Effect of Different Financial Incentive Structures for Promoting Physical Activity Among Adults: A Randomized Clinical Trial

Chethan Bachireddy, MD, MSc[1,2]
Andrew Joung, BA[3]
Leslie K. John, PhD[4]
Francesca Gino, PhD[4]
Bradford Tuckfield, PhD[3]
Luca Foschini, PhD[5]
Katherine L. Milkman, PhD[3]

[1] Virginia Commonwealth University School of Medicine, Richmond, VA, USA

[2] Leonard Davis Institute Center for Health Incentives and Behavioral Economics, University of Pennsylvania, Philadelphia, PA, USA

[3] Operations, Information and Decisions Department, The Wharton School of the University of Pennsylvania, Philadelphia, PA, USA

[4] Negotiation, Organizations and Markets Unit, Harvard Business School, Boston, MA, USA

[5] Evidation Health, Inc. Santa Barbara, CA, USA

Manuscript word count: 3860
Abstract word count: 341

Keywords: Physical Activity, Habit Formation, Behavioral Economics, Behavior Change, Incentives, Technology

Corresponding Author: Chethan Bachireddy

Corresponding Author's Email: chet86@gmail.com

Corresponding Author's Phone Number: 936-635-1559

Corresponding Author's Mailing Address: 1201 East Marshall Street, Richmond, VA, 23298

Date of Revision: 06/28/2019

KEY POINTS

Question

Is it more effective to disburse fixed total incentives at a constant, increasing or decreasing rate to encourage physical activity?

Findings

In this randomized controlled trial, financial incentives for physical activity were significantly more effective at motivating activity during and immediately following a payment period if they were offered at a constant rate rather than an increasing or decreasing rate.

Meaning

This finding has implications for the design and delivery of incentive programs to promote healthy behaviors.

ABSTRACT

Importance

Few adults engage in recommended levels of physical activity. Financial incentives can promote physical activity, but little is known about how their structure influences their effectiveness; for example, whether incentives are more effective if they are disbursed at a constant rate, versus increasing or decreasing rates.

Objective

To determine if it is more effective to disburse fixed total incentives at a constant, increasing or decreasing rate to encourage physical activity.

Design

Two-week, four-arm randomized controlled trial from June 2, 2014 – June 15, 2014. Data analyses finalized in 2018.

Setting

An online platform that automatically records daily steps of pedometer-wearing users and awards points redeemable for cash.

Participants

3,515 users of the online platform in the lower 70th percentile of steps taken among all users pre-treatment.

Intervention

Participants were randomized to either a control group or to one of three intervention groups over two weeks. Control participants received a constant daily rate of $0.00001/step. The three intervention groups received a 20-fold incentive increase ($0.00020/step) distributed differently over two weeks—at a constant, increasing, or decreasing rate. Reminder emails explaining incentive schedules were sent the day before the intervention and half-way through the two-week intervention.

Main Outcome and Measure

Change in mean daily steps during the two-week intervention and three weeks post-intervention. The study had 80% power to detect a difference of 280 steps/day during the intervention at $\alpha=0.05$.

Results
During the intervention, compared to control, constant incentives generated 306.7 more steps/day (95% CI [91.5,521.9]; p=0.005), decreasing incentives generated 96.9 more steps/day (95% CI [15.3,178.5]; p=0.020), and increasing incentives generated no change (1.5; 95% CI [-81.6,84.7]; p=0.971). One week post-intervention, compared to control, only constant incentives generated significantly more steps/day (329.5; 95% CI [20.6,638.4]; p=0.037). Two and three weeks post-intervention, there were no significant differences compared to control. Overall, for each dollar spent, constant incentives generated 475.5 more steps than increasing incentives and 429.4 more steps than decreasing incentives.

Conclusions and Relevance
Financial incentives for physical activity are more effective during a payment period if offered at a constant rather than an increasing or decreasing rate. However, this effectiveness dissipates shortly after the incentives are removed.

Trial Registration
Clinicaltrials.gov identifier: NCT02154256
https://clinicaltrials.gov/ct2/show/NCT02154256

INTRODUCTION

Physical inactivity has been implicated as a major risk factor for disability and death globally, on par with obesity and smoking, yet receives considerably less attention[1,2]. Inactivity accounts for 9% of premature mortality[2]. In the United States, inactive individuals older than 50 years of age would gain 1.3-3.7 years of life expectancy if they became active[3]. Activity alone can reduce the risk of developing diabetes, cardiovascular disease, colon and breast cancers, and improve bone and mental health; however, less than half of adults in the US engage in recommended levels of physical activity[4]. The benefits of activity and the costs of inactivity are clear, but motivating individuals to increase their activity is challenging.

Financial rewards are a useful tool for encouraging healthy behaviors including smoking cessation, eating nutritious foods, physical activity, and weight loss[5–14]. At least 15 state Medicaid programs and over 1/4 of large employers offer financial incentive-based health and wellness programs[15,16].

Recent literature suggests that principles from behavioral economics can be effectively harnessed to design and deliver incentives capable of changing health behaviors like physical activity. We understand that rewards conditioned on performing specific behaviors are more likely to be successful in promoting exercise, for instance, compared to unconditional rewards encouraging behavior change (such as free gym memberships) [17].

In a recent systematic review, Mitchell *et al* identify key principles of effective incentive design to promote physical activity including immediate incentives, realistic daily goals, and longer interventions (>24 weeks) among less active adults[18]. To have impact, the size of the incentive does not have to be large (~$1/day), and its effect can be multiplied through frequent and personalized feedback.

Still, much remains unknown. First, while we understand that small daily incentives over time can be effective, few studies have sought to assess how those small daily incentives should be distributed over time. A critical question is therefore how to disburse incentives for maximal impact. This is the primary research question of the present research.

Second, while financial incentives can encourage healthy behaviors including exercise, it is unclear how to create behavior change that is sustained after incentives are inevitably removed. Among the studies demonstrating the benefits of financial incentives, few have measured post-intervention behavior, and fewer still have demonstrated evidence of behavior change lasting beyond the time period when incentives were offered[7–9, 17, 18].

While maintaining the same financial incentive over time has the benefit of simplicity, it may not be the best way to foster sustained behavior change. Starting with a small incentive and increasing it over time may help individuals gradually build a habit by preventing the development of tolerance to a specific incentive value - just as patients may develop tolerance to medications and require an increased dosage to maintain the same effect [19]. Yet, starting with a large incentive may help motivate individuals to overcome inertia and initiate a new routine[20,21]. Gradually reducing incentives over time from an initially high level may then help diminish individuals' reliance on financial rewards for motivation to exercise, making it easier to transition to un-incentivized engagement[22–24].

Our primary objective was to compare three different two-week incentive programs with rewards for daily steps taken to determine which was the most effective for increasing steps both during and post-intervention. We build upon the existing literature and leverage an online platform with points-based daily financial incentives enhanced with frequent, personalized feedback to conduct this study. Each of the three programs offer the same total incentives but distributed differently: increasing, decreasing, or constant over time. In a four-arm randomized controlled trial, we compare the effectiveness of these incentive programs versus a control group.

METHODS

Study Design

We conducted a field experiment with Achievement (formerly called AchieveMint), an online platform (www.myachievement.com) owned by Evidation Health (the employer of study author LF) that automatically records the daily steps of users who wear pedometers and awards them points redeemable for cash. One step earns $0.00001 (i.e., 10,000 steps = $0.10). We tested whether and by how much offering incentives to users that are twenty times as large as usual over two consecutive weeks changes the steps taken during and after the intervention compared to a control group.

This study was approved and a waiver of informed consent was granted by the University of Pennsylvania Institutional Review Board. The study was registered with ClinicalTrials.gov (NCT02154256).[1]

Setting and Participants

Participants were users of the online platform who logged steps using a pedometer at least once between May 9, 2014 and May 22, 2014 (the date participant selection occurred). At the time of the study, the online platform did not routinely collect demographic information on its users; therefore we do not have demographic data on study participants.

To maximize the health impact of our intervention, we excluded the most active users and conducted our study among users whose logged steps were in the bottom 70th percentile of all users between May 9, 2014 and May 22, 2014. We calculated that a sample of 3,515 participants would allow us to detect a difference of 280 steps per day at α=0.05 with 80% power.

Based on the resources available and the study power calculations, we decided that a two week intervention would be sufficient to answer the study's key question regarding which incentive structure would most effectively promote physical activity.

Randomization and Interventions

Participants were stratified by one of nine pedometer brands (ActiveBeats, BodyMedia FIT, FitBug, Fitbit, Jawbone UP, MapMyWalk, Misfit Wearables, Moves, and Withings) and randomly assigned to one of four experimental conditions as outlined in **Figures 1 and 2**: a control condition (in which participants received incentives as usual: $0.00001 per step (i.e. $0.10 per

---

[1] Our clinical trial pre-registration was regrettably vague. The goal of the project was always to look at user step counts (as the goal users were trying to achieve was taking as many steps as possible per day), and because the intervention only lasted for 2 weeks and the effects dissipated after another week, we determined that a lengthier follow-up analysis than the one we present would not add value.

10,000 steps) or one of three treatment conditions. In the three treatment conditions, participants were offered an average of 20x their usual points per step (i.e. $2.00 per 10,000 steps) during the two-week intervention period. Thus, comparing the control condition to these three treatment conditions enabled us to test the effect of a 20x incentive increase on walking behavior. However, as described next, comparing the three treatment conditions enabled us to test the effect of incentive structure, our primary interest.

In the constant incentive condition, participants were offered $0.00020 per step every day. In the increasing incentive condition, they were initially offered $0.00005 per step (i.e. $0.50 per 10,000 steps); this increased by $0.00005 per step every two days up to a maximum of $0.00035 per step (i.e. $3.50 per 10,000 steps) on the last two days. In the decreasing incentive condition, participants were initially offered $0.00035 per step; this decreased by $0.00005 per step every two days down to a minimum of $0.00005 per step on the last two days. The schedule of incentives is detailed further in **Supplement eTable 1**.

Routinely, users of the online platform receive an update email on Sunday reflecting their weekly earnings in points and dollars. During the two-week intervention, these Sunday emails continued to be sent. Study-specific emails were also sent. The day before the intervention began (on Sunday, June 1, 2014), all study participants received an email describing the program designed to help them increase their physical activity (**Supplement eFigures 1-4**). In the treatment arms, participants received a precise schedule detailing the incentives they would receive for each step taken on each day over the subsequent two weeks. All study participants who wore pedometers on a given day and "synced" their pedometers with the online platform within seven days were recorded in our dataset, and others were recorded as missing observations, allowing for analyses accounting for missing observations in a variety of ways. On day seven of the fourteen-day intervention, all study participants received a reminder email encouraging them to be physically active (**Supplement eFigures 5-6**). The reminder email for treatment participants also included their specific incentive rate.

Outcomes

We report individual daily step counts for a period of eight weeks total: three weeks pre-intervention, two weeks during the two-week long intervention, and three weeks post-intervention. The primary outcome measure was change in daily steps taken, which was collected remotely through participants' pedometers. The intervention began on June 2, 2014 and concluded on June 15, 2014. Initial data analysis occurred in 2014. Based on peer feedback, a sensitivity analysis was conducted and all analyses finalized in 2018.

Statistical Analysis

Prior studies have demonstrated that pedometer daily step counts lower than 2,000 are unlikely to be reflective of true daily step count values; we define a missing data day as any day with fewer than 2,000 recorded steps[25]. To address the possibility that some participants walked without pedometers, we present all analyses in two different ways (the results of which converge on the same conclusion):

In our primary analysis, we use an intent-to-treat strategy in which we replace missing data with an average of a given participant's pre-intervention daily step counts greater than 2,000 steps, stratified by day of week to account for person-within-week differences in physical activity (i.e., a participant may routinely get more physical activity on Saturdays compared to Wednesdays). To further minimize the potential for bias, we conduct a sensitivity analysis in which we delete all daily step data recording fewer than 2,000 steps – an approach that would bias towards a null effect.

We use ordinary least squares regression to predict the overall and separate effects of our three treatment arms (*constant, increasing, and decreasing*) on participants' daily steps. We include person-by-day-of-week fixed effects and cluster standard errors by person-by-day-of-week to control for individual differences in steps and further for differences in participant routines that vary by day of week; these fixed effects also capture condition assignment. In addition, we include fixed effects by pedometer brand and for each day of the year to account for seasonal conditions that may influence step count. We employ Wald tests to assess differences between treatment conditions and conduct a cost-effectiveness analysis of additional steps taken per $1 paid to each treatment condition participant relative to control participant.

RESULTS

The sample (N = 3,515) was distributed randomly among the control (n=879), constant (n=879), increasing (n=881), and decreasing (n=876) conditions. In the three weeks pre-intervention, mean daily steps across all study participants was 6,804.48 (SD = 3,506.91). Pre-intervention, each day on average 83.4% of participants in the control, 82.8% in the increasing, 84.2% in the decreasing, and 84.1% in the constant incentive condition used their pedometers. During the intervention, each day on average 76.4% of participants in the control, 79.0% in the increasing, 79.1% in the decreasing, and 79.7% in the constant incentive condition used their pedometers. While there were significant differences in pedometer adherence between conditions in the pre-intervention and during intervention periods, the differences were small in magnitude. To address this difference and the possibility that some participants walked without pedometers, we employ an intent-to-treat approach in which we replace missing data as described above.

Effect of 20x increase in incentives

Participants collapsed across the three treatment arms took an estimated 135.0 additional daily steps relative to control during the intervention period (95% CI [41.0, 228.9]; p=0.005). In the three weeks following the intervention, there were no significant differences.

Figure 3 shows the unadjusted differences in mean steps taken by treatment participants compared to control participants for 3 weeks before, 2 weeks during, and 3 weeks after the intervention. Treatment participants experienced an increase in physical activity mid-way through the intervention (when the regular Sunday earnings update email was sent), and the increase was particularly large for those in the constant condition. In the more conservative sensitivity analysis in which we delete all step count data less than 2,000, we find qualitatively similar results.

Effect of incentive structure

During the intervention. **Table 1** presents the results of regressions and Wald tests comparing the effectiveness of each treatment arm relative to control and to each other treatment arm during the two-week intervention. Participants in the constant condition logged 306.7 additional daily steps (95% CI [91.5, 521.9]; p=0.005) relative to those in the control condition, 305.1 additional daily steps (95% CI [89, 521.2]; p=0.006) relative to those in the increasing condition, and 209.8 additional daily steps (95% CI [-5.7, 425.3]; p=0.056) relative to those in the decreasing condition. Participants in the decreasing condition demonstrated a small increase in daily steps relative to those in the control condition (96.9 additional daily steps; 95% CI [15.3, 178.5]; p=0.020) and relative to those in the increasing condition (95.3 additional daily steps; 95% CI [11.3, 179.3]; p=0.026). Participants in the increasing condition did not log significantly more steps than those in the control condition (1.5; 95% CI [-81.6, 84.7]; p=0.971). In the sensitivity analysis, we find similar results except there is no longer a statistically significant effect of decreasing incentives compared to control during the intervention period (80.5; 95% CI [-38.5, 199.4]; p=0.185).

After the intervention. **Table 2** presents the effectiveness of each treatment arm in the three weeks after the intervention. In the first week post-intervention, participants in the constant condition took 329.5 more daily steps (95% CI [20.6, 638.4]; p=0.037) than those in the control condition, 397.8 more daily steps (95% CI [89.2, 706.4]; p=0.012) than those in the increasing condition and 308.6 more daily steps (95% CI [0.1, 617.1]; p=0.050) than those in the decreasing condition. There were no significant differences between the increasing and decreasing conditions and the control condition (-68.3; 95% CI [-174.6, 38.1]; p=0.208 and 21.0; CI [-84.9, 126.8]; p=0.698, respectively).

In the second week post-intervention, participants in the constant condition logged significantly more daily steps than increasing condition participants (315.2 additional daily steps; 95% CI [6, 624.4]; p=0.046). Constant condition participants also logged more daily steps than those in the control and decreasing conditions, but these differences were not significant (213.5 additional daily steps; 95% CI [-94.8, 521.8]; p=0.175 and 297.1 additional daily steps; 95% CI [-10.9, 605.1]; p=0.059, respectively). There were no significant differences between the increasing and decreasing conditions and the control condition (-101.7; 95% CI [-209.2, 5.8]; p=0.064 and -83.6; 95% CI [-187.7, 20.6]; p=0.116, respectively)

In the third week post-intervention, there were no significant differences in steps taken between the constant condition and the increasing (53.6; 95% CI [-100.5, 207.7]; p=0.773), decreasing (-82.7; 95% CI [-233.8, 68.4]; p=0.177), or control conditions (-22.8; 95% CI [-177.3, 131.8]; p=0.271). There was, however, a significant increase of 136.3 daily steps in the decreasing condition compared to the increasing condition (95% CI [30.3, 242.3]; p=0.012).

In the sensitivity analysis, we find similar results except one week post-intervention there is no longer a statistically significant effect of constant incentives compared to control (485.4; 95% CI [-20.1, 990.9]; p=0.060)). Constant incentives demonstrate a sustained effect one-week post-intervention compared to the increasing (607.4; 95% CI [103.7, 1111.1]; p=0.013) and decreasing incentive conditions (515.4; 95% CI [12, 1018.8]; p=0.042). There is no statistically significant difference in steps one-week post-intervention between the constant and control conditions (485.4; 95% CI [-20.1, 990.0]; p=0.060). Two weeks post-intervention, there is a statistically significant effect of increasing incentives compared to control (-183.4; 95% CI [-351.5, -15.3]; p=0.033), decreasing incentives compared to control (-212.9; 95% CI [-376.2, -49.6]; p=0.011), and decreasing incentives compared to constant incentives (-532.0; 95% CI [-1029.4, -34.6]; p=0.036); and three weeks post-intervention, there is no longer a statistically significant effect of increasing incentives compared to decreasing incentives (-138.2; 95% CI [-309.3, 32.9]; p=0.113).

Cost-effectiveness

During the intervention, participants in the constant condition were paid an average of $15.48 per person compared to an average of $14.54 per person in the increasing condition, and an average of $14.67 per person in the decreasing condition.

Compared to control and including post-intervention effects, for each additional $1 paid, there were 582.4 additional steps per participant in the constant condition, 107.0 additional steps per participant in the increasing condition, and 153.1 additional steps per participant in the decreasing condition.

DISCUSSION

To our knowledge, this is one of the largest randomized controlled trials of financial incentives for physical activity. We tested the short-term impact of different incentive structures on physical activity. Incentive structure affected physical activity during the 2-week intervention: the constant incentives significantly increased physical activity relative to all other conditions – control, increasing incentives, and decreasing incentives. These effects held for one week after the incentives had been removed. These effects dissipated two to three weeks post-intervention. Indeed, similar to prior studies, after the withdrawal of incentives, physical activity tapered in all conditions[14,26].

In sum, we conclude that incentive structure – independent from incentive *size*, which was the same across our treatment groups – affects physical activity at least during the period when incentives are offered. Thus, in designing wellness programs, incentive-designers and policy-makers should consider not simply the *magnitude* of incentives, but also their structure[6,11].

It is important to recognize that the control effectively received an incentive at a constant rate, just 20-fold lower than the treatment conditions. Interestingly, the constant rate structure was so effective that during the intervention, the control performed equally as well as a 20-fold greater incentive delivered at an increasing rate and only marginally worse than a 20-fold greater incentive delivered at a decreasing rate.

Our results on the comparative effectiveness of constant versus decreasing incentives are consistent with findings from a working paper by Carrera *et al* directly comparing the impact of a constant incentive and a decreasing incentive on gym initiation and attendance over eight weeks among employees of a Fortune 500 company[26]. They find that among non-gym members the constant and decreasing incentives were equally effective in increasing gym join rates. However, among existing gym members, the constant incentive was significantly more effective than the decreasing incentive in motivating physical activity during and after the intervention. They complement a host of recent studies exploring different payment disbursement schemes for motivating physical activity[8–11,26–28].

Only a handful of studies providing financial incentives for exercise and physical activity have measured and demonstrated behavior change post-intervention[8–10,26,29]. These studies differ from the current study in a number of ways – almost all incentivized and measured gym attendance rather than step count, lasted four weeks or longer, provided an incentive with a daily expected value more than twice that of the current study ($1.40), and recruited samples fewer than 1000 participants[8–10,26,29].

Our findings raise the question of why incentives delivered at a constant rate were more effective than other disbursement strategies. One potential explanation is that the constant incentive was easier to remember and therefore more salient and impactful in promoting physical activity[30]. By contrast in the other disbursement strategies, getting paid different amounts for doing the same thing may have been confusing, or even felt unfair, potentially contributing to their relative ineffectiveness[31]. Further research exploring these and other possibilities would be valuable.

Prior work suggests a differential and often lower impact of financial incentives among those with existing exercise habits[9,26]. Users of the online platform we studied have higher daily step counts than the average US adult, which is why for our study we sampled from users in the bottom 70% of physical activity. As a result, our study findings reflect a population with similar baseline physical activity as the US population[32]. However, we cannot say as much about how our incentive conditions might impact those who are on the extremes of physical activity, including those who are sedentary.

Our study has several limitations. First we were dependent on participants' device-wearing behaviors. We could not detect steps if a participant did not wear the pedometer, resulting in missing data. Missing data is a common challenge when conducting experimental research in real-world settings. Prior studies have dealt with missing or partially recorded step data by excluding or replacing the data with a uniform step number. These approaches have their own shortcomings because deleting the data biases towards a null effect and replacing missing data with zeros biases towards finding an effect because of better observability in treatment groups (who are more incentivized to wear pedometers). Instead, as described in the methods section, we took a more conservative approach, replacing missing data with the average of pre-intervention steps greater than 2,000 and employed an intent-to-treat analytic strategy. This approach has a slight bias towards a null effect but is more balanced than prior approaches to the common occurrence of missing step data. Furthermore, all analyses are presented using an even more conservative approach of deleting all step data below a certain threshold, consistent with prior research[28].

Second, pedometers restricted us to step count, even though other metrics such as metabolic equivalents or minutes of moderate-vigorous physical activity might be more relevant to long-term health outcomes.

Third, despite randomization, pre-intervention mean daily steps were significantly higher in the increasing compared to the decreasing conditions. We attempted to minimize this bias through a focus on change in mean daily steps and inclusion of fixed effects to account for time-invariant differences among participants. Importantly, this limitation does not apply to comparisons with the constant condition since there were no significant differences in pre-intervention mean daily steps between the constant and increasing, decreasing, or control conditions.

Fourth, we do not have demographic data for the population which may have revealed insights and further strengthened our regression analyses. We attempt to address this limitation through an advanced analytic approach which includes fixed effects by person-day-of-week, pedometer, and day-of-year and clustered standard errors by person-day-of-week.

Fifth, compared to prior experiments on incentives for health behaviors, our intervention time period of two weeks was relatively short and our incentive relatively small. On the other hand, it is noteworthy that the incentives, in particular the constant ones, had an effect despite their size. Incentives delivered over a longer period of time may lead to greater behavior change during and after an intervention[7–10,26].

Sixth, the study was not well-powered to detect step differences long after the intervention. Nonetheless, we see significantly greater steps in the constant condition compared to the increasing condition in the first week post-intervention. While this experiment was designed to assess which incentive condition produced the most physical activity during and briefly after our intervention, it cannot answer another important and broader question, which is what incentive structure is optimal to promote long-term changes in physical activity.

In summary, to the best of our knowledge, this is one of the largest randomized controlled trials of financial incentives for physical activity. For the same possible earnings, daily incentives of constant value delivered over two weeks were more effective in promoting physical activity compared to incentives of increasing or decreasing value. These findings have implications for the psychology of behavior change and suggest that incentive *structure* should be a key design consideration in the delivery of health incentive programs. Future research should continue to explore strategies to improve health through incentives and remote technology with an eye towards building persistent behaviors that lead to habit formation.

ACKNOWLEDGMENT

Evidation Health and Humana provided funding for the study participant incentives but not for author salaries (other than author L.F. who works for Evidation Health) or data analysis. Evidation Health facilitated data collection through their online platform called Achievement. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. Author AJ had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Authors KLM, BT, CB, and AJ conducted data analysis. The collected data for this study contains personally identifying information that poses a risk of re-identification for study participants. As part of our efforts to ensure the welfare of participants and maintain their privacy, we have committed to the data provider that we will share data through a Data Use Agreement that asks third party researchers to 1) request the data explicitly from the authors: 2) pledge to not attempt participant re-identification; 3) establish safeguards for participant privacy; and 4) bear the liability of leaking the data in the public domain. We would like to thank Kevin Volpp, Maurice Schweitzer, Xuanming Xu, researchers at Behavior Change for Good, and participants at the 2015 Center for Health Incentives and Behavioral Economics Roybal Retreat for comments and feedback.

CITATIONS

1. Das P, Horton R. Rethinking our approach to physical activity. *The Lancet*. 2012 Jul 21;380(9838):189-90.
2. Lee IM, Shiroma EJ, Lobelo F, *et al.* Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The Lancet.* 2012 Jul 21;380(9838):219-29.
3. Franco OH, de Laet C, Peeters A, *et al.* Effects of physical activity on life expectancy with cardiovascular disease. *Arch Intern Med.* 2005 Nov 14;165(20):2355-60.
4. Centers for Disease Control and Prevention. Facts about physical activity. Accessed at https://www.cdc.gov/physicalactivity/data/facts.htm on 4 February 2018.
5. Halpern SD, French B, Small DS, *et al*. Randomized trial of four financial-incentive programs for smoking cessation. *N Engl J Med.* 2015 May 28;372(22):2108-17.
6. Volpp KG, Troxel AB, Pauly MV, *et al.* A randomized, controlled trial of financial incentives for smoking cessation. *N Engl J Med.* 2009 Feb 12;360(7):699-709.
7. Loewenstein G, Price J, Volpp K. Habit formation in children: Evidence from incentives for healthy eating. *J Health Econ.* 2016 Jan;45:47-54.
8. Acland D, Levy MR. Naiveté, projection bias, and habit formation in gym attendance. *Manag. Sci.* 2015 Jan 15;61(1): 146-160.
9. Charness G, Gneezy U. Incentives to exercise. *Econometrica*. 2009 May 21;77(3):909-931.
10. Royer H, Stehr M, Sydnor J. Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a Fortune-500 company. *Am Econ J Appl Econ.* 2015 July;7(3):51-84.

11. Patel MS, Asch DA, Rosin R, *et al.* Framing financial incentives to increase physical activity among overweight and obese adults: A randomized, controlled trial. *Ann Intern Med.* 2016 Mar 15;164(6):385-94.

12. Volpp KG, John LK, Troxel AB, *et al*. Financial incentive-based approaches for weight loss: a randomized trial. *JAMA*. 2008 Dec 10;300(22):2631-7.

13. Kullgren JT, Troxel AB, Loewenstein G, *et al*. A randomized controlled trial of employer matching of employees' monetary contributions to deposit contracts to promote weight loss. *Am J Health Promot*. 2016 Jul;30(6):441-52.

14. John LK, Loewenstein G, Troxel AB, *et al*. Financial incentives for extended weight loss: a randomized, controlled trial. *J Gen Intern Med*. 2011 Jun;26(6):621-6.

15. The Medicaid and CHIP Payment and Access Commission (MACPAC). The use of healthy behavior incentives in Medicaid: Issue brief. 2016 Aug. Accessed at https://www.macpac.gov/publication/the-use-of-healthy-behavior-incentives-in-medicaid/ on February 4, 2018.

16. Kaiser Family Foundation / Health Research & Education Trust. Employer health benefits survey. 2017 Sep 19. Accessed at https://www.kff.org/report-section/ehbs-2017-summary-of-findings/ on February 4, 2018.

17. Barte JCM, Wendel-Vos GCW. A Systematic Review of Financial Incentives for Physical Activity: The Effects on Physical Activity and Related Outcomes. *Behav Med*. 2017;43(2):79-90.

18. Mitchell MS, Orstad SL, Biswas A, et al. Financial incentives for physical activity in adults: systematic review and meta-analysis. *Br J Sports Med*. May 2019:bjsports-2019-100633.

19. Wood W, Neal DT. A new look at habits and the habit-goal interface. *Psychol Rev.* 2007 Oct;114(4):843-63.

20. Madrian BC, Shea DF. The power of suggestion: Inertia in 401(k) participation and savings behavior. *Q J Econ.* 2001 Nov 01;116(4):1149-1187.

21. Hengchen D, Milkman KL, Riis J. The fresh start effect: Temporal landmarks motivate aspirational behavior. *Manag Sci.* 2014 Jun 23;60(10):2563-2582.

22. O'Donoghue T, Rabin M. Doing it now or later. *Am Econ Rev.* 1999 Mar;89(1):103-124.

23. Frey BS, Jegen R. Motivation crowding theory. *J Econ Surv*. 2001 Dec;15(5):589-611.

24. Gneezy U, Rustichini A. A fine is a price. *J Leg Stud.* 2000 Jan;29(1):1-17.

25. Hirvensalo M, Telama R, Schmidt MD, *et al.* Daily steps among Finnish adults: Variation by age, sex, and socioeconomic position. *Scand J Public Health*. 2011 Sep;39(7):669-677.

26. Carrera M, Royer H, Stehr MF, Sydnor JR. The structure of health incentives: Evidence from a field experiment. National Bureau of Economic Research. Working Paper No 23188. 2017 Feb.

27. Patel MS, Volpp KG, Rosin R, *et al.* A randomized trial of social comparison feedback and financial incentives to increase physical activity. *Am J Health Promot.* 2016 Jul;30(6):416-24.

28. Patel MS, Asch DA, Rosin R, *et al.* Individual versus team-based financial incentives to increase physical activity: A randomized, controlled trial. *J Gen Intern Med.* 2016 Jul;31(7):746-54.

29. Mitchell MS, Goodman JM, Alter DA, *et al.* Financial incentives for exercise adherence in adults: Systematic review and meta-analysis. *Am J Prev Med.* 2013 Nov;45(5):658-67.

30. John LK, Milkman KL, Gino F, Tuckfield B, Foschini L. The role of incentive salience in habit formation. Harvard Business School Working Paper No 16-090. 2016 Feb.

31. Gneezy U, Meier S, Rey-Biel P. When and why incentives (don't) work to modify behavior. *J Econ Perspect.* 2011;25(4):191-210.

32. Tudor-Locke C. Steps to better cardiovascular health: How many steps does it take to achieve good health and how confident are we in this number? *Curr Cardiovasc Risk Rep.* 2010

Jul;4(4):271-276.

Table 1. Adjusted Mean Daily Step Counts by Experimental Condition During the 14-Day Intervention Period[2] [3]

	Main Model[4]				*Sensitivity Analysis*[5]			
	Control	Constant	Increasing	Decreasing	Control	Constant	Increasing	Decreasing
Daily Step Count	6,968.9	7,275.5	6,970.4	7,065.8	7,386.1	7,803.5	7,389.4	7,466.6
95% CI	[6912.4 , 7025.4]	[7075.5 , 7475.6]	[6910.6 , 7030.2]	[7008.1 , 7123.4]	[7301.7 , 7470.6]	[7498.8 , 8108.1]	[7303.7 , 7475.2]	[7384.3 , 7548.9]
Difference in Daily Step Count								
Relative to Control	-	306.7**	1.5	96.9*	-	417.3*	3.3	80.5
95% CI		[91.5 , 521.9]	[-81.6 , 84.7]	[15.3 , 178.5]		[91.7 , 742.9]	[-118.1 , 124.7]	[-38.5 , 199.4]
Relative to Increasing	-	305.1**	-	-	-	414.0*	-	-
95% CI		[89 , 521.2]				[88.4 , 739.6]		
Relative to Decreasing	-	-	-95.3*	-	-	-	-77.1	-
95% CI			[-179.3 , -11.3]				[-197 , 42.8]	
Relative to Constant	-	-	-	-209.8+	-	-	-	-336.9*
95% CI				[-425.3 , 5.7]				[-661.6 , -12.2]

2 + $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

3 Ordinary least square regression models used to generate these estimated step counts include fixed effects for person-by-day-of-week, day-of-year, and pedometer brand. Robust standard errors are clustered by participant-day-of-week. Between intervention arm differences were calculated using Wald tests.

4 The main model uses an ITT strategy and replaces missing data based on an average of pre-intervention step counts greater than 2000 steps, stratified by day of week – this would bias slightly towards a null effect. (N = 815,480; $R^2 = 0.38$)

5 The sensitivity analysis uses an ITT strategy and only includes step count data ≥ 2000 steps, stratified by day of week – this would bias more heavily towards a null effect. (N = 509,275; $R^2 = 0.26$)

Table 2. Adjusted Mean Daily Step Counts by Experimental Condition During the 21-Day Post-Intervention Period [6] [7]

	Main Model [8]				***Sensitivity Analysis*** [9]			
	Control	Constant	Increasing	Decreasing	Control	Constant	Increasing	Decreasing
1-Week Post-Intervention (Days 1 - 7)								
Daily Step Count	6,985.8	7,315.4	6,917.6	7,006.8	7,426.6	7,912.0	7,304.6	7,396.6
95% CI	[6910.7 , 7061]	[7023.4 , 7607.3]	[6843.4 , 6991.8]	[6933.3 , 7080.4]	[7304.5 , 7548.8]	[7434 , 8390]	[7189.4 , 7419.8]	[7282.6 , 7510.7]
Difference in Daily Step Count								
Relative to Control	-	329.5*	-68.3	21.0	-	485.4+	-122.0	-30.0
95% CI		[20.6 , 638.4]	[-174.6 , 38.1]	[-84.9 , 126.8]		[-20.1 , 990.9]	[-290.2 , 46.2]	[-197.4 , 137.5]
Relative to Increasing	-	397.8*	-	-	-	607.4*	-	-
95% CI		[89.2 , 706.4]				[103.7 , 1111.1]		
Relative to Decreasing	-	-	-89.2+	-	-	-	-92.0	-
95% CI			[-194.4 , 16]				[-254.5 , 70.5]	
Relative to Constant	-	-	-	-308.6*	-	-	-	-515.4*
95% CI				[-617.1 , -0.1]				[-1018.8 , -12]
2-Week Post-Intervention (Days 8 - 14)								
Daily Step Count	7,025.9	7,239.4	6,924.2	6,942.3	7,472.3	7,791.4	7,288.9	7,259.5
95% CI	[6952.3 , 7099.6]	[6947.7 , 7531.2]	[6846.9 , 7001.6]	[6869.6 , 7015.1]	[7354.6 , 7590.1]	[7319.1 , 8263.7]	[7169.4 , 7408.4]	[7147 , 7372]
Difference in Daily Step Count								
Relative to Control	-	213.5	-101.7+	-83.6	-	319.1	-183.4*	-212.9*
95% CI		[-94.8 , 521.8]	[-209.2 , 5.8]	[-187.7 , 20.6]		[-179.7 , 817.9]	[-351.5 , -15.3]	[-376.2 , -49.6]
Relative to Increasing	-	315.2*	-	-	-	502.5*	-	-
95% CI		[6 , 624.4]				[3.5 , 1001.5]		
Relative to Decreasing	-	-	-18.1	-	-	-	29.4	-
95% CI			[-125 , 88.8]				[-135.1 , 193.9]	
Relative to Constant	-	-	-	-297.1+	-	-	-	-532.0*
95% CI				[-605.1 , 10.9]				[-1029.4 , -34.6]

[6] + $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

[7] Ordinary least square regression models used to generate these estimated step counts include fixed effects for person-by-day-of-week, day-of-year, and pedometer brand. Robust standard errors are clustered by participant-day-of-week. Between intervention arm differences were calculated using Wald tests.

[8] The main model uses an ITT strategy and replaces missing data based on an average of pre-intervention step counts greater than 2000 steps, stratified by day of week – this would bias slightly towards a null effect. (N = 815,480; R^2 = 0.38)

[9] The sensitivity analysis uses an ITT strategy and only includes step count data ≥ 2000 steps, stratified by day of week – this would bias more heavily towards a null effect. (N = 509,275; R^2 = 0.26)

3-Week Post-Intervention (Days 15 - 21)								
Daily Step Count	6,981.7	6,959.0	6,905.4	7,041.7	7,423.0	7,350.7	7,268.6	7,406.8
95% CI	[6903.3 , 7060.1]	[6829 , 7088.9]	[6828 , 6982.8]	[6970.1 , 7113.2]	[7291.8 , 7554.3]	[7131.2 , 7570.3]	[7143.1 , 7394]	[7290.8 , 7522.8]
Difference in Daily Step Count								
Relative to Control	-	-22.8	-76.3	59.9	-	-72.3	-154.5+	-16.2
95% CI		[-177.3 , 131.8]	[-187.1 , 34.5]	[-46.8 , 166.7]		[-333.6 , 188.9]	[-336.4 , 27.4]	[-191.7 , 159.2]
Relative to Increasing	-	53.6	-	-	-	82.1	-	-
95% CI		[-100.5 , 207.7]				[-176.4 , 340.6]		
Relative to Decreasing		-	-136.3*	-	-	-	-138.2	-
95% CI			[-242.3 , -30.3]				[-309.3 , 32.9]	
Relative to Constant	-	-	-	82.7	-	-	-	56.1
95% CI				[-68.4 , 233.8]				[-197.5 , 309.7]

606