Incentivizing Behavioral Change: The Role of Time Preferences

Shilpa Aggarwal, Rebecca Dizon-Ross, & Ariel Zucker*

July 27, 2018

Abstract

How should the design of incentives vary with the time preferences of agents? We formulate predictions for two incentive contract variations that should increase efficacy for myopic agents relative to patient ones: increasing the frequency of incentive payments, and making the contract "dynamically non-separable" by only rewarding compliance in a given period if the agent complies in a minimum number of other periods. We test the efficacy of these variations, and their interactions with time preferences, using a randomized evaluation of an incentives program for exercise among 3,200 diabetics in India. On average, providing incentives increases daily walking by 1,300 steps or roughly 13 minutes of brisk walking, and decreases the health risk factors for diabetes. Increasing the frequency of payment does not increase effectiveness, suggesting limited impatience over payments. However, making the payment function dynamically non-separable increases cost-effectiveness. Consistent with our theoretical predictions, agent impatience over walking appears to play a role in non-separability's efficacy: both heterogeneity analysis based on measured impatience and a calibrated model suggest that the non-separable contract works better for the impatient.

*Aggarwal: Indian School of Business shilpa_aggarwal@isb.edu. Dizon-Ross: Chicago Booth School of Business rdr@chicagobooth.edu. Zucker: MIT adzucker@mit.edu. This study was funded by the Government of Tamil Nadu, the Initiative for Global Markets, J-PAL USI Initiative, the Chicago Booth School of Business, the Tata Center for Development, the Chicago India Trust, and the Indian School of Business. The study protocols received approval from the IRBs of MIT, Chicago, and IFMR. The experiment was registered on the AEA RCT Registry. We thank Ishani Chatterjee, Rupasree Srikumar, and Sahithya Venkatesan for their great contributions in leading the fieldwork, and Christine Cai for outstanding research assistance. We are grateful to Abhijit Banerjee, Esther Duflo, Pascaline Dupas, Seema Jayachandran, Devin Pope, Gautam Rao, and Frank Schilbach for helpful conversations and feedback, and to numerous seminar and conference participants for insightful discussions. All errors are our own.

1 Introduction

Incentive design is of core economic interest. While most classic contracting models assume that agents are relatively patient, there is growing evidence that many people are myopic. This raises an important question: What are the implications of agent myopia for the design of incentives? In this paper, we develop and test insights for how to tailor incentive design for impatient agents. We first formulate predictions for incentive contract variations that should be more effective for myopic agents. We then conduct a randomized controlled trial (RCT) to test the average efficacy of the contract variations, and to test our prediction that the variations are more effective for the impatient.

Our RCT focuses on incentives for behavioral change, which are increasingly prevalent in areas such as health (Morris et al., 2004; Thornton, 2008; Martins et al., 2009; Duflo et al., 2010), education (Fryer, 2011) and the environment (Davis et al., 2014; Jayachandran et al., 2017). Tailoring incentives for impatience may be particularly important in the behavioral-change domain, since the express purpose of many behavioral-change incentive programs is to address present bias.[1] Present-biased agents may fail to undertake behaviors with short-run costs but only long-run benefits (e.g., eating right or studying), even if those behaviors are in their own long-run self-interest. Behavioral-change incentive programs thus deliver small, short-run incentives in an attempt to better align agents' behavior with their long-run interests. However, even in this domain where impatience is a key rationale for using incentives, there is very little evidence on how to adjust incentive design for impatient agents.

We begin by developing predictions about how incentive contract design depends on time preferences. To do so, we distinguish between discount rates over consumption and over financial rewards. The recent time-preference literature emphasizes that time preferences over the receipt of financial rewards should reflect agents' borrowing and lending opportunities, which are distinct from their (true) time preferences over time-dated consumption (Augenblick et al., 2015). As a result, we model agents as having separate time preferences in the domains of consumption and financial rewards. We then identify two contract variations, one per domain, with increasing efficacy in agents' impatience in that domain.

In particular, we first test a novel prediction of our model: that making the contract "dynamically non-separable" can improve the relative efficacy of incentives for those with high discount rates over consumption (effort). By dynamically non-separable, we mean that the incentive paid for action in a given period depends on the actions in other periods. For example, the contract might pay people for walking at least 10,000 steps on a day if and only if they reach that step target on at least 5 days in the week. Second, we test the often-

[1] In contrast, other types of incentives (e.g., incentives for workers) often aim primarily to solve moral hazard issues instead of an "internality" like present bias.

posited but previously-untested prediction that for agents who are impatient over payments, providing more frequent payment can increase efficacy. Notably, both of these contract variations are predicted to improve the relative efficacy of incentives for agents with a variety of types of myopia, including those who are time-inconsistent and "sophisticated" (aware of their own time inconsistency), those who are time-inconsistent and "naive" (unaware of their time inconsistency), and those who are impatient but not time-inconsistent.

We test these two predictions through an RCT with roughly 3000 participants in an important setting: the exploding policy problem of diabetes in India. There are an estimated 60 million diabetics in India, with the number expected to surpass 100 million by 2025, or 7% of the population (Whiting et al., 2011; Anjana et al., 2011). Diagnosed diabetics often fail to follow the primary health recommendations for diabetes: exercising, eating right, and taking prescribed medicines. Since these behaviors involve short-run costs and long-run benefits, providing incentives is a promising approach to improve compliance and promote behavioral change. Our experiment delivers incentives to diabetics and prediabetics for walking, a key recommendation for diabetes management (Qiu et al., 2014; Zanuso et al., 2009).

The incentive program monitors participants' walking using pedometers, and, if they achieve a daily step target of 10,000 steps, provides them with small financial incentives in the form of mobile recharges (i.e. cell-phone credits). Within the program, we randomly vary (i) whether payment is a linear function of the number of days the agent meets the step target or is "dynamically non-separable," only rewarding step-target compliance on a given day if the agent meets the step target on a minimum number of other days that week, and (ii) the frequency with which incentives are paid. We also randomly assign some participants to a pure control group, and some to a "monitoring group" which receives pedometers but no incentives, allowing us to test for the overall effects of incentives on exercise and health.

Our analysis proceeds as follows. We first establish that our incentives program is highly effective at inducing exercise. Providing just 20 INR (0.33 USD) per day of compliance with a daily step target increases compliance by 21 percentage points (pp), off of a base of 30 percent. Average daily steps increase by 1300, or roughly 13 minutes of brisk walking.

We then use our experiment to explore the implications of time preferences for incentive design, presenting three main results. We begin with dynamic-nonseparability, examining both its average efficacy and whether it is more effective for those who are impatient over consumption. Our base incentive contract is separable – and, in particular, linear – in behavior across days, which means that the agent is paid separately for his behavior on each day (normally, 20 INR for each day of exercise). We examine the impact of making the contract dynamically non-separable by adding a "dynamic threshold" that requires a minimum compliance level before a reward is paid; we use two threshold levels, one that

only pays agents if they meet their step targets on at least 4 days in a week, and one that only pays them if they meet it on at least 5 days. (Note that the daily behavior incentivized in all contracts is to walk at least 10,000 steps in a given day; what we vary is whether there is a cross-day, or dynamic, threshold dictating the minimum number of days on which the 10,000 step target must be met in the week.)[2] Dynamic non-separability has several advantages and disadvantages that have been discussed before.

Our new theoretical insight is that dynamic non-separability interacts with impatience. Relative to linear contracts, certain dynamically non-separable contracts, including dynamic thresholds, should increase compliance for people who are impatient over future effort compared to for those who are patient.[3] This is because payments in dynamically non-separable contracts are a function of behavior on multiple days, thus linking the agent's decision about exercise over time. In the linear contract, the agent always compares the reward to the cost of walking *today* (which is not discounted), and so discount rates over walking/consumption do not matter.[4] In contrast, with dynamic threshold contracts, the agent compares the financial reward to the present *discounted* cost of walking in multiple periods – which will be lower for those who discount future walking more heavily. Intuitively, this type of contract takes advantage of the fact that impatient people discount their future effort spent walking.[5]

Our first main finding is that, on average, the "dynamically non-separable" threshold contract is more cost-effective than the linear contract. Introducing 4-day and 5-day dynamic thresholds does not change the percent of days on which people hit their step target. However, with the dynamic threshold, if agents do not meet the threshold number of days, they do not receive incentives for every day the daily target is reached as they do with the linear contract. As a result, the 4-day and 5-day threshold contracts cost roughly 10 and 15% less while generating the same amount of exercise. Thresholds have a potential downside, however: they generate more extreme outcomes, working better for some but worse for others, making it important to determine for whom they work well.

Our second finding is that, consistent with our theoretical predictions, the dynamically non-separable contract is more efficacious for the impatient. We first perform heterogeneity

[2]While dynamic thresholds interact with time preferences, this is not true of all thresholds. For example, any step target, such as 10,000 steps, involves a minimum "static threshold" required for payment. Since all steps have to be completed in a single period, however, the performance of static thresholds does not depend on time preferences, although the demand for static thresholds may (Kaur et al., 2015).

[3]See Section 2 for discussion of which types of non-separable contracts our predictions apply to.

[4]Note that this statement is conditional on discount rates over payment.

[5]This logic holds both for sophisticates and naives (as well as impatient time-consistents), although the logic plays out somewhat differently by type. For sophisticates, non-separability creates a commitment motive: agents exercise today to induce their future selves to exercise. For time-inconsistent naives, who are overoptimistic about their future desire to exercise, it creates an "option value" motive: they exercise today to give their future selves the opportunity to follow-through.

analysis based on a baseline measure of discount rates over exercise[6] to test our prediction that, relative to the linear contract, the dynamically non-separable will increase efficacy more for the impatient. The prediction holds: relative to linear contracts, non-separable contracts increase compliance 4 pp more for those whose impatience is above-median relative to those who are below, and 10pp more for those above the 75th percentile. These magnitudes are large relative to the sample-average effect of either contract (20 pp), especially given that there is no sample-average difference between the contracts. We also calibrate a model using experimental estimates of the distribution of daily walking costs; the results there also suggest that threshold contracts work considerably better for the impatient.

We next turn to the second dimension of contracts we varied – payment frequency. If agents are impatient over payment receipt, then more immediate payment should produce larger effects on behavior. This is the most intuitive prediction of impatience for incentive design, and, to our knowledge, the main prediction discussed previously in the literature (besides pre-commitment) for tailoring incentives to impatience,[7] with, for example, Cutler and Everett (2010) proclaiming "the more frequent the reward, the better." However, there are no empirical tests that compare different payment frequencies, and theoretical reasons to question both whether and what types of increases in payment frequency would matter. In particular, discount rates over money – and thus the overall responsiveness to payment frequency – could be small or large: if credit markets are perfect, discount rates over payment should only reflect the interest rate, but if not (as may be true in developing countries) they could be closer to discount rates over consumption. Furthermore, the types of increases in frequency that matter depend on which time-preference model holds in practice: in some models used in the literature (e.g., the "beta-delta" model), the gains to increasing frequency are limited unless payments can be made very frequently (e.g., every day), whereas in other models there could be large gains between, say, every month and every week.

We test the efficacy of three payment frequencies – monthly, weekly, and daily – to assess whether, and what type of, increases in payment frequency improve compliance. The three frequencies also allow us to explore which model of discounting best fits the data.

Our third main result is that increasing the frequency of incentive delivery has limited impact. Incentives delivered at daily, weekly, and monthly frequencies all have equally large impacts on walking, indicating that the discounting model that best fits our sample is one of patience over rewards. We find additional evidence in support of this conclusion: step-target compliance does not increase as the date of reward delivery approaches. Survey-based measured impatience over rewards is limited, and we find no evidence of heterogeneity in

[6] We use Andreoni and Sprenger (2012a)'s convex time budget method; people divide steps over time.

[7] O'Donoghue and Rabin (1999) examine how to optimally penalize time-inconsistent procrastinators for delays in a setting where delay is costly to the principal but task costs vary over time.

the effects of frequency by baseline measures of impatience over rewards. This null finding suggests that, in contrast with the conventional wisdom, increasing incentive frequency is not an effective way to adjust incentives for impatience. This result is consistent with Augenblick et al. (2015), who find limited time inconsistency in monetary choices, but is perhaps still surprising given the limited access to borrowing in our setting.

We conclude the paper with a program evaluation of our incentives program. We show that the large increases in walking induced by incentives cause moderate improvements in physical health and emotional wellbeing. Our sample has high rates of diabetes and hypertension; regular exercise can prevent complications from both. We find that incentives for walking improve an index of overall health risk, including measures of blood sugar, blood pressure, and BMI, by a moderately-sized, statistically significant amount. Incentives also improve mental health, suggesting a causal link between exercise and mental health.

This paper makes several contributions to the literature. First, it contributes to the literature on motivating time-inconsistent or impatient agents, and on contract design with non-standard time preferences. To date, the primary way that researchers have attempted to motivate time-inconsistent agents is to provide commitment devices or contracts that restrict the possible actions of their future selves (e.g., Ariely and Wertenbroch (2002); Kaur et al. (2015); Ashaf et al. (2006); Giné et al. (2010); Duflo et al. (2011); Schilbach (2017)). Although pre-commitment can be a very useful tool, it is not a panacea: take-up of commitment contracts is generally modest, as discussed in Laibson (2015). Indeed, commitment contracts are only effective for sophisticated time-inconsistents, but evidence suggests that a large share of the population is at least partially naive, and that commitment can in fact be harmful for partially naive agents (Augenblick and Rabin, 2017; Bai et al., 2017). In contrast, the predictions we test do not require sophistication: they work for multiple types of myopia, including naive time-inconsistency. Beyond the literature on pre-commitment, there is limited work, theoretical or empirical, on how to optimize incentive design for impatience.[8] O'Donoghue and Rabin (1999) is a notable exception, examining the theoretical implications of time-inconsistent procrastination for the design of "temporal incentive schemes," which reward agents based on when they complete tasks. In their setting, the focus is on how to avoid delay, which is costly to the principal. In contrast, in our setting, the period in which the agent completes the task is less important; rather the goal is to maximize the average level of compliance over time. Another exception is the literature showing that worker performance tends to improve toward the end of pay cycles, and attributing this effect to impatience. In particular, Clark (1994) finds extensive anecdotal

[8]Dellavigna and Malmendier (2004) study how the firm's profit-maximizing contract varies with consumer time preferences. Opp and Zhu (2015) study the implications of agent impatience for dynamically self-sustaining agreements when agents can renege on agreements, e.g., settings with upfront payment to workers.

evidence that 19th century factory workers often shirked at the beginning of pay cycles, Oyer (1998) finds dramatic spikes in sales among US salespeople near the end of the year when bonuses are calculated and paid, and Kaur et al. (2015) show that individuals paid at a piece-rate increase output as a randomly-assigned weekly pay-day approaches. While these papers suggest a potential role for high-frequency payment to improve performance,[9] ours is the first to directly test this suggestion.

Second, we build on the literature examining dynamic contracting (see for example Lazear (1979) and Prendergast (1999)). Many theoretical dynamic contracting papers yield the prediction that it is optimal to defer some component of current pay until the future, with receipt of the payment contingent on future effort. We make two main contributions to this literature. First, since our dynamically non-separable contract entails deferred compensation, our paper is the first to theoretically demonstrate that impatience over consumption can improve the efficacy of deferred compensation. Second, to our knowledge, our experiment represents the first empirical comparison of contracts with and without deferred compensation, albeit outside the workplace compensation setting on which the compensation literature focuses.

We build on a third body of literature that measures the shape of time preferences. The majority of the recent work has focused on distinguishing whether time preferences are time-consistent or time-inconsistent (Andreoni and Sprenger, 2012a; Andreoni et al., 2016; Augenblick et al., 2015). However, within those classes, there is large variation in the feasible shape and size of discount rates with important policy implications; to our knowledge, our paper is the first to test the policy implications of that variation.[10] Finally, we contribute to the growing literature on incentives for health, including incentives for weight loss (Volpp et al., 2008; Kullgren et al., 2013), disease monitoring (Labhardt et al., 2011), and physical activity (Charness and Gneezy, 2009; Finkelstein et al., 2008). Although there are several financial incentives trials for exercise for non-diabetics (Charness and Gneezy, 2009; Finkelstein et al., 2008), as well as one trial incentivizing 3-month blood sugar control (as measured by Hba1c), there is a lack of interventions that incentivize important daily habits among diabetics. Ours represents the first evaluation of incentives to diabetics for daily disease management, and the first trial of incentives for exercise in a developing country.

The rest of the paper proceeds as follows. Section 2 presents the theoretical predictions that motivate the experiment. Section 3 discusses the study setting and research design. In Section 4, we discuss the data. Section 5 presents the results on incentive design and

[9] Kaur et al. (2015) suggest this explicitly.

[10] Two previous papers have explored the shape of discounting *within* the time-consistent or time-inconsistent subclasses, both using lab-experimental preference measures in the monetary domain. Benhabib et al. (2008) test between hyperbolic, quasi-hyperbolic, and exponential models, but do not have the power to distinguish between them. Tanaka et al. (2010) reject that preferences are purely hyperbolic, quasi-hyperbolic, or exponential.

its relationship with time preferences. Section 6 presents the impacts on health outcomes. Section 7 concludes.

2 Predictions from a simple model of walking

We now present a simple model of walking to derive predictions for how two incentive contract features – dynamic-nonseparability and payment frequency – interact with time preferences. We consider a model with daylong periods. In each period, individuals experience a utility cost if they walk 10,000 steps (which might be negative) and receive utility from their other consumption in that period, which in our experiment will be consumption of mobile recharges:

$$U = \sum_{t=0}^{\infty} d_c(t) \left(c_t - e_t^{\mathbf{1}(w_t=1)} \right)$$

The term e_t is the utility cost from walking 10,000 steps, i.e, the cost of complying with the program exercise target; w_t is an indicator for compliance; c_t are the mobile recharges consumed on day t; and individuals discount the cost of walking and consumption k days in advance by $d_c(k)$. Because the amounts are small, we model utility as linear in payments/recharges for simplicity, but the model's qualitative predictions are the same if we relax this assumption. We assume that walking costs e_t are independently and identically distributed (i.i.d.) with cumulative distribution function $F(\cdot)$, are separable from other consumption, and are known in advance. The individual's problem is to choose c_t and w_t to maximize utility subject to a budget constraint. Individuals earn income by complying with the step target and so the budget constraint depends on the incentive contract mapping compliance to income.

We consider first a separable, linear incentive contract; this contract specifies that walking on each day t will be rewarded with an incentive of size m in k_t days. Denote the total value of recharges received in period t as m_t. The form of the budget constraint depends on the availability of borrowing/savings technology. We consider two main cases:

1. **No savings, borrowing, or storage**. In this case, consumption in a given period is equal to the total amount of recharges received in the period: $c_t = m_t$ in all periods. That would imply that in any given period t, an individual would choose to walk as long as the walking costs are less than the discounted value of consuming the mobile recharge reward k_t days in the future: $e_t \leq d_c(k_t)m$.

2. **Can borrow and save at an interest rate r.** The lifetime budget constraint becomes $\sum_{t=0}^{\infty} \left(\frac{1}{1+r}\right)^t c_t = \sum_{t=0}^{\infty} \left(\frac{1}{1+r}\right)^t m_t$. Thus, on day t, the value of receiving m in rewards k_t days in the future is $\left(\frac{1}{1+r}\right)^{k_t} m$, and so the individual chooses to walk as long as $e_t \leq \left(\frac{1}{1+r}\right)^{k_t} m$.

More broadly, one can accommodate both of these (and other)[11] cases in a reduced form way by defining a discount factor representing the amount by which individuals discount rewards received k periods in the future, which encompasses both their "primitive" discount rate and any financial frictions. We denote this discount factor as $d_m(k)$; in case 1, $d_m(k) = d_c(k)$, whereas in case 2, $d_m(k) = \left(\frac{1}{1+r}\right)^k$. Using this new notation, individuals choose to walk on day t as long as walking costs are less than the discounted value of the mobile recharges received for the walk: $e_t \leq d_m(k_t)m$. The probability of compliance with the step target on day t for a reward in k_t days is thus:

$$Pr\left(w_t = 1|\text{Linear}\right) = F\left(d_m(k_t)m\right). \tag{1}$$

Thus, in a separable, linear contract on which payments are received every X days for the previous X days' worth of walking (e.g., a weekly contract where payments are received on the last day of the week), the expected days of compliance per payment cycle is:

$$E\left[\sum_{t=1}^{X}(w_t)|\text{Linear}\right] = \sum_{t=1}^{X} F\left(d_m(X-1)m\right) \tag{2}$$

Predictions regarding dynamic non-separability We now explore the effect of making the contract dynamically non-separable. We focus on a specific form of non-separability: a "dynamic threshold" wherein payment is a function of the number of periods of compliance in a given time range (in our experiment, it will be total days of compliance within a week). The payment function has a minimum threshold for total compliance below which no incentive is received (e.g., 4 days); if they reach that threshold, they will receive m per day of walking that week. We focus on dynamic thresholds because they are a simple, implementable form of dynamic-non-separability, but the prediction we demonstrate about the interaction between dynamic non-separability and time preferences holds for a broader set of contracts that display a "dynamic complementarity" (i.e., a period in which the payment for effort is increasing in future effort).[12]

Dynamic non-separability make an individual's decision to walk considerably more complicated. While the decision to walk in linear contracts is separable across days, in dynamically non-separable, the reward for compliance – and hence the decision to walk – depends on compliance on others of the days in the payment period. For simplicity, we thus illus-

[11] For example, this approach also nests the case where there is no storage of recharges and time preferences over consumption are domain-specific, so $U = \sum_{t=0}^{\infty} d_m(t)m_t - d_c(t)e_t^{\mathbf{1}(w_t=1)}$.

[12] We conjecture that having a dynamic complementarity is a necessary condition for the prediction to hold. The prediction would thus not hold for contracts that only contain "dynamic substitutabilities" (e.g., paid for at most one day of walking in a week.) Note that dynamic complementarity might not be a sufficient condition.

trate how dynamic non-separability interacts with compliance and time preferences using a shorter payment cycle than used in our experiment; in particular, we consider a two-day threshold contract. If the individual complies with the step target on both days, she receives $2m$ on the second day. However, if she complies on only one day, she receives nothing. We restrict treatment here for presentational simplicity to the case where all costs are positive (i.e., there is no inframarginal walking), an assumption which does not affect the results, but simplifies the notation.[13]

Intuitively, the key difference between the threshold and linear contracts is that in linear contracts, in each period the individual compares the reward only with her (undiscounted) cost of effort today. Thus, conditional on discount factors over money, discount factors over consumption do not affect the decisions to comply;[14] she complies if $c_1 < d_m(1)m$ in period 1 and $c_2 < m$ in period 2. In the threshold contract, in contrast, discount factors over consumption matter. In particular, the individual in period 1 makes a joint decision about whether it is worth it to walk in both periods in order to get paid on the second day, and so compares the present discounted value of effort across both periods with the rewards. On the first day, she complies if the present discounted cost of walking on both days, $c_1 + d_c(1)c_2$, is less than the discounted value of the reward $d_m(1)2m$ and she knows she will follow through on the second day. She thus complies if both (i) $c_1 + d_c(1)c_2 < d_m(1)2m$, and (ii) $c_2 < 2m$. Importantly, condition (i) is more likely to be satisfied if agents discount future effort more. On the second day, the agent complies with the step target if she has already walked on the first day, and condition (ii) above holds. The agent's expected total compliance in the 2-Day threshold contract is thus

$$\begin{aligned} E\left[\sum_{t=1}^{2}(w_t)|\text{Threshold}\right] &= P(c_1 + d_c(1)c_2 < d_m(1)2m \text{ AND } c_2 < 2m) \\ &= 2\int_{-\infty}^{2m} F(d_m(1)2m - d_c(1)c)f(c)dc \end{aligned} \tag{3}$$

We can compare that to compliance in the linear, separable contract with a 2-day payment

[13] With negative costs, the addition to the decision problem below is that if her cost of compliance e_t is negative for either day, she will walk on that day for her own intrinsic enjoyment, and equation 3 becomes $E\left[\sum_{t=1}^{2}(w_t)|\text{2-Day Threshold}\right] = 2\int_{-\infty}^{2m} F(d_m(1)2m - d_c(1)c)f(c)dc + \int_{-\infty}^{0}[1 - F(d_m(1)2m - d_c(1)c)]f(c)dc + F(0)[1-2m]$

[14] As described in detail in Section 4, we measure both discount factors (over walking and over recharges) in our sample; the correlation between discount factors over recharges and consumption is low and not significant, and the sample-average discount rate over walking is much higher than over recharges, suggesting that the discount rate over recharges represents the interest rate in most cases.

period (note that this comparison holds payment frequency constant):

$$E\left[\sum_{t=1}^{2}(w_t)|\text{Linear}\right] = F\left(d_m(1)m\right) + F\left(m\right) \tag{4}$$

Whether total compliance with a 2-day threshold (Equation 3) is larger than the compliance without a threshold (Equation 4) depends on the distribution of walking costs; thus, the effect of adding a threshold to a linear contract on overall compliance with the step target is theoretically ambiguous.[15] However, we can show the following prediction:

Prediction 1. *Compliance in the dynamic threshold contract relative to in the linear contract is increasing in the discount rate over walking (i.e., decreasing in the discount factor over walking, $d_c(k)$).*

This follows directly from inspection of equations 3 and 4. As the discount factor decreases, the present discounted cost of walking on days 1 and 2 decreases, increasing the probability of walking. In other words, individuals who discount future walking heavily have a lower total discounted cost of reaching the threshold, and thus higher compliance.

Note that the prediction holds for both time consistent and time inconsistent time preferences. Although this might seem like an artifact of our focus on a 2-period model, that statement also holds in multi-period models. Within time-inconsistent time preferences, the finding holds for both naives and sophisticates; in fact, non-separability can in some cases work better even for naives, because their overoptimism about their future compliance makes them even more likely to comply today, as can be shown with a longer-period model.

In our experiment, we test prediction 1 by randomly varying whether the contract has a dynamic threshold, and testing for heterogeneity based on the discount rate over walking.

Predictions regarding payment frequency We now return to the linear contract setup to analyze the effects of changing the frequency of payment. Using equation 2, we can make three predictions, all quite intuitive.

Prediction 2. *For an agent who is "impatient" over the receipt of financial rewards (i.e., for whom $d_m(k) < 1$ and is decreasing in k), compliance is increasing in the payment frequency. For an agent who is patient ($d_m(k) \approx 1$), payment frequency does not affect compliance.*[16]

[15] In models that incorporate uncertainty over walking costs, overall efficacy will also depend on the degree of uncertainty and risk aversion.

[16] The prediction for patient agents relies on the linearity assumption: if utility were concave and there were no storage, then more frequent payments could still increase the likelihood of walking through a concavity channel, as higher frequency would mean the rewards were broken up into smaller tranches. However, linearity does not affect the important comparative static of payment frequency with respect to patience.

This follows from equation (2): the likelihood of walking is increasing in the discount factor over rewards $d_m(k)$, and increasing payment frequency weakly decreases the delay to payment k_t on each day t.

Prediction 3. *The quantitative effect of increasing the payment frequency depends not just on average discount rates but on the shape of the agent's discount factor over time.*

Figure 1 shows how discount factors might change over time for four different models of discounting used in the literature: Quasi-hyperbolic ("beta-delta"), hyperbolic, exponential impatient, and exponential patient (with the former two time-inconsistent and the latter two time-consistent). One can see that under the models where discount factors decay gradually over time (hyperbolic or time-consistent impatient), there could be large gains to switching from low-frequency (e.g., monthly) to medium-frequency (e.g., weekly) payments. In contrast, in a quasi-hyperbolic model, where the biggest difference is between "the present" and "the future," there would only be big gains to increasing frequency if payment could be made within the "beta window" (often modeled as 1 day, which would require daily payments.) Given that paying within the "beta window" could be costly or infeasible in some settings, it is important to distinguish between these scenarios.

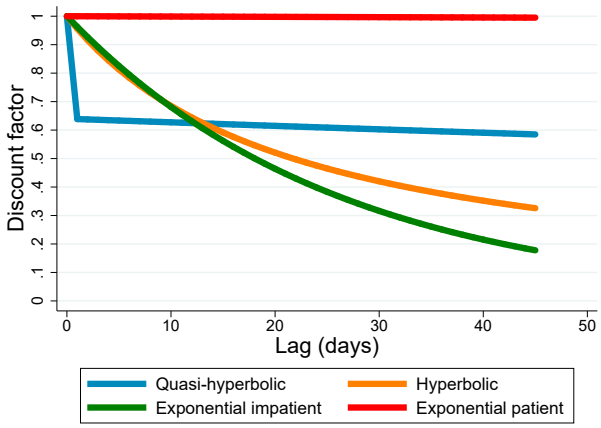


Figure 1: Hypothetical discount factors

Note: Figure displays hypothetical discount factors as a function of lag length under different models of discounting.

As a result, our experimental design will test the efficacy of three payment frequencies – monthly, weekly, and daily – in order to answer the question of whether and what type of increases in payment frequency improve compliance. Our three frequencies also allow us to explore which discount factor model for payments best fits the data, with the overall magnitude of frequency effects informing our understanding of the overall level of discounting,

and the relative effects of moving from monthly to weekly frequency, and from weekly to daily frequency informing our understanding of the shape of time-preferences. A final prediction allows us to use our experiment to shed further light on the model of discounting.

Prediction 4. *If the discount factor over payments is decreasing in k and agents are paid every X days with $X > 1$, then compliance will increase as the "payday" (e.g., the end of the week if agents are paid weekly) approaches.*

This follows from equation (1) since the time to payment decreases as the payment date approaches.

3 Study Setting and Experimental Design

India is facing a diabetes epidemic. In addition to 60 million diabetics, there are also 77 million pre-diabetics in the country (Whiting et al., 2011; Anjana et al., 2011). This has large economic and social implications: in 2010, diabetes imposed an estimated cost of $38 billion – 2 percent of India's GDP – on the healthcare system, and led to the death of approximately 1 million individuals (Tharkar et al., 2010).

There is widespread agreement that lifestyle changes are essential for managing the burden of diabetes, but existing strategies to promote change have had limited success. In particular, increased physical activity can prevent diabetes, and help the diagnosed avert serious (and expensive) long-term complications such as amputations, heart disease, kidney disease, and stroke. Recognizing this, the Indian government is piloting a National Programme on Prevention and Control of Diabetes, Cardiovascular diseases and Stroke (NPCDS), with a key objective of generating awareness of appropriate lifestyle changes. However, anecdotal evidence from physicians suggests that adoption of the lifestyle changes recommended by NPCDS is low, and new strategies are needed to encourage change.

3.1 Sample Selection and Pre-Intervention Period

Our project is a collaboration with the government of the state of Tamil Nadu to encourage lifestyle changes, specifically exercise, among those with or at risk of Type 2 diabetes. We selected our sample through a series of public screening camps in the city of Coimbatore, Tamil Nadu. In order to recruit diverse socioeconomic groups, the camps were held in locations ranging from the government hospital to markets, mosques, temples, and parks. During the camps, trained surveyors took health measurements; discussed each individual's risk for diabetes, hypertension, and obesity; and conducted a brief eligibility survey. In order to be included in the study, individuals were required to have elevated blood sugar or have been diagnosed with diabetes, have low risk of injury or complications from regular walking, be capable with a mobile-phone, and be able to receive personal rewards in the form of

mobile recharges.[17] Within a week of attending a screening camp, eligible individuals were contacted by phone and invited to participate in a program to encourage walking.

Surveyors visited the potential participants at their homes or workplaces in order to conduct an initial baseline health survey and enroll participants in a one-week phase-in period. During the baseline health survey, surveyors collected detailed health, fitness, and lifestyle information. Surveyors then prepared respondents for the phase-in period, which was designed to collect baseline walking data and to familiarize participants with pedometer-wearing and step-reporting. Surveyors first demonstrated how to properly wear and read a pedometer. Next, they demonstrated how to report steps to our database by either responding to an automated call or directly calling into the system, and how to check text messages sent by the reporting system (as explained in Section 3.3, we created this automated calling system for respondents, who typically lack internet access, to self-report their daily steps). After the demonstration, respondents were asked to consistently wear a pedometer, and to report their steps each day through the automated call system for the weeklong phase-in period.[18]

Following the phase-in period, surveyors again visited respondents to sync the data from the pedometers, and conducted a baseline time-preference survey.[19] In the time-preference survey, surveyors elicited time preferences with a series of choices in the two domains relevant for our intervention: walking and mobile recharges (the financial reward we used to incentivize walking). The choices follow the Convex Time Budget (CTB) methodology pioneered by Andreoni and Sprenger (2012a) and Andreoni and Sprenger (2012b), and adapted in many studies, including Augenblick et al. (2015); Carvalho et al. (2016); Andreoni et al. (2016); Augenblick and Rabin (2017); Giné et al. (2017).

Finally, the participants were randomly assigned to participate in one of two comparison groups (a monitoring group that received pedometers during the intervention period and a control group that did not), or to one of six incentive contracts for walking. All participants who withdrew or were found ineligible for the study prior to randomization were excluded from the sample, leaving a final experimental sample of 3192 individuals.

[17]The full list of eligibility criteria was that the respondent must: either be diabetic or have elevated Random Blood Sugar, or RBS, ($>$ 130 if haven't eaten, $>$ 150 if have eaten in previous 2 hours); be 30-65 years of age; have a prepaid mobile number which is used solely by them and without an unlimited calling pack; be literate in Tamil; be physically capable of walking half an hour; be currently living in Coimbatore city; not be pregnant; not be currently receiving insulin injections for diabetes; not be suffering from blindness, kidney disease or foot ulcers; not have had medical conditions such as stroke or heart attack; and not have been diagnosed with Type 1 diabetes.

[18]Respondents received a small cash reward of 50 INR at the end of the phase-in period for consistently wearing their pedometers and reporting their steps.

[19]Surveyors first used the Fitbit web application to automatically sync the actual walking data from the phase-in week to an online step database. They compared actual steps to reported steps, and reviewed the step-reporting processes as needed, before administering the time-preference survey.

3.2 Experimental Design

3.2.1 The Daily Step Target

Our interventions center around encouraging participants to walk at least 10,000 steps a day. We chose this daily step target to match exercise recommendations for diabetics. The choice of a daily target reflects the fact that research organizations like the Center for Disease Control (CDC) and American Diabetes Association (ADA) recommend daily exercise sessions with no more than two consecutive days of rest. The target choice of 10,000 steps approximates the number of steps that our average participant would take if he added the exercise routine recommended by the CDC and ADA to his existing behavior.[20] In addition, 10,000 steps per day is a widely quoted target among health advocates and a common benchmark in health studies, making our choice consistent with existing literature and standard advice.

3.2.2 Treatment Groups

Participants were randomized into the incentives group or one of two comparison (non-incentive) groups:

1. Incentives: Receive a pedometer and incentives to reach a daily step target of 10,000 steps.

2. Monitoring (often plays the role of control): Receive a pedometer but no incentive contract.

3. Control: Receive neither a pedometer nor an incentive contract.

Within the incentives group, we randomized participants into one of six incentive contracts for walking. All treatments are summarized in Figure 2 and further elaborated below. The randomization was stratified by baseline Hba1c (a measure of blood sugar control) and a simple survey-based measure of impatience, using a randomization list generated in Stata.[21] Treatment groups were not of equal size: the size of each treatment group was chosen to ensure power to detect health impacts of the pooled incentives treatments relative to the comparison treatment, and the interactions between particular baseline characteristics and incentive contract features.

[20] In particular, daily exercise recommendations for diabetics translate into approximately 3,000 steps of brisk walking per day (Marshall et al., 2009). In our sample, the average participant does not walk for exercise, but completes 7,000 steps per day. Our daily target is the sum of average daily pre-intervention steps plus the steps needed for daily recommended exercise.

[21] Specifically, participants were stratified into four cells according to whether their baseline Hba1c was greater than 8 mmol/mol, and whether the average of their answer to the question "On a scale of 1 to 10, how patient are you?" at screening and baseline is greater than 6.5.

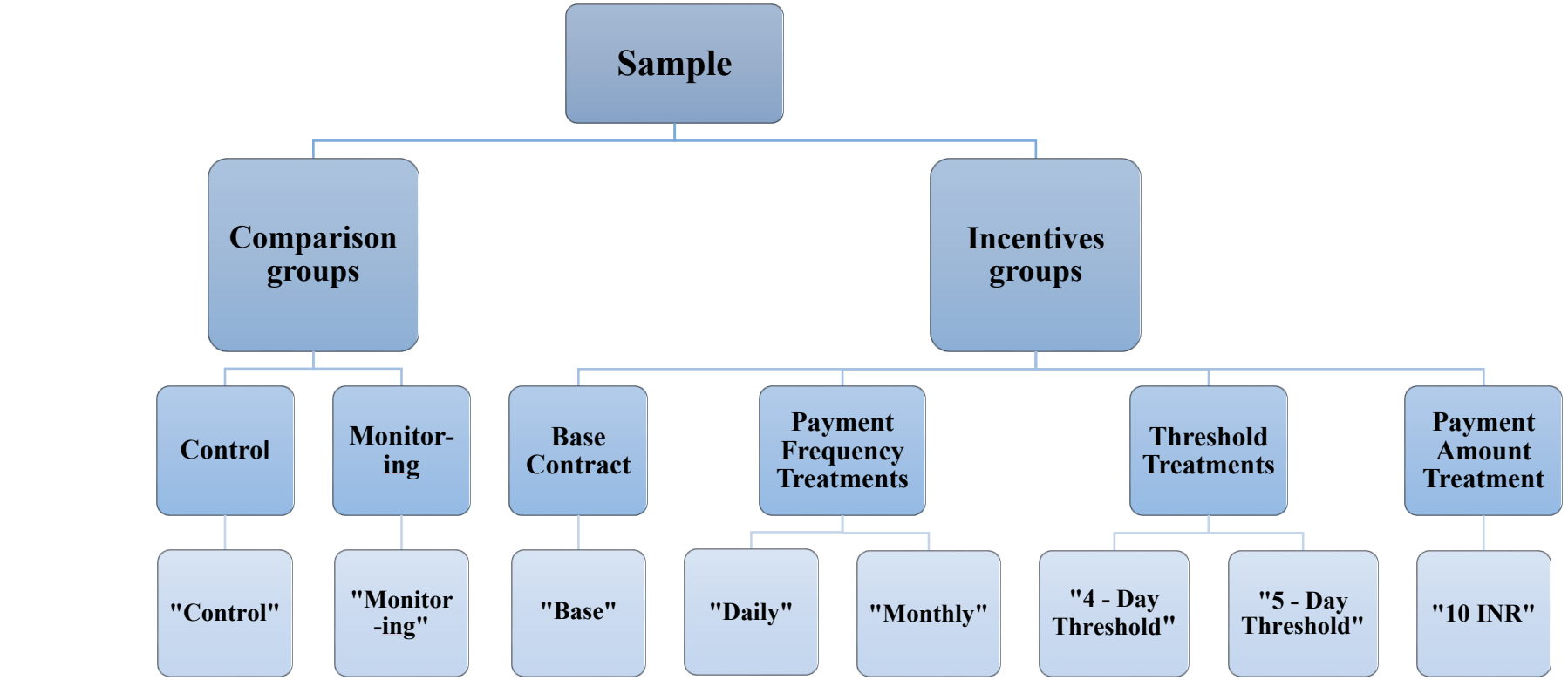


Figure 2: Experimental Design

Incentives Groups

All incentives groups were rewarded for accurately reporting steps above the daily 10,000 step target through the automated step-reporting system. As in the phase-in period, this step-reporting system called participants every evening (participants could choose a call time at the beginning of the intervention period), and prompted them to enter their daily steps as shown on the pedometer. Participants also had the option to call in their steps to a dedicated phone line at any time. The step-reporting system sent immediate text-message confirmations of each step report, and weekly text messages summarizing walking behavior.

During the explanation of the incentive contract, surveyors explained the step target to participants in the context of health recommendations, saying: "Remember that doctors recommend that you walk at least 10,000 steps a day, and more is always better! We recommend that you try to walk at least 10,000 steps a day and build up."

The threshold treatments implicitly gave participants a goal of how many days to walk per week. To control for these goal effects, surveyors verbally encouraged participants in all treatment groups to walk at least 4 or 5 days per week at contract launch.

The Base Case Incentives Group We vary three dimensions of the payment: frequency, linearity/non-separability, and amount. The base case incentives group serves as our "base contract" or comparison group for all other incentives groups. To assess the responses to variation on each dimension, we compare the base case incentives group to a treatment group differing only along that dimension.

The base case incentives group was offered a separable, linear incentive contract awarding them mobile recharges worth 20 INR for each day they reported complying with the daily 10,000 step target. Recharges were delivered at a weekly frequency for each day the participant complied with the step target in the previous week.

Our next treatment groups differ from the base case incentive group in one of the following two dimensions that we predict will interact with time preference: payment frequency and whether the contract has a dynamic threshold.

Payment Frequency Two other treatment groups, the daily and monthly groups, differed from the base case incentives group only by the frequency of incentive delivery. In the daily group, recharges were delivered at 1am the same night participants reported their steps. In the monthly group, recharges were delivered every four weeks for all days of compliance in the previous four weeks. Comparing walking behavior in these two groups with the base case incentives group allows us to assess the role of payment frequency and time preferences over mobile recharges in incentive effectiveness. See Figure 2.

Dynamic Thresholds Two other treatment groups, the 4-day threshold and the 5-day threshold groups, differed from the base case incentives group only by the minimum threshold of weekly step-target compliance required before an incentive was paid. The base case incentives group's contract was separable across days: participants received 20 INR for each day of compliance, while the threshold contracts were dynamically non-separable. The 4-day threshold group received mobile recharges worth 20 INR for each day of compliance if they exceeded the target at least 4 days in the weeklong payment period. So, a 4-day threshold participant who exceeded the step target on only three days in a payment period would receive no reward, but a participant who exceeded the step target on four days would receive mobile recharges worth 80 INR at the end of the week. Similarly, the 5-day threshold group received mobile recharges worth 20 INR for each day of compliance if they exceeded the target at least 5 days in the week.

Recall that, to control for goal effects, at the start of the intervention period, surveyors verbally encouraged participants in all treatment groups to walk at least 4 or 5 days per week. For those in the threshold groups, the target days-per-week was the same as their assigned threshold levels; for those in the other groups, the target days-per-week was randomly assigned in the same proportion as the threshold participants are divided between the 4- and 5-day threshold groups.

Payment Amount Finally, we included a small-payment treatment group that differed from the base case incentive group only by the amount of incentive paid. The 10-INR group was offered an incentive contract awarding them mobile recharges worth 10 INR, instead of the base-case 20INR, for each day they reported exceeding the daily step target. This treatment was included to help us learn about the distribution of walking costs, and to benchmark the magnitude of our other treatments effects (following for example Bertrand et al. (2005) and Kaur et al. (2015)).

Control Groups

We include two control groups in our experiment, a monitoring group and a pure control. In order to measure the overall health effects of the incentives program, we compare outcomes from endline surveys between the pooled incentives treatments to outcomes in the pure control treatments.[22]

Monitoring The monitoring group allows us to isolate the effects of incentives alone. The monitoring group was treated identically to the incentives groups, but for the fact that monitoring participants did not receive incentives. In particular, they received pedometers and were encouraged to wear the pedometers and report their steps every day through the

[22]Our experiment was not powered to detect differences in health outcomes between the control and monitoring groups, but we report these comparisons nonetheless.

step reporting system.[23] To control for the possibility that incentives may increase the salience of walking behavior, monitoring participants received daily confirmations of their step reports, and weekly text messages summarizing their walking behavior. In order to control for the effect of step goal-setting that an incentive for 10,000 daily steps may bring, monitoring and incentive treatment participants are given the same verbal step target of 10,000 daily steps at contract launch, and the same encouragement to walk at least 4 or 5 days per week.

Pure Control The pure control group allows us to measure the impact of all those aspects of the Incentive treatments that were necessary for operating a walking incentives program in our setting, excluding the incentives themselves. Participants in the pure control group returned their pedometers at randomization (after the one-week phase-in period), but, because we wanted to net out any effects due to survey visits related only to research needs, still received regular visits from the survey team at the same frequency of the pedometer sync visits. Thus, the difference between the pure control and incentives groups includes the effect of incentives bundled with the effect of receiving a pedometer, but excludes the effect of the regular survey visits. Because any feasible incentive program would bundle the "monitoring" effect of a pedometer with the effect of incentives, the pure control group is a useful benchmark from a policy perspective: the difference between the pure control group and the incentives groups measures the total effect of a walking incentives program, including the effects that come simply from participants utilizing a step monitoring technology.[24]

3.3 The Intervention Period and After

After randomization, all participants in the experiment were given a contract that detailed the specifics of the treatment group they had been assigned to, and also outlined the evaluation activities entailed for the rest of the study. A trained surveyor walked them through the contract and answered any of their questions to make sure it was clear.

In order to determine the number of steps taken, we gave those assigned to the "monitoring" and incentive groups Fitbit Zip pedometers for the duration of the intervention.[25]

[23]Participants in all incentives groups and the monitoring group received a cash bonus of 200 INR for regularly wearing the pedometer and reporting their steps at the endline survey. In addition, if participants did not report steps for a number of days, the system would send them messages asking them to please report their steps regularly.

[24]To accommodate a request from our government partners, we also cross-randomized one additional intervention a small sub-sample. In particular, 10% of the sample, cross-randomized across all other treatments, received the "SMS treatment," which consisted of weekly text-message-based reminders to engage in healthy behaviors for diabetes such as eating right and exercising, adapted from another SMS program that had been shown to be successful in the Tamil Nadu region for diabetes prevention (Ramachandran, 2013). We control for the presence of the SMS in our main regressions and the results of this treatment are shown in the online appendix.

[25]We chose Fitbit Zip pedometers due to their wearability, long memory, and relatively simple process for

Although these pedometers could be synced to a central database with an internet connection, most participants did not have regular internet access and so these data were not available in real time. Instead, we asked participants to report their daily step count to an automated calling system every evening. Incentive deliveries, i.e., mobile credits, were based on these reports. To verify the reports, we visited participants every two to three weeks to manually sync their pedometers and discuss any discrepancies with them. Anyone found to be chronically over-reporting was suspended from the program. All empirical analysis is based on the synced data from the Fitbits, not the reported data.

We visited all participants three times during the twelve-week intervention period. The primary purpose was to sync pedometers, but we also conducted short surveys to collect biometric and mobile phone usage data (we conducted these visits even with those participants who did not have a pedometer). We conducted a slightly longer midline survey at the second sync visit. Following the twelve-week intervention period, we conducted an endline survey. At endline, surveyors again collected detailed health, fitness, and lifestyle information. The timeline of the full intervention is outlined in Figure 3.

4 Data and Summary Statistics

4.1 Baseline Data: Health, Walking, and Time Preference

In the paper, we use three datasets of baseline characteristics: a baseline health survey, a week of baseline walking data, and a time-preference survey. The baseline health survey, conducted at the first household visit, contains information on respondent demographics, as well as health, fitness, and lifestyle information. Health measures include Hba1c, a measure of blood sugar control over the previous three months and the most commonly used measure of diabetes risk; random blood sugar, a measure of more immediate blood sugar control; BMI and waist circumference, two measures of obesity; blood pressure, a measure of hypertension; and a short mental health assessment. The baseline also includes two fitness measures (time to complete 5 stands from a seated position, and time to walk 4 meters), and lifestyle information including information on dietary, exercise, and substance use habits. Then, during the phase-in period between the baseline health survey and randomization, we collected one week of pedometer data, consisting of daily step counts.

Following the phase-in period, we conducted a baseline time-preference survey. The survey adapts the convex time budget (CTB) methodology of Andreoni and Sprenger (2012a) to measure time preferences in two domains, walking and mobile recharges, which correspond to $d_c(k)$ and $d_m(k)$ from Section 2. We asked participants to make a series of decisions allocating either recharges to be disbursed, or steps to be taken, on two dates: a "sooner"

syncing data to a central database.

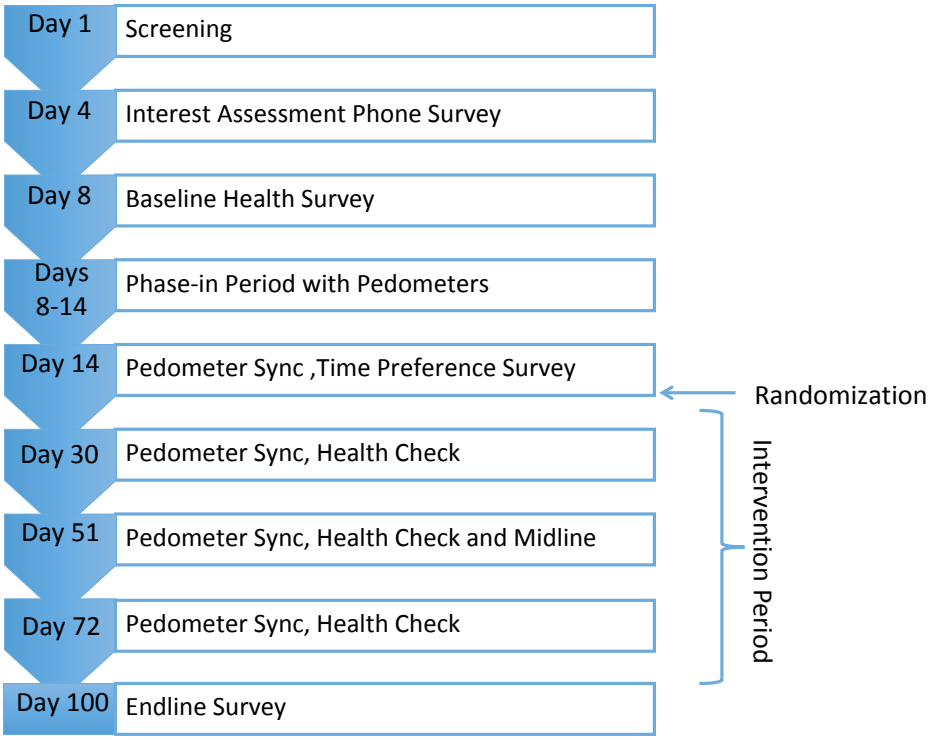


Figure 3: Experimental Timeline for a Sample Participant

Notes: This figure shows a representative experimental timeline for a participant in the experiment. Screening camps occurred throughout Coimbatore, Tamil Nadu from January 2016 to October 2017. In practice, visits were scheduled according to the availability of the respondent, leading to variation in the exact number of days between each visit. In addition, we intentionally introduced random variation into the timing of incentive delivery by randomly delaying the start of the intervention period by one day for selected participants. However, the intervention period was 12 weeks for all participants.

and "later" date. Each decision satisfies a budget set of the form

$$c_t + \frac{1}{r} c_{t+k} = m$$

where (c_t, c_{t+k}) are the chosen recharge amounts to be disbursed or steps to be taken on the sooner and later dates, respectively. The sooner date t, the time lag between the sooner and later date k, and the interest rate r vary across decisions. Within a domain for a given respondent, the total budget m is fixed across allocations.

In order to avoid potential biases in baseline time-preference measures, our CTB environment builds on a number of features from previous studies. First, the choices are made after the one-week phase-in period in which all participants have pedometers and report their daily steps, ensuring that participants are familiar with the costs of walking. This allows for meaningful allocations of steps between sooner and later dates. Second, to reduce bias from inattention, the responses are designed to be incentive compatible; all respondents were informed that we would implement their choice from a randomly selected survey question. We set the probabilities such that for most respondents, the randomly selected survey question was a multiple price list of lotteries over money (which measures risk preferences), but for a few, a CTB allocation was selected. Because the allocations might have interfered with any walking program offered, we excluded these respondents from the experimental sample. To ensure that participants complete the allocated steps, we offer a large cash completion bonus of 500 INR in the step domain if the allocation is selected to be implemented, and the steps are completed as allocated.

We take a number of precautions to avoid various potential confounds, including confounds reflecting fixed costs or benefits of taking an action, or confounds due to the time of day of measurement.[26] However, we were not able to fully address one potential confound

[26]In particular, to avoid confounds related to fixed costs or benefits, such as the effort of wearing a pedometer or the psychological benefit of receiving a free recharge, we include minimum allocations on both sooner and later days in each domain. The minimum allocations were chosen to be high enough that any fixed costs would be included (e.g. one could not easily achieve the minimums by simply shaking the pedometer), but low enough to avoid corner solutions. In the step domain, this required a novel modification of the CTB methodology: individual-specific minimum allocations. Our step allocations also featured individual-specific total step budgets m, which were chosen to be large enough that achieving them would require some effort beyond simply wearing the pedometer, but small enough that participants would certainly achieve them in exchange for the completion bonus. Specifically, minimum steps on each day are calculated as $\frac{X}{10}$, and the total step budget m is $X + 2\frac{X}{10}$, respectively, where $X \in \{3000, 4000, 5000\}$ is the element closest to the participant's average daily walking during the phase-in period. That is, minimum steps are one of 300, 400, or 500 on each day, and the total step budget is one of 3,600, 4,800, or 6,000. To avoid confounding impatience with the time of day that the baseline time-preference survey was administered (which could influence the desirability of walking and/or recharges delivered in the next 24 hours), as well as to capture heterogeneity in time preferences including any present-bias for very short beta-windows, we required that all walking on any date be conducted within a 2 hour period, which was chosen to start at the time immediately after the time-preference survey would end (e.g., if the survey ended at 4pm, the time period for any day's walking

to our estimates of time-preferences across individuals: variation across people in the cost of walking over time, or in the benefit of receiving a recharge over time. For example, an individual with a particularly busy week after the time-preference survey, and therefore relatively high costs to steps in the near-term relative to the distant future, will appear to be particularly impatient over steps in our data (he will wish to put off walking). An individual with a relatively free week just after the time-preference survey will instead appear particularly forward-looking (he will not wish to put off walking). The same concerns can also arise with recharges, but because recharges are storable and thus may be consumed on a different day than when they are received, we expect less variation (and heterogeneity) in the utility of recharge receipt over time.

We used a participant's decisions to construct two individual-specific structural parameter estimates of time-preference: one in the walking domain and one in the mobile recharge domain. In each domain we construct an estimate of the *Daily discount rate*, $\frac{1}{\delta} - 1$. This is a one-parameter estimate of the daily discount rate that is increasing both in time-consistent impatience and in present bias. Following Augenblick et al. (2015), our estimate is from a two-limit Tobit specification of the standard intertemporal Euler equation for an agent with an exponential daily discount factor δ, and concavity over recharges (or convexity over steps) α. Further details of the estimation methodology are described in Appendix A. [27,28] However, because the discount rate can only be estimated for individuals making interior choices, we cannot estimate our structural parameter for all individuals in our sample. As a result, we supplement the estimation with several survey-based measures of impatience and time preference taken from the psychology literature in order to demonstrate whether the effects we see in the structural parameter sample extend beyond it; note that these are not specific to the domain of walking but are meant to proxy for discount rates over

would be 5-7pm). The short window could potentially bias our overall measures of impatience downwards, as uncertainty about future schedules in a short time window could lead participants to want to get their walking done early when they had more certainty over their schedule. However, our primary purpose was to capture heterogeneity in time-preferences, and we considered the potential loss in validity of aggregate time preference estimates to be worth the ability to capture heterogeneity in time preferences in the time frames near to the present.

[27]Our predictions are generally about overall impatience, not about whether an individual is time-consistent, and so we want one summary measure capturing impatience over the time horizon. Estimating just one parameter has the advantage of avoiding overfitting, which is relevant in our setting because we use fewer CTB allocations than some of the US-based methodological papers on CTB.

[28]Other papers have also used the CTB data to estimate reduced-form measures (e.g., Giné et al. (2017) use the number of present-biased reversals). The structural measures have several advantages and so we choose to focus on them. First, the structural measure of impatience is increasing both in present-bias and in overall myopia, both of which are theoretically relevant to the performance of the contracts we offer. Second, whereas the standard reduced-form measures treat all preference reversals as equal regardless of magnitude, the structural measure takes into account the magnitude of impatience indicated by each decision. Third, by estimating the concavity (convexity) of preferences over recharges (steps), the structural measure avoids bias in the estimated discount rate from assuming linear utility (Andreoni and Sprenger, 2012a).

consumption.[29]

4.2 Summary Statistics

The baseline characteristics of the full experimental sample are reported in the first column of Table 1. Our sample is on average 49.42 years old, and has slightly more males than females. Their average annual household income is approximately 16,000 INR (about 200 USD) per month; for comparison, in 2015 the median urban household in India earned between 10,000 and 20,000 INR per month (Labor Bureau of India). Panel B shows that our sample is at high risk for diabetes and its complications: 67 % of the sample has been diagnosed with diabetes by a doctor, and 82 % have Hba1c levels which are strongly indicative of diabetes. The random blood sugar concentrations are also indicative of high diabetes risk. Note that Hba1c above 6.5 is considered diabetic, and RBS above 180 (even just after eating) is unlikely except among diabetic individuals; average Hba1c and RBS in our sample surpass both of these cut-offs. The sample also has high rates of common diabetes comorbidities: 41 % have hypertension (defined as systolic blood pressure above 140 or diastolic blood pressure above 90), and 61 % are overweight (defined as BMI above 25) at baseline.

Panel C shows that although baseline walking levels are below international daily walking recommendations of 10,000 steps per day, they are comparable to the average steps taken in many developed countries. On average, participants walked just under 7000 steps per day in the phase-in period. For comparison, Japanese adults also take approximately 7,000 steps per day, whereas adults in the United States take approximately 5,000 steps per day, and adults in western Australia take about 9,000 steps per day (Bassett et al., 2010).

Panel D of Table 1 reports measures of impatience measured using the CTB survey questions. First, we do not see evidence of impatience over recharges on aggregate. In particular, the average estimated daily discount rate over recharges is only 0.01, which is similar to monetary discount rates estimated using the CTB methodology in other settings (e.g. Andreoni and Sprenger (2012a) and Augenblick et al. (2015)). Second, individuals are quite impatient over steps: present-biased preference reversals are more common than future-biased reversals, and the average estimated daily discount rate over steps is 0.37. Our estimate of the average discount rate over steps is somewhat larger than effort discount rates estimated by Augenblick et al. (2015) using a similar CTB methodology. However, because the present time period in our CTB questions was only two hours whereas our future time periods were entire days, inflating the costs of present relative to future steps, it is likely that our measures of daily discount rates over steps are biased away from zero. In addition,

[29]Note that we only began measuring these latter measures partway through the data collection (at the point when we realized it was common for participants to choose non-interior solutions) and so the measures are available for only part of the sample.

the discount rates over steps show wide variation across individuals compared to Augenblick et al. (2015), suggesting that these estimates may be confounded heterogeneous differences in walking costs in the CTB allocation dates. Such a confound would add noise to our discount rate estimates over walking and attenuate the observed relationship between time preferences and incentive contract design.

Baseline health and time preferences are similar across treatment groups. Columns 1 and 2 of Table 1 show means for the pure control and monitoring groups, and Columns 5-10 show means separately for each incentive group, with standard deviations in parentheses. To explore whether randomization provided balance in these characteristics across the different groups, we test that all characteristics are jointly orthogonal to treatment assignment relative to the pure control group (Hansen and Bowers, 2008). We fail to reject that the coefficients on all characteristics are 0 in regressions of treatment assignment on characteristics, suggesting that balance was achieved.

4.3 Outcomes

Our outcomes are gathered from two datasets. The first is a time-series dataset of daily steps walked for each participant with a pedometer during the twelve-week intervention period. Because surveyors collect pedometers back from pure control participants after the phase-in period, we do not have daily steps for this group. Surveyors collect pedometer data at three separate "pedometer sync" visits during the intervention period and at the endline survey.[30]

A potential issue with the daily step data is that we only observe steps taken while participants wear the pedometer. Because participants in the incentives groups are rewarded for taking 10,000 steps in a day with the pedometer,[31] they have an additional incentive to wear the pedometer on days that they expect to walk more. This could lead to a potential selection issue: if the incentives group selectively makes an effort to wear the pedometer when they think they will walk more but the monitoring group does not, then we will see a spurious positive relationship between incentives and observed daily steps.

In order to minimize selective pedometer-wearing, we incentivize all monitoring and incentives participants to wear their pedometers even on days with few steps. We do this by offering a cash bonus of 200 INR (About 3 USD) if participants wear their pedometer (i.e., have non-zero recorded steps) on at least 70% of days in the intervention period. This

[30] In order to collect pedometer data, surveyors ask to see the pedometer, open the Fitbit web application on a wifi-enabled tablet computer, sign into a respondent-specific account, and upload the previous 30 days of daily pedometer step data to the Fitbit database. We later pull these data through the Fitbit application program interface (API) using a web application we designed for this study.

[31] Although incentives are delivered for steps reported, we cross-check step reports with actual pedometer data after every pedometer sync visit. Anyone found to be over-reporting is initially warned, and is eventually suspended from the program if the behavior continues.

Table 1: Baseline summary statistics in full sample and by treatment group.

	Averages of Baseline Characteristics by Treatment Group									
	Full Sample	Control	Monitoring	Incentives Pooled	Daily	Base Case	Monthly	4-Day TH	5-Day TH	10 INR
A. Demographics										
Age (from BL)	49.54	49.78	50.28	49.44	49.57	49.60	48.80	49.31	49.67	49.11
	(8.52)	(8.19)	(8.95)	(8.55)	(8.60)	(8.33)	(8.94)	(8.68)	(8.77)	(7.84)
Female (=1)	0.42	0.46	0.43	0.41	0.44	0.41	0.38	0.42	0.38	0.48
	(0.49)	(0.50)	(0.50)	(0.49)	(0.50)	(0.49)	(0.49)	(0.49)	(0.49)	(0.50)
Labor force participation	0.75	0.73	0.72	0.75	0.75	0.74	0.81	0.74	0.77	0.70
	(0.44)	(0.45)	(0.45)	(0.43)	(0.43)	(0.44)	(0.39)	(0.44)	(0.42)	(0.46)
Daily mobile usage (INR)	6.61	7.22	6.47	6.44	5.86	6.58	7.67	6.43	6.01	4.94
	(8.79)	(10.14)	(8.95)	(8.36)	(6.25)	(8.77)	(9.19)	(8.05)	(8.87)	(5.77)
Mobile balance (INR)	29.26	30.80	29.48	28.98	28.61	29.69	28.55	28.57	28.14	30.05
	(49.42)	(48.79)	(48.68)	(49.88)	(38.54)	(52.08)	(63.65)	(49.10)	(44.98)	(36.59)
Per capita income (INR/month)	4463	4488	4620	4447	4068	4477	4599	4454	4480	4341
	(3638)	(4483)	(3160)	(3447)	(2765)	(3496)	(3235)	(3590)	(3525)	(2615)
Private water source	0.67	0.65	0.68	0.67	0.66	0.69	0.63	0.65	0.70	0.67
	(0.47)	(0.48)	(0.47)	(0.47)	(0.48)	(0.46)	(0.48)	(0.48)	(0.46)	(0.48)
Household Size	3.91	3.94	3.82	3.91	3.92	3.89	3.74	3.96	3.96	3.58
	(1.62)	(1.54)	(1.51)	(1.64)	(1.45)	(1.70)	(1.59)	(1.64)	(1.68)	(1.29)
B. Health										
Diagnosed diabetic	0.67	0.67	0.68	0.66	0.62	0.68	0.62	0.67	0.68	0.59
	(0.47)	(0.47)	(0.47)	(0.47)	(0.49)	(0.47)	(0.49)	(0.47)	(0.47)	(0.50)
Hba1c (mmol/mol)	8.68	8.67	8.76	8.68	8.58	8.72	8.66	8.68	8.69	8.35
	(2.33)	(2.36)	(2.40)	(2.32)	(2.36)	(2.29)	(2.44)	(2.32)	(2.38)	(2.14)
RBS (mmol/L)	192.42	191.32	196.07	192.51	195.58	193.26	193.30	192.12	192.50	177.38
	(89.39)	(88.73)	(86.67)	(89.87)	(91.54)	(88.25)	(98.14)	(89.96)	(91.75)	(77.00)
Systolic BP (mmHg)	133.35	133.33	134.06	133.34	135.25	133.27	134.18	132.49	133.71	135.62
	(19.15)	(20.34)	(17.68)	(18.99)	(21.55)	(19.07)	(19.13)	(18.00)	(19.20)	(21.42)
Diastolic BP (mmHg)	88.47	88.54	88.53	88.46	89.30	88.19	88.60	88.23	89.01	90.00
	(11.11)	(11.50)	(10.10)	(11.09)	(12.79)	(10.75)	(10.10)	(10.73)	(11.96)	(13.19)
BL BMI	26.42	26.52	26.47	26.40	26.41	26.47	26.39	26.34	26.19	26.99
	(4.35)	(4.34)	(3.67)	(4.39)	(5.35)	(4.53)	(4.81)	(4.21)	(3.70)	(4.10)
HbA1c: Diabetic (=1)	0.82	0.82	0.81	0.82	0.77	0.84	0.79	0.81	0.82	0.77
	(0.38)	(0.38)	(0.39)	(0.38)	(0.42)	(0.36)	(0.41)	(0.39)	(0.38)	(0.42)
BP: Hypertensive (=1)	0.41	0.36	0.42	0.42	0.45	0.41	0.44	0.42	0.43	0.29
	(0.49)	(0.48)	(0.50)	(0.49)	(0.50)	(0.49)	(0.50)	(0.49)	(0.50)	(0.46)
Overweight	0.61	0.62	0.66	0.60	0.57	0.60	0.58	0.61	0.59	0.67
	(0.49)	(0.48)	(0.47)	(0.49)	(0.50)	(0.49)	(0.50)	(0.49)	(0.49)	(0.48)
C. Walking										
Pr(exceeded step)	0.25	0.25	0.24	0.25	0.25	0.23	0.27	0.26	0.25	0.27
target (phase-in)	(0.32)	(0.31)	(0.32)	(0.32)	(0.32)	(0.31)	(0.33)	(0.32)	(0.34)	(0.34)
Avgerage daily steps	6960	7015	6868	6960	6961	6795	7403	7069	6916	7018
(phase-in)	(4004)	(3972)	(3729)	(4038)	(4240)	(3984)	(3889)	(4047)	(4111)	(4195)
D. Time Preferences										
i. Mobile Recharges										
Discount rate	0.01	0.01	0.01	0.01	-0.01	0.01	0.02	0.01	0.01	0.01
	(0.09)	(0.08)	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.10)	(0.06)
ii. Steps										
Discount rate	0.36	0.21	0.28	0.41	0.18	0.47	0.27	0.49	0.29	0.11
	(3.76)	(0.71)	(1.03)	(4.34)	(0.56)	(4.96)	(1.02)	(5.29)	(1.66)	(0.47)
F-test for Joint Orthogonality										
F-statistic			0.87	1.17	1.19	1.05	1.16	1.13	1.16	0.88
Sample size										
Number of individuals	3,192	585	203	2404	166	902	164	794	312	66
Percent of sample	100.0	18.3	6.4	75.3	5.2	28.3	5.1	24.9	9.8	2.1

Notes: Hba1c (glycated haemoglobin) and RBS (random blood sugar) respectively measure medium- and short-term blood sugar concentrations. Systolic and Diastolic BP measure blood pressure. Hba1c, RBS, and weight are risk factors for developing diabetes and diabetes-related complications. High blood pressure and weight are cardiovascular risk factors. In each domain (mobile recharges and steps), the discount rate $\frac{1}{\delta_i} - 1$ is an individual-level measure of impatience estimated from a two-limit Tobit regression with the restriction that $\beta_i = 1$ and $\alpha_i = \alpha$. The F-statistic tests the joint orthogonality of all characteristics to treatment assignment relative to control.

approach largely addressed the issue: Figure 4 shows that the rates of pedometer-wearing are high and similar between treatment groups. However, despite this similarity, incentives group participants do wear their pedometers on a statistically significant 2% more days than monitoring group participants. To address this potential selection issue, we report Lee (2009) bounds when comparing pedometer data between the incentives and monitoring groups.[32]

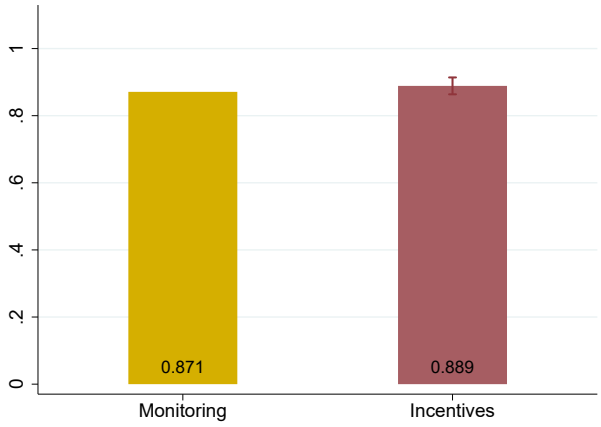


Figure 4: Fraction of days participants wore Fitbits

The second outcomes dataset – the endline survey – gathered health, fitness, and lifestyle information similar to the baseline health survey, as well as information about dietary and exercise behavior changes made during the intervention period. These data are available for participants in all treatment groups, including the pure control group. The primary purpose of these data is to allow us to study the impacts of the programs on health and healthy behaviors. Many of the measures (such as biometric health measures) have little margin for manipulation. There is no differential incentive for attending the endline survey across treatments, and although some respondents did withdraw from the study prior to the endline survey, Table B.1 shows that endline attrition rates are not statistically distinguishable between the pure control, monitoring, and incentives groups.

5 Results: Incentive Design

This section examines the effects of our incentive contract variations on exercise, and explores the implications of our results for incentive design in the presence of impatience. We begin by establishing that providing incentives increases exercise in this context; if it did

[32]Lee bounds assume that selection is monotonic, i.e. that being in the incentives group makes every participant weakly more likely to wear their pedometer on any given day or weakly less likely to do so. This assumption is likely justified in our setting, as being in the incentives group weakly increases the payoff to pedometer wearing by increasing the probability of reward.

not, then this would not be a good laboratory to explore the effects of varying the contract. We then explore the implications of time preferences for incentive design, first exploring the role of frequency and then of dynamic thresholds.

5.1 Incentives and exercise

We begin by establishing that our incentives program impacts exercise. This finding not only suggests that we can use the experiment to explore the design of incentives, but is also of independent policy interest: exercise has been shown to benefit health for diabetics (Hill, 2005; Praet and van Loon, 2009; Thomas et al., 2009; Zanuso et al., 2009; Shenoy et al., 2010; Manders et al., 2010; Qiu et al., 2014)and ours is the first randomized trial of exercise incentives in a developing country.

We use intent-to-treat (ITT) estimates to assess the impact of incentives for daily walking on exercise. In particular, we compare average exercise outcomes in the pooled incentives treatment groups to those in the monitoring treatment group. We focus here on the Fitbit exercise data as it is both less noisy than self-reported exercise and less prone to bias; as a result, we only evaluate the effect of incentives relative to monitoring, not control. Because monitoring may have an independent impact, this likely understates the policy impact of incentives overall, and these estimates should be interpreted as lower bounds on the effects of incentives relative to control. We return to exploring the effects of the monitoring group relative to the control group in Section 6.2.

For pedometer outcomes, which are measured at a daily frequency during the intervention period, we compare averages on the person-day level across treatment groups using regressions of the following form:

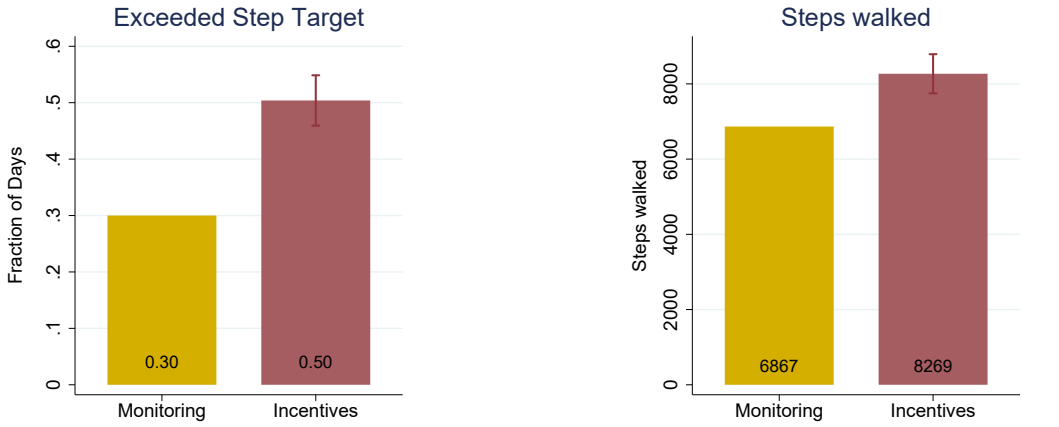
$$y_{it} = \alpha + \beta \times incentives_i + \boldsymbol{X}_i'\gamma + \varepsilon_{it}, \tag{5}$$

where y_{it} is either daily steps or an indicator for whether the individual surpassed the 10,000-step target, for individual i on day t during the intervention period; $incentives_i$ is an indicator for being in the incentives group; $\boldsymbol{X}_i$ is a vector of individual-specific controls;[33] and the standard errors ε_{it} are clustered at the individual level. The coefficient of interest, β, is the ITT effect of incentives relative to the monitoring group. The results are shown in Panel A of Table 2, and without individual-specific controls in Table B.2. The results are also shown graphically in Figure 5, where the confidence interval shown on the incentives bar is the 95% confidence interval for the gap between the incentives and monitoring groups (as is the case for all other graphs in this section).

We find that incentives have large and sustained impacts on walking during the inter-

[33]Individual-level controls are age, gender, weight, and phase-in walking.

vention period. Incentives increase the number of days that participants reach their 10,000 step target, and the size of the effect is large: Column 1 of Table 2 shows that incentivized participants exceed their step target on 20 % more days than those in the monitoring group (the Lee bounds are 19.1% to 23.8%). These effects are not simply a result of participants shifting steps from one day to another: Column 2 shows that incentives increase walking by 1291 steps per day (the Lee bounds are 934 to 1509 steps). This effect is equivalent to approximately 13 minutes of additional brisk walking daily, averaged across the intervention period.



(a) Step-target compliance

(b) Daily steps walked

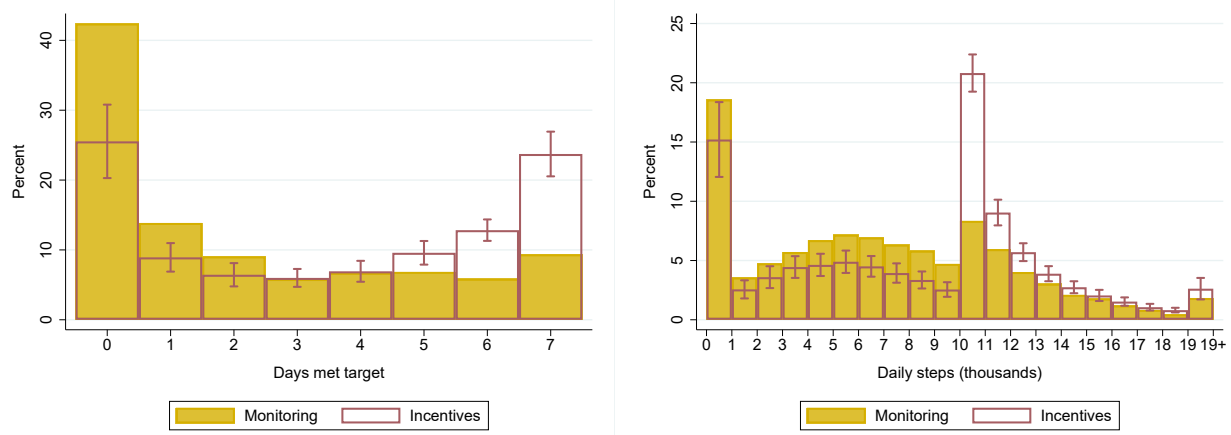
Figure 5: Incentives increase average walking

Note: Figure displays the impact of the pooled incentives treatments on steps during the intervention period, and the confidence interval for the test of equality between the incentives and monitoring groups. The dependent variable for the first panel is whether the participant met the daily step target of 10,000 steps on that day. The dependent variable in the second panel is the average steps walked.

To examine the impacts of incentives on walking routines, Figure 6a shows histograms of the number of days the step target was met per week (i.e., each data point is a respondent $\times$ week) in the monitoring and incentives groups. Relative to the monitoring group, the incentives group has a striking reduction in the number of weeks where the step target is never met and an equally striking increase in the number of weeks where the target is met on every day.

Figure 6b shows the impact of incentives on the distribution of daily steps. One can see that there is bunching at the 10,000 step level in both groups, but that the bunching in the incentive group is much more severe. Encouragingly, providing incentives also appear to shift the entire distribution of daily steps, rather than simply pushing marginal participants who would otherwise walk nearly 10,000 steps in a day over the 10,000-step target. There is less mass everywhere below the 10,000 step target, and more mass everywhere above the

target.



(a) Days per week exceeded step target

(b) Daily steps walked

Figure 6: Incentives shift the distributions of days walked per week and steps walked per day

Note: Figure displays the impact of the pooled incentives treatments relative to the monitoring group during the intervention period, with the confidence intervals representing the confidence interval for the test of equality between the incentives and monitoring groups.

The effect of incentives on walking also does not attenuate over time. Figure 7 shows that, if anything, total walking in the incentives groups increases over the course of the twelve-week program, a rare finding in the literature.[34]

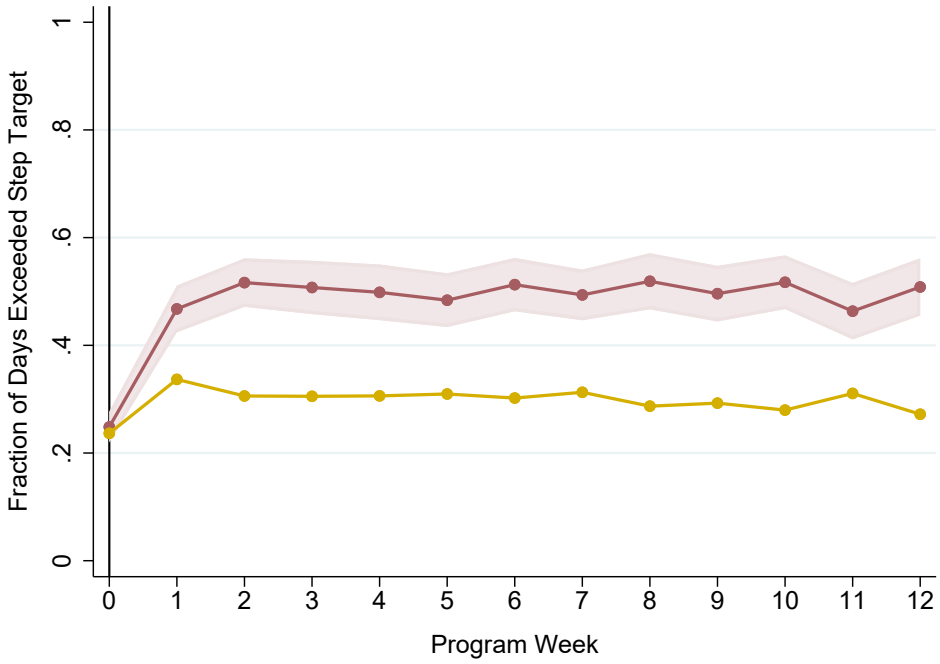
Having established that providing incentives affects behavior in this setting, we next use our experiment to explore the effectiveness of incentive contract variations designed to improve performance in the face of impatience over consumption and impatience over rewards, respectively.

5.2 Dynamic Thresholds

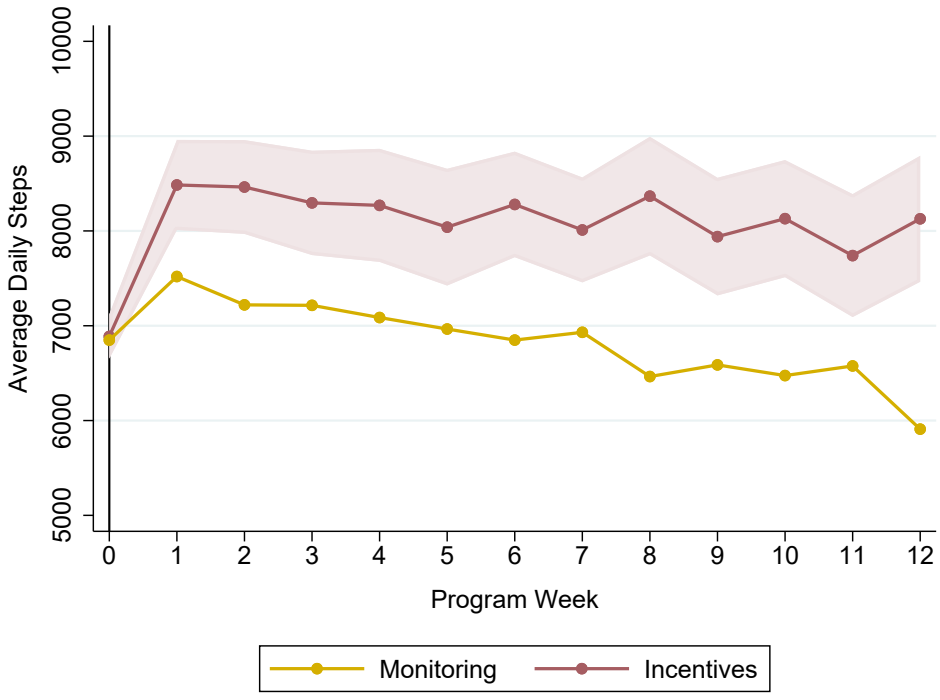
Our primary prediction for the effect of the dynamic threshold contracts is that relative to linear contracts, they should improve performance for those who are impatient over consumption; we do not have any predictions for the average effect. However, it is still useful to analyze the average effect of thresholds for several reasons: first, because our experiment

[34]Patel et al. (2016) find that physical activity drops steeply 5-7 weeks into a 12-week walking-incentive program. Regarding persistence after the intervention period ends, the literature finds mixed results: Although Charness and Gneezy (2009) find that the effects of a roughly 4-week incentive program incentivizing gym visits persists even 7 weeks after incentives are paid, Acland and Levy (2015) and Royer et al. (2015) find that the effects of similar interventions are indistinguishable from zero by 8 weeks after incentives are removed. Although we can only speculate, potential reasons why our program led to more sustained engagement include that we use pedometers which are hard to ignore; that participants needed to engage nightly with a reporting system that may have kept them engaged; and that our subjects may face fewer competing demands for their attention than participants in other studies.

Figure 7: The effects of incentives remain stable throughout the 12-week program



(a) Step-target compliance



(b) Daily steps walked

Notes: Panel A shows the average weekly probability of exceeding the step target over time for the monitoring and pooled incentives groups, and Panel B shows the steps walked per day averaged over each weekly period. Week 0 is the phase-in period, before randomization. The intervention period runs from Week 1 - Week 12.

Table 2: Impacts of incentives on exercise

	Pedometer Data (Intervention Period)		
	Proportion of Days Achieved 10K Steps	Daily Steps	Daily Steps (conditional on being > 0)
A. Pooled Incentives			
Incentives	0.197***	1291.3***	1159.8***
	[0.0182]	[207.8]	[186.2]
B. Unpooled Incentives			
Base Case	0.207***	1413.3***	1208.0***
	[0.0198]	[220.8]	[197.4]
Daily	0.201***	1149.7***	1190.3***
	[0.0305]	[324.5]	[272.5]
Monthly	0.180***	1284.3***	1210.6***
	[0.0284]	[304.6]	[264.3]
4-Day Threshold	0.189***	1214.1***	1113.0***
	[0.0204]	[228.2]	[202.5]
5-Day Threshold	0.210***	1312.8***	1226.7***
	[0.0250]	[261.6]	[230.0]
10 INR	0.129***	798.3**	528.5
	[0.0379]	[371.5]	[326.3]
Monitoring mean	0.302	6822.783	7986.135
Controls	Yes	Yes	Yes
P-value for Base Case vs			
Daily	.83	.34	.94
Monthly	.28	.61	.99
4-Day Threshold	.23	.18	.45
5-Day Threshold	.89	.61	.91
10 INR	.03	.06	.02
# Individuals	2,563	2,563	2,559
Observations	203,235	204,561	180,408

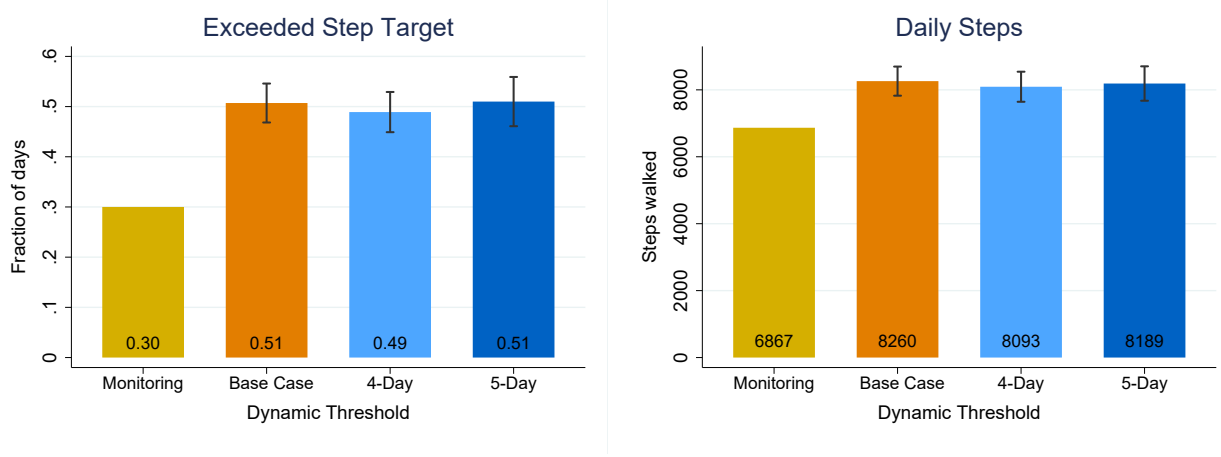
Notes: We report pooled incentive effects in Panel A, and separately by incentive treatment group in Panel B. The columns show coefficient estimates from regressions based on Equations 5 (Panel A) and 6 (Panel B), using daily panel data from pedometers during the intervention period. The sample includes the incentives and monitoring groups. Controls include age, gender, weight, and the average of the dependent variable during the phase-in period (before randomization). The omitted category in all columns is the monitoring group. Standard errors, in brackets, are clustered at the individual level.

represents (to our knowledge) the first randomized evidence comparing a dynamically non-separable contract to one without; and second, because understanding its effects is useful from a policy perspective. We thus begin by analyzing the average effect of threshold contracts relative to linear, before turning to test for heterogeneity in the impact of the threshold by time preferences.

5.2.1 Threshold Results: Average Effectiveness

We find that adding a dynamic threshold decreases the total amount of incentives delivered without sacrificing overall program effectiveness. Figure 8 and the fourth and fifth rows of Table 2) show that individuals in the 4-day threshold and 5-day threshold treatment groups exceed the 10,000 daily step target roughly as frequently as individuals in the base case incentive group (which has no threshold). For both threshold treatments, compliance with the step target is within 2 percentage points of compliance in the base case incentive (linear) treatment, with the difference statistically insignificant. The number of steps are similar across treatments as well.

However, individuals in the two threshold groups only receive a reward for exceeding the step target if they do so on at least 4 or 5 days in a weeklong payment cycle; when they walk on fewer than the threshold days, they are not rewarded. Because individuals with threshold contracts do not reduce overall walking, but are paid for a lower fraction of days walked, the threshold contracts we offer are more cost-effective than base case incentive contracts without a threshold.



(a) Probability Exceeded Step Target

(b) Average Daily Steps

Figure 8: Adding a dynamic threshold does not significantly affect average walking

Notes: Figures compare the effects of the dynamic threshold treatments with the "base case" (linear) incentive treatment. Panel A shows the average probability of exceeding the daily 10,000-step target during the intervention period; Panel B shows average daily steps walked during the intervention period. Confidence interval bars control for the phase-in value of the dependent variable, gender, age, and weight.

Table 3 quantifies the cost-effectiveness of the base case incentive, 10-INR, and threshold contracts in two ways. Column 5 shows the average incentive delivered on a day the participant exceeded the daily 10,000 step target. For the base case and 10-INR contracts, which pay out linearly according to the number of days the target is reached, this is just the incentive amount. However, as Column 4 shows, in the 4- and 5- day Threshold groups, participants are paid 91% and 86% of the days the achieve the step target, respectively. Thus, the incentive paid per day the target is reached is lower than in the base case (linear) group: 18 INR and 17 INR per day of compliance as compared to 20 INR. These cost savings of 8% and 14% are made while participants achieve nearly the same amount of walking. Column 6 Shows the average incentive cost per additional day the step target was reached above the monitoring group, which accounts for the possibility that the amount of walking that is inframarginal to the incentives might differ between the compliers for different contracts. According to this metric, the 4-day threshold and 5-day threshold achieve cost savings of 6.1% and 11.0%, respectively. For comparison, the incentive amount per day walked is mechanically lower in the 10-INR treatment group, but this comes at the cost of reduced walking overall.

Table 3: Cost Effectiveness of Monitoring and Incentive Treatments

	Cost-effectiveness of Incentive Contracts					
	Walking		Rewards		Cost-effectiveness	
	Compliance with Target	Treatment Effect	Incentive Amount (INR)	Proportion Compliance Incentivized	INR per Day Complied	INR per Day Complied above Monitoring
	(1)	(2)	(3)	(4)	(5)	(6)
Monitoring	30.2%	0	0	0	0	N/A
Daily	50.99%	20.8	20	1	20	49.04
Base Case	50.71%	20.52	20	1	20	49.44
Monthly	50.1%	19.9	20	1	20	50.34
4-Day Threshold	50.54%	20.34	20	.9	18.08	44.92
5-Day Threshold	51.7%	21.5	20	.85	17.07	41.04
10 INR	44.78%	14.58	10	1	10	30.71

One potential explanation for similar average walking in threshold and base case groups is that individuals simply do not notice the thresholds. However, the threshold contracts lead to markedly different walking patterns than the base case non-threshold group, showing

that individuals clearly understand and respond to the thresholds. Figure 9 shows that the threshold contracts have a large bimodal effect on walking: more individuals in the threshold contracts achieve their step target 7 days in a week or 0 days in a week. The bimodal treatment effect from thresholds is not simply a feature of behavior across weeks, but also appears across individuals. Figure 10 plots the density of each individual's probability of exceeding her step target, and mean daily steps, over the entire intervention period. The results across individuals mirror the results across weeks: the distribution of individual walking habits has thicker tails under the threshold treatments, with more people walking at the high and low ends. Appendix Table B.3 substantiates these conclusions using quantile regressions.[35] In sum, although thresholds do not work well for everyone, they work very well for some people, inducing them to walk with much more consistency across days, and for more steps per day, than non-threshold contracts.

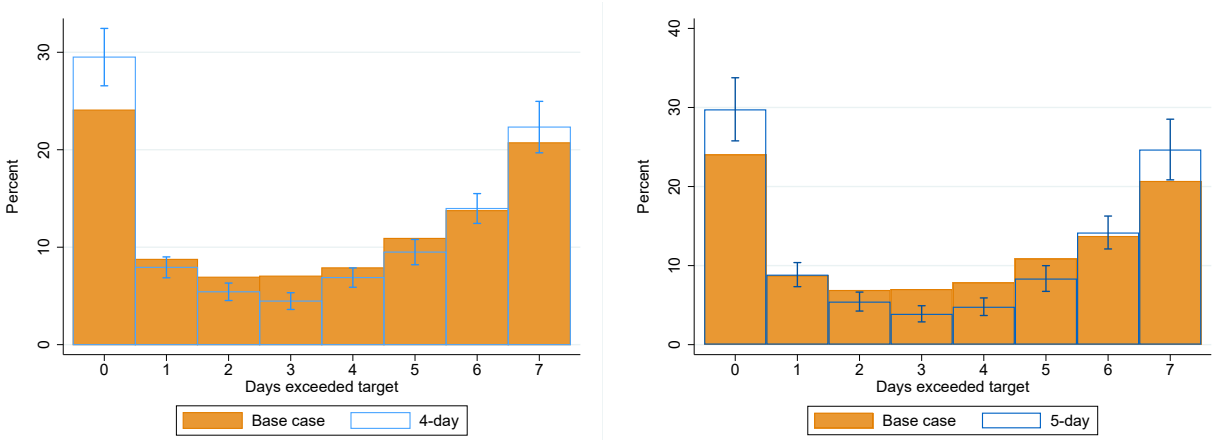


Figure 9: Days walked per week in threshold and base case (linear) contracts.

Notes: This figures shows the distribution of the number of days walked each week in the 4-day and 5-day threshold contracts and the base case (linear) contract during the intervention period. Effects control for average daily steps during the phase-in period, gender, age, and weight.

From a policy perspective, since threshold contracts create more extreme outcomes, we might be concerned if there are diminishing returns to behavior. In this setting, diminishing returns to exercise seem plausible, although the medical evidence is not definitive. If so, instituting a dynamic threshold creates a tradeoff: it decreases the cost per day of exercise induced, but perhaps also diminishes the health benefit per day of exercise induced.[36] The

[35]Specifically, we report a series of regressions of dummies for six average daily step quantiles, and six fraction-of-days-step-target-exceeded quantiles, on the different treatment groups. The fifth row of each panel shows the additional effect of the Threshold incentives contracts (pooled together) compared to the base case (linear) incentives. The coefficients are positive for the highest and lowest quintiles, but negative for intermediate quintiles, showing that the threshold treatments push more people to the extremes of walking behavior.

[36]On a similar note, one might think there would be greater value in inducing exercise among those who

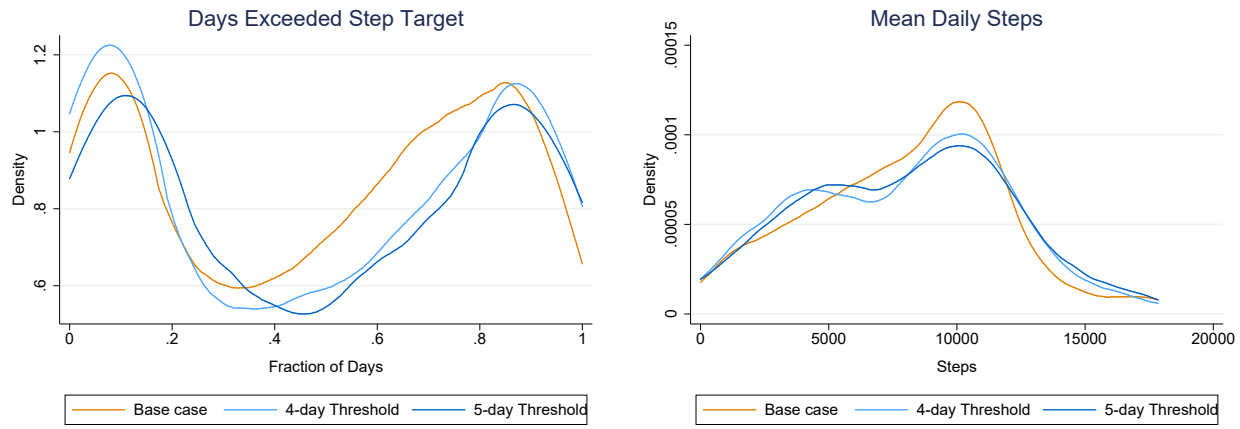


Figure 10: Fraction of days walked and average steps at the participant level, for base case vs. threshold

Notes: This figure shows the distribution of the fraction of days walked and average steps for each participant over the entire intervention period in each of the threshold contracts compared with the base case (linear) contract.

bimodal effects of thresholds also highlight the importance of understanding for whom they work best. We next proceed to test our theoretical prediction about one type of individual for whom the threshold contracts will work better: those who are impatient over walking.

5.2.2 Threshold Results: Time Preference Heterogeneity

Our theory from Section 2 predicts that the threshold will be more effective for those who discount future walking more heavily than current walking (Prediction 1). To test this, we regress an indicator for walking 10,000 steps (i.e., compliance with the incentivized activity) on our structural measure of impatience in the step domain, an indicator for being in a threshold treatment, and their interaction. We restrict the sample to only the base case incentive group and the 4- and 5-day threshold groups so that the only dimension that varies between groups is whether their contract has a dynamic threshold. Following our ex ante analysis plan, we pool the threshold treatments for power purposes (so *Threshold* is an indicator for being in either threshold group). The key coefficient of interest is on the interaction between the Threshold treatment and the measure of impatience over steps.

The evidence backs up our theory that impatience over steps play a role in the success of thresholds. Column 1 of Table 4 shows that those with higher impatience over exercise have higher compliance under the threshold treatment relative to the linear. Because the impatience measure (labeled "delta" in the table) often falls outside of the 0/1 range, to aid

have low levels of baseline exercise rather than those who already exercised a lot at baseline. Appendix Table B.5 and Figure B.1 explore heterogeneity by baseline walking, showing suggestive but weak evidence that thresholds increase exercise more for those who walk more at baseline, providing one other potential downside of threshold contracts.

Table 4: Thresholds are more effective for those who are more impatient over walking.

Dependent variable:	Met step target (×100)				
Impatience measure:	Delta	Above-median delta	Above-75th-perc. delta	Standardized self-control index	Correlates of self-control index
Sample:	Non-missing delta	Non-missing delta	Non-missing delta	Full	Full
Impatience × Threshold	0.389***	4.221	9.200**	5.292*	5.328***
	[0.134]	[3.368]	[3.804]	[2.999]	[1.998]
Impatience	-0.266**	0.928	-1.926	-5.754***	-3.889***
	[0.108]	[2.476]	[2.783]	[2.003]	[1.469]
Threshold	-1.518	-3.359	-3.690*	-1.865	-1.137
	[1.684]	[2.429]	[1.967]	[1.880]	[1.388]
Domain of Impatience	Steps	Steps	Steps	General	General
# Observations	110,932	110,932	110,932	85,246	156,351
Base Case mean	48.5	48.5	48.5	50.8	50.7

Notes: This table shows heterogeneity by time preferences in the effect of threshold contracts relative to linear contracts. The sample is restricted to the weekly groups – i.e., the base case incentive group, which has a linear contract, and the 2 threshold groups, 4-day threshold and 5-day threshold, pooled here together as "Threshold." The base case group is the omitted category. All columns control for gender, age, weight, and the baseline value of the dependent variable. The unit of observation is a respondent × day. Larger values of each impatience measure indicates more impatience. The first 3 columns all are based on the structural estimate of impatience in the step domain. Standard errors in brackets clustered at the respondent level. Significance levels: * 10%, ** 5%, *** 1%.

in interpretation of the magnitude, columns 2 and 3 show the results using indicators for whether impatience is above the 50th percentile and the 75th percentile, respectively. We lose some precision, but the columns show that the effect is large in magnitude: having above-median (above 75th percentile) increases compliance with the threshold contract relative to the base case contract by 4 percentage points (9 percentage points), which are large impacts relative to the average effect of incentives (20 percentage points). (Note that only the second coefficient is significant). Figure 11 shows the threshold effects visually by quartile of impatience: the effectiveness of the threshold seems to be relatively flat until the final quartile.

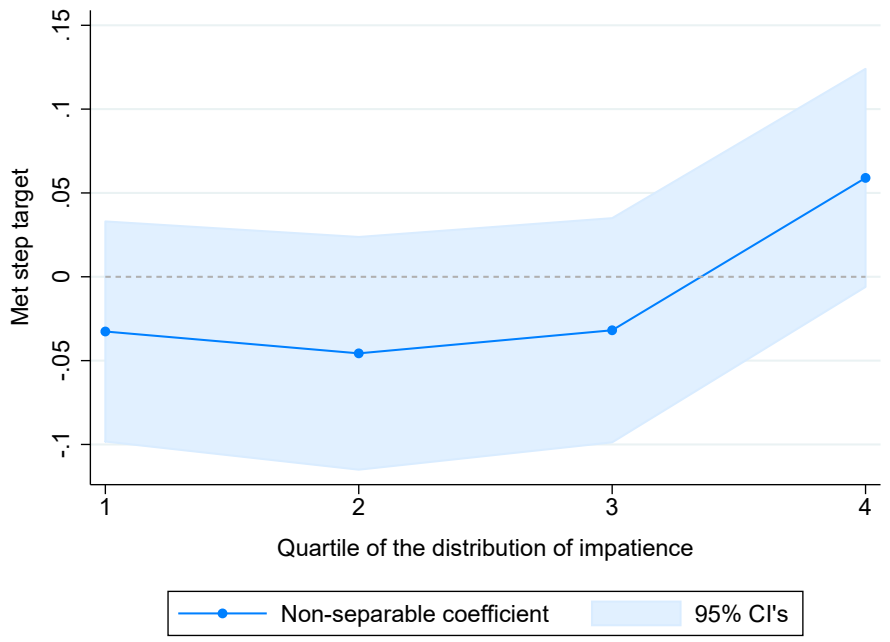


Figure 11: Threshold effect, by quartile of time preference

Notes: The sample is restricted to the weekly groups – i.e., the base case incentive group, which has a linear contract, and the 2 threshold groups, 4-day threshold and 5-day threshold, pooled here together as "Threshold." The chart shows the coefficient on a dummy for being in a threshold treatment from a regression of compliance on a dummy for being in a threshold treatment group, estimated separately by quartiles of the distribution of impatience. The base case group is the omitted category. All regressions estimated with control variables (gender, age, weight, and the baseline value of the dependent variable).

One concern with the structural impatience measure is that it can only be estimated for individuals making interior choices in the CTB allocation, and thus is missing for about 30% of the sample. We thus might be concerned that those who make interior choices are different than those who do not. Several weeks into the data collection for the project, we realized the prevalence of non-interior choices, and added to our baseline survey a set of questions on impatience and self-control from the psychology literature that proxy for impatience and

which we use to create a standardized index of impatience/self-control issues. Although these data are only available for participants who were enrolled after that point in the data collection (and thus precision is low), unlike for our structural measure, the sample with data available should be representative of the full sample. Reassuringly, we obtain a consistent finding using this standardized index, with the estimated coefficient in column 4 suggesting that the threshold works 5 pp better for those who are one standard deviation higher in the index. Finally, to get data from the full sample, we also create an index of procrastination-style questions that were included in the baseline survey from the beginning of data collection and that correlate with the index from column 4, and again we see consistent results (see column 5).

Of course, impatience is not randomly assigned and could correlate with other variables that influence the effectiveness of the threshold treatment. Luckily, Appendix Table B.6 shows that the impatience measure is not correlated with most other variables (e.g., risk aversion), somewhat assuaging this concern. We have two other approaches to address this concern. First, Appendix Table B.7 controls for other baseline covariates and their interactions with the threshold treatments. Reassuringly, the tables shows that the threshold interactions are relatively robust. Note that this test is in many senses too stringent ("over-controlling"), as some of these other control variables could also be downstream outcomes affected by impatience.

Second, we calibrate a model using the empirical distribution of walking costs to show that, in this setting, the performance of the threshold treatments should indeed increase meaningfully with impatience over exercise. We first extend the simple framework from Section 2 to cover a 7-day model with 4-day and 5-day thresholds. To calibrate the average compliance in the threshold and base case (linear) contracts, we need to estimate the distribution of walking costs, $F(\cdot)$. We do this by fitting a normal distribution to several moments from the data. In particular, the average walking in the monitoring treatment, in the 10 INR treatment, and in the 20 INR base case incentive group uncover $F(0)$, $F(10)$, and $F(20)$, respectively. We also use two additional moments: the probability of walking for the 4-day (5-day) threshold group when one had already walked 3 days (4 days) and it is the last day of the contract period to uncover $F(80)$ ($F(100)$).

We can then use this normal distribution to estimate how relative compliance in the base case (linear) and threshold contracts would vary with the discount rate over walking, d_c. The results are displayed visually in Figure 12, with the discount factor over walking on the x-axis, the gap between performance in the threshold and linear on the y-axis (shown separately for the 4-day and 5-day thresholds), and the figure shown separately for different scenarios of the discount rate over payments d_m. The figure confirms that, given the sample's distribution

of walking costs, increase in performance of the threshold contract as impatience increases should be quantitatively important. Note that the calibration overestimates the average effect of the threshold, likely at least in part because our simple model does not incorporate uncertainty over future walking costs and risk aversion, which would decrease the average performance of the dynamic threshold. However, these other factors should primarily affect the average effect of the dynamic threshold relative to the base case contract, and should not we believe meaningfully affect the heterogeneity by impatience, which is the main goal of this analysis.

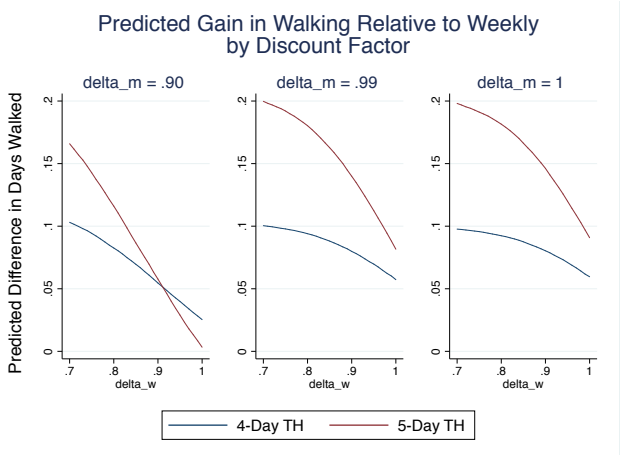


Figure 12: Calibration: Threshold compliance relative to the base case linear compliance, by the discount factor over walking

5.3 Payment Frequency

Motivated by Predictions 2-4 in Section 2, we conduct two primary tests:

1. Between-treatment: We compare average compliance between the daily, weekly (base case), and monthly payment groups. This allows us to both answer the policy question of how changing payment frequency changes compliance, and shed light on the shape of discount factors via Predictions 2 and 3.

2. Within-treatment: For each individual within the base case incentive group and monthly groups, we examine whether compliance increases as the payday approaches. Similar variation in worker effort as the payment date approaches has been used in previous studies, both in observational (Oyer, 1998) and experimental (Kaur et al., 2015) settings, to shed light on discount rates. (Prediction 4)

The approaches are complementary, each with its own advantages. The between-treatment approach allows us to directly answer the policy question of whether payment frequency matters, while the within-treatment approach exhibits higher statistical power and can shed light

on discount rates at a more detailed level. Note that since the shape of discount factors may vary across individuals, we test for heterogeneous effects both between- and within-treatment based on baseline measures of time preferences over recharges.

We begin with the between-treatment analysis in the full sample. Panel B of Table 2 evaluates the ITT effects of our incentive contract variations, estimating regressions of the following form:

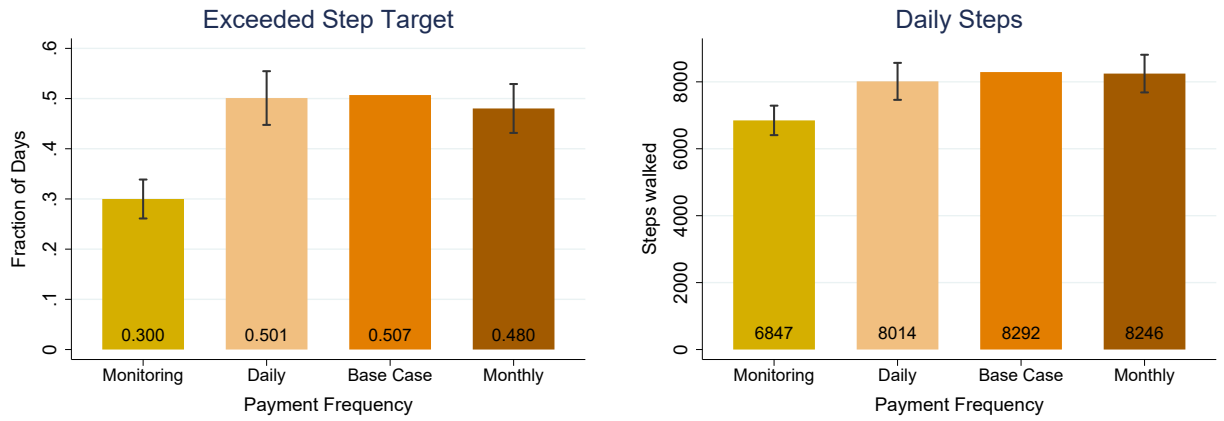
$$y_{it} = \alpha + \beta_j \times \left(incentives^j\right)_i + \boldsymbol{X}_i'\gamma + \varepsilon_{it}, \tag{6}$$

where y_{it} are daily walking outcomes and $\left(incentives^j\right)_i$ is an indicator for whether individual i is enrolled in incentive treatment group $j \in$ (daily, base case, monthly, 4-day threshold, 5-day threshold, 10 INR). Recall that all treatments besides the base case incentive vary from the base case contract on exactly 1 dimension (linearity, payment frequency, or payment amount); the bottom rows of the table thus show the p-values for testing for the significance of the difference between each treatment group and the base case incentive group.

We now compare average compliance between the daily, weekly (base case), and monthly groups. In addition to Panel B of Table 2, Figure 13 shows the compliance in the frequency treatments visually, with the confidence intervals showing the 95% confidence interval for a test of equality between the base case incentive group and each other treatment group. The monitoring group's compliance is also shown as a reference.

The impacts of the three frequency treatments on both the likelihood of exceeding the step target and on average steps walked are statistically indistinguishable, and the differences between the point estimates are relatively small. The treatment effects do not increase monotonically with frequency. Interpreting the magnitudes of the point estimates: participants receiving base case incentive delivery in fact walk the most, but the participants receiving daily incentive delivery walk nearly the same amount. The likelihood of achieving the step target is 2% lower in the monthly incentive delivery group than the daily group, but average steps are slightly higher.

We thus do not find any meaningful positive evidence that increasing payment frequency in the range from daily to monthly affects compliance, implying that the discount factor over financial rewards, $d_m(k)$, may be relatively stable over this range. However, there are two key caveats. First, our confidence intervals are relatively wide: we cannot rule out that daily has an effect 5 percentage points higher than the base case, or that monthly has one 8 percentage points lower. Second, although the point estimates for daily and base case are nearly identical, even the point estimate for the difference between monthly and the base case, 3 percentage points, is not trivial; for example, the effect of decreasing the payment level by 50% (which one can see by comparing the base case and 10 INR treatments in Table 2) is only 8 percentage points. We thus turn next to the within-treatment test, which has



(a) Probability Exceeded Step Target (b) Average Daily Steps

Figure 13: Payment frequency does not significantly impact walking.

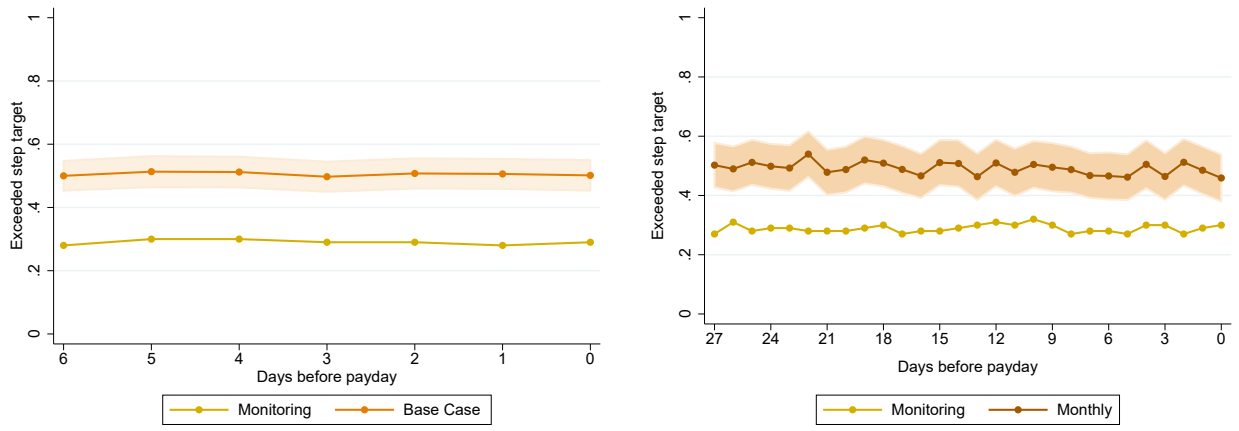
Notes: Panel A shows the average probability of exceeding the daily 10,000-step target during the intervention period for the 3 different frequency treatments (note that the "base case" treatment pays with weekly frequency); Panel B shows average daily steps walked during the intervention period. Confidence interval bars show tests for equality between each group and the base case incentive group, and come from regressions that control for the phase-in value of the dependent variable, gender, age, and weight.

somewhat higher statistical power, to confirm the suggestive evidence from this analysis that payment frequency does not meaningfully improve compliance.

Table 5: Frequency and Payday Effects

Dependent variable:	Met step target (×100)				
Payment Frequency:	Weekly		Monthly		
	(1)	(2)	(3)	(4)	(5)
Days before payday	0.135		0.0599		
	[0.0976]		[0.182]		
Payday		-0.811		-0.360	
		[0.585]		[1.093]	
Payweek					0.360
					[1.093]
# Observations	71,056	71,056	13,142	13,142	13,142
Sample mean	50.71	50.71	50.10	50.10	50.10

We next perform the within-treatment analysis. Figure 14 shows how compliance within the weekly (Panel A) and monthly (Panel B) treatments changes as the payment day approaches. The prediction of impatience over payments would be that compliance increases as the payday approaches. Instead, walking behavior is remarkably steady across the payment cycle, indicating that individuals do not differently discount rewards paid immediately and



(a) Weekly Payment Cycle

(b) Monthly Payment Cycle

Figure 14: The probability of exceeding the step target is stable over the payment cycle

Notes: Figures show the probability of exceeding the daily 10,000-step target among individuals receiving the base case, i.e., weekly, incentive (Panel A) and a monthly incentive (Panel B) relative to the monitoring group, according to days remaining until payday. Effects control for payday day-of-week fixed effects, gender, age, weight, and the phase-in value of the dependent variable. There is no evidence of a spike in walking on the day of incentive delivery for incentivized participants; in contrast, the slope of walking as the payday approaches is negative for participants in both the base case and monthly treatment groups.

paid one month in advance. Regression estimates of the slopes of each line are actually negative, although small in magnitude; for each day closer to the payday,base case participants are 0.2 percentage points less likely to comply, and monthly participants are 0.1 percentage points less likely. The 95% confidence intervals for those slopes rule out any positive discount rates, suggesting that aggregate discounting over payments is limited. Notably, given the between-treatment results, this means that monthly compliance is actually *lower* in the "payweek" (i.e., the last week leading up to the payday) than it is in the other weeks; this provides further evidence that the small negative point estimate of monthly relative to the base case seen earlier simply reflects noise, not meaningfully higher discounting for lags beyond 1 week.

Thus, consistent with the between-treatment evidence, the within-treatment evidence suggests that, on aggregate, the discounting model that best describes our participants is one of patience over mobile recharges within a 1-month time horizon; as a result, increasing frequency between daily and monthly does not have meaningful effects on average compliance. However, these are all average effects; it is possible that they mask heterogeneity across individuals. Appendix B tests for heterogeneity by discount rates over recharges in the between-treatment effects of payment frequency, as well as the within-treatment pattern as payday approaches. We do not find significant heterogeneity in either: even for those with high measured impatience over recharges, increasing payment frequency does not appear to

be effective.

6 Results: Program Evaluation

In this section, we evaluate the effects of the incentives and monitoring treatments on health and behavioral outcomes, and investigate whether monitoring alone affects exercise.

6.1 Health and Lifestyle Effects

The impacts of an incentives program on health and healthy behaviors are of independent policy interest, especially among a population at high risk for complications from non-communicable disease such as ours. Regular exercise such as walking can help prevent complications from diabetes, as well as hypertension. In addition, exercise may have coincident benefits for physical fitness and mental health. Finally, walking incentives programs may also impact other behaviors, either encouraging them (e.g., by increasing the salience of good health), or discouraging (e.g. if people substitute between healthy behaviors). We now assess the impacts of our programs on health and healthy behavior.

For looking at health outcomes – our primary outcomes of interest – our experiment was primarily powered to detect the difference between incentives groups (pooled) and the pure control group. Tables 6 through 9 report results from regressions of the following form:

$$y_i = \alpha + \beta_1 \times incentives_i + \beta_2 \times monitoring_i + \boldsymbol{X}_i'\gamma + \varepsilon_i \tag{7}$$

where y_i is a health or lifestyle outcome at endline for individual i; $incentives_i$ is an indicator for being in the incentives group; $monitoring_i$ is an indicator for being in the monitoring group; and $\boldsymbol{X}_i$ is a vector of controls, shown in the table notes. β_1 is the ITT effect of incentives relative to the pure control group, β_2 is the ITT effect of the monitoring relative to the pure control group, and, for benchmarking, β_3 is the ITT effect of being in the SMS reminders group.

This section reports ITT effects on outcomes in five categories: physical health, anaerobic fitness, mental health, diet, and addictive substance use. In order to maximize our power to detect overall effects on each category, we create a single index of all variables in each category by taking the simple average of each variable, standardized by the mean and standard deviation in the pure control group.[37] While we report regression estimates for each outcome individually, we focus on the category indices for inferring effectiveness.

6.1.1 Physical health and anaerobic fitness

Table 6 shows that the incentives program improves health indicators in this population. Column 1 presents the treatment effect on the "Health Risk Index", which averages the five

[37] We follow Kling et al. (2007), by imputing missings for each component using the sample mean.

health risk factors displayed in the table. We find a moderately sized statistically significant reduction in health risk of 0.06 SD's. Turning to the components of the index, we see average reductions in two measures of blood sugar at endline: Hba1c, a three-month weighted average of blood sugar levels; and random blood sugar (RBS), a measure of instantaneous blood sugar levels, although only the latter is significant, and only at the 10% level. Of the other 3 components, only waist circumference is individually significant (at the 10% level). The table also shows that monitoring alone did not seem to impact health.

To assess the size of our incentives treatment on health, we can also compare our effect sizes to the effects of other interventions in the literature. Although the Incentive treatment effects appear small, they are, in fact, relatively reasonable when compared to other interventions in terms of scalability, intensity, and cost. Appendix Table B.9 shows the effects sizes and intervention details of other SMS and exercise interventions. Note that the majority of studies that find larger effects on Hba1c utilize more intensive interventions that are both costly and are unlikely to be scalable.

Table 6: Impacts of incentives and monitoring on health risk factors.

	Health Risk Index	HbA1c	RBS	Mean Arterial BP	BMI	Waist Circum-ference
Incentives	-0.060**	-0.087	-6.12*	0.23	-0.065	-0.39*
	[0.026]	[0.071]	[3.42]	[0.43]	[0.049]	[0.24]
Monitoring	0.0038	-0.15	1.86	1.28*	0.034	-0.37
	[0.046]	[0.12]	[6.06]	[0.75]	[0.087]	[0.42]
Control mean	0.00	8.44	193.83	103.14	26.52	94.50
P-value: M = I	0.12	0.59	0.14	0.12	0.21	0.96
P-value: SMS = I	0.00	0.10	0.81	0.22	0.02	0.03
P-value: SMS = M	0.28	0.10	0.37	0.83	0.41	0.11
# Individuals	3,103	3,061	3,067	3,050	3,065	2,951

Notes: Standard errors in brackets. For the Health Risk Index, controls include age, gender, an SMS treatment dummy, and second order polynomials of all health variables underlying the index at baseline. We follow Kling et al. (2007) in constructing the index by imputing missings for each component using the sample mean. The Controls for all other outcomes include age, gender, SMS treatment, and a second order polynomial of dependent variable at baseline. The Health Risk Index is an index created by the average of endline Hba1c, RBS, MAP, BMI, and waist circumference standardized by their average and standard deviation in the control group. Hba1c is the average plasma glucose concentration (%), RBS is the blood glucose level (mg/dL), MAP is the mean arterial blood pressure (mm Hg), and BMI is the body mass index. The omitted category in all columns is the pure control group.

We next examine the ITT impacts of incentives on physical fitness. Our survey collected two measures of anaerobic fitness at baseline and endline: 4-meter timed walk, and standing five times from a sitting position (e.g., from a chair). Thus, a smaller value of either of these measures indicates greater anaerobic fitness. Table 7 shows that participants in the incentives groups are not meaningfully faster or slower at either the times walk or sit-stands, nor on our index of the two measures. Although is is surprising that walking does not have any detectable impacts on our measure of fitness, while having large impacts on exercise, this may partly be explained by the fact that our intervention motivated a low-intensity form of exercise, while we were only able to implement time trials of high-intensity, short-duration exercise activities in our surveys.

Table 7: Impacts of incentives and monitoring on fitness.

	Fitness Time Trial Index	Seconds to Walk 4m	Seconds for 5 Sit-Stands
Incentives	0.0064	0.015	-0.091
	[0.028]	[0.044]	[0.12]
Monitoring	0.025	0.040	-0.13
	[0.049]	[0.077]	[0.21]
Control mean	0.00	3.88	13.18
P-value: M = I	0.67	0.71	0.83
P-value: SMS = I	0.24	0.34	0.97
P-value: SMS = M	0.24	0.33	0.90
# Individuals	2,819	2,598	2,595

Notes: Standard errors in brackets. For the Fitness Time Trial Index, controls include age, gender, an SMS treatment dummy, and second order polynomials of both fitness variables underlying the index at baseline. Controls for all other outcomes include age, gender, SMS treatment, and a second order polynomial of dependent variable at baseline. A large value of Fitness Time Trial Index indicates low fitness: it is an index created by the average two trials of endline seconds to walk four meters, and the seconds to complete five sit-stands standardized by their average and standard deviation in the control group. The omitted category in all columns is the pure control group.

6.1.2 Mental health

We next turn to the ITT impacts of incentives on mental health. We measure mental health using seven questions adapted from the Rand 36-Item Short Form Survey (SF-36), a standard quality-of-life survey which has been validated for measuring emotional wellbeing in India (Sinha et al., 2013; Rajeswari et al., 2005). We selected questions related to emotional

health. Each question asks for the frequency of a feeling or event in the previous four weeks.[38] Answers are then recoded so that larger values indicate better mental health.

Table 8 shows that the incentives program significantly improve our index of mental health. In addition, the monitoring program has an even larger positive effect on mental health, although we cannot reject equality. Although many studies have found a positive association between exercise and mental health (Biddle, 2016), experimental evidence that exercise causes improvements in mental health is fairly scarce and mixed. Our result is novel experimental evidence that exercise can improve mental health among a general population.

Table 8: Impacts of incentives and monitoring on mental health.

	Mental Health Index	Felt Happy	Less Nervous	Peaceful	Energy	Less Blue	Less Worn	Less harm to Social Life
Incentives	0.073**	0.088**	0.026	0.055	0.061	0.015	0.090**	0.052*
	[0.032]	[0.045]	[0.044]	[0.047]	[0.048]	[0.043]	[0.039]	[0.030]
Monitoring	0.11*	0.072	0.12	0.092	0.035	0.13	0.17**	0.049
	[0.057]	[0.079]	[0.077]	[0.083]	[0.084]	[0.077]	[0.069]	[0.053]
Control mean	0.00	3.06	3.48	3.35	3.30	3.86	4.40	4.71
P-value: M = I	0.50	0.81	0.16	0.62	0.73	0.11	0.18	0.94
P-value: SMS = I	0.48	0.35	0.21	1.00	0.58	0.32	0.11	0.27
P-value: SMS = M	0.31	0.60	0.93	0.72	0.87	0.66	0.03	0.44
# Individuals	3,192	3,068	3,068	3,068	3,068	3,068	3,068	3,068

Notes: Standard errors in brackets. For the Mental Health Index, controls include age, gender, an SMS treatment dummy, and second order polynomials of all questions underlying the index at baseline. Controls for all other outcomes include age, gender, SMS treatment, and a second order polynomial of dependent variable at baseline (where available). The Mental Health Index is the average values of seven questions adapted from the Rand 36-Item Short Form Survey (SF-36), standardized by their mean and standard deviation in the pure control group. The omitted category in all columns is the pure control group.

6.1.3 Lifestyle

Finally, we examine the ITT impacts of incentives on two dimensions of a healthy lifestyle: diet, and the consumption of addictive goods. We do not find evidence that incentives lead to healthier diet choices or reduce consumption of addictive substances, as seen in Table 9, although most of the coefficients go in the hypothesized direction.

[38]For example, the "Felt happy" question asks: "In the previous four weeks, how often have you felt happy? All of the time, most of the time, a good bit of the time, some of the time, a little of the time, or none of the time?"

Table 9: Impacts of incentives and monitoring on lifestyle, relative to pure control.

Panel A. Healthy diet

	Healthy Diet Index	Wheat meals	Meals with vegeta-bles	Servings Fruit	-(Rice meals)	-(Junkfood pieces)	-(Spoons sugar in coffee)	-(Sweets yester-day)	Avoid un-healthy food
Incentives	0.051	0.027	0.060*	0.040	0.029	-0.017	-0.022	-0.028	0.0036
	[0.045]	[0.030]	[0.031]	[0.038]	[0.033]	[0.062]	[0.046]	[0.038]	[0.018]
Monitoring	0.021	0.018	0.078	0.060	-0.0075	0.13	-0.029	-0.048	-0.041
	[0.079]	[0.053]	[0.055]	[0.066]	[0.059]	[0.11]	[0.081]	[0.067]	[0.031]
Control mean	0.00	0.49	0.58	0.53	-2.34	-0.87	-1.08	-0.35	0.83
P-value: M = I	0.68	0.84	0.70	0.73	0.49	0.14	0.93	0.75	0.11
P-value: SMS = I	0.97	0.28	0.22	0.49	0.31	0.47	0.25	0.05	0.79
P-value: SMS = M	0.79	0.52	0.24	0.44	0.81	0.60	0.36	0.10	0.17
# Individuals	3,192	3,068	3,068	3,068	3,068	3,192	3,192	3,068	3,068

Panel B. Addictive consumption

	Addictive Good Consumption Index	Average Daily Areca	Average Daily Alcohol	Average Daily Cigarettes
Incentives	-0.010	0.035	-0.036*	-0.055
	[0.037]	[0.042]	[0.020]	[0.11]
Monitoring	-0.0058	0.013	-0.016	-0.019
	[0.065]	[0.074]	[0.036]	[0.19]
Control mean	0.00	0.13	0.11	1.02
P-value: M = I	0.94	0.75	0.54	0.83
P-value: SMS = I	0.50	0.89	0.10	0.60
P-value: SMS = M	0.65	0.73	0.44	0.59
# Individuals	3,192	3,068	3,068	3,068

Notes: Standard errors in brackets. For the two indices, controls include age, gender, an SMS treatment dummy, and second order polynomials of all questions underlying the index at baseline. Controls for all other outcomes include age, gender, SMS treatment, and a second order polynomial of dependent variable at baseline. The Healthy Diet Index is an index created by the average values of eight diet questions, standardized by their average and standard deviation in the control group. The Addictive Good Consumption Index is an index created by the average self-reported average daily consumption of areca, alcoholic drinks, and cigarettes, standardized by their average and standard deviation in the control group. A larger value indicates more consumption. A larger value indicates a healthier diet. The omitted category in all columns is the pure control group.

6.2 Monitoring and Exercise

The previous results suggest that the monitoring group had limited impact, although the results are somewhat imprecise. One may wonder whether this is because the monitoring treatment did not affect exercise, or whether the exercise impacts were too small to translate into measurable health impacts. In this section, we evaluate the impact of monitoring on walking, which is useful not just as a standalone exercise, but also to help us understand the impact of incentives on walking relative to control, since the Section 5.1 analysis only evaluated the effect of incentives relative to monitoring.

One potential evaluation strategy would be to compare endline self-reported walking data between the control and monitoring groups (note that we cannot use pedometer data as the control group did not have pedometers). However, using self-reported data is likely to bias the estimated monitoring effect upward since the monitoring treatment emphasized the importance of walking. Indeed, if we perform that analysis, the results appear very biased, and are hard to reconcile with the pedometer data. We thus do not discuss these results (presented for completeness in Online Appendix C).

Instead, we evaluate the effect of monitoring using a before-after design, comparing pedometer-measured walking in the monitoring group during the phase-in period (during which we had not given them a walking goal and just told them to walk the same as they normally do) to their behavior during the intervention period. This strategy will be biased either in the presence of within-person time trends in walking, or if the phase-in period directly effects walking behavior. We control for year-month fixed effects which helps address any time trends, but the latter concern is more difficult, as the phase-in period likely did increase walking above normal, either because of Hawthorne effects, or because participants received a pedometer and a step-reporting system, which are two of the elements of the monitoring treatment itself (the other three remaining that we can still evaluate are (a) a daily 10,000 step goal, (b) positive feedback for meeting the step goal through SMS messages and the step-reporting system, and (c) periodic walking summaries). Thus, we consider a pre-post comparison of walking in the monitoring group to be a lower bound of the monitoring program treatment effect.

One can visualize the variation used for our pre-post estimate in Figure 7. Walking increases immediately during the intervention period for the monitoring group, although the effects decay over time. We estimate the effect in Online Appendix Section C, controlling for time effects. The monitoring group achieves the 10,000 step target on approximately 7% more days in the intervention period than in the phase-in period, an effect significant at the 1% level and equal to roughly 35% of the estimated impact of incentives. However, the estimated effect on steps is only 62 steps (statistically indistinguishable from zero), which is

only 10% of the additional impact of incentives. The monitoring treatment thus appears to do more to make walking consistent across days than it does to increase total steps.

7 Conclusion

This paper investigates incentive design for impatient agents. From a model of agents with separate time preferences over consumption and financial rewards, we identify a feature of incentive contracts that will interact with impatience in each domain. In particular, our model predicts that the frequency of incentive delivery will interact with impatience over financial rewards, and whether the payment function is dynamically separable over time will interact with impatience over consumption.

In order to test our predictions, we implement an RCT to incentivize walking among approximately 3000 individuals with diabetes and prediabetes in India. Overall, the incentives program leads to a large increase in walking among the study population and leads to improvements in diabetes- and mental- health risk factors. This is encouraging evidence that exercise-incentives programs can be successful in a developing country setting.

We find evidence that individuals are impatient over consumption - that is, they prefer to put off the effort of walking. Consistent with our model, the dynamically non-separable contract works better for those who are more impatient over consumption. However, we find limited evidence of impatience over payments in our sample. Neither more frequent nor more immediate payment leads to increased walking behavior in our sample, and we cannot reject a null relationship between a survey-based measure of impatience over rewards and the effectiveness of more frequent payment. The finding that impatience is more prevalent in the consumption domain than the financial domain is consistent with previous experimental work (Augenblick et al., 2015). However, our finding that dynamically non-separable contracts can be used to motivate time-inconsistent individuals is a new and policy-relevant insight. The fact that these contracts have heterogeneous treatment effects opens up a key question, which we hope to address in future work: can we tailor incentive contracts to more cost-effectively encourage exercise?

References

Acland, D. and Levy, M. (2015). Naiveté, Projection Bias, and Habit Formation in Gym Attendance. *Management Science*, 61(1):146–160.

Andreoni, J., Callen, M., Khan, Y., Jaffar, K., and Sprenger, C. (2016). Using Preference Estimates to Customize Incentives: An Application to Polio Vaccination Drives in Pakistan. *NBER Working Paper*, 22019:1–66.

Andreoni, J. and Sprenger, C. (2012a). Estimating Time Preferences from Convex Budgets. *American Economic Review*, 102(7):3333–3356.

Andreoni, J. and Sprenger, C. (2012b). Risk Preferences Are Not Time Preferences. *American Economic Review*, 102(7):3357–3376.

Anjana, R. M., Pradeepa, R., Deepa, M., Datta, M., Sudha, V., Unnikrishnan, R., Bhansali, A., Joshi, S. R., Joshi, P. P., Yajnik, C. S., Dhandhania, V. K., Nath, L. M., Das, A. K., Rao, P. V., Madhu, S. V., Shukla, D. K., Kaur, T., Priya, M., Nirmal, E., Parvathi, S. J., Subhashini, S., Subashini, R., Ali, M. K., Mohan, V., and ICMRINDIAB Collaborative Study Group (2011). Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: Phase I results of the Indian Council of Medical ResearchINdia DIABetes (ICMRINDIAB) study. *Diabetologia*, 54(12):3022–3027.

Ariely, D. and Wertenbroch, K. (2002). Procrastination, Deadlines, and Performance: Self-Control by Precommitment. *Psychological Science*, 13(3):219–224.

Ashaf, N., Karlan, D. S., and Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, 121(2):635–672.

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working Over Time: Dynamic Inconsitency in Real Effort Tasks. *The Quarterly Journal of Economics*, 130(3):1067–1115.

Augenblick, N. and Rabin, M. (2017). An Experiment on Time Preference and Misprediction in Unpleasant Tasks. *Working Paper*, pages 1–81.

Bai, L., Handel, B. R., Miguel, E., and Rao, G. (2017). Self-Control and Demand for Preventive Health: Evidence from Hypertension in India. *Working Paper*.

Bassett, D. R., Wyatt, H. R., Thompson, H., Peters, J. C., and Hill, J. O. (2010). Pedometer-measured physical activity and health behaviors in U.S. adults. *Medicine and Science in Sports and Exercise*, 42(10):1819–1825.

Bertrand, M., Karlan, D. S., Mullainathan, S., Shafir, E., and Zinman, J. (2005). What's Psychology Worth? A Field Experiment in the Consumer Credit Market. *NBER Working Paper*, 11892.

Biddle, S. (2016). Physical activity and mental health: evidence is growing. *World Psychiatry*, 15(2):176–177.

Carvalho, L. S., Meier, S., and Wang, S. W. (2016). Poverty and Economic Decision - Making: Evidence from Changes in Financial Resources at Payday. *The American Economic Review*, 106(2):260–284.

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

Clark, G. (1994). Factory Discipline. *The Journal of Economic History*, 54(1):128–163.

Cutler, D. M. and Everett, W. (2010). Thinking Outside the Pillbox - Medication Adherence ans a Priority for Health Care Reform. *The New England Journal of Medicine*, 362(17):1553–1555.

Davis, L. W., Fuchs, A., and Gertler, P. J. (2014). Cash for coolers: Evaluating a large-scale appliance replacement program in Mexico. *American Economic Journal: Economic Policy*, 6(4):207–238.

Dellavigna, S. and Malmendier, U. (2004). Contract design and self-control: Theory and evidence. *The Quarterly Journal of Economics*, 119(2):135–146.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2010). Towards an Emissions Trading Scheme for Air Pollutants in India.

Duflo, E., Kremer, M., and Robinson, J. (2011). Nudging Farmers to Use Fertilizer : Theory and Experimental Evidence from Kenya . *American Economic Review*, 101(October):2350–2390.

Finkelstein, E. A., Brown, D. S., Brown, D. R., and Buchner, D. M. (2008). A randomized study of financial incentives to increase physical activity among sedentary older adults. *Preventive Medicine*, 47:182–187.

Fryer, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *Quarterly Journal of Economics*, 126(4):1755–1798.

Giné, X., Goldberg, J., Silverman, D., and Yang, D. (2017). Revising Commitments: Field Evidence on the Adjustment of Prior Choices. *Economic Journal*, Forthcomin.

Giné, X., Karlan, D., and Zinman, J. (2010). Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation. *American Economic Journal: Applied Economics*, 2(4):213–235.

Hansen, B. B. and Bowers, J. (2008). Covariate Balance in Simple, Stratified and Clustered Comparative Studies. *Statistical Science*, 23(2):219–236.

Hill, J. O. (2005). Walking and Type 2 Diabetes. *Diabetes Care*, 28(6):1524–1525.

Jayachandran, S., De Laat, J., Lambin, E. F., Stanton, C. Y., Audy, R., and Thomas, N. E. (2017). Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. *Science*, 357(6348):267–273.

Kaur, S., Kremer, M., and Mullainathan, S. (2015). Self-Control at Work. *Journal of Political Economy*, 123(6):1227–1277.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1):83–119.

Kullgren, J. T., Troxel, A. B., Loewenstein, G., Asch, D. A., Norton, L. A., Wesby, L., Tao, Y., Zhu, J., and Volpp, K. G. (2013). Individual- Versus Group-Based Financial Incentives for Weight Loss. *Annals of Internal Medicine*, 158:505–514.

Labhardt, N. D., Balo, J. R., Ndam, M., Manga, E., and Stoll, B. (2011). Improved retention rates with low-cost interventions in hypertension and diabetes management in a rural African environment of nurse-led care: A cluster-randomised trial. *Tropical Medicine and International Health*, 16(10):1276–1284.

Laibson, D. (2015). Why Don't Present-Biased Agents Make Commitments? *American Economic Review*, 105(5):267–272.

Lazear, E. P. (1979). Why Is There Mandatory Retirement? *Journal of Political Economy*, 87(6):1261–1284.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102.

Manders, R. J. F., Van Dijk, J. W. M., and Van Loon, L. J. C. (2010). Low-intensity exercise reduces the prevalence of hyperglycemia in type 2 diabetes. *Medicine and Science in Sports and Exercise*, 42(2):219–225.

Marshall, S. J., Levy, S. S., Tudor-Locke, C. E., Kolkhorst, F. W., Wooten, K. M., Ji, M., Macera, C. A., Ainsworth, B. E., Ming, J., Macera, C. A., and Ainsworth, B. E. (2009). Translating Physical Activity Recommendations into a Pedometer-Based Step Goal: 3000 Steps in 30 Minutes. *American Journal of Preventive Medicine*, 36(5):410–415.

Martins, N., Morris, P., and Kelly, P. M. (2009). Food incentives to improve completion of tuberculosis treatment: randomised controlled trial in Dili, Timor-Leste. *BMJ (Clinical research ed.)*, 339:b4248.

Morris, S. S., Flores, R., Olinto, P., and Medina, J. M. (2004). Monetary incentives in primary health care and effects on use and coverage of preventive health care interventions in rural Honduras: cluster randomised trial. *Lancet (London, England)*, 364(9450):2030–7.

O'Donoghue, T. and Rabin, M. (1999). Incentives for Procrastinators. *The Quarterly Journal of Economics*, 114(3):769–816.

Opp, M. M. and Zhu, J. Y. (2015). Impatience versus Incentives. *Econometrica*, 83(4):1601–1617.

Oyer, P. (1998). Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality. *The Quarterly Journal of Economics*, 113(1):149–185.

Patel, M. S., Asch, D. A., Rosin, R., Small, D. S., Bellamy, S. L., Eberbach, K., Walters, K. J., Haff, N., Lee, S. M., Wesby, L., Hoffer, K., Shuttleworth, D., Taylor, D. H., Hilbert, V., Zhu, J., Yang, L., Wang, X., and Volpp, K. G. (2016). Individual Versus Team-Based Financial Incentives to Increase Physical Activity: A Randomized, Controlled Trial. *Journal of General Internal Medicine*, 31(7):746–754.

Praet, S. F. E. and van Loon, L. J. C. (2009). Exercise therapy in Type 2 diabetes. *Acta Diabetol*, 46:263–278.

Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature*, 37(1):7–63.

Qiu, S., Cai, X., Schumann, U., Velders, M., Sun, Z., and Steinacker, J. M. (2014). Impact of walking on glycemic control and other cardiovascular risk factors in type 2 diabetes: A meta-analysis. *PLoS ONE*, 9(10).

Rajeswari, R., Muniyandi, M., Balasubramanian, R., and Narayanan, P. (2005). Perceptions of tuberculosis patients about their physical, mental and social well-being: a field report from south India. *Social Science & Medicine*, 60(8):1845–1853.

Ramachandran, A. (2013). Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India. *THE LANCET Diabetes & Endocrinology*, 1(3):191–198.

Royer, H., Stehr, M., and Sydnor, J. (2015). Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company. *American Economic Journal: Applied Economics*, 7(3):51–84.

Schilbach, F. (2017). Alcohol and Self-Control: A Field Experiment in India. *Mimeo*.

Shenoy, S., Guglani, R., and Sandhu, J. S. (2010). Effectiveness of an aerobic walking program using heart rate monitor and pedometer on the parameters of diabetes control in Asian Indians with type 2 diabetes. *Primary Care Diabetes*, 4(1):41–45.

Sinha, R., van den Heuvel, W. J. A., and Arokiasamy, P. (2013). Validity and Reliability of MOS Short Form Health Survey (SF-36) for Use in India. *Indian journal of community medicine : official publication of Indian Association of Preventive & Social Medicine*, 38(1):22–6.

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and time preferences: Experimental and household survey data from Vietnam. *American Economic Review*, 100(1):557–571.

Tharkar, S., Devarajan, A., Kumpatla, S., and Viswanathan, V. (2010). The socioeconomics of diabetes from a developing country: A population based cost of illness study. *Diabetes Research and Clinical Practice*, 89(3):334–340.

Thomas, D., Elliott, E. J., and Naughton, G. A. (2009). Exercise for type 2 diabetes mellitus (Review). *Cochrane Database of Systematic Reviews*, 2006(3).

Thornton, R. L. (2008). The Demand for, and Impact of, Learning HIV Status. *The American economic review*, 98(5):1829–1863.

Volpp, K. G., John, L. K., Troxel, A. B., Norton, L. A., Fassbender, J., and Loewenstein, G. (2008). Financial Incentive Based Approaches for Weight Loss: A Randomized Trial. *JAMA : the journal of the American Medical Association*, 300(22):2631–2637.

Whiting, D. D. R., Guariguata, L., Weil, C., and Shaw, J. (2011). IDF Diabetes Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Research and Clinical Practice*, 94(3):311–321.

Zanuso, S., Jimenez, A., Pugliese, G., Corigliano, G., and Balducci, S. (2009). Exercise for the management of type 2 diabetes: a review of the evidence. *Acta Diabetol*, 47:15–22.

A Baseline Time Preference Measurement

We follow earlier work and use a Convex Time Budget (CTB)methodology to estimate time preferences. In each CTB choice of the time-preference survey, the participant is asked to allocate a fixed budget of either steps or mobile recharges between a "sooner" and a "later" date using a slider bar. In particular, each choice allows the respondent to choose an allocation of consumption on the sooner and later dates, c_t, c_{t+k} that satisfies the budget constraint

$$c_t + \frac{1}{r} c_{t+k} = m \tag{8}$$

where the sooner date t, the later date $t + k$, the interest rate r, and the budget m change between each choice. A sample slider screen allowing for such choices is shown in Figure A.1.

Figure A.1: Sample decision screen for mobile recharges. In this example, the interest rate, r, is 1.25; the total budget, m, is 140; the "sooner" date is Today (0 days from today); and the "later" date decreases from 5 days from today in the first choice to 1 day from today in the final choice. The sliders are shown positioned at the choice ($c_t = 70, c_{t+k} = 82$).

We asked participants to make six allocations in the recharge domain, and eight allocations in the step domain, as summarized in Table A.1. We assume a time-separable and good-separable CRRA utility function with quasi-hyperbolic discounting. In the domain of recharges, individuals will then seek to maximize utility,

$$U\left(c_t, c_{t+k}\right) = \frac{1}{\alpha}\left(c_t - \omega\right)^{\alpha} + \beta\delta^k \frac{1}{\alpha}\left(c_{t+k} - \omega\right)^{\alpha} \tag{9}$$

and in the step domain, individuals will seek to minimize costs of effort

$$C\left(c_t, c_{t+k}\right) = \frac{1}{\alpha}\left(c_t + \omega\right)^{\alpha} + \beta\delta^{k}\frac{1}{\alpha}\left(c_{t+k} + \omega\right)^{\alpha} \tag{10}$$

The variation between consumption choices choices given different parameters of the budget constraint identify the daily discount factor δ, the present-bias parameter β, and the concavity or convexity of preferences α in each domain. We recover structural estimates of time preference and concavity parameters from the allocations, (c_t, c_{t+k}), using a two-limit Tobit specification of the intertemporal Euler condition following Augenblick et al. (2015).

Appendix Table A.1: This table summarizes the parameters of the six CTB allocations made over recharges, and the eight CTB allocations made over steps.

Summary of Convex Time Budget allocations					
Question no.	t	k	r	Recharge Domain	Step Domain
1	7	7	1	X	X
2	0	7	1	X	X
3	0	5	1.25	X	X
4	0	3	1.25	X	X
5	0	2	1.25	X	X
6	0	1	1.25	X	X
7	7	7	1.25		X
8	0	7	1.25		X

We estimate two time-preference measures from our CTB allocations in each domain: a reduced-form measure following Giné et al. (2017), and a structural measure following Augenblick et al. (2015). The reduced-form measure relies on allocation pairs with the same interest rate r and lag k, but where the "sooner" date of only one allocation is in the present $(t = 0)$, whereas for the other allocation the "sooner" date is in the future $(t > 0)$. If the allocation to the "sooner" date is larger when the "sooner" date is in the present, it is a present-biased preference reversal; if the allocation to the "sooner" date is smaller, it is a future-biased preference reversal. Our measure is the fraction of question pairs with present-biased reversals less the fraction with future-biased reversals in each domain. We call this measure net fraction of static present-biased preference reversals; its average in each domain is reported in Column 3 of Table A.2.

We also estimate the daily discount rate using two-limit Tobit specifications of the standard intertemporal Euler equation for an agent with an exponential daily discount factor δ, and concavity over recharges, or convexity over steps, α.[39]

[39]Details on the estimation strategy can be found in the Online Appendix of Augenblick et al. (2015).

$$\log\left(\frac{c_t}{c_{t+1}}\right) = \frac{\log(\delta)}{\alpha - 1}k + \frac{1}{\alpha - 1}\log(r) \tag{11}$$

Appendix Table A.2: Average fraction of present- and future-biased preference reversals, and net fraction of present-biased preference reversals, in the recharge and step domains.

Average Fraction of Preference Reversals			
	Recharges	Steps	Total
Fraction Present-Biased Reversals	0.185	0.313	0.249
Fraction Future-Biased Reversals	0.262	0.115	0.189
Net Fraction Present-Biased Reversals	-0.077	0.198	0.060
Observations	6433		

B Heterogeneity in frequency effects by impatience over recharges

This appendix explores the possibility that immediate incentive delivery is a driver of incentive effectiveness among the subset of more impatient participants. If so, we expect a positive interaction between more immediate incentive delivery and our measure of baseline impatience over mobile recharges. We test this interaction using both between-treatment and within-treatment variation in immediacy of payment.

Our first test is whether daily incentives are relatively more effective, and monthly relatively less effective than the base case of weekly payments, for those who display more impatience for recharges in their baseline CTB allocations. For simplicity, we restrict the sample to those who were in the daily, weekly, and monthly groups, and run the following regression:

$$\begin{aligned} y_{it} = &\alpha + \beta_0 Impatience_i + \beta_1 daily_i + \beta_2 monthly_i \\ &+ \beta_3 Impatience_i \times daily_i + \beta_4 Impatience_i \times monthly_i + \boldsymbol{X}_i'\gamma + \varepsilon_{it}, \end{aligned} \tag{12}$$

where y_{it} is a daily walking outcome; $Impatience_i$ is either the daily discount rate estimated using CTB allocations over recharges at baseline or an indicator for having above-median daily discount rate; and $daily_i$ and $monthly_i$ are indicators for being assigned to the daily and monthly treatments, respectively. β_1 and β_2 represent the effects of daily and monthly

relative to the base case weekly payment (respectively). The coefficients of interest are β_3 and β_4, showing whether the effects of daily or monthly relative to Weekly are differentially large for those who are more impatient. If impatience over recharges is a mechanism through which more immediate incentive delivery increases effectiveness, then we expect the daily treatment to be more effective ($\beta_3 > 0$) and the monthly treatment to be less effective ($\beta_4 < 0$) for more impatient individuals.[40] Our results are reported in Table A.3. We see no evidence that suggests that sooner payments work better for those with higher measured impatience, with the one marginally significant effect going the wrong direction.

Our second test is whether individuals who display more impatience for recharges in their baseline CTB allocations are more likely to increase step-target compliance on their payday. We perform this test among individuals in the base case incentive and monthly incentives groups. Following Kaur et al. (2015), we define individual-specific walking "payday effects" as the difference in the probability of exceeding 10,000 steps on paydays compared to all other days. The walking payday effect is a revealed-preference measure of impatience over rewards. We estimate the interaction between individual payday effects and our structural measure of baseline impatience over recharges using regressions of the following form:

$$\begin{aligned} y_{it} = &\alpha + \beta_0 \left(Impatience\, Measure\right)_i + \beta_1 \left(Payday\right)_{it} \\ &+ \beta_2 \left(Payday\right)_{it} \times \left(Impatience\, Measure\right)_i + \boldsymbol{X}_i'\gamma + \varepsilon_{it}, \end{aligned} \tag{13}$$

where y_i, $(Impatience\, Measure)$, and $\boldsymbol{X}_i$ are defined as in equation 12; and $(Payday)_{it}$ is an indicator for whether day t is a payday for individual i. To test whether more impatient individuals respond more to more immediate payment, we test whether $\beta_2 > 0$.

Our results are shown in Table A.4. Both measures of impatience predict larger payday effects, but the coefficients are not statistically different from zero. Thus, we find no strong evidence that even those individuals who are most impatient over rewards react to more immediate reward delivery over the payment cycle.

[40]Note that we do not have predictions for the interactions of the other incentive contracts with impatience over recharges; nonetheless, for completeness, Appendix Table B.4 shows regressions where the *Impatience* variable is interacted with all separate incentive treatments.

Appendix Table A.3: High-frequency treatments are not more effective for those who are impatient over recharges

Dependent variable:	Met step target	
Impatience measure:	Delta	Above-median delta
	(1)	(2)
Daily × Impatience	-0.275 [0.42]	-0.120* [0.07]
Monthly × Impatience	-0.0379 [0.41]	0.0286 [0.07]
Daily	-0.00396 [0.04]	0.0498 [0.05]
Monthly	-0.0254 [0.03]	-0.0406 [0.05]
Impatience	-0.111 [0.17]	-0.00398 [0.03]
Domain of Impatience	Recharges	Recharges
Methodology	Discount Rate	Discount Rate
Base Case: mean	.507	.507
# Daily NTH	92	92
# Base Case	483	483
# Monthly NTH	90	90
# Individuals	665	665
Observations	53,066	53,066

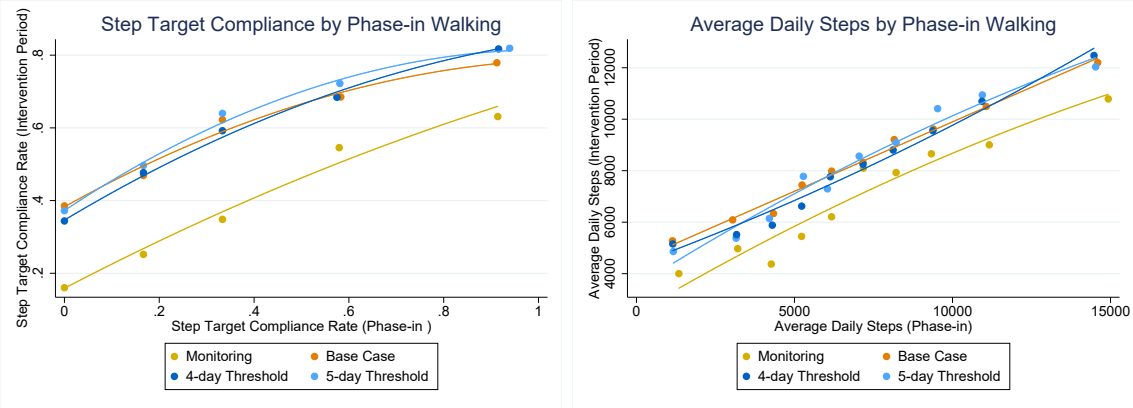
Notes: This table shows heterogeneity in the effect of the frequency subtreatments by treatment effects of each incentive non-threshold treatment, interacted with measures of impatience over steps; the base case incentive group is omitted. Standard errors clustered at the individual level in brackets. Controls include respondent's gender, age, and weight as well as the average phase-in period value of the dependent variable and month-year fixed effects. "Discount Rate" indicates a structural measure of the daily discount rate $\frac{1}{\delta_i} - 1$ estimated from a two-limit Tobit model of CTB allocations with individual discount-rate fixed effects, restricting the present-bias parameter β to be one. Larger values of each impatience measure indicates more impatience. The unit of observation is a respondent × day. Standard errors in brackets clustered at the respondent level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.4: Payday effects are not bigger for those with higher measured impatience over recharges

Dependent variable:	Met step target	
Impatience measure:	Delta	Above-median delta
	(1)	(2)
Impatience	-0.144	-0.00355
	[0.16]	[0.02]
Payday	-0.00559	-0.0110
	[0.01]	[0.01]
Impatience × Payday	0.0718	0.0127
	[0.09]	[0.02]
Controls	X	X
Domain of Impatience	Recharges	Recharges
Methodology	Tobit FE	Tobit FE
Dep. var. mean	0.51	0.51
# Base Case	484	484
# Monthly	90	90
# Individuals	574	574
Observations	45,979	45,979

Notes: This table shows heterogeneity in the "payday" effects for those in the base case incentive and the monthly incentive groups, by impatience in the recharge domain. Payday effects are defined as the difference in a daily exercise behavior on paydays compared to all other days. Standard errors clustered by individual are in brackets.

C Supplementary Tables and Figures



Appendix Figure B.1: Effects on Threshold by Quantiles of Phase-in Walking.

Appendix Table B.1: Endline attendance in full sample and by treatment group.

	Full Sample		Control	Monitoring	Incentives
A. Endline attendance by treatment status					
Attended EL	.97		.971	.97	.97
p-value: Control=Monitoring=Incentives			0.99		
B. Endline attendance by baseline characteristics					
Monitoring	-0.0005	0.003			
	[0.01]	[0.02]			
Incentives	-0.0009	-0.02			
	[0.008]	[0.01]			
Age (from BL)		0.0004	-0.002	0.0007	0.0005
		[0.0005]	[0.001]	[0.001]	[0.0006]
Female (=1), assume male if not female (from BL)		0.02**	0.005	-0.02	0.03**
		[0.01]	[0.02]	[0.02]	[0.01]
Labor Force Participation		0.02*	0.03	0.01	0.02
		[0.01]	[0.02]	[0.03]	[0.02]
Per capita income (1000 INR/month)		-0.003**	0.006*	-0.003	-0.004**
		[0.001]	[0.003]	[0.003]	[0.002]
Current Mobile Balance (INR)		0.0002**	0.00008	-0.00002	0.0003**
		[0.00010]	[0.0002]	[0.0002]	[0.0001]
Previous Day's Mobile Usage (INR)		-0.002***	0.0002	-0.007***	-0.002***
		[0.0005]	[0.0009]	[0.0009]	[0.0007]
Previously diagnosed diabetic [assume not if waiting for diagnosis] (=1)		-0.008	0.009	0.03	-0.01
		[0.010]	[0.02]	[0.02]	[0.01]
BP: Hypertensive (=1)		0.01	0.002	0.004	0.02
		[0.009]	[0.02]	[0.02]	[0.01]
Health risk index		-0.007	-0.0007	0.02	-0.01
		[0.008]	[0.01]	[0.02]	[0.01]
Share of days met step target in fitbit data during phase-in period		0.02	0.08	-0.02	0.01
		[0.03]	[0.06]	[0.06]	[0.03]
Average daily fitbit steps (/1000) during phase-in period		-0.000002	-0.005	0.006	0.0008
		[0.002]	[0.004]	[0.005]	[0.002]
Discount rate (recharges)		-0.02	0.0003	0.09	-0.007
		[0.05]	[0.1]	[0.1]	[0.07]
Discount rate (steps)		-0.002	0.01	0.005	-0.005
		[0.007]	[0.02]	[0.01]	[0.008]
Constant	1.0***	1.0***	1.0***	1.0***	0.9***
	[0.007]	[0.03]	[0.06]	[0.07]	[0.04]
Sample size					
Number of individuals	3,192	3,192	585	203	2,404

Appendix Table B.2: Impacts of incentives and monitoring on exercise outcomes, without baseline controls.

	Pedometer Data (Intervention Period)		Self-Reported Data (at Endline)	
	Fraction Days Achieved 10K Steps	Daily Steps	Fraction Days Exercised in Previous Week	Minutes Walked for Exercise Yesterday
A. Pooled Incentives				
Incentives	0.20***	1351.6***	0.057*	4.72**
	[0.023]	[260.7]	[0.034]	[2.34]
Pure Control			-0.13***	-7.94***
			[0.037]	[2.45]
B. Unpooled Incentives				
10 INR	0.15***	859.2*	-0.0043	2.91
	[0.049]	[517.6]	[0.065]	[4.89]
Daily	0.21***	1207.9***	0.027	6.24
	[0.035]	[386.5]	[0.047]	[3.89]
Weekly	0.21***	1367.4***	0.054	4.46*
	[0.024]	[276.1]	[0.036]	[2.52]
Monthly	0.20***	1572.0***	-0.029	4.48
	[0.035]	[392.8]	[0.048]	[5.09]
4-Day Threshold	0.20***	1342.2***	0.090**	6.19**
	[0.025]	[286.5]	[0.036]	[2.57]
5-Day Threshold	0.21***	1390.4***	0.053	1.37
	[0.030]	[335.0]	[0.041]	[2.89]
Pure Control			-0.13***	-7.94***
			[0.037]	[2.45]
Monitoring mean	0.30	6822.78	0.50	22.33
Controls	No	No	No	No
# Monitoring	199	200	195	195
# 10 INR	64	64	62	62
# Daily NTH	163	163	161	161
# Base Case	891	890	867	867
# Monthly NTH	163	163	160	160
# 4-Day TH	778	778	759	759
# 5-Day TH	305	305	293	293
# Control	0	0	568	568
# Individuals	2,563	2,563	3,065	3,065
Observations	203,235	204,561	3,065	3,065

Notes: Standard errors in brackets. The first two columns use daily panel data from pedometers, and standard errors are clustered at the individual level. The second two columns use a cross-section of self-reported data at endline. The omitted category in all columns is the monitoring group.

Appendix Table B.3: Impacts of incentives contracts, compared to the base case non-threshold contract, on the probability of being in 6 quantiles of average exercise outcomes.

	Differential Effects of Incentive Contracts on the Distribution of Exercise					
Outcome Quantile	1	2	3	4	5	6
A. Average Step-Target Compliance						
Incentives	-0.077***	-0.028	-0.10***	-0.030	0.021	0.23***
	[0.022]	[0.023]	[0.021]	[0.028]	[0.028]	[0.035]
Incentives X (10 INR)	-0.024	0.013	0.050	0.042	-0.042	-0.055
	[0.037]	[0.038]	[0.034]	[0.046]	[0.047]	[0.057]
Incentives X (Daily NTH)	0.019	-0.016	0.0081	0.038	-0.011	-0.031
	[0.024]	[0.025]	[0.022]	[0.031]	[0.031]	[0.038]
Incentives X (Monthly NTH)	-0.025	0.014	0.034	0.019	-0.076**	0.0036
	[0.024]	[0.025]	[0.023]	[0.031]	[0.031]	[0.038]
Incentives X (4- or 5-Day TH)	0.020	0.00036	0.012	-0.0025	-0.042***	0.012
	[0.013]	[0.013]	[0.012]	[0.016]	[0.016]	[0.020]
Monitoring mean	0.17	0.13	0.16	0.18	0.16	0.16
Controls	Yes	Yes	Yes	Yes	Yes	Yes
B. Average Daily Steps						
Incentives	-0.049*	-0.061**	-0.087***	-0.019	0.13***	0.072**
	[0.026]	[0.025]	[0.022]	[0.027]	[0.034]	[0.030]
Incentives X (10 INR)	0.037	0.0063	0.064*	-0.058	-0.043	-0.036
	[0.042]	[0.041]	[0.037]	[0.044]	[0.056]	[0.049]
Incentives X (Daily NTH)	0.042	-0.023	0.028	-0.0095	-0.065*	0.0091
	[0.028]	[0.027]	[0.024]	[0.029]	[0.037]	[0.032]
Incentives X (Monthly NTH)	-0.022	0.025	0.029	-0.013	-0.084**	0.038
	[0.028]	[0.027]	[0.024]	[0.029]	[0.037]	[0.032]
Incentives X (4- or 5-Day TH)	0.030**	0.021	-0.0016	-0.034**	-0.041**	0.025
	[0.015]	[0.014]	[0.013]	[0.015]	[0.020]	[0.017]
Monitoring mean	0.17	0.17	0.17	0.17	0.17	0.17
Controls	Yes	Yes	Yes	Yes	Yes	Yes
# Monitoring	200	200	200	200	200	200
# Daily NTH	163	163	163	163	163	163
# Base Case	888	888	888	888	888	888
# Monthly NTH	163	163	163	163	163	163
# 4-Day TH	778	778	778	778	778	778
# 5-Day TH	305	305	305	305	305	305
# 10 INR	64	64	64	64	64	64
Observations	2,561	2,561	2,561	2,561	2,561	2,561

Notes: Standard errors are in brackets. Controls include age, gender, weight, and the average daily steps taken during the phase-in period (before randomization). The omitted category in all columns is the monitoring group, and the omitted interaction with incentives is the base case incentive treatment group.

Appendix Table B.4: Differential Incentive Effects according to Impatience over Recharges and Steps

Dependent variable:	Met step target	Average daily steps	Met step target	Average daily steps
	(1)	(2)	(3)	(4)
Incentives	0.241***	1779.1***	0.219***	1655.0***
	[0.03]	[273.05]	[0.02]	[260.32]
10 INR	-0.0427	-478.2	-0.0759*	-581.2
	[0.05]	[466.86]	[0.05]	[425.50]
Daily	-0.00526	-387.2	-0.0197	-424.6
	[0.04]	[380.15]	[0.03]	[321.66]
Monthly	-0.0252	-90.06	-0.0530*	-426.5
	[0.03]	[366.63]	[0.03]	[294.34]
4-Day TH	-0.0212	-350.6*	-0.0199	-212.3
	[0.02]	[205.84]	[0.02]	[189.42]
5-Day TH	0.0140	-190.8	-0.0280	-398.5*
	[0.03]	[267.71]	[0.03]	[232.86]
Impatience	-0.0901	-367.6	-0.000600	-228.4
	[0.17]	[1567.33]	[0.02]	[194.81]
10 INR × Impatience	0.374	-821.9	-0.0774	-438.1
	[0.66]	[5694.07]	[0.05]	[498.41]
Daily × Impatience	-0.318	-8444.8**	0.0335	346.6
	[0.43]	[3721.21]	[0.06]	[538.16]
Monthly × Impatience	-0.0701	608.1	-0.0232	442.3
	[0.40]	[4733.30]	[0.04]	[618.17]
4-Day TH × Impatience	-0.184	-2180.7	0.0193	254.6
	[0.26]	[2370.02]	[0.03]	[256.54]
5-Day TH × Impatience	0.144	1823.5	0.104**	1090.2**
	[0.35]	[3453.62]	[0.04]	[427.41]
Controls	X	X	X	X
Domain of Impatience	Recharges	Recharges	Steps	Steps
Methodology	Discount Rate	Discount Rate	Discount Rate	Discount Rate
Base Case: mean	0.51	8190.18	0.51	8190.18
# Monitoring	97	98	144	145
# 10-INR	36	36	47	47
# Daily NTH	92	92	119	119
# Base Case	483	483	623	624
# Monthly NTH	90	90	116	116
# 4-day TH	428	428	558	558
# 5-day TH	159	159	218	218
# Individuals	1,385	1,386	1,825	1,827
Observations	109,893	110,709	145,093	146,119

Notes: This table shows the treatment effects of each Incentive treatment, interacted with measures of impatience over recharges; the base case incentive group is omitted. Standard errors clustered at the individual level in brackets. Controls include respondent's gender, age, and weight as well as the average phase-in period value of the dependent variable. "Discount Rate" indicates a structural measure of the daily discount rate $\frac{1}{\hat{\delta}_i} - 1$ estimated from a two-limit Tobit model of CTB allocations with individual discount-rate fixed effects, restricting the present-bias parameter β to be one. Larger values of each impatience measure indicates more impatience.

Appendix Table B.5: Heterogeneity in threshold impacts by baseline walking

	Exceeded Daily Step Target		Average Daily Steps	
	(1)	(2)	(3)	(4)
4- or 5-day TH × Walking Measure	0.061	0.0000040	748.8*	0.042
	[0.04]	[0.00]	[444.80]	[0.04]
4- or 5-day TH	-0.028	-0.041	-372.0**	-472.0
	[0.02]	[0.03]	[185.99]	[309.29]
Walking Measure	0.45***	0.000028***	5462.4***	0.47***
	[0.03]	[0.00]	[337.58]	[0.04]
Walking Measure	Step Target Compliance	Average Steps	Step Target Compliance	Average Steps
# Individuals	2,561	2,560	2,561	2,560

Notes: This table shows the treatment effects of the 4- and 5-day threshold treatments, interacted with measures of walking at baseline; the base case incentive group is the omitted group. The sample is limited to the base case, 4-day threshold, and 5-day threshold treatment groups. Standard errors clustered at the individual level in brackets. Controls include respondent's gender, age, and weight as well as the average phase-in period value of the dependent variable and its square.

Appendix Table B.6: The structural measure of impatience over exercise is not significantly correlated with other variables

Dependent variable:	Impatience measure: Delta	
Specification:	Separate regressions	Pooled regression
Age	0.009	0.004
	[0.006]	[0.006]
Female (=1)	0.066	0.078
	[0.161]	[0.168]
Previously diagnosed diabetic (=1)	0.158	0.196
	[0.123]	[0.175]
HbA1c	-0.020	-0.036
	[0.028]	[0.040]
Mean blood pressure	0.001	0.001
	[0.006]	[0.006]
Risk aversion (higher means more risk loving)	0.014	0.018
	[0.063]	[0.066]
Discount rate (recharges)	0.285	0.364
	[0.331]	[0.390]
Above-median baseline steps (=1)	0.057	0.076
	[0.158]	[0.152]

Notes: The unit of observation is a respondent. Robust standard errors in brackets. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table B.7: Time preference heterogeneity robust to including other controls

Dependent variable:	Met step target (×1000)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Above-75th-perc. delta × Threshold	86.36**	87.01**	86.53**	86.82**	86.98**	88.78**	84.60**	87.74**	72.77*	76.91**	81.82**
	[38.51]	[38.25]	[38.59]	[38.51]	[38.56]	[38.43]	[38.58]	[38.78]	[37.59]	[37.75]	[37.02]
Above-75th-perc. delta	-19.10	-19.44	-19.02	-19.27	-19.46	-21.23	-17.86	-19.70	-12.72	-18.63	-24.56
	[28.01]	[27.98]	[28.01]	[28.01]	[28.05]	[27.90]	[28.18]	[28.34]	[27.25]	[27.26]	[26.18]
Threshold	-32.13	-170.3*	-28.91	-40.91	16.20	-67.04***	20.64	-23.55	-54.76**	-66.44	-22.79
	[19.79]	[96.67]	[23.91]	[32.52]	[48.90]	[22.38]	[39.20]	[23.23]	[27.04]	[45.97]	[48.93]
Threshold × Covariate		2.791	-7.552	12.68	-5.528	0.340***	-16.12	0.0238	57.79*	0.00653	-0.000567
		[1.939]	[34.69]	[36.76]	[5.123]	[0.122]	[10.00]	[0.0352]	[33.07]	[0.00645]	[0.0107]
Covariate		1.772	4.716	-4.463	5.603	-0.423***	7.725	-0.00441	152.7***	0.0391***	0.0653***
		[1.398]	[25.19]	[26.25]	[5.120]	[0.0639]	[7.260]	[0.0258]	[27.85]	[0.00567]	[0.00484]
Threshold × Covariate^2											3.88e-08
											[0.000000553]
Covariate^2											-0.00000123***
											[0.000000125]
Covariate used	–	Age	Female	Prev. diagnosed diabetic	HbA1c	Mean arterial blood pressure	Risk aversion	Discount rate (recharges)	Above-median baseline steps	Baseline steps	Baseline steps
# Observations	110,932	110,932	110,932	110,932	110,932	110,932	110,932	110,932	110,932	110,932	110,932
Base case mean	507.12	507.12	507.12	507.12	507.12	507.12	507.12	507.12	507.12	507.12	507.12

Notes: The sample is restricted to the weekly groups – i.e., the base case (linear) group, and the 2 threshold groups, 4-day threshold and 5-day threshold, pooled here together as "Threshold." All columns control for the baseline value of the dependent variable. The unit of observation is a respondent × day. Standard errors in brackets clustered at the respondent level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table B.8: Time preference heterogeneity robust to including other controls

Dependent variable:	Met step target (×1000)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
A. Delta											
Delta × Threshold	3.648***	3.232**	3.689***	3.631***	3.646***	3.600***	3.701***	3.822***	2.086	1.792	1.392
	[1.360]	[1.367]	[1.360]	[1.364]	[1.357]	[1.347]	[1.354]	[1.369]	[1.373]	[1.233]	[1.184]
Delta	-2.263**	-2.292**	-2.295**	-2.254**	-2.262**	-2.210**	-2.458**	-2.295**	-1.990*	-2.092**	-1.884**
	[1.133]	[1.120]	[1.130]	[1.136]	[1.130]	[1.117]	[1.124]	[1.130]	[1.178]	[0.994]	[0.938]
Threshold	-11.87	-148.3	-10.06	-17.97	36.32	-45.85**	43.03	-3.552	-38.80	-51.10	-8.264
	[17.07]	[96.71]	[21.93]	[30.75]	[48.31]	[20.02]	[37.35]	[21.57]	[25.46]	[46.14]	[48.88]
Threshold × Covariate		2.763	-4.020	8.940	-5.499	0.338***	-16.91*	0.0223	60.96*	0.00703	0.000574
		[1.943]	[34.72]	[36.83]	[5.128]	[0.124]	[10.02]	[0.0352]	[33.13]	[0.00639]	[0.0108]
Covariate		1.764	6.320	-2.914	5.558	-0.417***	8.406	-0.00404	152.3***	0.0389***	0.0648***
		[1.399]	[25.20]	[26.28]	[5.124]	[0.0690]	[7.241]	[0.0256]	[27.87]	[0.00559]	[0.00487]
Threshold × Covariate^2											5.15e-09
											[0.000000561]
Covariate^2											-0.00000122***
											[0.000000126]
B. Above 75th Percentile Delta											
Above-75th-perc. delta × Threshold	86.36**	87.01**	86.53**	86.82**	86.98**	88.78**	84.60**	87.74**	72.77*	76.91**	81.82**
	[38.51]	[38.25]	[38.59]	[38.51]	[38.56]	[38.43]	[38.58]	[38.78]	[37.59]	[37.75]	[37.02]
Above-75th-perc. delta	-19.10	-19.44	-19.02	-19.27	-19.46	-21.23	-17.86	-19.70	-12.72	-18.63	-24.56
	[28.01]	[27.98]	[28.01]	[28.01]	[28.05]	[27.90]	[28.18]	[28.34]	[27.25]	[27.26]	[26.18]
Threshold	-32.13	-170.3*	-28.91	-40.91	16.20	-67.04***	20.64	-23.55	-54.76**	-66.44	-22.79
	[19.79]	[96.67]	[23.91]	[32.52]	[48.90]	[22.38]	[39.20]	[23.23]	[27.04]	[45.97]	[48.93]
Threshold × Covariate		2.791	-7.552	12.68	-5.528	0.340***	-16.12	0.0238	57.79*	0.00653	-0.000567
		[1.939]	[34.69]	[36.76]	[5.123]	[0.122]	[10.00]	[0.0352]	[33.07]	[0.00645]	[0.0107]
Covariate		1.772	4.716	-4.463	5.603	-0.423***	7.725	-0.00441	152.7***	0.0391***	0.0653***
		[1.398]	[25.19]	[26.25]	[5.120]	[0.0639]	[7.260]	[0.0258]	[27.85]	[0.00567]	[0.00484]
Threshold × Covariate^2											3.88e-08
											[0.000000553]
Covariate^2											-0.00000123***
											[0.000000125]
Covariate used	–	Age	Female	Prev. diagnosed diabetic	HbA1c	Mean arterial blood pressure	Risk aversion	Discount rate (recharges)	Above-median baseline steps	Baseline steps	Baseline steps
# Observations	110,932	110,932	110,932	110,932	110,932	110,932	110,932	110,932	110,932	110,932	110,932
Base case mean	507.12	507.12	507.12	507.12	507.12	507.12	507.12	507.12	507.12	507.12	507.12

Notes: The sample is restricted to the weekly groups – i.e., the base case (linear) group, and the 2 threshold groups, 4-day threshold and 5-day threshold, pooled here together as "Threshold." All columns control for the baseline value of the dependent variable. The unit of observation is a respondent × day. Standard errors in brackets clustered at the respondent level. Significance levels: * 10%, ** 5%, *** 1%

Appendix Table B.9: Other studies of interventions to improve health outcomes among diabetics

Author (year)	Intervention	N	Country	Health Effects	Intervention Intensity	Implementation Cost	Scalable
A. Diabetes – Management Studies:							
Wing, et al (1988)	Aerobic Exercise Training Program	25	NR	-0.20	high	high	no
Arora, et al (2013)	SMS encouragement and reminders	128	United States	-0.45	low	low	yes
Bjorgaas, et al (2004)	12-week Exercise Program	29	Norway	**-0.50***	high	high	no
Sigal, et al (2007)	Aerobic and Resistance Training Program	251	Canada	**-0.51**	high	high	no
Islam, et al (2015)	SMS encouragement and reminders	236	Bangladesh	**-0.66**	low	low	yes
Balducci, et al (2004)	Aerobic and Resistance Training Program	120	NR	**-1.21***	high	high	no
Yoon, et al (2008)	Monitoring and SMS based recommendations	51	South Korea	**-2.13**	low	high	no
B. Diabetes – Prevention Studies:							
Ramachandran, et al (2013)	Monitoring and SMS based recommendations	537	Chennai	**0.64**^{h}	low	low	yes
Wong, et al (2013)	Monitoring and SMS based recommendations	104	Hong Kong	**0.35**^{r}	low	low	yes

Notes: A value of "NR" for country indicates that the study location was not reported. Effects in bold are statistically significant.

Panel A reports health effects in terms of hba1c (%) and panel B reports health effects in terms of incidence of diabetes.

[h] Result is reported in terms of hazard ratio.

[r] Result is reported in terms of relative risk.

[*] Significance reported for pre-post test within the treatment group only.

Appendix Table B.10: Frequency and Payday Effects, Benchmarked against Monitoring

Dependent variable:	Met step target (×100)				
Payment Frequency:	Weekly		Monthly		
	(1)	(2)	(3)	(4)	(5)
Days before payday	0.194		0.192		
	[0.165]		[0.173]		
Weekly	20.88***	20.70***			
	[2.068]	[1.979]			
Days before payday X Weekly	-0.0717				
	[0.180]				
Payday		-0.604		-0.268	
		[0.930]		[1.007]	
Payday X Weekly		-0.252			
		[0.978]			
Monthly NTH			17.93***	17.38***	16.71***
			[2.926]	[2.850]	[3.036]
Days before payday X Monthly			-0.214		
			[0.236]		
Payday X Monthly				-0.671	
				[1.257]	
Payweek X Monthly					0.671
					[1.257]
# Observations	86,899	86,899	28,985	28,985	28,985
Sample mean	46.97	46.97	39.22	39.22	39.22

Supplemental Online Appendix for:

Incentivizing Behavioral Change:
The Role of Time Preferences

Table of Contents

C. SMS Treatment Impacts

We here present the effects of the SMS treatment, which we included in the experiment to appease our government partners who were interested in its efficacy. We thus estimate regressions of the following form, using the same outcome variables as in Section 6:

$$y_i = \alpha + \beta_1 \times incentives_i + \beta_2 \times monitoring_i + \beta_3 \times SMS_i + \boldsymbol{X}_i'\gamma + \varepsilon_i \qquad (14)$$

where y_i is a health or lifestyle outcome at endline for individual i; $incentives_i$ is an indicator for being in the incentives group; $monitoring_i$ is an indicator for being in the monitoring group; SMS is an indicator for being in the SMS treatment, and $\boldsymbol{X}_i$ is a vector of controls, shown in the table notes. β_1 is the ITT effect of incentives relative to the pure control group, β_2 is the ITT effect of the monitoring relative to the pure control group, and, for benchmarking, β_3 is the ITT effect of being in the SMS reminders group.

Table C.1 contains the health impacts. SMS treatments have been to satisfy our partly to benchmark the effect sizes of the incentives group, as that treatment has been effective in diabetes prevention in other contexts, although it has not been tested for diabetes management as we are examining here (Ramachandran, 2013). Surprisingly, the SMS treatment effect here is, if anything, positive; recall that here positive coefficients are associated with *worse* health. To probe further on why this might be the case, in Table C.2 we estimate a model including all of the interaction effects between the SMS treatment, and the monitoring or incentives treatments. Although these estimates should be interpreted as suggestive since (a) we did not plan to run this specification ex ante, and (b) the interaction effects are only marginally significant, it appears that one reason for the SMS Treatment's marginally negative average effect, is that the SMS treatment does not interact well with incentives, especially for their effect on blood sugar. We find no evidence that the SMS treatment affected any other behavioral outcomes either (see tables C.3 - C.6).

Online Appendix Table C.1: Impacts of incentives, monitoring, and SMS treatments on health risk factors.

	Health Risk Index	HbA1c	RBS	Mean Arterial BP	BMI	Waist Circum-ference
Incentives	-0.060**	-0.087	-6.12*	0.23	-0.065	-0.39*
	[0.026]	[0.071]	[3.42]	[0.43]	[0.049]	[0.24]
Monitoring	0.0038	-0.15	1.86	1.28*	0.034	-0.37
	[0.046]	[0.12]	[6.06]	[0.75]	[0.087]	[0.42]
SMS Treatment	0.064*	0.10	-4.77	1.08*	0.12*	0.47
	[0.033]	[0.091]	[4.40]	[0.55]	[0.063]	[0.31]
Controls	X	X	X	X	X	X
Control mean	0.00	8.44	193.83	103.14	26.52	94.50
# Control	566	560	563	559	563	538
P-value: M = I	0.12	0.59	0.14	0.12	0.21	0.96
P-value: SMS = I	0.00	0.10	0.81	0.22	0.02	0.03
P-value: SMS = M	0.28	0.10	0.37	0.83	0.41	0.11
# Individuals	3,103	3,061	3,067	3,050	3,065	2,951

Notes: Standard errors in brackets. For the Health Risk Index, controls include age, gender, and second order polynomials of all health variables underlying the index at baseline. Controls for all other outcomes include age, gender, and a second order polynomial of dependent variable at baseline. The Health Risk Index is an index created by the average of endline Hba1c, RBS, MAP, BMI, and waist circumference standardized by their average and standard deviation in the control group. Hba1c is the average plasma glucose concentration (%), RBS is the blood glucose level (mg/dL), MAP is the mean arterial blood pressure (mm Hg), and BMI is the body mass index. The omitted category in all columns is the pure control group.

Online Appendix Table C.2: Impacts of incentives, monitoring, SMS treatments, and their interactions on health risk factors

	Health Risk Index	HbA1c	RBS	Mean Arterial BP	BMI	Waist Circum-ference
Incentives	-0.072***	-0.15**	-8.15**	0.18	-0.068	-0.31
	[0.027]	[0.074]	[3.61]	[0.45]	[0.052]	[0.25]
Monitoring	0.015	-0.17	2.50	1.20	0.068	-0.26
	[0.048]	[0.13]	[6.39]	[0.80]	[0.092]	[0.44]
SMS Treatment	-0.017	-0.35*	-19.1*	0.65	0.12	1.15
	[0.077]	[0.21]	[10.0]	[1.25]	[0.14]	[0.71]
Incentives × SMS	0.12	0.59**	19.9*	0.51	0.036	-0.81
	[0.086]	[0.23]	[11.2]	[1.41]	[0.16]	[0.79]
Monitoring × SMS	-0.11	0.20	-7.35	0.75	-0.33	-1.10
	[0.15]	[0.42]	[20.2]	[2.48]	[0.29]	[1.40]
Controls	X	X	X	X	X	X
Control mean	0.00	8.44	193.83	103.14	26.52	94.50
# Control	566	560	563	559	563	538
P-value: M = I	0.04	0.85	0.06	0.15	0.10	0.90
P-value: SMS = I	0.46	0.30	0.25	0.70	0.18	0.03
P-value: SMS = M	0.70	0.42	0.05	0.69	0.76	0.07
# Individuals	3,103	3,061	3,067	3,050	3,065	2,951

Notes: Standard errors in brackets. All definitions follow Table 6. The omitted category in all columns is the pure control group.

Online Appendix Table C.3: Impacts of incentives, monitoring, and SMS treatments on fitness.

	Fitness Time Trial Index	Seconds to Walk 4m	Seconds for 5 Sit-Stands
Incentives	0.0064	0.015	-0.091
	[0.028]	[0.044]	[0.12]
Monitoring	0.025	0.040	-0.13
	[0.049]	[0.077]	[0.21]
SMS Treatment	-0.047	-0.053	-0.099
	[0.037]	[0.058]	[0.15]
Controls	X	X	X
Control mean	0.00	3.88	13.18
# Control	505	465	459
P-value: M = I	0.67	0.71	0.83
P-value: SMS = I	0.24	0.34	0.97
P-value: SMS = M	0.24	0.33	0.90
# Individuals	2,819	2,598	2,595

Notes: Standard errors in brackets. For the Fitness Time Trial Index, controls include age, gender, and second order polynomials of both fitness variables underlying the index at baseline. Controls for all other outcomes include age, gender, and a second order polynomial of dependent variable at baseline. A large value of Fitness Time Trial Index indicates low fitness: it is an index created by the average two trials of endline seconds to walk four meters, and the seconds to complete five sit-stands standardized by their average and standard deviation in the control group. The omitted category in all columns is the pure control group.

Online Appendix Table C.4: Impacts of incentives, monitoring, and SMS treatments on mental health.

	Mental Health Index	Felt Happy	Less Nervous	Peaceful	Energy	Less Blue	Less Worn	Less harm to Social Life
Incentives	0.073**	0.088**	0.026	0.055	0.061	0.015	0.090**	0.052*
	[0.032]	[0.045]	[0.044]	[0.047]	[0.048]	[0.043]	[0.039]	[0.030]
Monitoring	0.11*	0.072	0.12	0.092	0.035	0.13	0.17**	0.049
	[0.057]	[0.079]	[0.077]	[0.083]	[0.084]	[0.077]	[0.069]	[0.053]
SMS Treatment	0.035	0.020	0.11**	0.055	0.018	0.085	-0.011	-0.0018
	[0.042]	[0.058]	[0.056]	[0.061]	[0.061]	[0.056]	[0.050]	[0.039]
Controls	X	X	X	X	X	X	X	X
Control mean	0.00	3.06	3.48	3.35	3.30	3.86	4.40	4.71
# Control	585	563	563	563	563	563	563	563
P-value: M = I	0.50	0.81	0.16	0.62	0.73	0.11	0.18	0.94
P-value: SMS = I	0.48	0.35	0.21	1.00	0.58	0.32	0.11	0.27
P-value: SMS = M	0.31	0.60	0.93	0.72	0.87	0.66	0.03	0.44
# Individuals	3,192	3,068	3,068	3,068	3,068	3,068	3,068	3,068

Notes: Standard errors in brackets. For the Mental Health Index, controls include age, gender, and second order polynomials of all questions underlying the index at baseline. Controls for all other outcomes include age, gender, and a second order polynomial of dependent variable at baseline (where available). The Mental Health Index is the average values of seven questions adapted from the Rand 36-Item Short Form Survey (SF-36), standardized by their mean and standard deviation in the pure control group. The omitted category in all columns is the pure control group.

Online Appendix Table C.5: Impacts of incentives, monitoring, and SMS treatments on Healthy Diet, relative to pure control.

	Healthy Diet Index	Wheat meals	Meals with vegeta-bles	Servings Fruit	-(Rice meals)	-(Junkfood pieces)	-(Spoons sugar in coffee)	-(Sweets yester-day)	Avoid un-healthy food
Incentives	0.051	0.027	0.060*	0.040	0.029	-0.017	-0.022	-0.028	0.0036
	[0.045]	[0.030]	[0.031]	[0.038]	[0.033]	[0.062]	[0.046]	[0.038]	[0.018]
Monitoring	0.021	0.018	0.078	0.060	-0.0075	0.13	-0.029	-0.048	-0.041
	[0.079]	[0.053]	[0.055]	[0.066]	[0.059]	[0.11]	[0.081]	[0.067]	[0.031]
SMS Treatment	0.048	-0.025	-0.0018	-0.0026	-0.025	0.056	0.063	0.091*	0.011
	[0.058]	[0.039]	[0.040]	[0.048]	[0.043]	[0.080]	[0.059]	[0.049]	[0.023]
Controls	X	X	X	X	X	X	X	X	X
Control mean	0.00	0.49	0.58	0.53	-2.34	-0.87	-1.08	-0.35	0.83
# Control	585	563	563	563	563	585	585	563	563
P-value: M = I	0.68	0.84	0.70	0.73	0.49	0.14	0.93	0.75	0.11
P-value: SMS = I	0.97	0.28	0.22	0.49	0.31	0.47	0.25	0.05	0.79
P-value: SMS = M	0.79	0.52	0.24	0.44	0.81	0.60	0.36	0.10	0.17
# Individuals	3,192	3,068	3,068	3,068	3,068	3,192	3,192	3,068	3,068

Notes: Standard errors in brackets. For the Healthy Diet Index, controls include age, gender, and second order polynomials of all questions underlying the index at baseline. Controls for all other outcomes include age, gender, and a second order polynomial of dependent variable at baseline. The Healthy Diet Index is an index created by the average values of eight diet questions, standardized by their average and standard deviation in the control group. A larger value indicates a healthier diet. The omitted category in all columns is the pure control group.

Online Appendix Table C.6: Impacts of incentives, monitoring, and SMS treatments on Addictive Substance Consumption, relative to pure control.

	Addictive Good Consumption Index	Average Daily Areca	Average Daily Alcohol	Average Daily Cigarettes
Incentives	-0.010	0.035	-0.036*	-0.055
	[0.037]	[0.042]	[0.020]	[0.11]
Monitoring	-0.0058	0.013	-0.016	-0.019
	[0.065]	[0.074]	[0.036]	[0.19]
SMS Treatment	0.031	0.045	0.018	-0.15
	[0.047]	[0.054]	[0.026]	[0.14]
Controls	X	X	X	X
Control mean	0.00	0.13	0.11	1.02
# Control	585	563	563	563
P-value: M = I	0.94	0.75	0.54	0.83
P-value: SMS = I	0.50	0.89	0.10	0.60
P-value: SMS = M	0.65	0.73	0.44	0.59
# Individuals	3,192	3,068	3,068	3,068

Notes: Standard errors in brackets. For the Addictive Good Consumption Index, controls include age, gender, and second order polynomials of all questions underlying the index at baseline. Controls for all other outcomes include age, gender, and a second order polynomial of dependent variable at baseline. The Addictive Good Consumption Index is an index created by the average self-reported average daily consumption of areca, alcoholic drinks, and cigarettes, standardized by their average and standard deviation in the control group. A larger value indicates more consumption. The omitted category in all columns is the pure control group.

D. Monitoring treatment impacts on walking

We first present the estimates of the pre-post monitoring effect. In order to increase the precision of our estimated year-month fixed effects, we include the incentives group in the regression as well since that group is much larger. We thus estimate the following difference-in-differences regression using data from both the intervention and phase-in periods for the incentives and monitoring groups:

$$y_{it} = \alpha + \beta_1 Intervention\,Period_{it} + \beta_2 incentives_i + \beta_3 (Intervention\,Period_{it} \times incentives_i) + \boldsymbol{X}_i'\gamma + \boldsymbol{\mu}_m + \varepsilon_{it}, \quad (15)$$

where y_{it} are daily pedometer outcomes measured during both the phase-in and the intervention period, $Intervention\,Period_{it}$ is an indicator for whether individual i has been randomized into their contract at time t, $incentives_i$ is an indicator for whether i is in an incentives treatment group, $\boldsymbol{X}_i$ is a vector of individual-specific controls, and $\boldsymbol{\mu}_m$ is a vector of month fixed effects. The coefficient β_1 - the coefficient of interest - is the pre-post dif-

ference in pedometer outcomes within the monitoring group (controlling for aggregate time effects).

Online Appendix Table D.1: Impacts of monitoring (pre-post) and incentives (difference-in-differences) on exercise outcomes.

	Achieved 10K Steps			Daily Steps		
Incentives	0.013	0.014	0.0091	73.6	76.0	62.3
	[0.024]	[0.024]	[0.011]	[270.8]	[270.2]	[114.0]
Intervention Period	0.065***	0.081***	0.073***	-25.7	216.2	62.0
	[0.020]	[0.021]	[0.020]	[238.1]	[241.4]	[234.2]
Intervention Period X Incentives	0.19***	0.19***	0.19***	1278.0***	1263.1***	1224.2***
	[0.021]	[0.021]	[0.021]	[248.7]	[249.4]	[243.4]
Monitoring phase-in mean	0.24	0.24	0.24	6848.51	6848.51	6848.51
Year-month FEs	No	Yes	Yes	No	Yes	Yes
Individual controls	No	No	Yes	No	No	Yes
# Monitoring	203	203	203	203	203	203
# Incentives	2,404	2,404	2,401	2,404	2,404	2,402
# Individuals	2,607	2,607	2,604	2,607	2,607	2,605
Observations	218,812	218,812	218,573	220,169	220,169	220,040

Notes: This table shows coefficient estimates from regressions of the form specified in Equation 15. The outcomes are from daily panel data from the pedometers. Standard errors, in brackets, are clustered at the individual level. Individual controls include age, gender, weight, and the average of the dependent variable during the phase-in period (before randomization). The omitted category in all columns is the monitoring group in the phase-in period. The coefficient in the second row, on $Intervention\,Period_{it}$, corresponds to the pre-post estimate of the monitoring effect.

Taken together, the evidence from our experiment suggests that the monitoring treatment improved the likelihood of reaching the daily 10,000 step target by at least 7.3% during the intervention period (accounting for 30% of the total impact of incentives relative to control), and increased the self-reported likelihood of walking for exercise at endline – after the intervention period – by 12% (65% of the total impact of incentives). Thus, combining monitoring, goal-setting, and feedback - without incentives - appears to be an effective intervention per se to encourage individuals to exercise on a regular. However, the impacts of the monitoring treatment on average daily steps during the intervention period cannot be distinguished from 0, suggesting that the monitoring treatment may have induced intertemporal substitution of steps between days in order to regularly achieve the 10,000 step target.

Although we believe the self-reported exercise data was biased by the monitoring treatment, for completeness, we also compare self-reported exercise at endline across treatment groups using regressions of the following form:

$$y_i = \alpha + \beta_1 \times incentives_i + \beta_2 \times control_i + \boldsymbol{X}_i'\gamma + \varepsilon_i \tag{16}$$

where y_i is the outcome at endline for individual i, and $control_i$ is an indicator for being in the pure control group. The coefficient β_1 represents the ITT effect of incentives relative to

the monitoring group, while β_2 is the ITT effect of the pure control relative to the monitoring group. The results of these regressions are shown in Table D.2.

Online Appendix Table D.2: Impacts of incentives and monitoring on self-reported exercise.

	Self-Reported Data (at Endline)	
	Fraction Days Exercised Last Week	Minutes Walked for Exercise Yesterday
Incentives	0.0627*	4.968**
	[0.0320]	[2.257]
Pure Control	-0.124***	-7.626***
	[0.0353]	[2.365]
Monitoring mean	0.497	22.333
Controls	Yes	Yes
Observations	3,063	3,063

Notes: We report pooled incentive effects. The columns show coefficient estimates from regressions of the form specified in Equation 16, using a cross-section of self-reported exercise data at endline, after the intervention period is over. The sample includes the Pure control, incentives, and monitoring groups. Controls include age, gender, weight, SMS treatment, and the average daily steps and fraction of days step target was exceeded during the phase-in period (before randomization). The omitted category in all columns is the monitoring group. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

The results show that self-reported exercise was higher in the incentives group than the monitoring group, and the monitoring group than the pure control group. Monitoring participants report substantially higher levels of exercise than the control group. However, the results give reason for doubt, as the estimated effect of monitoring relative to control seems unreasonably large, and the effect of incentives relative to monitoring unreasonably small; in particular, the magnitudes of these estimates do not line up with our pedometer data.[41] This suggests that self-reported exercise during the intervention period may be biased upwards relative to the other groups, and makes us question the reliability of these analysis.

[41] In the self-reported data, the estimated monitoring effect is 50-90% *larger* than the estimated effect of incentives relative to monitoring. Given that, using pedometer data, the monitoring mean for days exceeded target was 30% and the incentive effect relative to monitoring was 20%, these would imply that the control group should have been exceeding their target on around 0% of days, which seems implausible given that people exceeded their step target on 25% of days in the phase-in period before the interventions were launched.