 1 and the corresponding figures in groups 2 and 3 were 1,327 and 1,815, respectively.[27] We also tested whether the difference in the number of steps between dropouts and completers changed over time, but we could not reject the null hypothesis that this difference remained the same over time.[28]

[27] There was no significant difference between the groups in these estimates, but the point estimates goes in the direction of a larger difference between dropouts and non-dropouts in the step contest groups. It is possible that since there are benefits for people to stay in the step contest groups in terms of the potential to win the contest, only those individuals that really hate it will drop out. The benefits of staying in the study for the control group are lower, so the marginal person who drops out may be someone who is more active.

[28] We ran a linear regression on all subjects dropping out testing if the effect (the difference in steps between dropouts and completers) was significantly related to either a linear time trend or a dummy variable for the second half of the study (controlling for the treatment groups). The time

That dropouts walked fewer steps than non-dropouts suggests that we have underestimated the effect of the step contest, as the rate of dropouts was substantially higher in group 1. The results in Table 5 can be used to make a rough adjustment to the step per day results in Table 2 for completers by imputing steps per day for individuals dropping out.[29] The adjustment for dropouts would imply that the difference in steps per day between group 1 and 2 would increase by about 200 steps and the difference between groups 1 and 3 would increase by about 100 steps and the difference between groups 2 and 3 would increase by about 100 steps. The effects on the estimates resulting from adjustment for dropouts would thus be relatively small, but they suggest that our estimates in Table 2 on the effect of the step contest (group 2 vs group 1) are conservative (lower bounds).

There is likely to be measurement error in the number of reported steps and the noise could potentially differ between the three treatment groups. For example, measurement errors could be larger in group 1 because participants in this group did not participate in the contest and therefore might have cared less about reporting accurately. To get an indication of whether measurement errors differed across groups, we calculated the standard deviation and coefficient of variation of the number of daily reported steps for each participant that completed the study. Both the standard deviation and the coefficient of variation is the lowest in group 1, but the only difference between groups that is statistically significant at the 10 percent level is that the standard deviation is lower in group 1 compared to group 2 (the *p*-value is 0.01).[30]

4.3 BMI

In the questionnaire we included questions about height and weight to measure BMI. It should be emphasized that this data is based on self-reported data, which is an imperfect measure of actual BMI (Cawley 2000). In Table 6 we include data on the change in BMI during the study; this data is only for the individuals that

---

variable was not significant in either of these regressions (p = 0.92 with a linear time trend and p = 0.69 with a dummy variable for the second half of the study).

[29] This can be done by imputing steps per day for dropouts based on the difference between dropouts and non-dropouts in Table 5 (i.e. individuals in group 1 that dropped out were assigned 1020 steps less per day during the study than completers in group 1; individuals in group 2 that dropped out were assigned 1327 steps less per day during the study than completers in group 2; and individuals in group 3 that dropped out were assigned 1815 steps less per day during the study than completers in group 3).

[30] To test if the differences were statistically significant we ran OLS regressions with each measure of variation for each individual as the dependent variable and dummy variables for the three groups as explanatory variables. We then tested whether the group dummy estimates differed based on the clustered standard errors (at the level of randomization units).

completed the study and filled out the questionnaire at the end of the study. We also included data for two questions about exercise in Table 6. The first question was about the hours of exercise in a normal week and the second question was about hours of exercise in the preceding week. To measure a change in exercise due to the study, the second question is more appropriate. The question is phrased in terms of the hours of exercise that lead to a high pulse rate and sweating. As we can see from Table 6, there was no significant difference between the groups in the change in BMI or exercise. However, there was a tendency in all groups towards increased exercise and decreased BMI. When outliers were excluded the decrease in BMI during the study was significant in all groups. However, as the BMI measure was based on self-reported data it has to be interpreted very cautiously. To use a self-reported BMI measure may be particularly worrying in this case as people knew that they were part of an effort to get people to be more active and exercise more.

Table 6. Change in BMI and exercise.*

	Group 1	Group 2	Group 3	*p*-value of difference**		
				Group 1 vs 2	Group 1 vs 3	Group 2 vs 3
Change in BMI	-0.022 (0.583)	-0.145 (0.237)	-0.144 (0.012)	0.337	0.077	0.994
Change in BMI (excluding >10 kg changes)	-0.064 (0.015)	-0.096 (<0.001)	-0.104 (0.004)	0.369	0.347	0.853
Change in BMI (excluding >5 kg changes)	-0.063 (0.009)	-0.068 (0.004)	-0.091 (0.016)	0.864	0.518	0.602
Change in exercise, hours per week	0.141 (0.087)	0.181 (<0.001)	0.233 (<0.001)	0.673	0.332	0.451
Change in exercise, hours the last week	0.468 (<0.001)	0.534 (<0.001)	0.540 (<0.001)	0.609	0.533	0.964

* *p*-values within parentheses. All changes were measured as the value at the end of the study minus the value at the beginning of the study and only data for subjects that completed the study were included. One participant reported 900 kg at the second occasion and was dropped in the comparison of BMI.

** These *p*-values were based on OLS regressions with dummy variables for the three experimental groups using clustered standard errors at the level of randomization units.

5. Discussion

Recent work has shown that cash incentives have the potential to increase physical exercise and promote health (Charness and Gneezy 2009). Our study complements those findings by showing that contests with symbolic rewards can also serve as an incentive to exercise. The study was designed to answer two main hypotheses about the effects of step contests. The first was if the competitive aspect per se increased the number of steps and thereby physical activity. We found support for this hypothesis. Our point estimates in our different estimations ranged between an effect of 731 and 1,175 steps depending on whether we controlled for background characteristics or not and depending on whether we incorporated any information about the steps of dropouts. If we included the available information about steps for dropouts the point estimates ranged between 908 and 1,175 steps per day.

The second hypothesis that we tested was whether adding a step goal would further increase the effect of the step contest. We failed to find support for this hypothesis. The point estimate of the number of steps in group 3 (with the 7,000 step per day step goal) was actually lower than in group 2, although the difference was not statistically significant. This is in contrast to the results of the meta-analysis by Bravata et al. (2007) who found that using a 10,000 step per day step goal significantly increased the number of steps. The reason for the lack of an effect of the step goal in our study is possibly that the step goal of 7,000 steps per day was too low; if we pool our data for all our three experimental groups only 6.3 percent of the subjects walked less than 7,000 steps per day on average.[31] At the same time it should be noted that the average baseline number of steps in the meta-analysis by Bravata et al. (2007) was 7,473. It is possible that the low step goal signaled to the participants that once they had reached this modest goal they had exercised sufficiently. Experimental work in psychology and economics suggest that reference points and framing may have important behavioral effects (Tversky and Kahneman 1974, 1981; McNeil et al. 1982; Ariely et al. 2003). The step goal provided a relatively low reference point and may therefore have decreased physical activity. To investigate if there was bunching of steps at or just above 7,000 steps in group 3, we plotted the distributions for the average number of steps during each week of the study (as the step goal was tied to the weekly lotteries), but there was no sign that the step goal caused people to report steps close to 7,000.[32] Our reason for the low step goal was that we wanted to try to

[31] 1,425 of the 1,521 (93.7 percent) individuals who filled in any steps averaged above 7,000 steps and 1,110 of the 1,158 (95.9 percent) individuals that completed the study averaged above 7,000 steps.

[32] The fraction of individuals with 7,000 to 7,500 steps per day in group 3 was 4.69 percent in the first week, 1.88 percent in the second, 2.58 percent in the third and 1.64 percent in the final week

provide an incentive for the most inactive individuals to increase their physical activity. An important lesson from our results is that if a step goal is used it should probably be set at a relatively high level like 10,000 steps per day.

Our estimates have to be interpreted cautiously due to the high rate of dropouts, especially in the control group. As a fraction of the individuals initially randomized to the study, the rate of dropout was almost 50 percent in the control group and about 20 percent in the two treatment groups. In their systematic review, Bravata et al. (2007) reported that nine out of the 26 studies on pedometer use had a zero dropout rate and that the average dropout rate in the remaining 17 studies was 20 percent. Those studies were, however, typically carried out on much more selected populations with small sample sizes (110 participants on average). Our study included a relatively unselected population compared to previous clinical trials (Bravata et al. 2007); out of the total number of employees at the hospital about 40 percent registered for the study. Our sample size of 1,689 is also much larger than the average sample size of 38 in the previous clinical trials (Bravata et al. 2007).

Our analysis of dropouts suggests that individuals that did not complete the study on average walked fewer steps than individuals that did. This suggests that our estimates of the effect of the step contest are lower bounds of the true effect. It should also be borne in mind that when we compared the estimates between dropouts and non-dropouts, it was based on the number of steps using a pedometer. It is likely that those individuals that dropped out of the study stopped using the pedometer and this may in itself have had a negative effect on physical activity (Bravata 2007). One important effect of the step contest was therefore that it greatly increased the likelihood that the individuals would continue in the study and thereby continue using the pedometer.

A limitation of our study is that it is based on the number of self-reported steps, i.e. that individuals honestly report their steps. We cannot directly verify that steps are honestly reported, but we investigated if the number of steps correlated with our measures of BMI, exercise (hours per week), and self-assessed physical activity relative to other individuals.[33] We found a significant correlation for all these measures, suggesting that the number of steps is related to physical activity.[34] The results of the meta-analysis by Bravata et al. (2007) also showed

of the study. The corresponding proportions in group 1 and 2 were slightly lower in the first week, but higher in the following three weeks.

[33] BMI and exercise were measured both at the beginning and at the end of the study, and self-assessed physical activity at the beginning of the study.

[34] The Pearson correlations between the average number of steps for completers and BMI before and after the study were -0.12 and -0.10 ($p < 0.001$). The Pearson correlations between number of steps and numbers of hours exercised where 0.38 (both before and after the study, $p < 0.001$). The Spearman correlation between the number of steps and self-assessed physical activity was 0.30 ($p < 0.001$).

that the 30 percent increase in the number of self-reported steps from using a pedometer was associated with significant health improvements (a reduction in BMI and systolic blood pressure). This suggests that the number of self-reported steps is a valid and reliable measure of exercise. A difference compared to our study is that having a contest with a symbolic reward may introduce an additional incentive to misrepresent the number of steps. However, a growing literature in experimental economics suggest that many individuals have a psychic cost of lying (Ellingsen and Johannesson 2004; Gneezy 2005; Lundquist et al. 2009), which may counteract the incentive to lie about the number of steps.

The intervention in our study, the step contest, only lasted four weeks. This is a brief intervention period for a physical activity intervention and a limitation of our study. It is especially difficult to detect any effects on health measures such as BMI after such a short intervention period. It would thus have been ideal with a longer intervention period. Even if the intervention as such, the step contest, only lasted four weeks it would have been interesting to follow-up the physical activity and the BMI level in the groups after the end of the study. Charness and Gneezy (2009) for instance also studied a physical intervention that lasted for only one month, but continued to follow up the health effects after the end of the intervention period and detected significant improvements in BMI and other health measures four months after the end of the intervention period.

Our estimated effects of the step contest suggest that the step contest led to an increase in about 1,000 steps per day. To interpret the size of this effect it can be compared with the average effect of pedometer use in the meta-analysis by Bravata et al. (2007). They reported a mean increase of 2,491 steps for the 26 studies included in their overview. The effect of the step contest according to our study is thus about 40 percent of the effect of using the pedometer per se. This is a sizeable additional effect on physical activity. The fraction of individuals that completed the study was also about 25 percentage units higher in the step contest groups compared to the control group. This is an important result in itself given that using a pedometer per se appears to be associated with a sizeable increase in physical activity (Bravata et al. 2007). Taken together these results suggest that step contests can potentially achieve important improvements in physical activity and health and be a useful public health tool. The potential for this as a desirable public health tool is furthered by the low cost per participant (≈$50). However, as in all studies using pedometers, it is unclear what the long term effects are on physical activity and health and this is important to investigate further.

References

Alevizos A., Lentzas J., Kokkoris S., Mariolis A., Korantzopoulos P., 2005, "Physical activity and stroke risk", *International Journal of Clinical Practice* 59, 922–930.

Andreoni J., Bernheim B. D., 2009, "Social image and the 50-50 norm: a theoretical and experimental analysis of audience effects", *Econometrica* 77, 1607–1636.

Ariely D., Bracha A., Meier S., 2009, "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially", *American Economic Review* 99, 544–555.

Ariely D., Loewenstein G., Prelec D., 2003, "'Coherent arbitrariness': stable demand curves without stable preferences", *Quarterly Journal of Economics* 118, 73–105.

Bénabou R., Tirole J., 2006, "Incentives and prosocial behavior", *American Economic Review* 96, 1652–1678.

Besley T., Ghatak M., 2008, "Status incentives", *American Economic Review Papers and Proceedings* 98, 206–211.

Booth A. L., Nolen P., 2009, "Choosing to compete: how different are girls and boys?". University of Essex, Discussion Paper Series, No. 673.

Branca F., Nikogosian H., Lobstein T. (eds.), 2007, *The challenge of obesity in the WHO European Region and the strategies for response*. Copenhagen: WHO Regional Office for Europe.

Bravata D. M., Smith-Spangler C., Sundaram V., Gienger A. L., Lin N., Lewis R., Stave C. D., Olkin I., Sirard J. R., 2007, "Using pedometers to increase physical activity and improve health", *Journal of the American Medical Association* 298, 2296–2304.

Brown D. R., 1992, "Physical activity, ageing, and psychological well-being: an overview of the research", *Canadian Journal of Sport Science* 17, 185–193.

Cawley J., 2000, "Body weight and women's labor market outcomes", *NBER Working Paper* 7841.

Cawley J., Price J. A., 2009, "Outcomes in a program that offers financial rewards for weight loss", *NBER Working Paper* 14987.

Charness G., Gneezy U., 2009, "Incentives to exercise", *Econometrica* 77, 909–931.

Colditz G. A., 1999, "Economic costs of obesity and inactivity", *Medicine & Science in Sports & Exercise* 31, S663–S667.

Cox E.R., Kortt M.A., Langley P.C., 1998, "A review of cost-of-illness studies on obesity", *Clinical Therapeutics* 20, 772–779.

Cutler D. M., Glaeser E. L., Shapiro J. M., 2003, "Why have Americans become more obese?", *Journal of Economic Perspectives* 17, 93–118.

Davidson R., MacKinnon J. G., 2000, "Bootstrap tests: how many bootstraps?", *Econometric Reviews* 19, 55–68.

Dreber A., von Essen E., Ranehill E., 2010, "Outrunning the gender gap: boys and girls compete equally", *SSE/EFI Working Paper Series in Economics and Finance* 709.

Efron B., Tibshirani R. J., 1993, "An introduction to the Bootstrap", *Monographs on Statistics and Applied Probability* 57. New York: Chapman and Hall.

Ellingsen T., Johannesson M., 2004. "Promises, threats and fairness", *Economic Journal* 114, 397–420.

Ellingsen T., Johannesson M., 2008, "Pride and prejudice: the human side of incentive theory", *American Economic Review* 98, 990–1008.

Fiebelkorn I. C., Finkelstein E. A., Wang G., 2003, "National medical spending attributable to overweight and obesity: how much, and who's paying?", *Health Affairs* W3, 219–226.

Finkelstein E. A., Linnan L. A., Tate D. F., Birken B. E., 2007, "A pilot study testing the effect of different levels of financial incentives on weight loss among overweight employees", *Journal of Occupational and Environmental Medicine* 49, 981–989.

Frank R. H., 1985, *Choosing the right pond: human behavior and the quest for status*. New York: Oxford University Press.

Frey B. S., 2006, “Giving and receiving awards”, *Perspectives on Psychological Science* 1, 377–388.

Frey B. S., 2007, “Awards as compensation”, *European Management Review* 4, 6–14.

Gneezy U., 2005, “Deception: the role of consequences”, *American Economic Review* 95, 384–394.

Gneezy U., Niederle M., Rustichini A., 2003, “Performance in competitive environments: gender differences”, *Quarterly Journal of Economics* 118, 1049–1074.

Gneezy U., Rustichini A., 2004, “Gender and competition at a young age”, *American Economic Review Papers and Proceedings* 94, 377–381.

Guh D. P., Zhang W., Bansback N., Amarsi Z., Birmingham C. L., Anis A. H., 2009, “The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis”, *BMC Public Health* 9:88.

Harrison G. W., List J. A., 2004, “Field experiments”, *Journal of Economic Literature* 42, 1009–1055.

Hirsch F., 1976, *The social limits to growth*. London: Routledge & Kegan Paul.

Keeler E. B., Manning W. B., Newhouse J. P., Sloss E. M., Wasserman J., 1989, “The external costs of a sedentary life-style”, *American Journal of Public Health* 79, 975–981.

Kosfeld, M., Neckermann, S., (2010), “Getting more work for nothing? Symbolic awards and worker performance”, *IZA Discussion Paper* 5040.

Lacetera N., Macis M., 2008, “Social image concerns and pro-social behavior”, *IZA Discussion Paper* 3771.

Lee C., Blair S., Jackson A., 1999, “Cardiorespiratory fitness, body composition, and all-cause and cardiovascular disease mortality in men”, *American Journal of Clinical Nutrition* 69, 373–380.

Lundquist T., Ellingsen T., Gribbe E., Johannesson M., 2009, “The aversion to lying”, *Journal of Economic Behavior & Organization* 70, 81–92.

Miller T. D., Balady G. J., Fletcher G. F., 1997, "Exercise and its role in the prevention and rehabilitation of cardiovascular disease", *Annals of Behavioral Medicine* 19, 220–229.

McNeil B., Pauker S. G., Sox H. Jr., Tversky A., 1982, "On the elicitation of preferences for alternative therapies", *New England Journal of Medicine* 306, 1259–1262.

Neckerman S., Frey B. S., 2007, "Awards as incentives". Institute for Empirical Research in Economics, University of Zurich, Working Paper No. 334.

Niederle M., Segal C., Vesterlund, L., 2010, "How costly is diversity? Affirmative action in light of gender differences in competitiveness". Mimeo, Stanford University.

Niederle M., Vesterlund L., 2007, "Do women shy away from competition? Do men compete too much?", *Quarterly Journal of Economics* 122, 1057–1101.

Speck B. J., Looney S. W., 2001, "Effects of a minimal intervention to increase physical activity in women: daily activity records", *Nursing Research* 50, 374–378.

Stajkovic A. D., Luthans F., 2003, "Behavioral management and task performance in organizations: conceptual background, meta-analysis, and test of alternative models", *Personnel Psychology* 56, 155–194.

Sutter M., Rutzler D., 2010, "Gender differences in competition emerge early in life". University of Innsbruck, Working Papers in Economics and Statistics, 2010–14.

Tudor-Locke C., Williams J. E., Reis J. P., Pluto D., 2002, "Utility of pedometers for assessing physical activity: convergent validity", *Sports Medicine* 32, 795–808.

Tversky A., Kahneman D., 1974, "Judgment under uncertainty: heuristics and biases", *Science* 185, 1124–1131.

Tversky A., Kahneman D., 1981, "The framing of decisions and the psychology of choice", *Science* 211, 453–458.

Volpp K. G., John J. K., Troxel A. B., Norton L., Fassbender J., Loewenstein G., 2008, "Financial incentive-based approaches for weight loss: A randomized trial", *Journal of the American Medical Association* 300, 2631–2637.

Volpp K. G., Troxel A. B., Pauly M. V., 2009, "A randomized, controlled trial of financial incentives for smoking cessation", *New England Journal of Medicine* 360, 699–709.

Whitlock G. et al., 2009, "Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies", *Lancet* 373, 1083–1096.

Wooldridge J. M., 2003, "Cluster-sample methods in applied econometrics", *American Economic Review Papers and Proceedings* 93, 133–138.