# Incentivizing Behavioral Change: The Role of Time Preferences


Shilpa Aggarwal — Indian School of Business

Rebecca Dizon-Ross — University of Chicago

Ariel Zucker* — UC Berkeley


April 23, 2019


**Abstract**

How should the design of incentives vary with the time preferences of agents? We formulate two incentive contract variations designed to increase efficacy for impatient agents relative to patient ones: 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, and increasing the frequency of payment. 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. Making the contract dynamically non-separable meaningfully increases efficacy among the impatient relative to the patient, providing an effective way to tailor incentive design for impatience. In contrast, increasing payment frequency has limited efficacy on average or for the impatient. On average, incentives increase daily walking by 1,300 steps (roughly 13 minutes of brisk walking) and improve health.



*Aggarwal: Indian School of Business shilpa_aggarwal@isb.edu. Dizon-Ross: Chicago Booth School of Business rdr@chicagobooth.edu. Zucker: University of California, Berkeley ariel.d.zucker@gmail.com.


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, Yashna Nandan, and Emily Zhang for outstanding research assistance. We are grateful to Abhijit Banerjee, Marianne Bertrand, Esther Duflo, Pascaline Dupas, Rick Hornbeck, Seema Jayachandran, Anett John, Supreet Kaur, Rohini Pande, Devin Pope, Canice Prendergast, 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. Most classic contracting models pay limited attention to the role of agent patience. However, growing evidence that many people are "impatient" (i.e., that they discount the future heavily) raises an important question: What are the implications of agent impatience for the design of incentives? In this paper, we develop predictions about contract variations that should improve the relative efficacy of incentives for impatient agents relative to patient ones. Using a randomized controlled trial (RCT) incentivizing exercise among 3,200 diabetics in India, we then implement the variations to test for the quantitative importance of adjusting incentives for impatience.

We design our contract variations to be effective for multiple types of impatient people: present-biased naifs (who are unaware of their own present bias), present-biased sophisticates (who are aware of it), and time-consistent impatients (who have high discount rates but are not present-biased). The efficacy for naifs is important since naivete is common (Augenblick and Rabin, 2017; Bai et al., 2017), but naifs have typically been hard to motivate.

We also distinguish between how agents discount future consumption and how they discount financial payments. As emphasized in the recent time preferences literature, when agents consider future financial payments, their discount rate depends on their borrowing and lending opportunities; in contrast, when they consider future consumption or effort, their discount rate is the "primitive" or structural discount rate from their utility functions (Augenblick et al., 2015). We consider the implications of this idea for contract design. We identify two contract variations, one whose efficacy increases with discount rates over consumption/effort, and one whose efficacy increases with discount rates over payments.

Our first prediction is that making contracts "dynamically non-separable" can make them more effective for those with high discount rates over effort ("impatient over effort"). By dynamically non-separable, we mean that the incentive paid for action in a given period depends on actions taken in other periods. For example, the contract might pay people for taking an action on a day if and only if they take that action on at least 5 days in the week. Dynamic non-separability has several advantages and disadvantages that have been discussed before. Our new theoretical insight is that, relative to dynamically separable contracts, certain dynamically non-separable contracts should increase compliance for those who are impatient over future effort compared to for those who are patient.[1] Why? Dynamically separable contracts pay agents separately for their behavior in each period. As a result, agents always compare the financial incentive to the cost of effort *this period* (which is not discounted) and so their discount rates over future effort do not matter. In contrast, dynamically non-separable contracts bundle current effort with future effort. Agents thus

[1] Section 2 discusses which types of dynamically non-separable contracts this prediction applies to.

compare the incentive to the present *discounted* cost of exerting effort in multiple periods; since some of those periods are in the future, the cost is lower for those who discount future effort more. Intuitively, the contract leverages the fact that, because impatient people discount their future effort, it is "cheaper" to buy their future effort than current effort.[2]

Our second prediction is that, if agents are impatient over payments, providing more frequent payment increases efficacy. This is the main idea (besides pre-commitment) the literature discusses for tailoring incentives to impatience,[3] with for example Cutler and Everett (2010) proclaiming "the more frequent the reward, the better." However, there is limited evidence on the effect of payment frequency, and theoretical reasons to question both whether and what types of frequency increases would matter. First, whether payment frequency matters depends on whether discount rates over money are small or large. If credit markets are perfect, discount rates over payment should be small (equal to the market interest rate) and increasing payment frequency may have limited impact. If instead credit markets are imperfect, as may be especially true in developing countries, discount rates over payment could approach those over consumption. Second, even if discount rates are high, there is still an open question regarding what types of frequency increases matter. The answer depends on the shape of discount rates over time: if discount rates decline very quickly with lag (as with for example the "quasi-hyperbolic" or "beta-delta" model used in the literature), 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 evaluate three payment frequencies – monthly, weekly, and daily – to assess whether, and what type of, increases improve compliance. The three frequencies also allow us to explore which model of payment discounting best fits the data.

Our experiment uses incentives for "behavioral change" (i.e., incentives encouraging people to change their behavior in ways that ostensibly benefit their own well-being) to evaluate our contract variations, for two main reasons. First, such incentives are increasingly common.[4] Second, such incentives are often motivated by present bias itself. Present bias can cause underinvestment in behaviors with short-run costs but long-run benefits (e.g., eating right, studying), which justifies the use of incentives to better align behavior with long-run self-interest. However, even in this domain where present bias is a key rationale for the use of incentives, there is little evidence on how to adjust incentives for present bias.

[2]This logic holds both for sophisticates and naifs (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.

[3]O'Donoghue and Rabin (1999a) 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.

[4]e.g., Duflo et al. (2010); Fryer (2011); Martins et al. (2009); Morris et al. (2004); Thornton (2008).

We incentivize exercise among diabetics in India. Lifestyle diseases like diabetes and hypertension are exploding problems in both developing and developed countries. The estimated cost of diabetes is 1.8% of global GDP (Bommer et al., 2017), and 2% of national GDP in India (Tharkar et al., 2010). The key to decreasing the burden of chronic disease is to encourage lifestyle changes such as better exercise and diet. These behaviors involve short-run costs and long-run benefits, making incentives a promising approach. We incentivize walking, a key component of diabetes management and a major public health priority (Qiu et al., 2014), among diabetics and prediabetics. We monitor participants' walking using pedometers and, if they achieve a daily step target of 10,000 steps, provide them with small financial incentives in the form of mobile recharges (i.e. cell-phone credits).

The experiment then randomly varies (i) the frequency of incentive payment, and (ii) 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 (we use two minimum compliance levels: 4 days and 5 days). 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 20 percentage points (pp) off of a base of 30%. Average daily steps increase by 1267 or roughly 13 minutes of brisk walking.

We then explore the implications of time preferences for incentive design, presenting three main results. First, we find that, consistent with our theoretical predictions, making the contract dynamically non-separable meaningfully increases relative efficacy for those who are impatient over effort. Heterogeneity analysis using a baseline measure of impatience shows that, relative to linear contracts, non-separable contracts increase compliance with the step target by 6pp more for people with above-median impatience relative to those with below-median, a large difference relative to the sample-average effect of either contract (20pp). We also calibrate a model using experimental estimates of the distribution of walking costs and find consistent results: projected compliance in the dynamically non-separable contract relative to the linear increases by 5pp for each 10pp decrease in the discount factor.

To complete our analysis of dynamic non-separability, we explore its average efficacy, presenting to our knowledge the first empirical comparison of a dynamically non-separable contract with a dynamically separable one. Our second result is that, on average, making the contract dynamically non-separable improves cost-effectiveness: the percent of days on which people hit their step target does not change, but if agents do not meet the step target on at least 4 or 5 days in the week, they do not receive incentives for every day the step target is reached like they do with the linear contract. As a result, the 4-day and 5-day non-separable contracts cost roughly 10 and 15% less while generating the same amount of exercise. Dynamic non-separability has a potential downside, however: it generates more extreme outcomes, working better for some but worse for others. This variation in efficacy makes it important to determine for whom dynamically non-separable contracts work well, highlighting the significance of our finding that they work better for the impatient.

Our third result is that increasing the frequency of incentive delivery has limited impact in our setting. Incentives delivered at daily, weekly, and monthly frequencies have equally large impacts on walking, indicating that the model that best fits our sample is one of patience over financial payments. We find additional evidence in support of this conclusion: step-target compliance does not increase as the date of payment delivery approaches. This null finding suggests that, in contrast with the conventional wisdom, increasing incentive frequency is not an effective way to adjust incentives for impatience in all settings. This result is consistent with Augenblick et al. (2015) who find limited impatience 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 incentive scheme. Our sample has high rates of diabetes and hypertension; regular exercise can prevent complications from both. We show that the large increases in walking induced by incentives cause moderate improvements in physical health and emotional wellbeing. Incentives improve an index of health risk, including blood sugar and body mass index, by a moderate amount. Incentives also improve mental health. These effects are important for policy, suggesting that incentives may be a cost-effective way to decrease the burden of chronic disease in India and beyond.

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 impatience. 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); Ashraf et al. (2006); Giné et al. (2010); Duflo et al. (2011); Schilbach (2017); Royer et al. (2015)). 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 harm partially naive agents (Augenblick and Rabin, 2017; Bai et al., 2017; John, 2019). In contrast, we introduce a novel contract variation that is effective for multiple types of impatience, including naifs. Beyond the literature on pre-commitment, there is limited other work, theoretical or empirical, adjusting incentive design for impatience; we discuss the few exceptions in Section 1.1. Our work also relates to papers examining optimal policy given behavioral biases, such as sin taxes or optimal payments for loss-averse agents (Hossain and List, 2012; O'Donoghue and Rabin, 2006; Rabin and O'Donoghue, 2003).

Second, we build on the dynamic contracting literature (see for example Lazear (1979) and Prendergast (1999)). Many theoretical papers yield the prediction that it is optimal to make contracts dynamically non-separable by deferring some current pay until the future, making it contingent on future effort. To our knowledge, however, our paper is the first to compare dynamically non-separable and dynamically separable contracts empirically,[5] as well as the first to examine the interaction between dynamic non-separability and impatience.

We build on a third body of literature that measures the shape of time preferences. The majority of the recent work focuses on distinguishing whether consumption and payment discount rates are time-consistent or time-inconsistent (Andreoni and Sprenger, 2012; 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 full policy implications of that variation.[6]

Finally, we contribute to the growing literature on incentives for health, such as weight loss (Kullgren et al., 2013; Volpp et al., 2008) and disease monitoring (Labhardt et al., 2011).[7]

The paper proceeds as follows. Section 1.1 discusses the literature on contract design and impatience. Section 2 presents the predictions motivating the experiment. Sections 3 and 4 discuss the study setting, design, and data. In Section 5, we present results on incentive design and impatience. Section 6 presents impacts on health outcomes. Section 7 concludes.

## 1.1 Related literature: Contract design and impatience

We now review the theoretical and empirical literature studying how (besides pre-commitment) one can improve incentive design for impatient agents.[8] We first review the two papers that study solutions other than increasing payment frequency. O'Donoghue and Rabin (1999a) outline the theoretical implications of time-inconsistent procrastination for the design of

[5] Carrera et al. (2017) and Bachireddy et al. (2018) evaluate *separable* contracts which vary the size of the per-period payments over time.

[6] Two papers test whether discount rates over money measured in lab experiments are best fit by hyperbolic, quasi-hyperbolic, or exponential models. Benhabib et al. (2008) test between models but do not have power to distinguish them. Tanaka et al. (2010) reject that discount rates purely follow any one model.

[7] There are several evaluations of incentives for exercise among non-diabetics (e.g., Charness and Gneezy (2009); Finkelstein et al. (2008)), and one incentivizing 3-month blood sugar control among diabetics; ours represents the first trial of incentives for daily disease management among diabetics and of incentives for exercise in a developing country.

[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.

"temporal incentive schemes," which reward agents based on when they complete tasks. Their focus is on avoiding delay; they find that optimal incentives for procrastinators typically involve an increasing punishment for delay as time passes. Carrera et al. (2017) work in a setting where there are one-time "startup costs" for compliance; they show that, in such a setting, offering larger incentives at the beginning of an incentive contract does not empirically decrease procrastination. These studies both differ from ours in their theoretical goals and environments: our work focuses on maximizing the average level of compliance over time rather than on avoiding delay or overcoming startup costs.

Regarding payment frequency, several papers show that worker performance improves toward the end of pay cycles, and attribute this effect to impatience (Clark, 1994; Kaur et al., 2015; Oyer, 1998).[9] These papers suggest a potential role for high-frequency payment to improve average performance, but the evidence on whether changing payment frequency actually affects efficacy is scant. Chung et al. (2010) show that moving from annual to quarterly payments in a pay-for-performance scheme for physicians has no effect; this could either be because payment frequency does not matter or because even their higher-frequency payment (quarterly) was not frequent enough to be motivating. In contemporaneous work in the psychology literature, Gardiner and Bryan (2017) randomly vary whether incentives to consume fruit and vegetables were paid daily or at the end of the 3-week intervention period. They find that daily incentives are more effective. Because their design bundles payment and positive feedback, the patterns could be due to monetary discounting or to a salience/reminder effect. Our study's contribution relative to this work is to isolate the effect of payment frequency as it operates through monetary discounting as opposed to salience; we hold fixed the positive feedback frequency across arms. Another contribution is to use multiple treatment arms with varying payment frequency to trace out the "shape" of discounting, that is, to distinguish whether payment discount rates take on an exponential, quasi-hyperbolic, or other present-biased shape. Finally, somewhat different in spirit, Casaburi and Macchiavello (2019) and Brune and Kerwin (2019) examine the effects of payment frequency not on incentive efficacy but rather on consumption patterns, showing that infrequent payments can be beneficial as a commitment device and/or way to ease savings constraints.

# 2 Theoretical predictions

We now present a simple model of incentives to derive predictions for how two incentive contract features – dynamic-nonseparability and payment frequency – interact with time

[9] Clark (1994) finds anecdotal evidence that 19th century factory workers often shirked at the beginning of pay cycles, Oyer (1998) finds sales spike among US salespeople near the end of the year when bonuses are paid, and Kaur et al. (2015) show that piece-rate workers increase output as the weekly pay-day approaches.

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 walking 10,000 steps; $c_t$ is consumption on day $t$; and individuals discount the cost of walking and consumption $k$ days in advance by $d_c(k)$. Because the consumption amounts are small, we model utility as linear in recharges $c_t$ 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 are part of an incentive program where they earn payments for complying with the step target; their budget constraints thus depends on the incentive contract mapping exercise compliance to income. We consider first a separable, linear incentive contract; this contract specifies that walking on each day $t$ will be rewarded with a financial 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. **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 $e_t \leq \left(\frac{1}{1+r}\right)^{k_t} m$.

2. **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. As a result, in any given period $t$, an individual chooses 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: $e_t \leq d_c(k_t)m$.

More broadly, one can accommodate both of these (and other)[10] cases in a reduced form way by defining a reduced form discount factor parameter representing the amount by which

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

individuals discount financial 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, $d_m(k) = \left(\frac{1}{1+r}\right)^k$, whereas in case 2, $d_m(k) = d_c(k)$. 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:

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

**Predictions regarding dynamic non-separability** We now explore the effect of making the contract dynamically non-separable. We focus on a specific form of dynamic non-separability: a "dynamic threshold" wherein payment is a function of the number of periods of compliance in a given time range and there is a minimum threshold level of compliance below which no incentive is received. If agents achieve that threshold, they 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 non-separable contracts that display a "dynamic complementarity" (i.e., a period in which the payment for effort is increasing in future effort).[11] Note that the daily behavior incentivized in all contracts is to comply with a "static threshold" which asks participants to walk at least 10,000 steps in a given day; what differentiates the dynamic threshold contract 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.[12]

Dynamic non-separability makes an individual's decision to walk considerably more com-

[11] 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 is not a sufficient condition.

[12] While dynamic thresholds interact with time preferences, this is not necessarily true of static 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).

plicated, as the reward for compliance – and hence the decision to walk – depends on compliance on multiple of the days in the payment period. For simplicity, we illustrate how dynamic non-separability interacts with compliance and time preferences using a shorter payment cycle than used in our experiment: a two-day payment period. For the two-day threshold contract, if the individual complies with the step target on both days, she receives $2m$ on the second day; if she complies on only one day, she receives nothing.

Intuitively, the key difference between the dynamic 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 $d_m(t)$, discount factors over consumption $d_c(t)$ do not affect the decisions to comply;[13] she complies in period 1 if $e_1 < d_m(1)m$ and in period 2 if $e_2 < m$. The expected total days of compliance $w_t$ over the payment period is thus:

$$E\left[\sum_{t=1}^{2}(w_t)|\text{Linear}\right] = F\left(d_m(1)m\right) + F\left(m\right) \tag{3}$$

In the dynamic threshold contract, in contrast, discount factors over consumption and effort 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, $e_1 + d_c(1)e_2$, is less than the discounted value of the reward $d_m(1)2m$ and she knows she will follow through on the second day. Restricting to the case where costs are positive (this does not affect the results but simplifies notation), she thus complies if both (i) $e_1 + d_c(1)e_2 < d_m(1)2m$, and (ii) $e_2 < 2m$.[14] 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. Assuming the agent is "sophisticated" about her own time preferences (which we relax and discuss below), the agent's expected

[13] 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 mainly represents the interest rate here.

[14] With negative costs, she also walks in period $t$ if $e_t < 0$ regardless of whether the other conditions are satisfied.

total compliance in the 2-Day dynamic threshold contract is thus:

$$E\left[\sum_{t=1}^{2}(w_t)|\text{Dynamic Threshold}\right] = 2P(e_1 + d_c(1)e_2 < d_m(1)2m \text{ AND } e_2 < 2m)$$
$$= 2\int_{-\infty}^{2m}\int_{-\infty}^{d_m(1)2m-d_c(1)e_2} f(e_1)f(e_2)de_1de_2$$
$$= 2\int_{-\infty}^{2m} F(d_m(1)2m - d_c(1)e)f(e)de \tag{4}$$

Whether total compliance with a 2-day dynamic threshold (Equation 4) is larger than the compliance in the linear contract (Equation 3) 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. 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 dynamic threshold, and thus higher compliance.

In our experiment, we test for the quantitative importance of prediction 1 in two ways. First, we randomly vary whether the contract has a dynamic threshold and test for heterogeneity based on the discount rate over walking. Second, we use data from our experiment to calibrate a model and see how much expected compliance in the threshold relative to linear contract varies with discount rates over walking.

Prediction 1 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 longer models.

Equation 4 assumes that agents are "sophisticated" about the fact that the relative value of their future effort compared with their future incentive payment may be different from the point of view of their period 2 selves than it is for their period 1 selves. In particular, from period 1's perspective, the agent would want her period 2 self to comply if $e_2 < \frac{d_m(1)}{d_c(1)}2m$ whereas in period 2, the agent will in fact comply if $e_2 < 2m$; these are only equivalent if $d_m(1) = d_c(1)$, for example if there is no borrowing.

However, even if agents were "naive" and assumed that their period 2 relative tradeoff between effort and payoffs would be the same as in period 1, Prediction 1 would still hold: equation 4 would become $E\left[\sum_{t=1}^{2}(w_t)|\text{Dynamic Threshold}\right] = P(e_1 + d_c(1)e_2 < d_m(1)2m \text{ AND } e_2 < 2m) + P(e_1 + d_c(1)e_2 < d_m(1)2m)$ which is also decreasing in $d_c(1)$. Interestingly, not only does Prediction 1 hold for naifs, but it holds for them even more strongly than for sophisticates: with the dynamically non-separable contract, for a given set of discounting parameters $d_m$ and $d_c$, the naif's compliance is higher than the sophisticate's as her overoptimism about future compliance makes her more likely to comply today.[15] Naifs will thus also have a relatively higher gap between non-separable and separable than sophisticates, as sophistication and naivete do not affect behavior with separable contracts.

Prediction 1 was written with a dynamic threshold contract with threshold level 100%; however, for many cost distributions, the prediction also holds for thresholds below 100% (e.g., if the agent has to comply at least 5 days out of a 7-day payment period to receive payment, as in our experiment).[16] The intuition is the same as above: those who discount future walking cost still have a lower discounted total cost to achieve the threshold level. Again, the prediction holds for both time-consistents and time-inconsistents. It also again holds for both sophisticates and naifs, although for thresholds less than 100% it no longer necessarily holds "more strongly" for naifs; rather the prediction is that, for both sophisticates and naives, relative compliance in the non-separable contract decreases in $d_c(k)$).[17]

**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.** *If agents are "impatient" over the receipt of financial rewards (i.e., if $d_m(k) < 1$ and is decreasing in $k$), compliance is increasing in the payment frequency. If agents are patient ($d_m(k) \approx 1$), payment frequency does not affect compliance.*[18]

[15] Note that we are following O'Donoghue and Rabin (1999b) in defining a sophisticated agent as one who "knows exactly what her future selves' preferences will be" and a naive one as one who "believe(s) her future selves' preferences will be identical to her current self's."

[16] We are in the process of characterizing the cost functions for which the prediction holds.

[17] Interestingly, one might expect that, when the threshold level decreases, the dynamic thresholds would stop working for naives because they would start to procrastinate. However, in simulations and analytical examples, this does not appear to be the case. The reason is as follows. The classic situation in which naives procrastinate is when agents have to complete just one task; in that case, current effort is always a substitute with future effort and so naivete (which increases perceived future effort) always decreases current effort. In contrast, with dynamic thresholds, in some periods current effort is a substitute with future effort, but in many periods it is a complement with future effort; indeed it is the fact that dynamically non-separable contracts generate periods where current effort is a complement with future effort that makes them work better for the impatient. When it is a complement, naives do better; when it is a substitute, sophisticates do better. In simulations, these two forces often cancel, leading dynamic thresholds to work similarly for naives and sophisticates.

[18] 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 with the lag length for four different potential shapes 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.

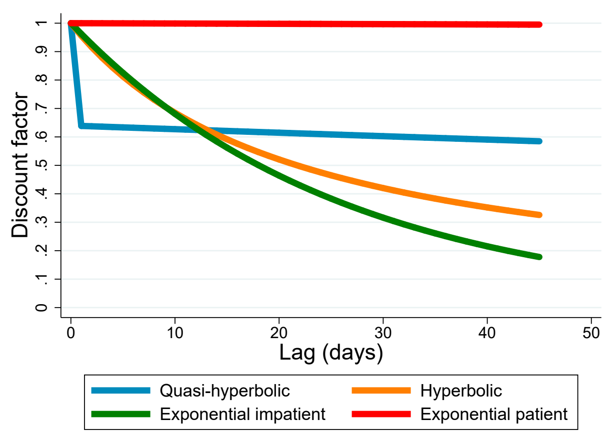



Figure 1: Hypothetical discount factors

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. 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. The presence of 60 million diabetics and 77 million pre-diabetics in the country 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. Despite the Indian government's extensive current efforts to address the epidemic, adoption of the recommended lifestyle changes is low according to physicians. Our study was conducted in partnership with and partially funded by the Government of Tamil Nadu, one of India's southern states, who were interested in evaluating and scaling up effective strategies to promote disease management among diabetics.

## 3.1 Sample Selection and Pre-Intervention Period

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

[19] 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-

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.[20]

Following the phase-in period, surveyors again visited respondents to sync the data from the pedometers, conduct a baseline time-preference survey, and then (after all baseline data was collected) tell participants what treatment group they had been randomly-assigned to for the intervention period.[21] Participants were randomly assigned to 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.

## 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

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.

[20] 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.

[21] 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.

added the exercise routine recommended by the CDC and ADA to his existing behavior.[22] 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: 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.[23] 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.

[22] 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.

[23] 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.

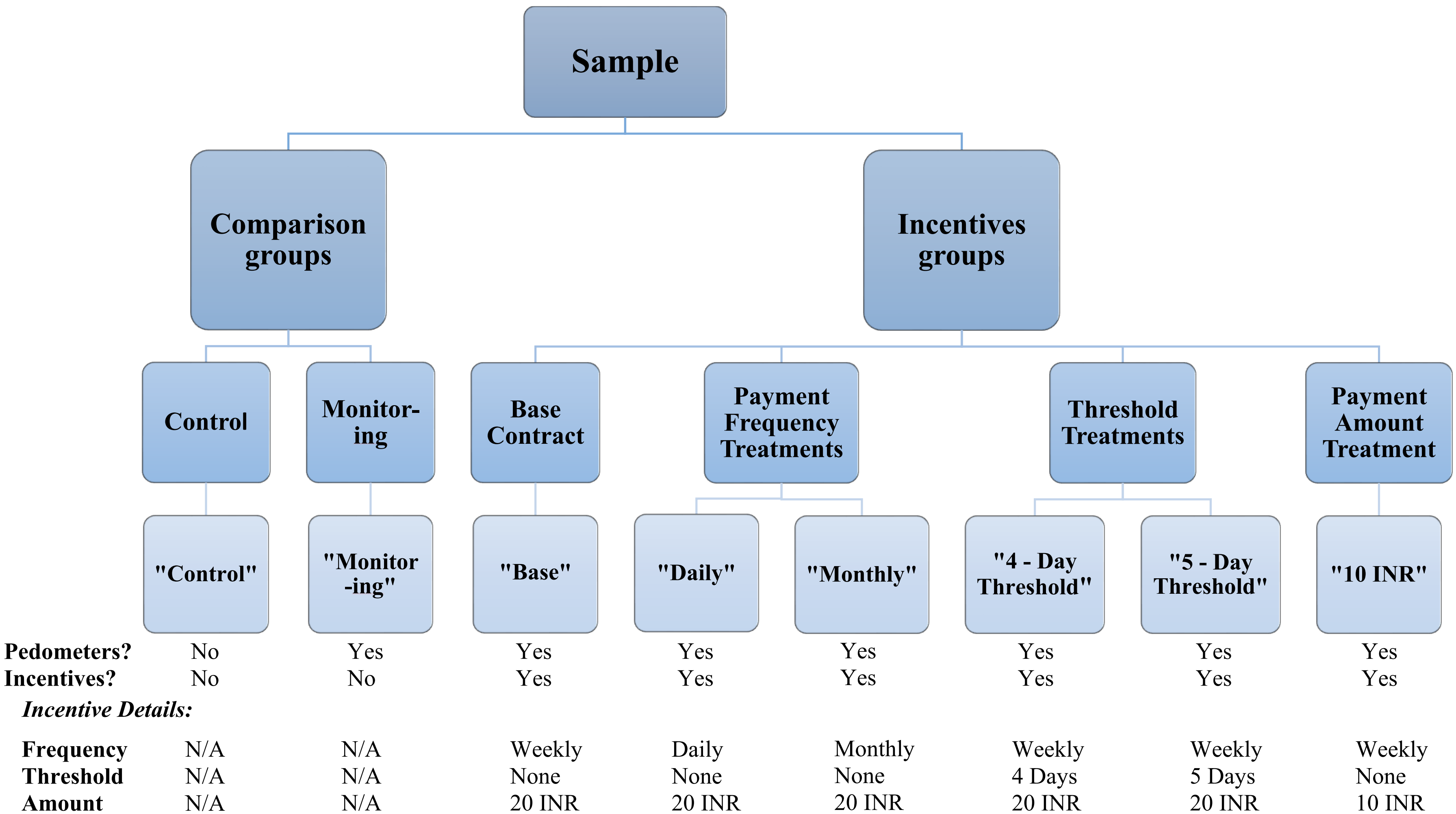



Figure 2: Experimental Design

**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 (including, if applicable, the payment earned and the payment date), and weekly text messages summarizing walking behavior and total payments earned.

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.

**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. 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 20 INR, 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.[24]

**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

[24]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.[25] To control for the possibility that incentives may increase the salience of walking behavior, monitoring participants received the same daily confirmations of their step reports and weekly text messages summarizing their walking behavior that incentives participants did. 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.[26]

## 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.[27] Although

[25] 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.

[26] To accommodate a request from our government partners, we also cross-randomized one additional intervention in 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 show the effects of this treatment in the online appendix.

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

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 bio-metric 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.

Finally, to assess the sustainability of treatment effects from incentives, we continued collecting data from a subset of participants for 12 weeks after the intervention period had ended ("post-endline measurement" group). In particular, we gave pedometers to all post-endline participants (including those originally assigned to the control group) so that we could measure the steps they took. However, no group received any incentives. Field officers simply returned every four weeks to sync pedometers and conduct health measurements. Though participants in all treatment groups received weekly SMS reminders to walk, they were not specifically encouraged to meet any daily step target, and no longer reported their steps daily.

# 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

syncing data to a central database.

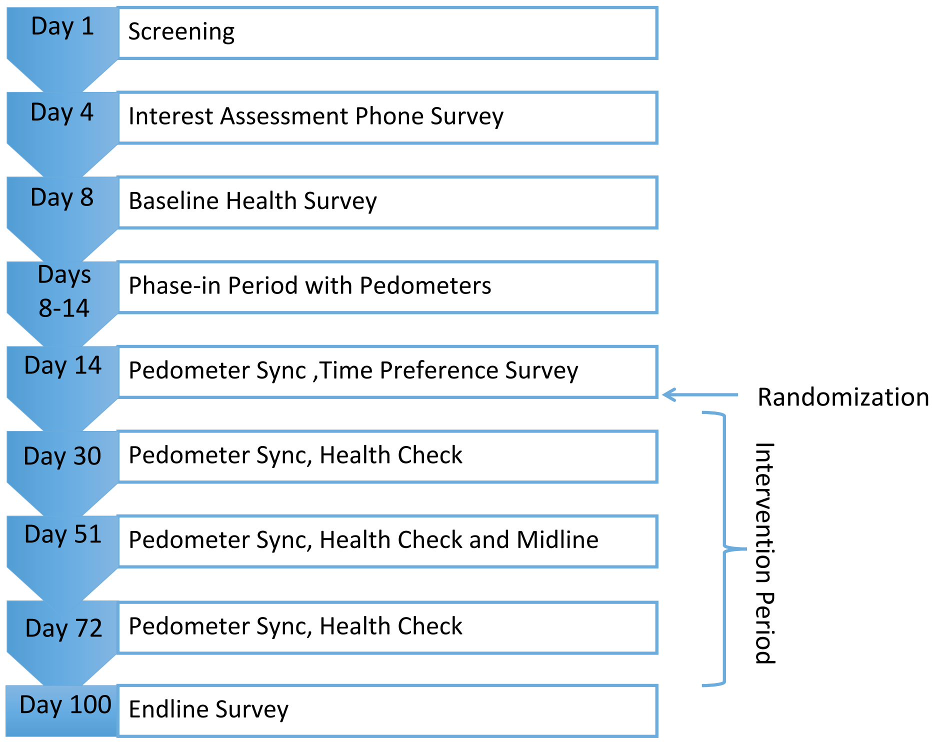



Figure 3: Experimental Timeline for a Sample Participant

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.

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. 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.[28] As highlighted in Kremer et al. (2019), "time preferences are difficult to measure, and the literature has not converged on a broadly accepted and easily implementable approach." Since our sample is somewhat elderly and not screen-savvy, we wanted to ensure we had some simple measures that everyone could follow. Our primary measure of impatience is thus an index of survey-based measures of impatience and procrastination taken from the psychology literature, which we use as a proxy for impatience in the consumption/effort domains. The questions are a subset of the Tuckman (1991) and Lay (1986) scales, with the specific question subset chosen by our field team as being most appropriate for our setting. The questions ask respondents to respond on a Likert scale to a series of questions such as 'I'm continually saying "I'll do it tomorrow"' (See Appendix A for full list of questions). These measures have two key benefits. First, they are simple to understand. Second, the psychology literature has validated that these measures can predict real behaviors well, such as poor academic performance (Kim and Seo, 2015; Özer et al., 2014).[29] Reassuringly, in our setting, the measures also correlate well with behavior; for example, those with higher levels of impatience according to the index have lower levels of baseline walking and worse diets (Appendix Table A.4). Note that the questions are tilted towards procrastination-style behaviors and hence may better detect naive time-inconsistent impatience than other types of impatience, thus potentially meaning that our empirical heterogeneity tests will primarily focus on testing whether the contracts are effective for naifs; since we think efficacy for naives is a key feature of our contract variations we do not find this too problematic but worth acknowledging.

We only began measuring these measures partway through the data collection[30] and so the measures are available for only the second half of the sample. To use the full sample for analysis, we predict the index using lasso based on other measures included in the baseline from the beginning that are similar in spirit (e.g., "In the past week, how many times have

[28]This survey was split temporally from the baseline survey both to avoid survey fatigue and because it was easier to measure time preferences over walking with participants who had used pedometers already so had a sense of what steps mean.

[29]The "money now vs. money later" questions more standard in the economics literature also have this benefit but require discount rates measured over money to not diverge from those over consumption, which can be reasonable in some settings (Andreoni et al., 2018) but (according to our CTB implementation described next) does not seem to be in ours.

[30]In particular, after our field implementation challenges for implementing the Andreoni and Sprenger (2012) measures became clear, as described next.

you found yourself exercising less than you had originally planned?"); see Appendix A.2 for full list of predictors.

One downside of these is that they focus on impatience in the consumption and effort domains, not the payment domain. To get at discounting in multiple domains, we also tried to adapt the convex time budget (CTB) methodology of Andreoni and Sprenger (2012) to measure time preferences over walking and mobile recharges, as described in Online Appendix C. However, these measures can be difficult to implement in the field (Kremer et al., 2019), and we had logistical challenges; for example, it was difficult to get respondents to understand the paradigm, especially in the walking domain.[31] As a result, our implementation did not appear to be successful, and our estimates using this methodology do not correlate in the right direction with any expected behaviors (Appendix Table A.4). We thus do not use these measures in our upfront analysis.

## 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 monthly 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: 65% of the sample has been diagnosed with diabetes by a doctor, and 81% 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: 49% 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

[31] Other challenges include: we were only able to include a fraction as many questions as used in lab implementations and so our estimates do not converge for much of the sample (for example, in the consumption domain, the estimates do not converge for roughly 44% of the sample when we allow the parameter governing utility function convexity to vary at the individual level, and 30% of the sample when we restrict it to be the same); and respondent follow-through on the activities they committed to was low, which is problematic for interpretation.

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 0-0.001, which is similar to monetary discount rates estimated using the CTB methodology in other settings (e.g. Andreoni and Sprenger (2012) 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 0-0.004. 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. This could reflect that people are more impatient over walking than other effort.

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 come 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.[32]

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,[33] 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

[32] 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.

[33] 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) | 0.49 | 0.46 | 0.51 | 0.49 | 0.53 | 0.49 | 0.51 | 0.48 | 0.50 | 0.45 |
| | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) |
| 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 - Phase-in*** | | | | | | | | | | |
| Pr(exceeded step target) | 0.25 | 0.25 | 0.24 | 0.25 | 0.25 | 0.23 | 0.27 | 0.26 | 0.25 | 0.27 |
| | (0.32) | (0.31) | (0.32) | (0.32) | (0.32) | (0.30) | (0.33) | (0.32) | (0.34) | (0.34) |
| Avgerage daily steps | 6999 | 7066 | 6892 | 6998 | 7046 | 6810 | 7449 | 7128 | 6950 | 7018 |
| | (3980) | (3946) | (3697) | (4014) | (4195) | (3969) | (3857) | (4015) | (4087) | (4195) |
| ***D. Time Preferences*** | | | | | | | | | | |
| ***i. Mobile Recharges*** | | | | | | | | | | |
| Discount rate | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.00) |
| ***ii. Steps*** | | | | | | | | | | |
| Discount rate | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 |
| | (0.06) | (0.06) | (0.07) | (0.06) | (0.05) | (0.06) | (0.07) | (0.07) | (0.06) | (0.06) |
| ***F-tests for Joint Orthogonality*** | | | | | | | | | | |
| P-value (relative to control) | N/A | N/A | 0.73 | 0.32 | 0.25 | 0.50 | 0.44 | 0.35 | 0.34 | 0.50 |
| P-value (relative to monitoring) | N/A | 0.73 | N/A | 0.90 | 0.57 | 0.96 | 0.57 | 0.88 | 0.90 | 0.36 |
| P-value (relative to base case) | N/A | 0.50 | 0.96 | N/A | 0.10 | N/A | 0.60 | 0.52 | 0.91 | 0.55 |
| ***Sample size*** | | | | | | | | | | |
| Number of individuals | 3192 | 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: High blood pressure and weight are cardiovascular risk factors. In each domain (mobile recharges and steps), the discount rate $\frac{1}{\hat{\delta}_i} - 1$ 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 the comparison group. The F-statistic is obtained by running regressions with each treatment group seperately.

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. Figure 4 shows that the rates of pedometer-wearing are high and the difference between treatment groups small in magnitude (85% in monitoring vs. 88% in incentives); however, the difference is statistically significant with a p-value of 0.047. To address this, we report Lee (2009) bounds accounting for missing data due to not wearing pedometers when comparing the incentives and monitoring groups.[34]

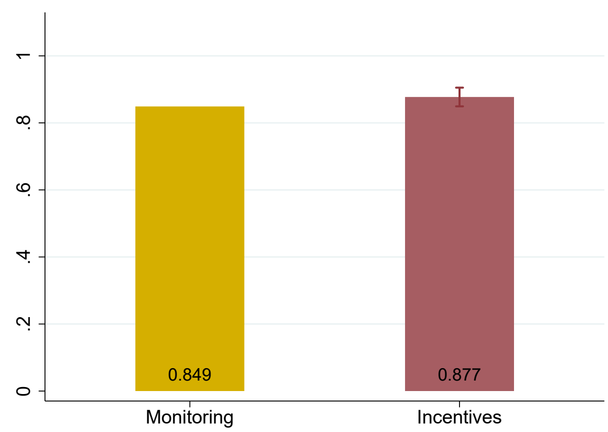



Figure 4: Fraction of days participants wore Fitbits

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. Table E.1 shows that endline attrition rates are not statistically distinguishable between the pure control,

[34] A smaller source of missing fitbit data is that, on 6% of days, we simply do not have data from respondents' fitbits, for example because the fitbit did not sync successfully or because the person withdrew from the intervention. Online Appendix Table E.2 shows that not having fitbit data is balanced across incentive and monitoring groups (col. 2), but that one specific source of missing fitbit data – in particular, withdrawals – is imbalanced (col. 4); however, the imbalance is small in magnitude (2%) and Online Appendix Table E.3 shows that our results are robust to Lee bounds calculated to account for withdrawals specifically. In our main specifications, we condition on having fitbit data but not on the participant wearing the fitbit, and then show that our results are robust to alternate specifications in Online Appendix Table E.3, including: Lee bounds accounting for all sources of missing data jointly; Lee bounds calculated only using one form of missing data or the other (i.e., where we only count data as missing if there is "no fitbit data" vs. only counting it as missing if the person did not wear their fitbit); and Lee bounds focused only on data missing due to mid-intervention withdrawals.

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 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 dynamic thresholds and then of frequency.

## 5.1 Incentives and exercise

We begin by establishing that our incentives program impacts exercise. This finding is of independent policy interest, as exercise has been shown to benefit health for diabetics (Hill, 2005; Manders et al., 2010; Praet and van Loon, 2009; Qiu et al., 2014; Shenoy et al., 2010; Thomas et al., 2009; Zanuso et al., 2009).

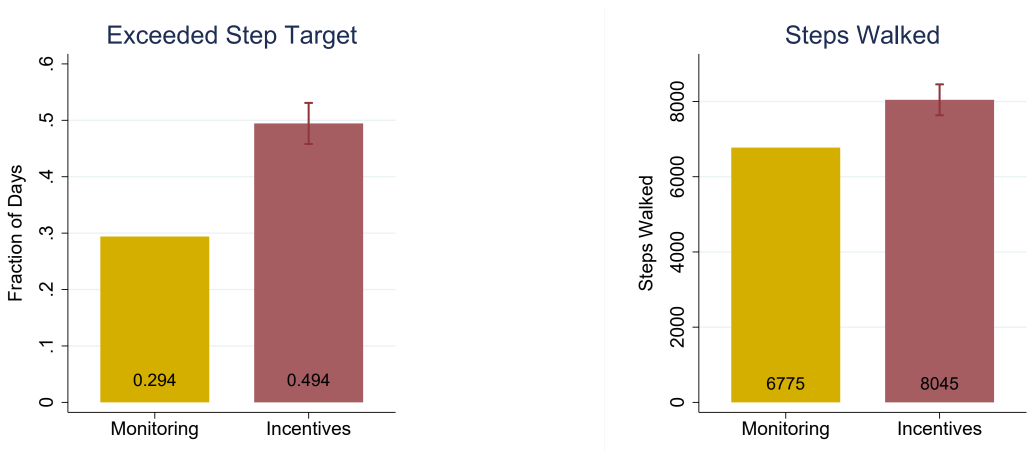
We use intent-to-treat (ITT) estimates to assess the impact of incentives on exercise. In particular, we compare average exercise outcomes in the pooled incentives treatment groups to those in the monitoring treatment group. We focus here on the Fitbit exercise data as it is both less noisy than self-reported 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 outcomes the person-day level across treatment groups using regressions of the following form:

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

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; and $\boldsymbol{X}_i$ and $\boldsymbol{X}_{it}$ are vectors of individual-level and day-level controls, respectively, described in the notes to Table 2. The standard errors $\varepsilon_{it}$ are clustered at the individual level. The coefficient of interest, $\beta$, is the ITT effect of incentives relative to the monitoring group. The results are shown in Panel A of Table 2, and without controls in Table C.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 intervention 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. These effects are not simply a result of participants shifting steps from one day to another: Column 2 shows that incentives increase walking by 1267 steps per day. This effect is equivalent to approximately 13 minutes of extra brisk walking daily on average. Online App. Table E.3 shows that Lee bounds accounting for missing data are significantly different than 0.




(a) Step-target compliance

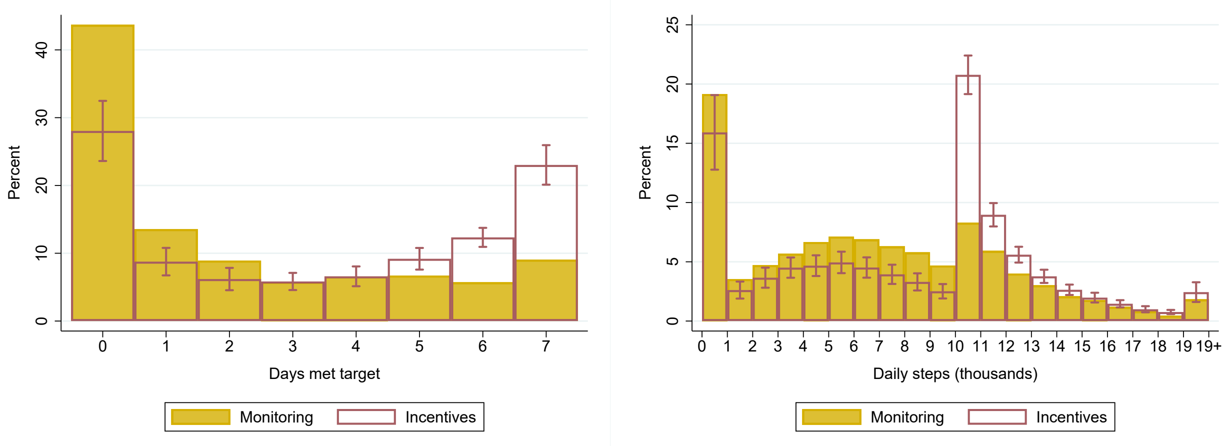
(b) Daily steps walked

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 impact 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. There is bunching at 10,000 steps in both groups, but the bunching in the incentive group is 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.




(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.

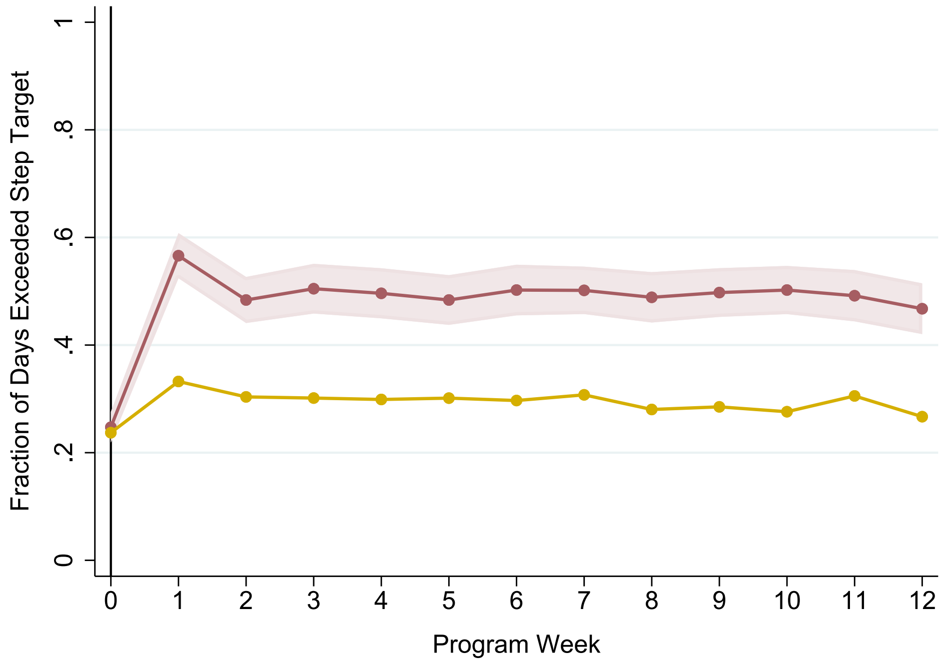
Figure 7 shows that, after an initial spike at week 1, the effect of incentives on walking also does not attenuate over time, a rare finding in the literature with, for example, Patel et al. (2016) finding that physical activity drops steeply 5-7 weeks into a 12-week walking-incentive program.

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

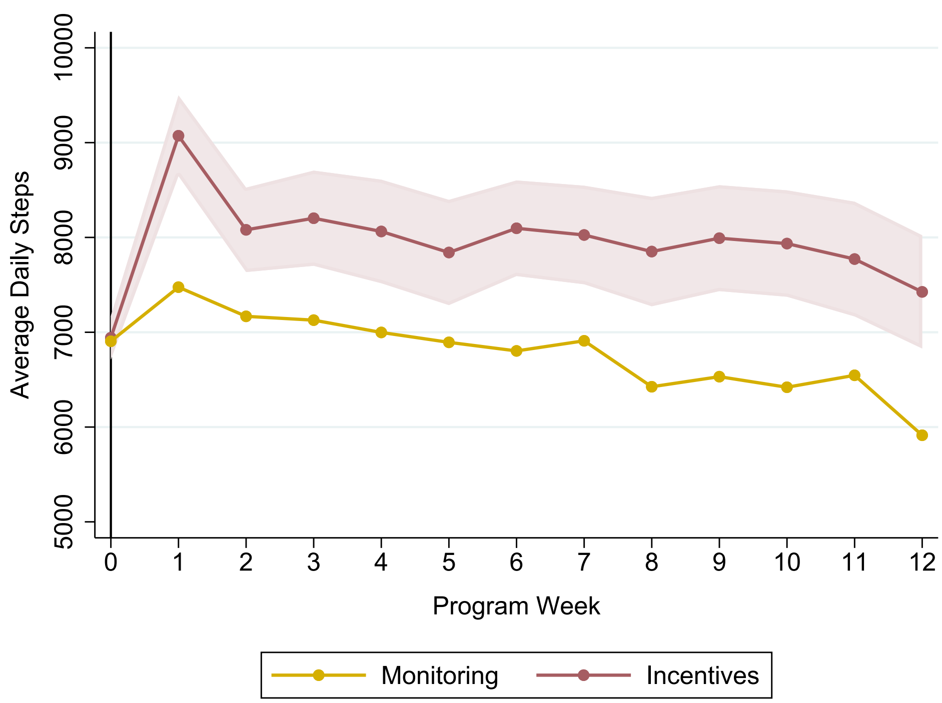
## 5.2 Dynamic Thresholds

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, most notably because our experiment represents (to our knowledge) the first empirical comparison between a dynamically separable and a dynamically non-separable contract. We thus begin by analyzing the average effect of dynamic threshold contracts relative to linear, before turning to test for heterogeneity in their impacts by time preferences.

Figure 7: The effects of incentives persist 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. The the confidence intervals represent the confidence interval for the test of equality between the incentives and monitoring groups from a regression controlling for the same control variables as in Table 2.

Table 2: Impacts of incentives on exercise

| | Pedometer Data (Intervention Period) | | |
|---|---|---|---|
| | Fraction Days Achieved 10K Steps | Daily Steps | Daily Steps (conditional on positive) |
| | (1) | (2) | (3) |
| ***A. Pooled Incentives*** | | | |
| Incentives | 0.197*** | 1266.5*** | 1156.4*** |
| | [0.0179] | [209.5] | [186.3] |
| ***B. Unpooled Incentives*** | | | |
| Base Case | 0.208*** | 1388.1*** | 1196.2*** |
| | [0.0196] | [222.9] | [197.3] |
| Daily | 0.202*** | 1124.4*** | 1202.5*** |
| | [0.0301] | [331.2] | [273.0] |
| Monthly | 0.180*** | 1280.5*** | 1211.1*** |
| | [0.0281] | [307.3] | [263.9] |
| 5-Day Threshold | 0.208*** | 1306.9*** | 1228.4*** |
| | [0.0250] | [264.4] | [229.9] |
| 4-Day Threshold | 0.189*** | 1182.1*** | 1114.8*** |
| | [0.0203] | [230.6] | [203.0] |
| Small Payment | 0.124*** | 717.9* | 499.7 |
| | [0.0382] | [385.5] | [327.2] |
| Monitoring mean | 0.294 | 6774.522 | 7985.931 |
| Controls | Yes | Yes | Yes |
| *P-value for Base Case vs* | | | |
| Daily | .84 | .35 | .98 |
| Monthly | .26 | .67 | .94 |
| 4-Day Threshold | .97 | .68 | .85 |
| 5-Day Threshold | .21 | .17 | .52 |
| Small Payment | .02 | .05 | .02 |
| # Individuals | 2,559 | 2,559 | 2,557 |
| #Observations | 205,732 | 205,732 | 180,018 |

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 pedometer data during the intervention period. The sample includes the incentives and monitoring groups. Individual-level controls include a second order polynomial of age and weight, gender, and the average of the dependent variable during the phase-in period (before randomization). Day-level controls include fixed effects for the month-year and day-of-week. The Small Payment group received 10 INR instead of 20. The omitted category in all columns is the monitoring group. Standard errors, in brackets, are clustered at the individual level. The number of individuals is smaller from Table 1 because 48 people in the monitoring and incentives groups withdrew immediately. We remain balanced across groups, as the difference between respondent's likeliness to withdraw immediately by group is not significant (p-value $> 0.7$).

#### 5.2.1 Average Effectiveness

Panel B of Table 2 evaluates the ITT effects of all of our incentive contract variations, estimating regressions of the following form:

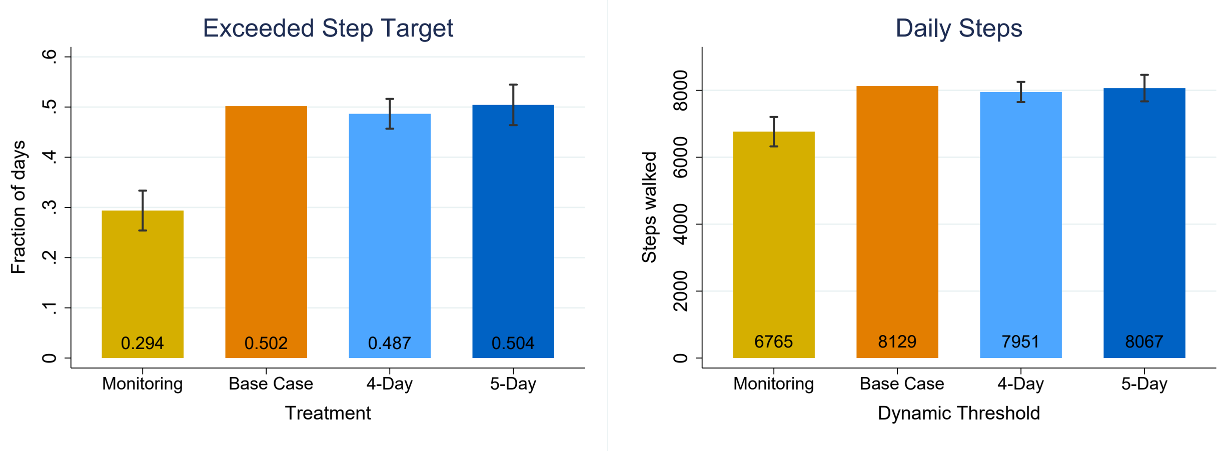
$$y_{it} = \alpha + \beta_j \times \left(incentives^j\right)_i + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\theta + \varepsilon_{it}, \tag{6}$$

where $y_{it}$ are daily walking outcomes and $(incentives^j)_i$ 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 (dynamic separability, payment frequency, or payment amount); the bottom rows of the table thus show the p-values for the significance of the difference between each treatment group and the base case group.

We find that adding a dynamic threshold does not affect the level of exercise. 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 not statistically significant. The number of steps are similar across treatments as well.

However, the dynamic threshold contracts have an advantage. 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.

Table 3 quantifies the cost-effectiveness of all contracts in two ways. Column 5 shows the average incentive delivered on a day the participant exceeded the daily 10,000 step target. For all contracts other than the 4- and 5-day thresholds, which all pay out linearly, this is by definition 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 10% and 15% are made while participants achieve the same amount of walking. Column 6 shows the average incentive cost per additional day the step target was reached above the monitoring group. According to this metric, the 4-day threshold and 5-day threshold achieve cost savings of 8% and 16%, respectively. For comparison, the incentive amount per day walked is mechanically lower in the 10-INR treatment group, but




(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. The confidence intervals show the 95% confidence interval for a test of equality between the base case incentive group and each other treatment group, controlling for the same things as Table 2.

this comes at the cost of reduced walking overall.

One potential explanation for similar average walking in threshold and base case groups is that individuals 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 C.3 substantiates these conclusions using quantile regressions. In sum, although thresholds do not work well for everyone, they work very well for some people, inducing them to walk with more consistency across days than non-threshold contracts.

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,

Table 3: Cost Effectiveness of Monitoring and Incentive Treatments

| | Cost-effectiveness of Incentive Contracts | | | | | |
|---|---|---|---|---|---|---|
| | Walking | | Rewards | | Cost-effectiveness | |
| | Compliance - Proportion of days met step target | Treatment effect relative to monitoring | Incentive Amount (INR) | Proportion Compliance Incentivized | INR per Day Complied | INR per Day Complied above Monitoring |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Monitoring | 0.29 | N/A | 0 | 0 | 0 | N/A |
| Daily | 0.5 | 0.21 | 20 | 1 | 20 | 48.69 |
| Base Case | 0.5 | 0.21 | 20 | 1 | 20 | 48.36 |
| Monthly | 0.49 | 0.2 | 20 | 1 | 20 | 49.71 |
| 4-Day Threshold | 0.5 | 0.2 | 20 | 0.9 | 18.09 | 44.26 |
| 5-Day Threshold | 0.51 | 0.22 | 20 | 0.85 | 17.07 | 40.41 |
| Small Payment | 0.44 | 0.15 | 10 | 1 | 10 | 30.18 |

Notes: INR per day complied represents Incentive Amount $\times$ Proportion compliance incentivized (i.e., Column (5) = Column(3) $\times$ Column(4)). INR per day complied above monitoring represents the total INR paid to a person (INR per day complied $\times$ Compliance) divided by the non-inframarginal compliance (Compliance - Monitoring Avg.) (i.e., Column(6) = Column(5) $\times$ Column(1)/(Column(1) - 0.3)).

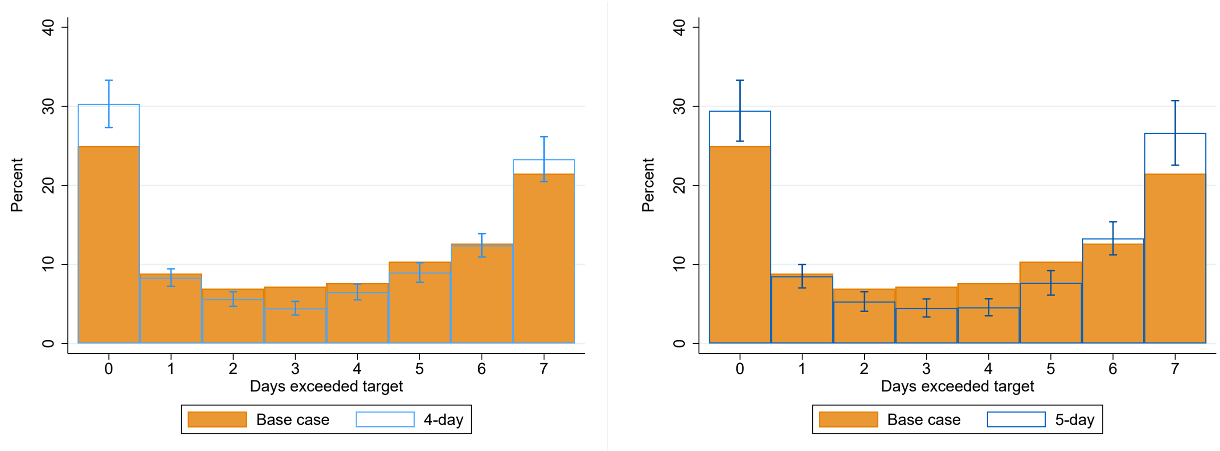



Figure 9: Days walked per week in threshold and base case (linear) contracts.

Notes: These figures show 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. All data is at the respondent × week level. Orange bars show raw base case means; blue bars add on the coefficient on a dummy for being in the 4-day or 5-day threshold treatment from a regression where the dependent variable is a dummy for whether the days on which the participant exceeded the step target is equal to the value on the x-axis. Regressions control for the same things as Table 2. Confidence intervals are for the test of equality between the base case and 4-day or 5-day treatment from the same regression.

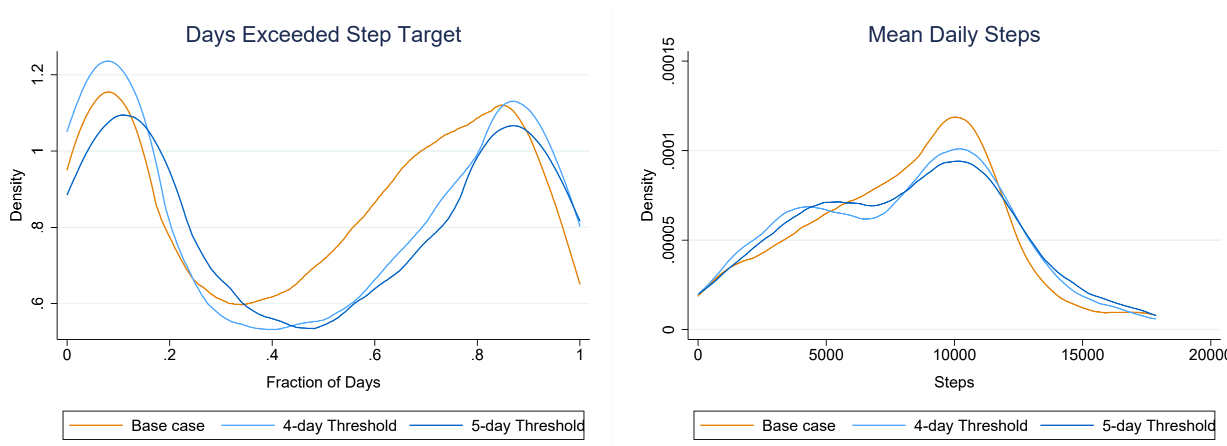



Figure 10: Fraction of days walked and average steps at the participant level, for base case vs. threshold

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.

instituting a dynamic threshold creates a tradeoff, decreasing the cost per day of exercise induced, but perhaps also diminishing the health benefit per day of exercise induced.[35] The 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 Heterogeneity in Dynamic Threshold Effects by Time Preferences

To test for heterogeneity in the effects of dynamic thresholds by impatience, we use a regression of the following form:

$$y_{it} = \alpha + \beta_1 impatience_i \times threshold_i + \beta_2 threshold_i + \beta_3 impatience_i + \boldsymbol{X}'_i \pi + \boldsymbol{X}'_{it} \theta + \varepsilon_{it}, \quad (7)$$

where $y_{it}$ is an indicator for whether the individual surpassed the 10,000-step target, for individual $i$ on day $t$ during the intervention period. Following our ex ante analysis plan, we pool the threshold treatments for power purposes, so $threshold_i$ is an indicator for being in either threshold group (see Appendix Table C.4 for disaggregated results), and $impatience_i$ is a measure of agent impatience.

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. The key coefficient of interest is $\beta_1$, which captures how the effect of the threshold (relative to the base case) varies with impatience.

Table 4 shows that, consistent with our theoretical predictions, thresholds work better for those with higher impatience over steps. Column 1 uses the impatience index as the measure of impatience (i.e., our standardized index of questions on impatience and self-control from the psychology literature). Although our power is low because we only have these measures for the second half of our sample (as described in Section 4.1), the coefficient is marginally statistically significant and the magnitudes meaningful, suggesting that having a 1 standard deviation higher value of the impatience index increases the average performance of the threshold contracts relative to linear contracts by 6pp, a large increase relative to the sample-average effect of either contract (20pp). To improve power, column 2 uses a lasso prediction of the impatience index available for the full sample; there we find similar results. Finally, to aid in interpretation, column 3 shows the results using a dummy for whether a participant is above the median on the predicted impatience index. The results show that (relative to

[35] On a similar note, one might think there would be greater value in inducing exercise among those who have low levels of baseline exercise rather than those who already exercised a lot at baseline. Appendix Table C.6 and Figure E.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: Dynamic thresholds increase walking more for those who are more impatient.

| Dependent variable: | Met step target (×100) | | |
|---|---|---|---|
| Impatience measure: | Impatience index | Predicted impatience index | Above median predicted index |
| Sample: | Late | Full | Full |
| | (1) | (2) | (3) |
| Impatience × Threshold | 5.59* | 3.087*** | 5.56** |
| | [3.013] | [1.740] | [3.634] |
| Threshold | -1.512 | -1.216 | -3.412** |
| | [1.919] | [1.763] | [2.353] |
| Impatience | -5.436*** | -1.989*** | -4.047** |
| | [2.009] | [1.231] | [2.620] |
| # Individuals | 1075 | 1969 | 1969 |
| Base case mean | 50.4 | 50.2 | 50.2 |

Notes: This table shows heterogeneity by impatience in the effect of threshold contracts relative to linear contracts. The impatience measure changes across columns. For Columns (1) and (2), the units are standard deviations on the index. Because columns (2) and (3) use an impatience measure that was predicted using lasso, we bootstrap standard errors and present the bootstrapped standard errors in brackets, with the stars indicating significance levels determined by the bootstrap confidence intervals. The sample includes the base case and dynamic threshold incentive groups only. Control variables include gender, age, weight, time fixed effects, and the baseline value of the dependent variable. Standard errors in brackets clustered at the respondent level. Significance levels: * 10%, ** 5%, *** 1%.

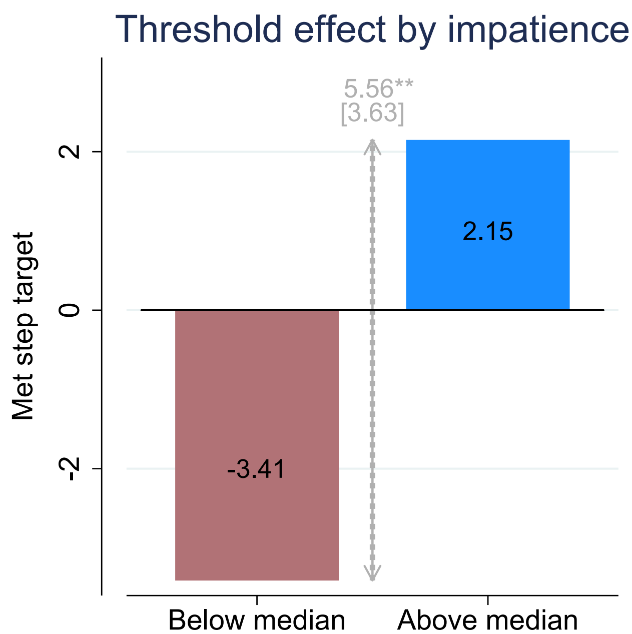



Figure 11: Dynamic thresholds increase walking more for the impatient

Notes: The sample includes the base case and dynamic threshold incentive groups only. The chart plots the total average compliance in the threshold contracts relative to the base case contract, estimated separately for those with below-median predicted impatience (left bar) vs. above-median predicted impatience (right bar). The vertical arrow between the columns shows the coefficient and standard error for the difference between the treatment effects, taken from column 3 of Table 4. The base case (linear) group is the omitted category. Control variables include all incentive subtreatment dummies, gender, age, weight, time fixed effects, and the baseline value of the dependent variable.

the base case linear contract), the threshold works 5.6pp better for those with above-median predicted impatience than those below. Figure 11 is a visualization of column 3, showing that (relative to a linear contract), adding the threshold actually decreases performance for those who are more patient, while increasing it for those who are more impatient, with the difference statistically significant.

Of course, impatience is not randomly assigned and could correlate with other variables that influence the effectiveness of the threshold treatment, thus biasing our estimate of $\beta_1$ in Table 4. Note that, if those correlations are likely to hold in other settings (e.g., because the correlated variables are downstream outcomes of impatience or fundamentally linked with impatience) then the results presented in Table 4 might be the most relevant ones from a policy perspective. For example, if policymakers were to assign participants to threshold vs. linear contracts based on individual-level impatience, what they would care about is the predictive relationship shown in Table 4.

However, as a purer test of our theory, we are also interested in the causal impact of impatience on threshold performance, holding other factors constant. We have two approaches to shed light on this relationship. First, Appendix Table C.7 controls for other baseline covariates and their interactions with the threshold treatments. The table shows that the threshold interactions are robust to controls, suggesting that it is likely impatience itself (and not its correlates) driving the estimated relationship. Note that another potential confound is habit formation: if individuals vary in their propensity to form habits, those with a higher propensity could perform better in the threshold contract, which could be a confound if correlated with discount rates. However, Appendix Table C.8 suggests that the propensity to form habits is not correlated with impatience, as impatience does not predict the persistence of incentive effects after payments stop.

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. Average walking in the monitoring treatment uncovers $F(0)$; average payday walking in the 10 INR treatment uncovers $F(10)$; and average walking in the base case group on payday and in the daily group uncover $F(20)$. 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 uncovers $F(80)$ ($F(100)$); these final moments improve fit, but do involve an assumption that costs are IID across people.

We then use this normal distribution to estimate how relative compliance in the base case and threshold contracts would vary with the discount rate over walking, $d_c$.[36] 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 base case 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, the increase in performance of the threshold contract as impatience increases should be quantitatively important. 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, and should not we believe meaningfully affect the heterogeneity by impatience, which is our main interest here.

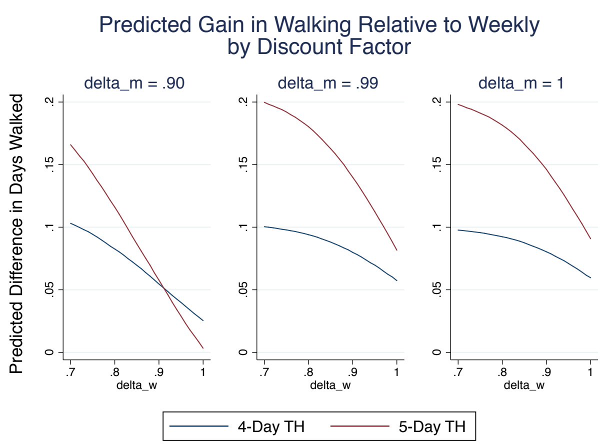



Figure 12: Calibration: Threshold compliance relative to the base case linear compliance, by the discount factor over walking

Notes: "delta w" represents here the discount factor over walking and "delta m" the discount factor over financial payments.

## 5.3 Payment Frequency

Motivated by Section 2, we conduct two primary tests:

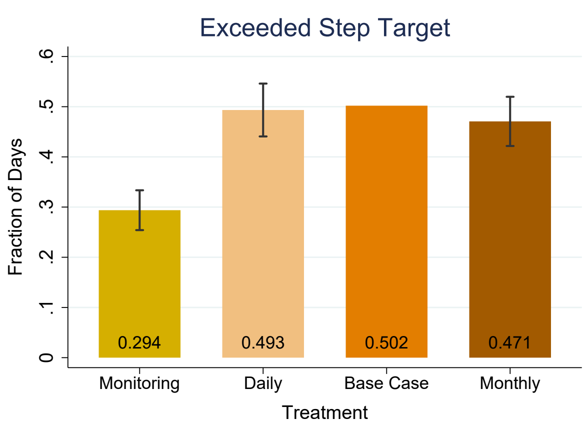
1. Between-treatment: We compare average compliance between the daily, weekly (base case), and monthly payment groups to assess how payment frequency affects compliance and shed light on the level and shape of discount rates via Predictions 2 and 3.

[36] We assume sophistication; we are in the process of adding naivete.

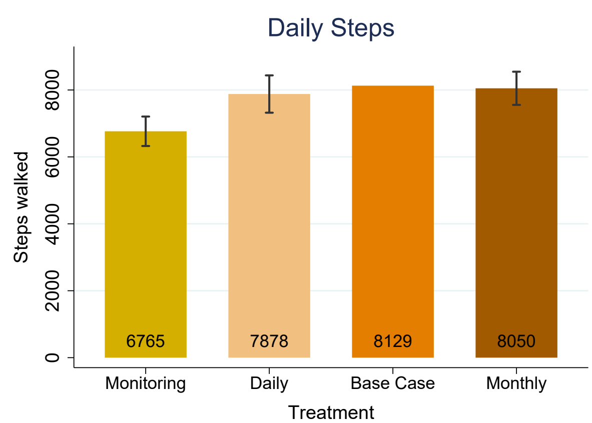
2. Within-treatment: Within the base case and monthly groups, we examine whether compliance increases as the payday approaches to shed light on discount rates via Prediction 4. Similar variation has been used in previous studies to shed light on discount rates (Kaur et al., 2015; Oyer, 1998).

The approaches are complementary. The between-treatment approach directly answers the policy question of whether payment frequency matters, while the within-treatment approach has higher statistical power and can shed additional light on the shape of discounting.

We begin with the between-treatment comparisons. In addition to Panel B of Table 2, Figure 13 shows the compliance in the frequency treatments visually. The impacts of the three frequency treatments on both the likelihood of exceeding the step target and on average steps walked are statistically indistinguishable. In addition, the differences between the point estimates are relatively small, and the effects not monotonic with frequency: the weekly (base case) group has slightly higher steps and compliance than the other groups, while daily is ranked second when compliance is the outcome and monthly ranked second when steps are the outcome.




(a) Probability Exceeded Step Target

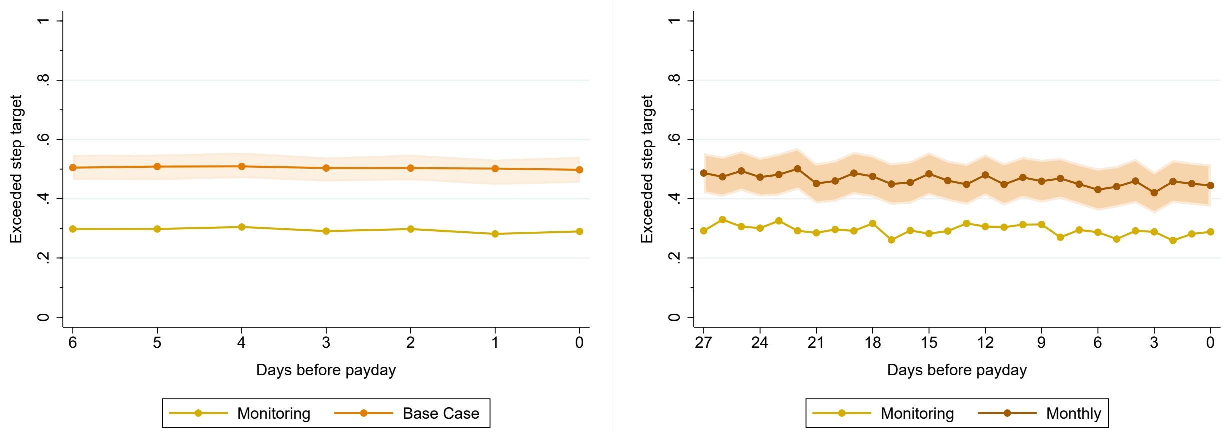



(b) Average Daily Steps

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 same things as Table 2. We also control for different contract period start dates completed surveys.

We thus do not find any meaningful evidence that increasing payment frequency in the range from daily to monthly affects compliance, implying that the discount rate over financial payments may be relatively small over this range. The implication for discount rates




(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, as well as the same things in Table 2. 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.

depends on whether there are other confounds besides discount rates that also affect individual responsiveness to payment frequency; however, most of the potential confounds one can think of (e.g., concavity of utility in payment, higher salience of higher frequency payments) would also bias higher frequency payments towards working better, thus making it unlikely that confounds (as opposed to lack of discounting of payments) are driving the null result. However, note that precision is a caveat to these conclusions: we cannot rule out that daily has an effect 4 percentage points higher than the base case, or that monthly has one 8 percentage points lower. We thus turn next to the within-treatment test, which has somewhat higher statistical power, to confirm the evidence from this analysis that discount rates over payment are low and that payment frequency does not meaningfully improve compliance.

Figure 14 shows how compliance within the base case 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. Table 5 estimates the slopes of each line as the payment day approaches, conditional on day of week fixed effects.[37] The estimates are not significant and suggest that, if anything, the slopes decrease as the

[37]There is variation in day of week relative to payday since the intervention launch visits were done on all days of the week. Since the launch visit dates were endogenous to participants' schedules, survey day of week may also be endogenous; as a result, we also randomized the delay between the survey date and the intervention start date. We thus also control for fixed effects for the day-of-week relative to survey date.

payment day approaches: for each day closer to the payday, base case (weekly) participants are 0.1% less likely to comply, and monthly participants are 0.08% less likely. Our confidence intervals are also tighter here: if we assume linearity of compliance in lag to payment, the bottom end of the confidence interval for the slope of the base case weekly treatment can rule out that weekly payment efficacy decreases by 0.09 percentage points with each day further from payment, which would correspond to ruling out that daily payments are on average more than 0.36 percentage points more effective than weekly. Although these results are consistent with recent work by Augenblick et al. (2015), the absence of payday spikes conflicts with Kaur et al. (2015); an open question for further work is to further explore the reasons for the differences (e.g., whether they reflect different countries, different payment amounts, or different settings in health vs. workplace).

Table 5: Walking does not vary significantly across the paycycle

.

| Dependent variable: | Met step target (×100) | | | | |
|---|---|---|---|---|---|
| Payment Frequency: | Weekly | | Monthly | | |
| | (1) | (2) | (3) | (4) | (5) |
| Days before payday | 0.0993<br>[0.0953] | | 0.0760<br>[0.0500] | | |
| Payday | | -0.596<br>[0.572] | | 0.0757<br>[1.051] | |
| Payweek | | | | | -0.0757<br>[1.051] |
| # Observations | 71,822 | 71,822 | 13,373 | 13,373 | 13,373 |
| Sample mean | 50.15 | 50.15 | 49.23 | 49.23 | 49.23 |

Notes: Columns show the effect of paydays on the probability of meeting the step target, in the weekly and monthly frequency groups. The sample in columns (1) and (2) is restricted to the weekly treatment group and the sample in columns (3) and (4) is restricted to the monthly treatment group. All regressions control for the same things as Table 2 along with a day-of-contract-period time trend. Standard errors in brackets clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Thus, the 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, and that increasing frequency between daily and monthly does not have meaningful effects on average compliance. These are all average effects. Appendix B tests for heterogeneity by discount rates over recharges. We do not find significant heterogeneity in either within-treatment or between-treatment effects. Even for those with high measured impatience over recharges, increasing payment frequency does not appear to be effective in this context.

# 6 Results: Program Evaluation

In this section, we evaluate the effects of the incentives and monitoring treatments on health and behavioral outcomes, investigate whether monitoring alone affects exercise, and check if the impact of incentives on walking persists after we stop paying respondents.

## 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. Table 6 reports results from regressions of the following form:

$$y_i = \alpha + \beta_1 \times incentives_i + \beta_2 \times monitoring_i + \boldsymbol{X}_i' \gamma + \varepsilon_i \tag{8}$$

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.

We report ITT effects on outcomes in five categories: physical health (our primary outcomes), 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.[38] While we report regression estimates for each outcome individually, we focus on the category indices for inferring effectiveness.

Panel A of Table 6 suggests that the incentives program improves health indicators in this population.[39] Column 1 presents the treatment effect on the "Health Risk Index", which

[38] We follow Kling et al. (2007), by imputing missings (for individuals who have non-missing responses to at least one component of the index) for each component using the sample mean.

[39] All physical health outcome variables are trimmed according to the WHO flexible exclusion method for dealing with biologically implausible values for health outcomes; specifically, for each measure, observations for which z-scores were greater than 4 units from the mean z-score are trimmed. Note that the WHO guidelines are designed for infants and children; the WHO does not provide guidelines for adults. Results are robust to other methods of trimming outliers (e.g., winsorizing at the 1st percentile).

averages the five health risk factors displayed in the table. We find a moderately sized change in health risk of -.05 SD's, significant at the 10% level. 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. The table also shows that monitoring alone did not seem to impact health. Since the effects of the incentives may vary with health status, Table 7 tests for heterogeneity in the impacts by baseline blood sugar, suggesting that the impacts may be more on the BMI margin for those with lower baseline blood sugar and on the blood sugar margin for those with higher baseline blood sugar.

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. Online Appendix Table E.6 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 not scalable.

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. Panel B of Table 6 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, 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 in our surveys.

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 (Rajeswari et al., 2005; Sinha et al., 2013). We selected questions related to emotional health. Each question asks for the frequency of a feeling or event in the previous four weeks.[40] Answers are then recoded so that larger values indicate better mental health.

Panel C of Table 6 shows that the incentives program significantly improves our index of mental health. Although many studies have found a positive association between exercise

[40]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?"

and mental health (Biddle, 2016), experimental evidence that exercise causes improvements in mental health is scarce and mixed. Our result is novel experimental evidence that exercise can improve mental health, although we cannot rule out that the channel is an income effect from the incentive payments themselves rather than an exercise effect.

Finally, Table C.1 examines effects two dimensions of a healthy lifestyle: diet, and addictive good consumption. We do not find significant incentive effects on either.

## 6.2 Monitoring and Exercise

The previous results suggest that the monitoring group had limited impact, although the results are somewhat imprecise. Did the monitoring treatment not affect exercise, or were the exercise impacts too small to translate into measurable health impacts? In this section, we evaluate the impact of monitoring on walking.

Because we do not have pedometer walking data from the control group and the self-reported exercise data appears to be biased (see Online Appendix C for more detail), 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 participants 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 to help address 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 E, controlling for date 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 39 steps (statistically indistinguishable from zero), which is only 3% 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.

Table 6: Impacts of incentives and monitoring on health.

| A. Health risk factors | Health Risk Index | HbA1c | Random Blood Sugar | Mean Arterial BP | Body Mass Index | Waist Circumference |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Incentives | -0.047* | -0.072 | -5.92* | 0.099 | -0.052 | -0.19 |
| | [0.025] | [0.071] | [3.42] | [0.42] | [0.042] | [0.27] |
| Monitoring | 0.023 | -0.14 | 1.74 | 1.19 | 0.067 | 0.017 |
| | [0.045] | [0.13] | [6.06] | [0.75] | [0.074] | [0.48] |
| Control mean | 0.00 | 8.44 | 193.83 | 103.02 | 26.45 | 94.44 |
| P-value: M = I | 0.08 | 0.56 | 0.16 | 0.10 | 0.07 | 0.63 |
| # Individuals | 3,192 | 3,066 | 3,067 | 3,056 | 3,058 | 3,059 |

| B. Fitness | Fitness Time Trial Index | Seconds to Walk 4m | Seconds for 5 Sit-Stands |
|---|---|---|---|
| | (1) | (2) | (3) |
| Incentives | 0.040 | 0.043 | -0.11 |
| | [0.028] | [0.043] | [0.12] |
| Monitoring | 0.076 | 0.086 | -0.088 |
| | [0.050] | [0.076] | [0.20] |
| Control mean | 0.00 | 3.88 | 13.18 |
| P-value: M = I | 0.42 | 0.53 | 0.92 |
| # Individuals | 3,192 | 2,825 | 2,793 |

| C. Mental Health | Mental Health Index | Felt Happy | Less Nervous | Peaceful | Energy | Less Blue | Less Worn | Less harm to Social Life |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Incentives | 0.073** | 0.089** | 0.026 | 0.055 | 0.065 | 0.015 | 0.091** | 0.052* |
| | [0.032] | [0.045] | [0.044] | [0.047] | [0.048] | [0.043] | [0.039] | [0.030] |
| Monitoring | 0.11* | 0.075 | 0.12 | 0.091 | 0.037 | 0.12 | 0.18** | 0.050 |
| | [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 | 0.51 | 0.84 | 0.16 | 0.62 | 0.71 | 0.12 | 0.17 | 0.97 |
| # Individuals | 3,192 | 3,068 | 3,068 | 3,068 | 3,068 | 3,068 | 3,068 | 3,068 |

Notes: Standard errors in brackets. Controls for all outcomes are the same as Table 2, along with second order polynomials of relevant health variables at baseline. We follow Kling et al. (2007) in constructing the index by imputing missings for each component using the relevant sample mean per group. The Health Risk Index is created by averaging the 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) and MAP is the mean arterial blood pressure (mm Hg). A large value of Fitness Time Trial Index indicates low fitness. The Mental Health Index averages the values of seven questions adapted from the Rand 36-Item Short Form Survey (SF-36). The omitted category in all columns is the pure control group.

Table 7: Incentives may affect different health margins depending on blood sugar levels.

| | Health risk index | HbA1c | Random blood sugar | Mean arterial BP | Body mass index | Waist circumference |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Incentives X Low Hba1c | -0.021 | -0.031 | -2.70 | 0.29 | -0.14** | -0.22 |
| | [0.035] | [0.097] | [4.56] | [0.59] | [0.058] | [0.38] |
| Incentives X High Hba1c | -0.079** | -0.15 | -12.1** | -0.15 | 0.064 | -0.089 |
| | [0.036] | [0.10] | [4.78] | [0.61] | [0.061] | [0.40] |
| Monitoring X Low Hba1c | 0.014 | -0.075 | -5.90 | 2.31** | 0.079 | -0.74 |
| | [0.063] | [0.17] | [8.22] | [1.06] | [0.10] | [0.68] |
| Monitoring X High Hba1c | 0.015 | -0.24 | 6.40 | -0.16 | 0.072 | 0.81 |
| | [0.063] | [0.18] | [8.32] | [1.06] | [0.11] | [0.69] |
| Control Mean | 0.00 | 8.44 | 193.83 | 103.02 | 26.45 | 94.44 |
| Control SD | 0.98 | 2.36 | 94.42 | 13.27 | 4.22 | 10.41 |
| # Individuals | 3186 | 3061 | 3062 | 3051 | 3053 | 3054 |
| P-value: I high Hba1c = I low Hba1c | 0.25 | 0.38 | 0.15 | 0.60 | 0.02 | 0.81 |

Notes: Standard errors in brackets. Controls for all outcomes are the same as Panel A of Table 6 and all indexes are constructed in the same way. We also control for the effect of pure incentives, monitoring and high or low blood sugar in the regression but only report heterogeneity terms. Interaction terms are generated from baseline Hba1c levels. The omitted category in all regression columns is the pure control group.

## 6.3 Persistence of Treatment Effects

Do the effects of offering incentives persist after the payment period ends? Results from the 12-week "post-endline" measurement period suggests that they do. Table 8 summarizes the average treatment effects of incentives during the 12-week intervention period, and then for the 12 week post-endline period.[41] Though the treatment effect declines in the post-endline period, there are still meaningful positive impacts: the average effect of receiving any type of incentive for walking in the contract period increases an individual's daily steps by 633 on average in the post-endline period, and his or her probability of meeting a 10,000 daily step target by 9 pp in the post-endline period, relative to those who received no types of incentives in the contract period. These post-endline treatment effects are 54% and 42% as large as the intervention period treatment effects, respectively. Figure 15 shows that these differences are significant at the 5% level, and Panel C of Figure 15 shows that the differences persist through the end of the post-endline period. These results suggest that incentives may

[41] Note that individuals wore their pedometers less in the post-endline measurement period than the contract period, with average wearing rates declining from around 87% to 70% (see Online Appendix Table E.4). However, reassuringly, wearing rates were balanced between the incentives and monitoring groups (Online Appendix Table E.5. For easier comparability with the intervention period effects, we show upfront the results conditional on wearing the pedometer and include the unconditional results in Online Appendix Figure E.2 and Table E.8; regardless of the specification we see substantial persistence.

Table 8: Percentage persistence of treatment effects from incentives

| | Outcomes | |
|---|---|---|
| Dependent variable: | Compliance with 10,000 steps | Daily Steps |
| Treatment effect **during** intervention | 0.217 | 1173 |
| Treatment effect **after** intervention | 0.092 | 633 |
| Percentage retained after intervention | 42% | 54% |

Note: Table summarizes the average treatment effect of incentives during the intervention period and post-endline period, conditional on individuals wearing their pedometers. A respondent was considered to have worn their pedometer if their pedometer recorded a step count > 0, conditional on non-missing data. We use the same control variables as in Table 2.

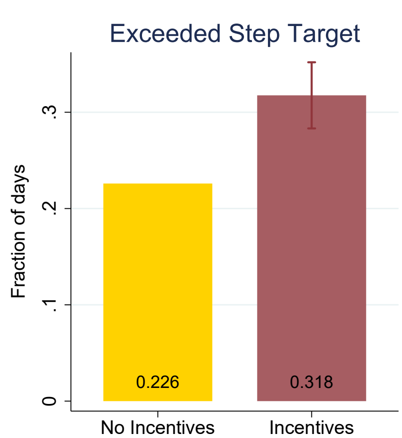
lead to habit formation, and are promising for the cost-effectiveness of the program.

# 7 Conclusion

This paper investigates incentive design for impatient agents. Starting from a standard model where agents discount consumption and financial rewards differently, we identify incentive contracts variations 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 that the dynamic separability of the contract will interact with impatience over consumption.

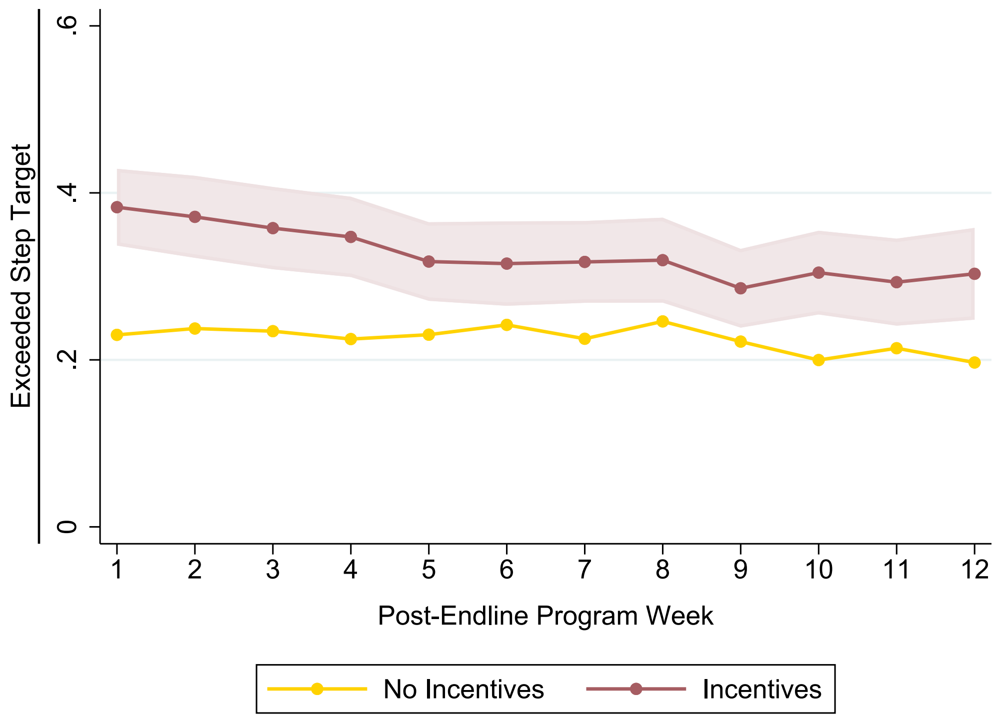
In order to test our predictions, we implement an RCT to incentivize walking among approximately 3000 individuals with diabetes and prediabetes in India. Overall, the incentives program leads to a large increase in walking among the study population and leads to improvements in diabetes- and mental-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

Regarding the interaction of incentive design and time preferences, we find evidence that individuals are impatient over consumption - that is, they prefer to put off the effort of walking. Moreover, consistent with our model, the dynamically non-separable contract works better for those who are more impatient over consumption. In the payment domain, we find limited evidence of impatience over payments. As a result, neither more frequent nor




(a) Step-target compliance


(b) Daily steps walked




(c) Week-by-week probability of walking 10,000 steps each day

Figure 15: Treatment effects from incentives persist in the 12 week post-endline period

Note: Figure displays the impact of the pooled incentives treatments on steps during the post-endline period, conditional on wearing a pedometer. 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. The the confidence intervals represent the confidence interval for the test of equality between the incentives and non-incentives groups from a regression controlling for the same control variables as in Table 2.

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


Andreoni, J., Gravert, C., Kuhn, M. A., Saccardo, S., and Yang, Y. (2018). Arbitrage or Narrow Bracketing? On Using Money to Measure Intertemporal Preferences. *National Bureau of Economic Research Working Paper Series*, No. 25232.

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

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

Ashraf, 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.

Bachireddy, C., Joung, A., John, L. K., Gino, F., Tuckfield, B., Foschini, L., and Milkman, K. L. (2018). Comparing Financial Incentive Structures for Promoting Physical Activity: A Randomized Controlled Trial. *Working Paper*.

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.

Bommer, C., Heesemann, E., Sagalova, V., Manne-Goehler, J., Atun, R., Bärnighausen, T., and Vollmer, S. (2017). The global economic burden of diabetes in adults aged 2079 years: a cost-of-illness study. *The Lancet Diabetes and Endocrinology*, 5(6):423–430.

Brune, L. and Kerwin, J. (2019). Income Timing and Liquidity Constraints: Evidence from a Randomized Field Experiment. *Journal of Development Economics*, 138:294–308.

Carrera, M., Royer, H., Stehr, M. F., and Sydnor, J. R. (2017). The Structure of Health Incentives: Evidence from a Field Experiment. *NBER Working Paper Series*, page 49.

Casaburi, L. and Macchiavello, R. (2019). Demand and Supply of Infrequent Payments as a Commitment Device: Evidence from Kenya. *American Economic Review*, 109(2):523–555.

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

Chung, S., Palaniappan, L., Wong, E., Rubin, H., and Luft, H. (2010). Does the frequency of pay-for-performance payment matter? Experience from a randomized trial. *Health Services Research*, 45(2):553–564.

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.

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.

Gardiner, C. K. and Bryan, A. D. (2017). Monetary Incentive Interventions Can Enhance Psychological Factors Related to Fruit and Vegetable Consumption. *Annals of Behavioral Medicine*, 51:599–609.

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.

Hossain, T. and List, J. A. (2012). The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations. *Management Science*, 58(12).

John, A. (2019). When Commitment Fails - Evidence from a Field Experiment. *Management Science*, Forthcomin.

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

Kim, K. R. and Seo, E. H. (2015). The relationship between procrastination and academic performance: A meta-analysis. *Personality and Individual Differences*, 82:26–33.

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

Kremer, M., Rao, G., and Schilbach, F. (2019). Behavioral Development Economics. *Handbook of Behavioral Economic*, 2(Forthcoming).

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.

Lay, C. H. (1986). At last, my research article on procrastination. *Journal of Research in Personality*, 20:474–495.

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. (1999a). Incentives for Procrastinators. *The Quarterly Journal of Economics*, 114(3):769–816.

O'Donoghue, T. and Rabin, M. (1999b). Incentives for Procrastinators. *The Quarterly Journal of Economics*, 10(2):130–153.

O'Donoghue, T. and Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics*, 90(10-11):1825–1849.

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.

Özer, B. U., Saçkes, M., and Tuckman, B. W. (2014). Psychometric Properties of the Tuckman Procrastination Scale in a Turkish Sample. *Psychological Reports*, 113(3):874–884.

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).

Rabin, M. and O'Donoghue, T. (2003). Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes. *American Economic Review*, 93(2):186–191.

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.

Tuckman, B. W. (1991). The Development and Concurrent Validity of the Procrastination Scale. *Educational and Psychological Measurement*, 51:473–480.

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.

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.

# Appendix for:

## Incentivizing Behavioral Change: The Role of Time Preferences

# Table of Contents

## Appendix



## Supplemental Online Appendix

**Note:** The Online Appendix is a separate document that can be found at:
http://faculty.chicagobooth.edu/rebecca.dizon-ross/research/papers/IncentiveDesignOnlineApp.pdf

# A Impatience Measurement

Appendix Table A.1: Impatience index questions

| Questions: |
| --- |
| 1. I am continually saying "I'll do it tomorrow" |
| 2. I usually accomplish all the things I plan to do in a day |
| 3. I postpone starting on things I dislike to do |
| 4. I'm on time for appointments |
| 5. I often start things at the last minute and find it difficult to complete them on time |

Note: Table shows the questions from the Tuckman (1991) and Lay (1986) scales that we included in the time preferences survey to measure impatience. We standardize these questions and take the average to create our "impatience index."

Appendix Table A.2: Questions used to predicted the impatience index using LASSO

| Questions: |
| --- |
| 1. Do you worry that if you kept a higher balance on your phone, you would spend more on talk time? |
| 2. In the past week, how many times have you found yourself exercising less than you had planned? |
| 3. In the past 24 hours, how many times have you found yourself eating foods you were trying to avoid? |
| 4. Would you like to spend more of your free time on exercise than you currently do? |

Note: Table shows baseline variables that we use to predict the impatience index using LASSO.

Appendix Table A.3: LASSO coefficients: predicted impatience index

| | Coefficient |
|---|---|
| Higher mobile balance = higher talk time | 0.167 |
| In the past week, how many times did you not stick to your exercise plans | 0.099 |
| In the last 24 hours, # times have you eaten foods you were trying to avoid | 0.006 |
| Would you like to spend more of your free-time exercising | -0.048 |

Note: Table shows coefficients from the LASSO model that we use to predict the impatience index using the predictors outlined in Section A.2.

Appendix Table A.4: Our impatience index correlates with real behaviors

| Covariate type: | Exercise | | Baseline Indices | | | |
|---|---|---|---|---|---|---|
| Baseline variable: | Daily steps | Daily exercise (min) | Health index | Negative vices index | Healthy diet index | # Individuals |
| ***A. Impatience Index Measures*** | | | | | | |
| Impatience index | -0.080*** | -0.070*** | -0.017 | -0.052 | -0.185*** | 1760 |
| *Incentives* | -0.094*** | -0.073*** | -0.019 | -0.036 | -0.165*** | 1313 |
| *Monitoring* | 0.013 | -0.038 | 0.104 | -0.179* | -0.060 | 111 |
| *Control* | -0.046 | -0.068 | -0.042 | -0.082 | -0.297*** | 316 |
| Predicted impatience index | -0.009 | -0.026 | -0.052*** | 0.011 | -0.024 | 3232 |
| *Incentives* | -0.008 | -0.029 | -0.066*** | 0.025 | -0.022 | 2404 |
| *Monitoring* | 0.048 | 0.045 | 0.053 | 0.000 | -0.008 | 203 |
| *Control* | -0.043 | -0.023 | -0.049 | -0.047 | -0.039 | 585 |
| ***B. CTB Measures*** | | | | | | |
| Individual alpha estimates | | | | | | |
| *Discount rate* | 0.037 | -0.020 | -0.025 | -0.007 | 0.028 | 1820 |
| *Discount rate percentile* | -0.017 | 0.017 | 0.033 | 0.012 | -0.035 | 1820 |
| *Beta* | 0.046* | 0.000 | -0.012 | -0.017 | 0.021 | 1801 |

Notes: This table reports correlations between various impatience measures and baseline behavioral measures. The health index includes an individual's measures of Hba1c, random blood sugar, blood pressure, body mass index, and waist measurement. The vices index includes an individual's daily cigarette, alcohol, and areca nut usage. The healthy diet index includes an individual's daily number of wheat meals, vegetable meals, rice meals, spoonfuls of sugar, and fruit, junk food, and sweets intake, as well as whether a respondent goes out of his or her way to avoid unhealthy foods. Significance levels: * 10%, ** 5%, *** 1%.

# 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 simplicity, we restrict the sample to those who were in the daily, weekly, and monthly groups, and run the following regression:

$$\begin{aligned} y_{it} = &\alpha + \beta_0 Impatience_i + \beta_1 daily_i + \beta_2 monthly_i \\ &+ \beta_3 Impatience_i \times daily_i + \beta_4 Impatience_i \times monthly_i + \boldsymbol{X}'_i\gamma + \varepsilon_{it}, \end{aligned} \tag{9}$$

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. $\beta_1$ and $\beta_2$ represent the effects of daily and monthly relative to the base case weekly payment (respectively). The coefficients of interest are $\beta_3$ and $\beta_4$, 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.[42] Our results are reported in Table B.1. 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 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:

$$\begin{aligned} y_{it} = &\alpha + \beta_0 \left(Impatience\,Measure\right)_i + \beta_1 \left(Payday\right)_{it} \\ &+ \beta_2 \left(Payday\right)_{it} \times \left(Impatience\,Measure\right)_i + \boldsymbol{X}'_i\gamma + \varepsilon_{it}, \end{aligned} \tag{10}$$

where $y_i$, $(Impatience\,Measure)$, and $\boldsymbol{X}_i$ are defined as in equation 9; and $(Payday)_{it}$ is an indicator for whether day $t$ is a payday for individual $i$. To test whether more impatient

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

Appendix Table B.1: High-frequency treatments are not more effective for those who are more impatient

| Dependent variable: | Met step target | |
|---|---|---|
| Impatience measure: | Impatience index | Predicted impatience index |
| | (1) | (2) |
| Daily × Impatience | 0.0744 | -0.000629 |
| | [0.06] | [0.03] |
| Monthly × Impatience | 0.0453 | 0.0264 |
| | [0.05] | [0.02] |
| Daily | 0.00929 | -0.00651 |
| | [0.04] | [0.03] |
| Monthly | -0.0485 | -0.0278 |
| | [0.03] | [0.02] |
| Impatience | -0.0534*** | -0.0200** |
| | [0.02] | [0.01] |
| Base case mean | 0.50 | 0.50 |
| # Individuals | 1,397 | 2,559 |
| Observations | 112,215 | 205,732 |

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; the base case incentive group is omitted. Standard errors clustered at the individual level in brackets. Controls are the same as Table 2. 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%.

individuals respond more to more immediate payment, we test whether $\beta_2 > 0$.

Our results are shown in Table B.2. 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.

Appendix Table B.2: Payday effects are not bigger for those with higher measured impatience

| Dependent variable: | Met step target | |
|---|---|---|
| Impatience measure: | Impatience index | Predicted impatience index |
| | (1) | (2) |
| Impatience × Payday | 0.00590 | -0.00162 |
| | [0.01] | [0.01] |
| Payday | 0.0285 | 0.0288 |
| | [0.03] | [0.02] |
| Impatience | -0.0488*** | -0.0168* |
| | [0.02] | [0.01] |
| Controls | X | X |
| Base case mean | 0.50 | 0.50 |
| # Base Case | 481 | 890 |
| # Monthly | 93 | 163 |
| # Individuals | 574 | 1053 |
| Observations | 46241 | 85005 |

Notes: This table shows heterogeneity in the "payday" effects for those in the base case incentive and the monthly incentive groups, by impatience. 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 Table C.1: Impacts of incentives and monitoring on diet and addictive consumption.

| **A. Healthy diet** | Healthy Diet Index | Wheat meals | Meals with vegetables | Servings fruit | Negative of rice meals | Negative of junk-food pieces | Negative of spoons sugar in coffee | Negative of sweets yesterday) | Avoid unhealthy food |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Incentives | 0.026 | 0.026 | 0.055* | 0.040 | 0.029 | -0.022 | -0.023 | -0.028 | 0.0045 |
| | [0.029] | [0.030] | [0.031] | [0.038] | [0.033] | [0.064] | [0.047] | [0.038] | [0.018] |
| Monitoring | 0.021 | 0.014 | 0.071 | 0.060 | -0.0085 | 0.13 | -0.028 | -0.046 | -0.038 |
| | [0.051] | [0.053] | [0.054] | [0.066] | [0.059] | [0.11] | [0.083] | [0.067] | [0.031] |
| Control mean | 0.00 | 0.49 | 0.58 | 0.53 | -2.34 | -0.91 | -1.12 | -0.35 | 0.83 |
| P-value: M = I | 0.91 | 0.81 | 0.73 | 0.74 | 0.47 | 0.13 | 0.95 | 0.76 | 0.12 |
| # Individuals | 3,192 | 3,068 | 3,068 | 3,068 | 3,068 | 3,068 | 3,068 | 3,068 | 3,068 |

| **B. Addictive consumption** | Addictive Good Consumption Index | Average Daily Areca | Average Daily Alcohol | Average Daily Cigarettes |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Incentives | -0.0019 | 0.036 | -0.035* | -0.060 |
| | [0.024] | [0.042] | [0.020] | [0.11] |
| Monitoring | -0.00048 | 0.016 | -0.015 | -0.026 |
| | [0.042] | [0.074] | [0.036] | [0.19] |
| Control mean | 0.00 | 0.13 | 0.11 | 1.02 |
| P-value: M = I | 0.97 | 0.76 | 0.53 | 0.84 |
| # Individuals | 3,192 | 3,068 | 3,068 | 3,068 |

Notes: Standard errors in brackets. For the two indices, controls are the same as Table 2, along with second order polynomials of all questions underlying the indices at baseline. We follow Kling et al. (2007) in constructing the indices by imputing missings (for individuals who have non-missing responses to at least one component of the index) for each component using the relevant sample mean per group. Controls for all other outcomes are the same as Table 2. 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.

Appendix Table C.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 |
| | (1) | (2) | (3) | (4) |
| ***A. Pooled Incentives*** | | | | |
| Incentives | 0.21*** | 1337.6*** | 0.057* | 4.72** |
| | [0.022] | [261.1] | [0.034] | [2.34] |
| Pure Control | | | -0.13*** | -7.94*** |
| | | | [0.037] | [2.45] |
| ***B. Unpooled Incentives*** | | | | |
| 10 INR | 0.15*** | 820.5 | -0.0043 | 2.91 |
| | [0.049] | [524.0] | [0.065] | [4.89] |
| Daily | 0.21*** | 1202.7*** | 0.027 | 6.24 |
| | [0.034] | [389.5] | [0.047] | [3.89] |
| Weekly | 0.21*** | 1356.6*** | 0.054 | 4.46* |
| | [0.024] | [277.0] | [0.036] | [2.52] |
| Monthly | 0.20*** | 1568.7*** | -0.029 | 4.48 |
| | [0.035] | [393.8] | [0.048] | [5.09] |
| 4-Day Threshold | 0.20*** | 1321.2*** | 0.091** | 6.19** |
| | [0.025] | [287.7] | [0.036] | [2.57] |
| 5-Day Threshold | 0.22*** | 1380.8*** | 0.053 | 1.37 |
| | [0.030] | [336.8] | [0.041] | [2.89] |
| Pure Control | | | -0.13*** | -7.94*** |
| | | | [0.037] | [2.45] |
| Monitoring mean | 0.29 | 6774.52 | 0.50 | 22.33 |
| Controls | No | No | No | No |
| # Monitoring | 200 | 200 | 195 | 195 |
| # 10 INR | 64 | 64 | 62 | 62 |
| # Daily NTH | 163 | 163 | 161 | 161 |
| # Base Case | 890 | 890 | 867 | 867 |
| # Monthly NTH | 163 | 163 | 160 | 160 |
| # 4-Day TH | 775 | 775 | 757 | 757 |
| # 5-Day TH | 304 | 304 | 293 | 293 |
| # Control | 0 | 0 | 568 | 568 |
| # Individuals | 2,559 | 2,559 | 3,063 | 3,063 |
| Observations | 205,732 | 205,732 | 3,063 | 3,063 |

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. 48 people in the monitoring and incentives group withdrew at the start of the contract period, but there is no statistically significant difference in likeliness to immediately withdraw given randomized group (p-value > 0.7).

Appendix Table C.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 | -0.087*** | -0.061** | -0.10*** | -0.019 | 0.022 | 0.24*** |
| | [0.022] | [0.025] | [0.020] | [0.028] | [0.029] | [0.035] |
| Incentives X (10 INR) | -0.023 | 0.061 | 0.0091 | 0.061 | -0.050 | -0.067 |
| | [0.036] | [0.042] | [0.033] | [0.046] | [0.048] | [0.057] |
| Incentives X (Daily NTH) | 0.014 | -0.022 | -0.0053 | 0.053* | -0.022 | -0.036 |
| | [0.024] | [0.027] | [0.022] | [0.031] | [0.031] | [0.038] |
| Incentives X (Monthly NTH) | -0.024 | 0.040 | 0.042* | 0.022 | -0.080** | 0.0063 |
| | [0.024] | [0.027] | [0.022] | [0.031] | [0.031] | [0.038] |
| Incentives X (4- or 5-Day TH) | 0.011 | 0.017 | 0.0081 | 0.0010 | -0.054*** | 0.0067 |
| | [0.013] | [0.015] | [0.012] | [0.016] | [0.017] | [0.020] |
| Monitoring mean | 0.17 | 0.17 | 0.16 | 0.17 | 0.17 | 0.16 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| ***B. Average Daily Steps*** | | | | | | |
| Incentives | -0.059** | -0.058** | -0.090*** | -0.0090 | 0.13*** | 0.066** |
| | [0.025] | [0.025] | [0.022] | [0.027] | [0.034] | [0.030] |
| Incentives X (10 INR) | 0.029 | 0.019 | 0.054 | -0.028 | -0.051 | -0.032 |
| | [0.041] | [0.041] | [0.036] | [0.045] | [0.056] | [0.049] |
| Incentives X (Daily NTH) | 0.046* | -0.011 | -0.0044 | 0.023 | -0.067* | 0.0061 |
| | [0.027] | [0.027] | [0.024] | [0.030] | [0.037] | [0.032] |
| Incentives X (Monthly NTH) | -0.013 | 0.029 | 0.044* | -0.0075 | -0.083** | 0.042 |
| | [0.027] | [0.027] | [0.023] | [0.030] | [0.037] | [0.032] |
| Incentives X (4- or 5-Day TH) | 0.026* | 0.017 | -0.00052 | -0.040** | -0.040** | 0.027 |
| | [0.014] | [0.014] | [0.012] | [0.016] | [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 | 890 | 890 | 890 | 890 | 890 | 890 |
| # Monthly NTH | 163 | 163 | 163 | 163 | 163 | 163 |
| # 4-Day TH | 775 | 775 | 775 | 775 | 775 | 775 |
| # 5-Day TH | 304 | 304 | 304 | 304 | 304 | 304 |
| # 10 INR | 64 | 64 | 64 | 64 | 64 | 64 |
| Observations | 2,559 | 2,559 | 2,559 | 2,559 | 2,559 | 2,559 |

Notes: This table presents a series of regressions of dummies for six average daily step quantiles, and six fraction-of-days-step-target-exceeded quantiles. The fifth row of each panel shows the additional effect of the threshold incentives contracts compared to the base case contracts. 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. Standard errors are in brackets. Controls are the same as Table 2. The omitted interaction with incentives is the base case incentive treatment group.

Appendix Table C.4: Differential Incentive Effects according to Impatience

| Impatience Measure: | Impatience index | | Predicted impatience index | |
|---|---|---|---|---|
| Dependent variable: | Met step target | Average daily steps | Met step target | Average daily steps |
| | (1) | (2) | (3) | (4) |
| Incentives | 0.218*** | 1380.0*** | 0.208*** | 1403.6*** |
| | [0.03] | [318.15] | [0.02] | [228.34] |
| 10 INR | -0.145*** | -894.4* | -0.0821** | -652.0* |
| | [0.05] | [525.15] | [0.04] | [340.89] |
| Daily | 0.00929 | -194.2 | -0.00651 | -288.8 |
| | [0.04] | [382.23] | [0.03] | [283.45] |
| Monthly | -0.0485 | -202.1 | -0.0278 | -107.0 |
| | [0.03] | [360.76] | [0.02] | [252.57] |
| 4-Day TH | -0.0176 | -152.5 | -0.0179 | -195.2 |
| | [0.02] | [204.07] | [0.02] | [151.19] |
| 5-Day TH | -0.00805 | -92.41 | 0.000811 | -95.88 |
| | [0.03] | [286.12] | [0.02] | [199.47] |
| Impatience | 0.0594 | 317.2 | -0.00397 | -103.0 |
| | [0.04] | [459.60] | [0.02] | [205.27] |
| Incentives × Impatience | -0.113*** | -598.7 | -0.0160 | -92.96 |
| | [0.04] | [508.13] | [0.02] | [227.04] |
| 10 INR × Impatience | 0.0679 | -20.73 | 0.0570* | 619.4** |
| | [0.07] | [747.55] | [0.03] | [293.60] |
| Daily × Impatience | 0.0744 | 628.0 | -0.000629 | -175.2 |
| | [0.06] | [544.28] | [0.03] | [280.45] |
| Monthly × Impatience | 0.0453 | 560.9 | 0.0264 | 336.7 |
| | [0.05] | [590.45] | [0.02] | [263.53] |
| 4-Day TH × Impatience | 0.0579* | 426.5 | 0.0345** | 327.5** |
| | [0.03] | [325.74] | [0.01] | [149.27] |
| 5-Day TH × Impatience | 0.0526 | 422.5 | 0.0204 | -34.34 |
| | [0.05] | [495.23] | [0.02] | [202.76] |
| Controls | X | X | X | X |
| Base Case: mean | 0.50 | 8131.12 | 0.50 | 8131.12 |
| # Monitoring | 108 | 108 | 199 | 199 |
| # 10-INR | 36 | 36 | 64 | 64 |
| # Daily NTH | 82 | 82 | 161 | 161 |
| # Base Case | 480 | 480 | 887 | 887 |
| # Monthly NTH | 92 | 92 | 162 | 162 |
| # 4-day TH | 424 | 424 | 768 | 768 |
| # 5-day TH | 166 | 166 | 304 | 304 |
| # Individuals | 1,397 | 1,397 | 2,559 | 2,559 |
| Observations | 112,215 | 112,215 | 205,732 | 205,732 |

Notes: This table shows the treatment effects of each Incentive treatment, interacted with measures of impatience over recharges. The dummy for the base case incentive sub treatment is omitted; the "Incentives" coefficient along with other incentive sub treatment dummies are interpreted relative to the base case contract. Standard errors clustered at the individual level in brackets. Controls are the same as Table 2. "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 $\beta$ to be one. Larger values of each impatience measure indicates more impatience.

Appendix Table C.5: Heterogeneity in the effects of incentives by baseline walking

| Dependent Variable: | Achieved 10k Steps | | | Daily Steps | | | Wore Fitbit | | |
|---|---|---|---|---|---|---|---|---|---|
| Sample: | Full | Above Median Steps | Below Median Steps | Full | Above Median Steps | Below Median Steps | Full | Above Median Steps | Below Median Steps |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Incentives × Baseline | -0.058 | -0.14* | 0.24 | 0.015 | 0.13 | 0.086 | -0.0020 | 0.19 | -0.011 |
| | [0.05] | [0.08] | [0.39] | [0.06] | [0.10] | [0.19] | [0.13] | [0.23] | [0.14] |
| Incentives | 0.21*** | 0.27*** | 0.19*** | 1166.6** | -70.7 | 1072.1 | 0.037 | -0.16 | 0.051 |
| | [0.02] | [0.05] | [0.02] | [461.52] | [1082.34] | [872.48] | [0.12] | [0.23] | [0.13] |
| Baseline Walking | 0.55*** | 0.47*** | 0.15 | 0.55*** | 0.41*** | 0.43** | 0.34*** | 0.13 | 0.30** |
| | [0.05] | [0.08] | [0.37] | [0.06] | [0.10] | [0.18] | [0.12] | [0.19] | [0.13] |
| # Individuals | 2545 | 1289 | 1256 | 2545 | 1289 | 1256 | 2545 | 1289 | 1256 |
| Monitoring mean | 0.24 | 0.46 | 0.01 | 6876.39 | 9700.79 | 4108.02 | 0.94 | 0.98 | 0.90 |

Notes: This table shows the effects of incentives, interacted with measures of walking at baseline. The monitoring group is the omitted group. The sample is restricted to all individuals with phase-in step data and further divided into those that are above and below the median daily phase-in step count. Standard errors clustered at the individual level in brackets. Controls are the same as Table 2.

Appendix Table C.6: 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 | 0.052 | 0.0000043 | 635.5 | 0.044 |
| | [0.04] | [0.00] | [469.63] | [0.04] |
| 4- or 5-day TH | -0.027 | -0.044 | -345.7* | -497.2 |
| | [0.02] | [0.03] | [190.82] | [321.17] |
| Walking Measure | 0.46*** | 0.000029*** | 5651.3*** | 0.49*** |
| | [0.03] | [0.00] | [356.05] | [0.04] |
| Walking Measure | Step Target Compliance | Average Steps | Step Target Compliance | Average Steps |
| # Individuals | 2,545 | 2,545 | 2,545 | 2,545 |

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 are the same as Table 2, along with the average phase-in period value of the dependent variable and its square.

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

| Dependent variable: | Met step target (×100) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| ***A. Predicted impatience index*** | | | | | | | | | | | |
| Predicted index × Threshold | 3.09** | 3.23** | 3.09** | 3.12** | 3.15** | 3.17** | 3.16** | 3.08** | 3.18** | 3.04** | 3.04** |
| | [1.37] | [1.37] | [1.37] | [1.37] | [1.37] | [1.36] | [1.37] | [1.37] | [1.35] | [1.36] | [1.35] |
| Predicted index | -1.99** | -2.04** | -1.99** | -1.98** | -1.98** | -2.09** | -1.99** | -1.92** | -1.95** | -2.07** | -2.11** |
| | [0.97] | [0.97] | [0.97] | [0.97] | [0.97] | [0.96] | [0.97] | [0.97] | [0.96] | [0.97] | [0.97] |
| Threshold | -1.22 | -12.9 | -0.70 | -1.80 | 1.53 | -5.23*** | 2.49 | 1.45 | -1.19 | -5.70* | -5.51 |
| | [1.40] | [8.10] | [1.80] | [2.52] | [5.45] | [1.83] | [3.00] | [6.11] | [1.38] | [3.45] | [4.03] |
| Threshold × Covariate | | 0.24 | -1.25 | 0.86 | -0.31 | 0.039*** | -1.13 | -0.74 | -0.015 | 0.00069 | 0.00098 |
| | | [0.16] | [2.88] | [3.03] | [0.60] | [0.013] | [0.80] | [1.59] | [0.017] | [0.00044] | [0.00085] |
| Covariate | | 1.45* | 0.14 | -2.03 | -0.17 | 0.0060 | 0.38 | 1.76 | 16.1*** | 0.0037*** | 0.0059*** |
| | | [0.88] | [2.11] | [2.21] | [0.43] | [0.058] | [0.58] | [1.15] | [1.94] | [0.00039] | [0.00042] |
| Threshold × Covariate$^2$ | | | | | | | | | | | -0.000000036 |
| | | | | | | | | | | | [0.000000041] |
| Covariate$^2$ | | | | | | | | | | | -0.00000011*** |
| | | | | | | | | | | | [0.000000012] |
| ***B. Impatience index*** | | | | | | | | | | | |
| Impatience index × Threshold | 5.59* | 5.74* | 5.58* | 5.63* | 5.61* | 5.93** | 5.66* | 5.18* | 6.18** | 6.28** | 5.61* |
| | [3.01] | [3.00] | [3.01] | [3.01] | [3.01] | [3.02] | [3.02] | [3.04] | [2.95] | [3.01] | [3.00] |
| Impatience index | -5.44*** | -5.49*** | -5.44*** | -5.46*** | -5.46*** | -5.73*** | -5.44*** | -4.97** | -5.12*** | -4.81** | -4.37** |
| | [2.01] | [1.99] | [2.01] | [2.01] | [2.01] | [2.01] | [2.01] | [2.04] | [1.97] | [2.03] | [2.00] |
| Threshold | -1.51 | -10.3 | -1.10 | -3.12 | -1.81 | -5.95** | -1.38 | 6.95 | -1.71 | -8.07* | -7.64 |
| | [1.92] | [11.1] | [2.48] | [3.29] | [7.09] | [2.34] | [4.12] | [8.09] | [1.88] | [4.72] | [5.50] |
| Threshold × Covariate | | 0.18 | -0.97 | 2.46 | 0.038 | 0.043*** | -0.075 | -2.31 | 0.0042 | 0.0010 | 0.0015 |
| | | [0.22] | [3.91] | [4.04] | [0.77] | [0.016] | [1.10] | [2.11] | [0.015] | [0.00061] | [0.0012] |
| Covariate | | 1.37 | -1.19 | -1.95 | -0.15 | -0.091 | -0.13 | 1.99 | 18.0*** | 0.0034*** | 0.0057*** |
| | | [1.20] | [2.93] | [3.00] | [0.56] | [0.079] | [0.79] | [1.44] | [2.72] | [0.00054] | [0.00056] |
| Threshold × Covariate$^2$ | | | | | | | | | | | -0.000000066 |
| | | | | | | | | | | | [0.000000062] |
| Covariate$^2$ | | | | | | | | | | | -0.00000010*** |
| | | | | | | | | | | | [0.000000014] |
| Covariate used | – | Age | Female | Prev. diagnosed diabetic | HbA1c | Mean arterial blood pressure | Risk aversion | Scheduling certainty | Above-median baseline steps | Baseline steps | Baseline steps |
| # Observations | 86,215 | 86,215 | 86,215 | 86,215 | 86,215 | 86,215 | 86,215 | 86,215 | 86,215 | 86,215 | 86,215 |
| Base case mean | 50.19 | 50.19 | 50.19 | 50.19 | 50.19 | 50.19 | 50.19 | 50.19 | 50.19 | 50.19 | 50.19 |

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 C.8: Heterogeneity in Post-Endline Persistence by Impatience

| Impatience measure: | Impatience index | | Predicted impatience index | |
|---|---|---|---|---|
| Sample: | Late | | Full | |
| | (1) | (2) | (3) | (4) |
| ***A. Met Step Target (×100)*** | | | | |
| Impatience × Incentives | 1.593 | -0.449 | 0.941 | -0.305 |
| | [2.062] | [2.143] | [1.277] | [1.308] |
| Impatience | -1.719 | -0.391 | -1.523 | -0.630 |
| | [1.744] | [1.715] | [1.134] | [1.103] |
| Incentives | -27.50*** | 8.029*** | -27.24*** | 8.393*** |
| | [2.257] | [1.857] | [2.253] | [1.869] |
| Baseline Steps | 0.00126*** | 0.00316*** | 0.00124*** | 0.00313*** |
| | [0.000220] | [0.000212] | [0.000216] | [0.000209] |
| Intervention Steps | 0.00423*** | | 0.00422*** | |
| | [0.000237] | | [0.000236] | |
| ***B. Average Daily Steps*** | | | | |
| Impatience × Incentives | 267.4 | -39.05 | 140.5 | -46.25 |
| | [347.9] | [351.0] | [215.1] | [216.7] |
| Impatience | -326.8 | -127.4 | -220.5 | -86.58 |
| | [313.9] | [298.7] | [198.7] | [190.1] |
| Incentives | -4758.2*** | 573.6** | -4687.7*** | 652.5** |
| | [341.5] | [285.2] | [341.9] | [288.2] |
| Baseline Steps | 0.192*** | 0.476*** | 0.189*** | 0.473*** |
| | [0.0311] | [0.0301] | [0.0308] | [0.0300] |
| Intervention Steps | 0.634*** | | 0.633*** | |
| | [0.0315] | | [0.0315] | |
| # Individuals | 1,112 | 1,112 | 1,122 | 1,122 |
| Base Mean | 5144.6 | 5144.6 | 5144.6 | 5144.6 |

Notes: This table shows heterogeneity by time preferences in persistence of treatment effects. The sample includes everyone who walked in the post-endline period. Control variables include gender, age, weight and time fixed effects. The Base contract is the omitted group, and individual group level dummies are not reported. Because we have no intervention step data for the control group, regressions that include intervention steps only include only treatment groups and the monitoring group. We add a missing intervention period dummy to prevent the control group from dropping out of the sample. 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%.

Appendix Table C.9: Minute by minute summaries

| | Incentives | Monitoring | I - M | P-value: I=M |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| ***A. Activity (by minute)*** | | | | |
| Average daily activity | 213 | 197 | 16 | 0.002 |
| Average steps per minute | 41 | 38 | 3 | 0.001 |
| ***B. Time of Day*** | | | | |
| Average start time | 7:11 | 7:15 | 4 | 0.489 |
| Average end time | 20:50 | 20:49 | 1 | 0.672 |
| ***C. High step counts per minute (share)*** | | | | |
| Steps > 242 | 0 | 0 | 0 | - |
| Steps > 150 | $1.32\times10^{-6}$ | 0 | $1.32\times10^{-6}$ | - |
| # Individuals: | 2368 | 201 | | |

Notes: This table presents various statistics at the respondentXminute level. High step counts thresholds (242 and 150) were determined based on the average number of steps an individual takes when running at 5 mph and 8 mph, respectively. Only one individual's minute-by-minute data coincides with jogging at a pace greater than 5 miles per hour, and only for a total of 15 minutes over one day in the contract period.