Incentivizing Behavioral Change: The Role of Time Preferences

Shilpa Aggarwal
Indian School of Business

Rebecca Dizon-Ross
University of Chicago

Ariel Zucker*
UC Berkeley

December 23, 2020

Click here for latest version

Abstract

Many of the people whom incentives aim to affect are impatient. We develop a prediction for how to make incentives work particularly well for those who are impatient over effort: implement "time-bundled" contracts that make the payment for future effort increase in current effort. We test this prediction using a randomized evaluation of an incentive program for exercise (walking) among diabetics in India and find empirical support for the prediction. In addition, we find that increasing the frequency of payment – which should be effective if individuals are impatient over payment – has no effect, suggesting limited impatience over payments. On average, incentives increase daily steps by 1300 (13 minutes of brisk walking) and improve health.

*Aggarwal: Indian School of Business, shilpa_aggarwal@isb.edu. Dizon-Ross: University of Chicago Booth School of Business, rdr@chicagobooth.edu. Zucker: University of California, Berkeley, adzucker@berkeley.edu. A previous version of this working paper was released as NBER Working Paper No. 27079. 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. We also appreciate support from the National Science Foundation (Dizon-Ross through Award # 1847087). 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, Ted O'Donoghue, 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 contracting models pay limited attention to the role of agent patience. However, growing evidence that many people are "impatient" (i.e., 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 derive predictions about contract variations that should improve the efficacy of incentives for impatient agents relative to patient ones. We then implement the variations in a randomized controlled trial (RCT) that incentivizes exercise among 3,200 diabetics and pre-diabetics in India and assess the quantitative importance of adjusting incentives for impatience.

When formulating our predictions, we distinguish between discount rates over effort and over financial payments. The literature has long emphasized that while agents use "primitive" discount rates from their utility functions to make intertemporal decisions about effort and consumption, their intertemporal decisions about financial payments should instead be driven by the available borrowing and saving opportunities (Cubitt and Read, 2007). For example, with perfect credit markets, even the most impatient utility-maximizing agents discount future payments at only the market interest rate. While this stark prediction requires that people exploit all arbitrage opportunities, which they may not do in practice (Andreoni et al., 2018b), empirical evidence suggests that individuals do often discount effort differently than financial payments (Augenblick et al., 2015). In light of this, we develop two contract variations, one whose efficacy increases with the discount rate over effort and a second whose efficacy increases with the discount rate over payments.

Our first contract variation is a "time-bundled" contract that makes the payment for future effort increase in current effort; we show theoretically that this variation induces more effort from people with high discount rates over effort. To illustrate the intuition, imagine you need a worker to perform two days of work. Consider first a time-bundled contract that pays a lump sum if and only if she works both days. For the contract to induce two days of work, the total payment must exceed the worker's present discounted cost of effort.[1] For example, if her daily cost of effort is \$10, and she discounts future effort by 50%, the payment only needs to be \$15: \$10 for the first day plus a discounted \$5 for the second. In contrast, if you pay her separately for each day of work, the minimum payment to induce two days of work must be higher, at \$20: \$10 per day of effort. Time-bundled contracts thus exploit the fact that, when individuals have high effort discount rates, it is "cheaper" to buy their future (discounted) effort than their current effort.

One advantageous feature of time-bundled contracts is that we predict that they will induce extra effort from all types of people with high discount rates over effort, notably

[1] We assume a zero short-run interest rate on payments for simplicity.

including “naive” time-inconsistents — a common type that are traditionally difficult to motivate (e.g., Bai et al., 2020).[2] Time-bundled contracts also induce extra effort from “sophisticated” time-inconsistent individuals and those who are time-consistent but impatient.

The fact that time-bundled contracts are effective for a broad range of impatient people differentiates them from the standard approach of offering commitment contracts to motivate time-inconsistent people. Under the commitment approach, people can choose to limit their own future options in order to encourage or lock their future selves into taking a particular action.[3] Take-up thus requires sophistication about the differences between one’s preferences and discount rates in the future relative to the present-day. In contrast, time-bundled contracts directly leverage *present-day* discount rates, which even those who are not sophisticated (i.e., naifs) understand. High present-day discount rates (over effort) make future work attractive. Time-bundled contracts offer better (i.e., higher-paid) opportunities for future work to those who work today, thereby motivating all those with high discount rates – even naifs – to work today to access the better future opportunities. Since time-bundled contracts are broadly effective, our empirical analyses pool all people with high discount rates over effort.

Our second contract variation is to increase the frequency of payment, motivated by the (less novel) prediction that if individuals are impatient over payments, more frequent payment increases efficacy. Scholars have long theorized that because people are impatient, “the more frequent the reward, the better” (Cutler and Everett, 2010). However, there are reasons to question whether frequency increases will matter much in practice. One reason is that impatience over payments may be limited even if impatience over effort is not, since the discount rate over payment should only equal the market interest rate for individuals with access to borrowing and saving. However, if individuals irrationally ignore financial arbitrage opportunities (Andreoni et al., 2018b), or if access to credit and liquidity is limited (Carvalho et al., 2016), the discount rate over payment may approach that over consumption.

After presenting our theoretical predictions, we evaluate time-bundled contracts and payment frequency using an experiment offering incentives for behavior change, a particularly apt setting for introducing these contract variations. Incentives are increasingly being used by policymakers to encourage behavior that may be in an individual’s or a society’s best interest. The motivation is often present bias itself, as present bias can cause underinvestment in

[2] Naive time-inconsistent people are unaware of their own time-inconsistency, while sophisticates are aware.

[3] For example, in the incentive domain, a commitment contract approach might offer workers a future contract that pays less than the benchmark contract for low effort but no more than the benchmark contract for high effort (Kaur et al., 2015). The prediction is that time-inconsistent sophisticates might choose the commitment contract – even though it is dominated – to encourage their future selves to exert more effort. Unlike our time-bundled contracts, classic commitment contracts are not normally more effective for impatients conditional on selection.

behaviors with short-run costs and long-run benefits (such as exercise, diet, and studying). Incentives can mitigate this underinvestment by better aligning behaviors with long-run self-interest. The use of incentives to address present bias makes it particularly important to understand how to tailor incentives for present bias – an issue on which the evidence is thin.

Our incentives are designed to encourage walking among diabetics and prediabetics. Lifestyle diseases like diabetes are exploding problems in both developing and developed countries. The estimated cost of diabetes is 0.9% of global GDP and 4.5% of GDP in India. There is widespread agreement that promoting lifestyle changes, such as better exercise and diet, is essential to address the growing economic and health burdens of diabetes (International Diabetes Federation, 2019). However, a large portion of diabetes patients fail to adopt recommended lifestyle changes, and the existing evidence-based interventions promoting lifestyle change are intensive and prohibitively expensive (Howells et al., 2016). Governments are thus interested in scalable interventions to promote lifestyle change among diabetics, and the government of Tamil Nadu, one of the southern states of India, supported and partially funded this study in an effort to develop such an intervention.

Our program monitors participants' walking using pedometers and, if they achieve a daily step target of 10,000 steps, provides them with small financial incentives in the form of mobile phone credits. We randomly assign participants to an "incentive" group that receives both pedometers and walking incentives, a "monitoring only" group that receives pedometers but no incentives, or a control group that receives neither pedometers nor incentives.

Within the incentive group, we randomly implement our two contract variations: time-bundled contracts and more-frequent payment. First, we randomize whether payment is a linear function of the number of days the participant meets the 10,000 step target ("step-target compliance"), or whether payment is instead a time-bundled function that only rewards step-target compliance if the step target is met a minimum number of days that week. We use two minimum compliance thresholds: four days and five days. The variation in time-bundling allows us to explore its average efficacy and to test our core prediction: that it will have heterogeneous impacts by impatience over effort. Second, we randomize three payment frequencies: monthly, weekly, and daily. We use this variation to assess the impact of payment frequency and to investigate the shape of payment discount rates over time.

We design our experiment to assess the quantitative importance of our theoretical predictions and present three main empirical results. Our first result is that, consistent with our theoretical prediction, making the contract time-bundled 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, time-bundled contracts increase compliance with the step target by 6 percentage points (pp) more for people with above-median impatience than for those with below-median impatience, a large difference relative to the sample-average effect of either contract (20 pp). We also calibrate a model using experimental estimates of the distribution of walking costs and find consistent results: projected compliance in the most effective time-bundled contract increases by 3 pp relative to the linear for each 10 pp decrease in the discount factor.

We also explore the overall efficacy of our time-bundled contracts; our second result is that, while thresholds do not increase compliance, they generate more extreme outcomes. Relative to the base case, thresholds increase the variance of walking, increasing outcomes at the top of the distribution but decreasing outcomes at the bottom. The fact that thresholds cause poor outcomes for some makes it important to determine for whom the contracts work well, highlighting the significance of our finding that they work well for the impatient.

Our third result is that increasing payment frequency has limited efficacy in our setting, apparently because individuals have low discount rates over the contract payments. 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: there is little stated demand for high-frequency payments, and step-target compliance does not increase as the date of payment delivery approaches. We thus find that in contrast with the conventional wisdom, increasing incentive frequency is not always an effective way to adjust incentives for impatience. This result is consistent with Augenblick et al. (2015), who find limited impatience in monetary choices among American college students, but our finding is perhaps surprising in light of the prevalence of liquidity constraints in settings like ours and prior evidence that liquidity constraints can lead to impatience over financial payments (Carvalho et al., 2016).

We supplement our primary analysis with a program evaluation of the incentive scheme. Our sample has high rates of diabetes and hypertension; regular exercise can prevent complications from both. We find that incentives are highly effective at inducing exercise. Providing just 20 INR (0.33 USD) per day of compliance with the daily step target increases compliance by 20 pp off of a base of 30%. Average daily steps increase by 1,266 — roughly 13 minutes of brisk walking. The large increases in walking induced by incentives boost mental health and moderately improve an index of health risk that includes blood sugar and body mass index. Much of the effect of incentives on exercise also persists after the intervention ends. These impacts are important for policy, suggesting incentives may be a cost-effective way to decrease the burden of chronic disease in India and beyond.

1.1 Contributions to the Literature

This paper contributes to three strands of literature: on contract design for impatience, nonlinear incentives, and incentives for health behaviors.

Our primary contribution is to the literature on contract design for impatient agents: we develop and validate time-bundled contracts as a novel strategy for motivating a wide range of people with impatient or time-inconsistent discount rates over effort.

Researchers have previously motivated impatient and time-inconsistent agents primarily with commitment devices (e.g. Ashraf et al., 2006; Royer et al., 2015; Kaur et al., 2015);[4] commitment is a useful tool, but it is not a panacea. Take-up of commitment devices is modest (Laibson, 2015) and often reflects errors in judgement (Carrera et al., 2020b), which undermines their use as an effective policy solution. Moreover, commitment devices are only predicted to be effective for sophisticated time-inconsistents; they are less effective — and can even be harmful — for naifs (Bai et al., 2020), who make up a large share of individuals (Augenblick and Rabin, 2019). One reason the standard commitment approach is only effective for sophisticates is that its efficacy relies on take-up: conditional on take-up, most commitment devices do not induce more effort from impatients than others. In contrast, our time-bundled contracts work better for impatient people *conditional* on selection, enabling them to be effective for all types of impatience, notably including partial and full naivete.[5]

While other papers have also examined which incentive contracts work best for impatient people conditional on selection, they have had different goals than ours. O'Donoghue and Rabin (1999b) and Carrera et al. (2020a) both examine ways to help time-inconsistent procrastinators avoid delay in completing a single task; in contrast, our objective to to maximize average effort over time.[6] Andreoni et al. (2018a) customize incentives using time preference estimates with the goal of having agents exert the same effort on two different days, again an objective that is very distinct from ours. Our theoretical insights about time-bundled contracts are also related to theoretical work by Jain (2012), who shows that firms can increase productivity by offering multi-period quotas to salespeople who are present-biased over both payments and effort.[7]

[4]For example, Kaur et al. (2015) offer workers a dominated incentive contract that pays lower amounts for low output but no higher amount for high output, showing that some workers select the dominated contract and work more as a result.

[5]Of course this approach raises the question of how to effectively assign individuals to contracts. Targeting based on observables is one option.

[6]In particular, O'Donoghue and Rabin (1999b) examine how to adjust "temporal incentive schemes" that reward agents based on when they complete a single task. They find that, to to avoid delay among time-inconsistent procrastinators, the optimal incentive typically involves an increasing punishment for delay over time. Carrera et al. (2020a) examine whether they can help time-inconsistent procrastinators overcome startup costs by offering higher incentives upfront in a separable contract; they find this approach is ineffective empirically.

[7]Jain's (2012) starting point is that people are present-biased over both payment and effort, assuming

Our paper is also one of the first papers to study the implications of domain-specific discounting for contract design, and the first to examine this distinction empirically. Although many papers show that discount rates over payment and effort should in theory be different (Cubitt and Read, 2007), and Augenblick et al. (2015) provide evidence of an empirical distinction, the vast majority of dynamic contracting models use the same discount rate for both payment and effort (e.g., Lazear, 1981; Chassang, 2013).[8] Our work studies whether allowing these discount rates to differ has implications for contract design and shows that it does.

We also contribute to a better understanding of the role of payment frequency in contracts for impatient agents. We point out a natural implication of the distinction between discount rates over money and consumption: increasing payment frequency is only effective if people are impatient over *payments*, which even those with high primitive discount rates may not be. We then evaluate the effect of increasing payment frequency. Most of the previous evidence on frequency is indirect: several papers show that worker performance improves at the end of pay cycles (Oyer, 1998; Kaur et al., 2015), suggesting that payment frequency *could* increase effort. We perform a direct test by randomizing payment frequency, holding the frequency of feedback constant across treatment arms to isolate the payment discounting channel. This test complements Gardiner and Bryan's (2017) work in the psychology literature, which finds that simultaneously increasing payment frequency and feedback frequency improves efficacy. Since we isolate the payment frequency channel and find no effect, Gardiner and Bryan's (2017) findings could reflect the salience effect of receiving frequent feedback.

Our contribution to the literature on nonlinear contracting is empirical: we experimentally compare the efficacy of contracts with linear and nonlinear incentive structures. Other experiments comparing linear and nonlinear contracts focus on the selection effects (Larkin and Leider, 2012; Kaur et al., 2015). In contrast, we examine the effect of thresholds conditional on selection and detect a potential pitfall: thresholds do not work well for everyone and so create dispersion in performance. This finding complements other work examining the advantages and disadvantages of contract nonlinearities, especially a rich theoretical literature starting with Lazear (1981) showing that many optimal dynamic contracts display nonlinearities over time, and an empirical literature showing that in practice nonlinearities

that people discount payment and effort identically. In contrast, we allow for different discount rates over payment and effort and demonstrate that the efficacy of time-bundled contracts for the impatient is driven by high discount rates over effort, not present-biased time preferences per se.

[8]The one exception, Edmans et al. (2012), derives optimal CEO contracts in the presence of savings, which introduces a wedge between the payment and effort discount rates. Our work contrasts with theirs in three ways. First, their optimality result is sensitive to the exact model environment. We depart from optimality to examine how variations to a benchmark linear contract impact performance in a way that is more environmentally robust and has broader empirical relevance. Second, they assume that the payment discount rate is the interest rate, while we allow it to be flexible. Third, we empirically test our results.

often suboptimally distort behavior and promote cheating (e.g. Jacob and Levitt, 2003). Our work adds evidence on both sides of the ledger, documenting a new disadvantage of nonlinear schemes (excess dispersion) as well as a new advantage of some *dynamic* or time-bundled nonlinear schemes (effectiveness when effort discount rates are high).

Finally, we make several contributions to the growing literature on incentives for health, such as exercise (e.g., Royer et al., 2015) and weight loss (e.g., Volpp et al., 2008). Prior work has examined incentives for other health behaviors among diabetics without success (e.g., Long, 2012). We are the first to implement walking incentives among diabetics and prediabetics and the first trial of incentives for exercise in a developing country. While previous work generally finds that incentives increase walking among non-diabetic populations (Bachireddy et al., 2019; Burns and Rothman, 2018; Finkelstein et al., 2016; Patel et al., 2016), our incentives increased walking by more — and at less cost — than previously studied walking incentive interventions. Moreover, while many previous studies of walking incentives do not find health impacts, our program led to moderate gains in cardiovascular wellness and mental well-being.

The paper proceeds as follows. Section 2 presents our theoretical predictions. Sections 3 and 4 discuss the study setting and design. Section 5 presents empirical results on incentive design and impatience. Section 6 shows the overall program impacts. Section 7 concludes.

2 Theoretical Predictions

In this section, we show how the "effectiveness" of two features of incentive contracts — time-bundling and payment frequency — depends on time preferences. Taking the perspective of a policymaker whose objective is to maximize compliance subject to a budget constraint, we define effectiveness as the average compliance for a given payout.[9]

The setup is as follows. Each day, an individual chooses whether to complete a binary action. Define w_t as an indicator for whether the individual "complies" (i.e., completes the action) on day t. In our experiment, w_t is an indicator for walking 10,000 steps on day t. The incentive contracts we consider pay individuals based on compliance over a sequence of T days. We call this sequence of days the payment period and index its days $t = 1, .., T$. Payments are delivered on day T.

We consider two types of incentive contracts:

1. **Separable contracts** have payment functions that are separable across days. That is, payment for w_t depends only on w_t and not on any $w_{t'}$ for $t' \neq t$. We assume these

[9] Our approach is analogous to the standard contract theory approach of maximizing effort subject to incentive and budget constraints. While there is a question of whether this is the socially optimal objective function, we discuss its appropriateness for this and other settings in Section 5.4.

contracts pay m per day of compliance. Total payment is thus:

$$\text{Payment} = m \sum_{t=1}^{T} w_t. \tag{1}$$

2. **Time-bundled contracts** have payment functions that are not separable across days. Their defining feature is that the payment function displays at least one "dynamic complementarity" (i.e., a day on which the payment for future compliance is increasing in current compliance). We focus on a type of "threshold" time-bundled contract, where there is a minimum threshold level of compliance C below which no incentive is received, and above which payment is a linear function of the number of days of compliance, $\sum_{t=1}^{T} w_t$:

$$\text{Payment} = \begin{cases} m \sum_{t=1}^{T} w_t & \text{if } (\sum_{t=1}^{T} w_t \geq C) \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

We first specify the agent's problem and solve for compliance under a "base case" separable contract. We then examine two variations to the base case contract. The first variation makes the contract time-bundled while maintaining the same payment period length. The second maintains the separable payment function from the base case and instead varies the payment period length T. We pay particular attention to how the efficacy of these variations depends on agent discount rates over effort and payments.

2.1 Utility and Discounting

To solve for compliance, we consider the following simple specification of agent utility:

$$U = \sum_{t=0}^{\infty} \delta(t) \left(c_t - e_t \times w_t \right), \tag{3}$$

where e_t is the utility cost of complying on day t, c_t is consumption, and $\delta(t)$ represents the structural discount factor over effort and consumption: individuals discount effort costs and consumption t days in advance by $\delta(t)$, with $\delta(t) \leq 1$. We assume utility is linear in c_t, which is likely a good approximation in our setting, as payment amounts are small relative to overall consumption.[10] We also assume at first for simplicity that e_t is weakly positive and is known in advance. However, we show in the appendix that our main prediction is robust to stochastic costs. The individual's problem is to choose c_t and w_t to maximize utility subject to a budget constraint.

Discounting of payments vs. effort We differentiate the discount rate over payments k days in advance, which we denote as $d_m(k)$, from the discount rate over effort and con-

[10]The model's qualitative predictions are also robust to relaxing this assumption.

sumption, $\delta(k)$. While $\delta(k)$ comes directly from the utility function, the discount rate over payments depends on the availability of borrowing and savings in the budget constraint. In perfect credit markets with borrowing and saving at interest rate r, individuals should discount future payments at the interest rate by financial arbitrage arguments.[11] At the opposite extreme, with no savings or borrowing, day k payments are immediately consumed and future payments are discounted by the consumption discount factor $\delta(k)$. We accommodate both these (and other)[12] cases by defining $d_m(k)$ as the reduced-form payment discount factor that encompasses both the "primitive" discount factor and any financial frictions. With perfect credit markets, $d_m(k) = \left(\frac{1}{1+r}\right)^k$, whereas with no savings or borrowing, $d_m(k) = \delta(k)$.

Sophistication Individuals will have time-inconsistent preferences if either $\delta(k)$ or $d_m(k)$ are non-exponential functions of k or if $\delta(k) \neq d_m(k)$. We follow O'Donoghue and Rabin (1999a) and define a sophisticate as one who is fully aware of her own discount factors (over both effort and money) and a naif as one who "believe(s) her future selves' preferences will be identical to her current self's."

2.2 Compliance under a Separable Contract (the Base Case)

We now solve for compliance under the base case separable contract defined in Equation 1. On day t, the individual complies if the present discounted value of being paid m on day T outweighs the effort cost of compliance:

$$w_t|^{\text{Separable}} = \mathbb{1}\{e_t < d_m(T-t)m\}. \tag{4}$$

This compliance decision holds for both naifs and sophisticates since it never involves forecasting future behavior. Total compliance within a payment period is thus:

$$\sum_{t=1}^{T} w_t|^{\text{Separable}} = \sum_{t=1}^{T} \mathbb{1}\{e_t < d_m(T-t)m\}. \tag{5}$$

2.3 Variation 1: Time-Bundled Contracts

We now examine the effect, relative to the base case, of making the contract time-bundled while maintaining the same payment period length. Time-bundling makes an individual's decision more complicated, as the marginal payment for compliance on a given day — and hence the decision to comply — depends on compliance on other days in the payment period. For simplicity, we examine a threshold time-bundled contract with a two-day payment period

[11] Define m_t as income on day t. With perfect credit markets, the lifetime budget constraint is $\sum_{t=0}^{\infty}\left(\frac{1}{1+r}\right)^t c_t = \sum_{t=0}^{\infty}\left(\frac{1}{1+r}\right)^t m_t$. The value of a payment of size m delivered in k days is thus $\left(\frac{1}{1+r}\right)^k m$.

[12] For example, our approach nests domain-specific time preferences: $U = \sum_{t=0}^{\infty} d_m(t)c_t - \delta(t)e_t^{\mathbf{1}(w_t=1)}$.

and a two-day minimum threshold of compliance (i.e., $T = 2$ and $C = 2$ in Equation 2), which are smaller than those used in our experiment. The individual thus receives $2m$ if she complies with the step target on both days; otherwise she receives nothing.

The effect of adding a time-bundled threshold on average compliance with the step target is theoretically ambiguous, depending on several factors, such as the distribution of e_t.

However, we can derive an unambiguous prediction regarding *heterogeneity* in the performance of the time-bundled threshold by discount rates over effort. The prediction is also empirically relevant: under many cost distributions, the discount rate over effort is pivotal to the relative performance of the threshold and linear contracts, with the time-bundled threshold having higher average compliance for some values of $\delta(1)$ and lower for others.

Prediction 1. *Holding all else equal, average compliance in the time-bundled threshold contract relative to the separable contract is weakly decreasing in the discount factor over effort,* $\delta(k)$.

Proof. See Appendix B.1 for full proof. We provide a sketch here.

The result stems from the fact that while compliance in the separable contract is independent of $\delta(k)$, compliance in the time-bundled threshold contract is decreasing in $\delta(k)$.

To see that compliance in the separable contract is independent of $\delta(k)$ (conditional on $d_m(k)$), consider our two-period example. The individual will comply on day 1 if $e_1 < d_m(1)m$ and on day 2 if $e_2 < m$, neither of which depend on $\delta(1)$.

In contrast, with the time-bundled threshold, compliance depends on $\delta(1)$. On day 1, the individual decides if it is worth it to comply on *both* days in order to be paid, comparing the present *discounted* cost of effort on both days, $e_1 + \delta(1)e_2$, with the value of the payment, $d_m(1)2m$. She wants to comply on both days if the costs are low enough, that is, if:

$$e_1 + \delta(1)e_2 < d_m(1)2m. \tag{6}$$

Prediction 1 is driven by the fact that equation 6 is more likely to hold if $\delta(1)$ is small. Holding all else constant, the lower is $\delta(1)$, the less costly are two days of effort from the day 1 perspective since day 2 effort is discounted. Importantly, this statement holds for both sophisticates and naifs, underlying the broad efficacy of time-bundled thresholds for impatients. For both types, the lower $\delta(1)$, the higher the compliance in a threshold contract, because those with a lower $\delta(1)$ have a lower discounted cost of reaching the threshold. □

To illuminate how the effect of time-bundled thresholds depends on discount rates, we compare the minimum payment needed to achieve *full* compliance under two contracts: first, a separable contract with variable payments across days (m_1 for day 1 and m_2 for day 2);

and second, a threshold contract paying $m_1 + m_2$ for compliance on both days:

$$\underset{m_1,m_2}{\operatorname{argmin}}\, m_1 + m_2 \text{ s.t. } (w_1 = 1)\&(w_2 = 1) = \begin{cases} \frac{e_1}{d_m(1)} + e_2 & \text{if separable} \\ \frac{e_1}{d_m(1)} + \frac{\delta(1)}{d_m(1)} e_2{}^{13} & \text{if threshold.} \end{cases}$$

If $\frac{\delta(1)}{d_m(1)} < 1$, as Augenblick et al. (2015) suggest is common, then the threshold contract achieves compliance at a lower cost than the separable contract because it allows the principal to buy day 2 effort at the discounted rate of $\frac{\delta(1)}{d_m(1)} e_2$ (the agent's "day 1-valued" day 2 effort cost) rather than having to pay the larger "day 2-valued" day 2 effort cost, e_2.

Time-bundled thresholds thus leverage the fact that it is cheaper to buy future effort than current effort from effort-impatient agents.

2.3.1 Option, Commitment, and Other Time-Bundled Contracts

Appendix B.2 explores whether Prediction 1 holds for other 2-day time-bundled contracts besides thresholds. We show that Prediction 1 holds for contracts in which, for some values of $\delta(1)$, agents believe on day 1 that day 1 compliance is pivotal for day 2 compliance. The day 1 compliance decision then becomes a decision of whether to comply on *both* days, which is what allows the principal to effectively buy day 2 effort at the "discounted" $\frac{\delta(1)}{d_m(1)} e_2$ rate.

Interestingly, the exact contract features that make time-bundled contracts work better for impatient agents than patient ones are different for naifs and sophisticates. Time-bundled contracts are effective for naifs when day 1 compliance generates lucrative "options" for day 2 compliance, i.e., when day 1 compliance is pivotal to the day 1 self *wanting* her day 2 self to comply. Concretely, and assuming for simplicity throughout this subsection that $d_m = 1$, this means that the payment for day 2 compliance must be greater they $\delta(1)e_2$ if and only if the agent complies on day 1. In contrast, contracts that are effective for sophisticates generate "commitment," wherein day 1 compliance is pivotal to whether the day 2 self *actually* complies. In this case, the payment for day 2 compliance must be greater than e_2 (not $\delta(1)e_2$) if and only if the agent complies on day 1.

Some contracts, like thresholds, are "effective" (i.e., have compliance decreasing in $\delta(1)$) for both naifs and sophisticates because they feature both option and commitment. That is, day 1 compliance is pivotal to both whether the day 1 self *wants* her day 2 self to comply and whether the day 2 self *actually* complies. This requires that the payment for day 2 compliance must be greater than e_2 if the agent complies, and less than $\delta(1)e_2$ if the agent does not comply, as it is with the threshold contract (for certain cost realizations, at least). Other contracts are only effective for one type. Contracts that generate commitment

[13] This assumes for simplicity that $e_2 < e_1 + \delta e_2$; if not, the payment would be $\max\{\frac{e_1}{d_m(1)} + \frac{\delta(1)}{d_m(1)} e_2, 2m\}$.

without option (i.e., day 1 compliance is pivotal to whether the day 2 self *actually* complies but not whether the day 1 self *wants* her day 2 self to comply) only generate extra effort from impatient sophisticates, while contracts that generate option without commitment only generate extra effort from impatient naifs.[14]

Reassuringly, no two-day time-bundled contracts produce "procrastination" among naifs, wherein naifs delay compliance from day 1 to day 2 and then fail to follow through on day 2. Procrastination occurs when day 1 effort is a substitute with day 2 effort (e.g., when people are paid for the completion of one task only between both days). This can cause naifs, who are overoptimistic about future effort, to decrease day 1 effort. In contrast, in two-day *time-bundled* contracts, day 1 effort and day 2 effort are complements by definition.

2.3.2 Robustness of Prediction 1

Section B.3 shows that, under reasonable assumptions about the distribution of e_t, the prediction also holds when the agent is uncertain about future effort cost realizations but knows the distribution of effort costs. Section B.4 relaxes the assumption that time-bundled threshold contracts require 100% compliance to earn payment (i.e., we allow for $C < T$). In a simple model where costs each day can either be high or low and the agent has to comply at least two days out of a three-day payment period to receive payment, we show that compliance in the threshold contract relative to the non-threshold contract is weakly higher for those with $\delta < 1$ than those with $\delta = 1$. The intuition is the same as above: those who discount future walking costs still have a lower discounted total cost to achieve the threshold. Again, the prediction holds for both naifs and sophisticates, time-consistents and time-inconsistents.[15]

2.4 Variation 2: Payment Frequency

We now return to the base case separable linear contract from Equation 1 and analyze compliance under different payment frequencies by changing the length of the payment period T. To simplify notation, we assume effort costs e_t are independently and identically distributed (i.i.d.) across t, with cumulative distribution function $F(\cdot)$.

Recall that individuals in separable contracts comply on day t as long as the discounted value of the payment outweighs effort costs (i.e., if $e_t < d_m(T-t)m$). The probability of

[14] Option without commitment means that the payment for day 2 compliance is between $\delta(1)e_2$ and e_2 if the agent complies, and $\leq \delta(1)e_2$ otherwise. Commitment without option means that the payment for day 2 compliance is $> e_2$ if the agent complies, and between $\delta(1)e_2$ and e_2 otherwise.

[15] Note that Section B.4 examines a threshold contract where C/T is relatively high. Thresholds where C/T is very low may not always be better for impatient naifs than patient people because they include more days where current and future effort are substitutes and hence where naifs may procrastinate.

compliance on day t is thus:

$$Pr\left(w_t = 1|\text{Separable}\right) = F\left(d_m(T-t)m\right), \tag{7}$$

and the average fraction of days complied is:

$$E\left[\frac{1}{T}\sum_{t=1}^{T}(w_t)|\text{Separable}\right] = \frac{1}{T}\sum_{t=1}^{T}F\left(d_m(T-t)m\right). \tag{8}$$

Using these equations, we can make two intuitive predictions.

Prediction 2. *If agents are impatient over the receipt of financial payments (i.e., if $d_m(k) < 1$ and $d'_m(k) \leq 0$), then average compliance is increasing in the payment frequency. If agents are sufficiently patient ($d_m(k) \approx 1$), then payment frequency does not affect compliance.*[16]

Proof. This proof follows from Equation (8). The average fraction of days complied is increasing in the discount factor over payment $d_m(T-t)$. If agents are "impatient" and $d'_m(T-t) \leq 0$, then the the discount factor is (weakly) decreasing in the delay to payment $T-t$. Increasing payment frequency weakly decreases the delay to payment $T-t$ on each day t, which weakly increases average compliance for impatients. If agents are patient, then the discount factor is nearly 1 irrespective of the delay to payment $T-t$ and increasing payment frequency has no detectable effect on average compliance. □

The quantitative importance of Prediction 2 depends not only on average discount factors over payment but also on the shape of $d_m(k)$ for the specific range over which T varies between contracts. For example, say $d_m(k)$ has a quasi-hyperbolic or "beta-delta" shape, with a large one-time decrease between $d_m(0)$ and $d_m(1)$, but is relatively flat for further increases in k. Then increasing frequency would only meaningfully increase compliance if payments were made daily. In contrast, if $d_m(k)$ decays more gradually with k, then more intermediate increases in frequency — say from monthly to weekly — could also be quantitatively important.

Prediction 3. *If the discount factor over payments $d_m(k)$ is decreasing in k, then average compliance increases as the payday approaches.*

Proof. This proof follows from Equation (7): as the payment date approaches, the time to payment $T-t$ decreases, and so the probability of compliance increases. □

[16] Although linear utility is necessary for the stark prediction for patient agents, it is not necessary for the prediction that impact of higher-frequency payments is increasing in the discount rate over payments.

2.5 Empirical Tests

We design our experiment in light of the predictions above. We assess the quantitative importance of Prediction 1, that compliance with the time-bundled threshold relative to the linear contract is decreasing in the discount factor over effort, in two ways. We first randomly vary whether the contract has a time-bundled threshold or is a separable linear contract and test for heterogeneity in threshold effectiveness based on a baseline measure of impatience over effort (we address potential confounds to this impatience measure in Section 5.2.2). Second, we calibrate a model using our experiment data and examine how predicted compliance in the threshold relative to linear contract varies with effort discount rates.

To shed light on our predictions regarding impatience over payments, our experiment randomizes three payment frequencies: monthly, weekly, and daily. We compare the average compliance across these treatments to understand whether varying payment frequency has a quantitatively important impact, and thereby (per Prediction 2) also understand if agents are meaningfully impatient over payments. In addition, we compare the change in compliance when moving from monthly to weekly with the change when moving from weekly to daily to understand both whether intermediate increases in payment frequency (from monthly to weekly) can be effective and whether discount rates over payment decay quickly. Finally, Prediction 3 allows us to use within-treatment variation to shed further light on the shape of the discount factor, as in Kaur et al. (2015).

3 Experimental Design

3.1 Sample Selection and Pre-Intervention Period

We conducted our experiment in an urban area of South India. India is facing a diabetes epidemic, and prevalence is higher both in southern than in northern states and in urban than in rural areas. We selected our sample through a series of public screening camps in the city of Coimbatore, Tamil Nadu. To recruit diverse socioeconomic groups, we held the camps in locations ranging from the government hospital to markets, religious institutions, and parks. During the camps, trained surveyors took health measurements; discussed each individual's risk for diabetes and hypertension; and conducted an eligibility survey. To be eligible for the study, individuals needed to have a diabetes diagnosis or elevated blood sugar, have low risk of injury from regular walking, be capable with a mobile phone, and be able to receive payments in the form of "mobile recharges".[17] After screening, we contacted eligible

[17] The full list of eligibility criteria was: must be diabetic or have elevated random blood sugar (> 130 if has not or > 150 if has eaten in previous two hours); be 30–65 years old, physically capable of walking 30 minutes, literate in Tamil, and not pregnant or on insulin; have a prepaid mobile number used solely by

individuals by phone and invited them to participate in a program encouraging walking.

Surveyors visited the participants at their homes or workplaces to conduct a baseline health survey, deliver lifestyle modification advice, and enroll them in a one-week phase-in period designed to collect baseline walking data and to familiarize participants with program procedures. Surveyors demonstrated how to properly wear a pedometer, report steps, and check text messages from our reporting system (described in Section 3.3). Surveyors asked respondents to wear the pedometer and report their steps each day of the phase-in period.[18]

At the end of the phase-in period, surveyors visited respondents to sync the data from the pedometers, conduct a baseline time-preference survey, and then (after all baseline data were collected) tell participants what treatment group they had been randomly assigned to for the intervention period. To do so, they walked participants through a contract describing their assigned treatment group. We exclude from the sample all participants who withdrew or were found ineligible prior to randomization, leaving a final experimental sample of 3,192 individuals. The sample represents 41% of the screened, eligible population (see Table A.1 for the share of people dropped in each stage of the enrollment process).

3.2 Experimental Design and Contract Launch

Our interventions encouraged participants to walk at least 10,000 steps a day. We chose this daily step target to match exercise recommendations for diabetics; it is also a widely quoted target among health advocates and a common benchmark in health studies.

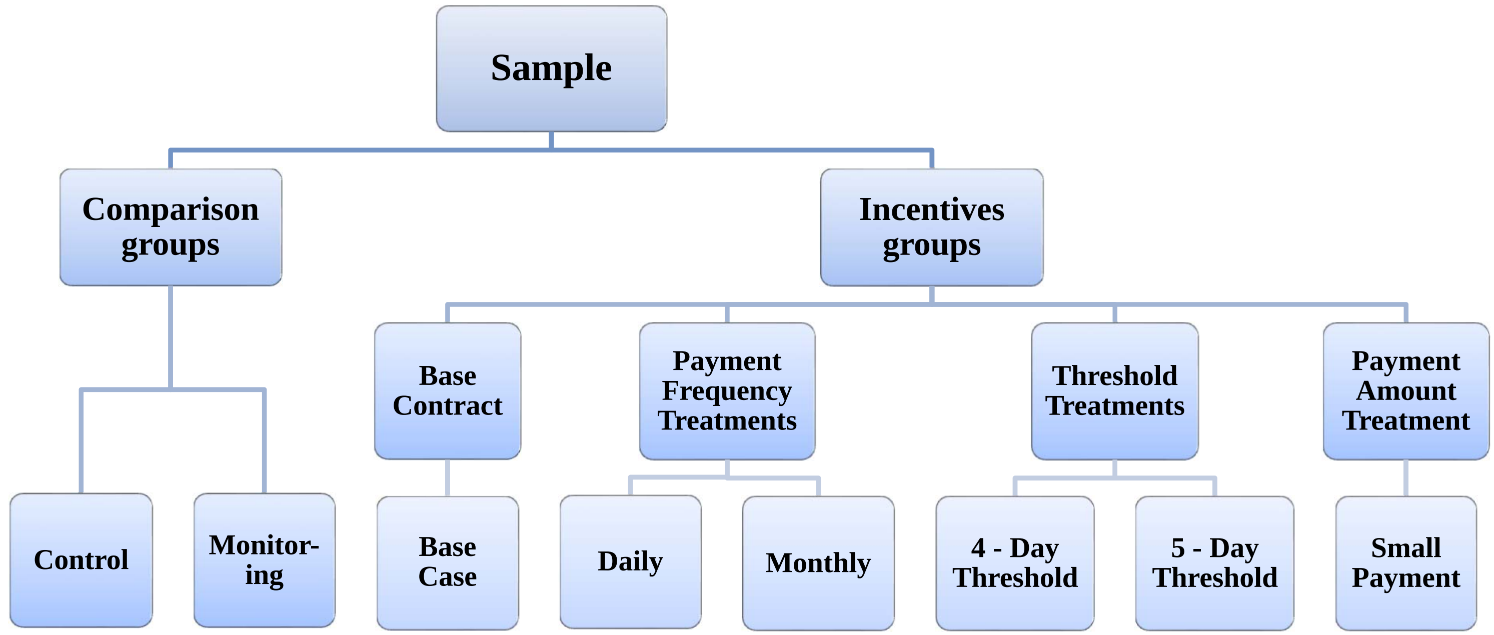
We randomized participants into the incentive group or one of two comparison groups.

1. **Incentive:** Receive a pedometer and incentives to reach a daily target of 10,000 steps.
2. **Monitoring:** Receive a pedometer but receive no incentive contract.
3. **Control:** Receive neither a pedometer nor an incentive contract.

Within the incentive group, we randomized participants into one of six incentive contracts for walking, as shown in Figure 1 and described next. Participants also answered survey questions testing their understanding of the contracts they received. Statistics presented in Table A.2 indicate that a vast majority of participants did indeed understand them.

them, without unlimited calling; reside in Coimbatore; not have blindness, kidney disease, type 1 diabetes, or foot ulcers; not have had major medical events such as stroke or heart attack.

[18] Respondents received 50 INR for consistently wearing the pedometer and reporting steps in this period.



	Control	Monitoring	Base Case	Daily	Monthly	4 - Day Threshold	5 - Day Threshold	Small Payment
Pedometers	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Incentives	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Incentive Details								
Frequency	N/A	N/A	Weekly	Daily	Monthly	Weekly	Weekly	Weekly
Threshold	N/A	N/A	None	None	None	4 Days	5 Days	None
Amount (INR)	N/A	N/A	20	20	20	20	20	10
Sample Sizes	585	203	902	166	164	794	312	66

Figure 1: Experimental Design

3.2.1 Incentive Groups

All incentive groups received payments for accurately reporting steps above the daily 10,000 step target through the automated step-reporting system. We delivered all incentive payments as mobile recharges (credits to the participant's mobile phone account).[19]

After reporting steps, participants immediately received text-message confirmations of their step report, payment earned, and the payment date. We also sent participants weekly text messages summarizing their walking behavior and total payments earned.

When surveyors explained the incentive contract to participants, they explained the step target 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."

Within the incentive group, we randomly assigned participants to one of six groups. Each group received a different incentive contract, with three dimensions of variation: whether the contract was separable or time-bundled, the payment frequency, and the payment amount.

The Base Case This group received a separable, linear contract paying 20 INR per day of compliance with the 10,000 step target. Payments were made at a weekly frequency.

We call this the *base case* contract because all other contracts differ from it in exactly one dimension: separability, payment frequency, or payment amount. We can compare any other group to the base case group to assess the effect of changing a single contract dimension.

Our next treatment groups differ from the base case group in one of the two dimensions that we predict will interact with time preferences.

Payment Frequency Two groups, the *daily* and *monthly* groups, differ from the base case only in the payment frequency. In the daily group, recharges were delivered at 1:00 am 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.

Receiving payments more frequently could increase the salience of step target compliance and trust in the payment system. To hold salience and trust in the payment system constant, all incentive groups both received daily feedback on step target compliance and received a test payment of 10 INR the night before their incentive contract launched, respectively.

Time-Bundled Threshold Contracts The *threshold* treatment groups differ from the base case incentive group only in separability. The base case is a separable linear contract,

[19] The relevant payment discount rate is therefore over mobile recharges, which could be higher, lower, or the same as that over cash (e.g., it could be the same for people whose baseline daily mobile usage is higher than the payment amount: payment would decrease money spent on recharges and increase cash on hand).

paying out 20 INR for each day of compliance. In contrast, the threshold contracts use time-bundled threshold payment functions. The *4-day threshold group* received 20 INR in payment for each day of compliance only if they met the target at least four days in the week-long payment period. So, a 4-day threshold participant who met the step target on only three days in a payment period would receive no payment, while one who met it on five days would receive $5 \times 20 = 100$ INR. Similarly, the *5-day threshold group* received 20 INR in payment for each day of compliance if they met the target at least five days in the week.

The threshold contracts implicitly gave participants a goal of how many days to walk per week. To control for goal effects, surveyors verbally encouraged all incentive groups to walk at least four or five days per week when initially explaining the contracts. For those in the threshold groups, the target days-per-week was the same as their assigned threshold level; for those in the other groups, it was randomly assigned in the same proportion as the threshold groups are divided between the 4- and 5-day groups.

Following our *ex ante* analysis plan, to maximize statistical power, we pool the *4-day threshold* and *5-day threshold* treatment groups for our main analyses.[20] We sometimes also show the results for the two thresholds separately as exploratory analyses.

Payment Amount Finally, we included a treatment group, the *small payment group*, that differs from the base case group only by the amount of incentive paid. This group received 10 INR, instead of the base case 20 INR, for each day of compliance. We included this treatment to learn about the distribution of walking costs and to benchmark the size of our other treatments effects.

Incentive Group Sample Sizes We determined the relative sizes of the incentive groups through power calculations. Since the base case group serves as the reference group for all other contracts, we made it the largest group. Recall that our theoretical predictions for thresholds regard heterogeneity, whereas for frequency they regard main effects. As a result, we allocated larger sample sizes to the threshold treatments than to the frequency treatments to be able to detect heterogeneous effects of the threshold treatments. For the frequency treatments, our analysis is powered instead for main effects.

3.2.2 Comparison Groups

The incentive program could affect behavior because it provides incentive payments or simply because it monitors behavior. We include two control groups in our experiment, a

[20] The reason we included two different groups when we intended *ex ante* to pool them was to maximize the chance that we would be able to test our prediction about the heterogeneity in threshold efficacy by discount rate. Testing this prediction requires that we have at least one threshold contract that does not "fail" (i.e., have extremely low compliance), for example because the threshold was too high to be attainable or so low that it was inframarginal.

monitoring group and a pure control, to allow us to shed light on these two channels.

Monitoring Monitoring group participants were treated identically to the incentive groups except that they did not receive incentives. They received pedometers and were encouraged to wear the pedometers and report their steps every day. They also received the same daily step report confirmation texts and weekly text message summaries that the incentive groups received. Finally, during the upfront explanation of the contract, surveyors also delivered to the monitoring group the same verbal step target of 10,000 daily steps and the same encouragement to walk at least four or five days per week.

Pure Control The pure control group received neither pedometers nor incentives during the intervention period (they returned their pedometers at the end of the phase-in period). Because most incentive programs bundle the "monitoring" effect of a pedometer with the effect of incentives, the pure control group is a useful benchmark from a policy perspective.[21]

3.3 The Intervention Period and After

To measure steps, we gave monitoring and incentive group participants Fitbit Zip pedometers for the duration of the intervention. Although these pedometers could be synced to a central database with an internet connection, most participants did not have regular internet access, so these data were not available in real time. Instead, we asked participants to report their daily step count to an automated calling system, which called participants every evening and prompted them to enter their daily steps from the pedometer. Incentive payments were based on these reports. To verify the reports, we visited participants every two to three weeks to manually sync their pedometers, cross-check the pedometer data against the reported data, and discuss any discrepancies. Anyone found to be chronically overreporting was suspended from the program. All empirical analysis is based on the synced data from the Fitbits, not the reported data.[22]

We visited all participants three times during the 12-week intervention period. The primary purpose was to sync pedometers, but we also conducted short surveys to collect biometric and mobile phone usage data (we conducted these visits even with pure control group participants who did not have a pedometer in order to hold survey visits constant across participants). At the end of the 12-week intervention period, we conducted an endline

[21] To accommodate a request from our government partners, we also tested one additional intervention. Ten percent of the sample, cross-randomized across all other treatments, received the "SMS treatment," which consisted of weekly text message reminders to engage in healthy behaviors such as eating right and exercising. We control for the SMS treatment in our main regressions and test its effects in the Online Appendix (available at faculty.chicagobooth.edu/-/media/faculty/rebecca-dizon-ross/research/incentivedesignapp.pdf.)

[22] Appendix C contains detailed statistics on misreporting. Misreporting rates are similar across monitoring and incentive groups, suggesting misreports were primarily accidental.

survey. Figure A.1 shows the intervention timeline.

Finally, to assess the persistence of our treatment effects on exercise, we gave pedometers to the final 1,171 participants enrolled in our experiment (including control group participants) for 12 weeks after the intervention period had ended. Participants no longer reported steps daily, but surveyors still returned every four weeks to sync their pedometers.

4 Data and Summary Statistics

4.1 Baseline Data: Health, Walking, and Time Preference

We use three baseline datasets: 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, health, fitness, and lifestyle. Health measures include Hba1c, a measure of blood sugar control over the previous three months; random blood sugar (RBS), a measure of more immediate blood sugar control; body mass index (BMI) and waist circumference, two measures of obesity; blood pressure, a measure of hypertension; and a short mental health assessment. The baseline also includes two fitness measures (time to complete five stands from a seated position, and time to walk four meters), diet, and substance use. 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 conduct a baseline time-preference survey to measure impatience over effort in order to test Prediction 1. As highlighted in Kremer et al. (2019), "time preferences [over effort and consumption] 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 has difficulty with the more complicated screen-based measures used in the literature, we included simple measures that the full sample could comprehend.

Impatience over effort Our primary measure of impatience over effort and consumption is an index of survey-based measures of impatience and procrastination taken from the psychology literature. The questions, listed in Panel A of Table A.3, are a subset of the Tuckman (1991) and Lay (1986) scales, with the specific subset chosen *ex ante* by our field team as being most appropriate for our setting. The questions ask respondents to respond on a Likert scale of agreement with statements such as "I'm continually saying 'I'll do it tomorrow'." We construct the index by standardizing all question responses and taking the average, following our initial analysis plan when we included the questions in the survey.

The questions in the index are tilted toward procrastination-style behaviors and hence may better detect naive time-inconsistent impatience than other types of impatience. Our empirical heterogeneity tests using this measure may thus tilt toward testing whether the contracts are effective for naifs in particular. Since naifs and partial naifs appear to constitute a large share of impatient individuals (Augenblick and Rabin, 2019; Bai et al., 2020), and since we consider the efficacy for naifs to be a nice advantage of our time-bundled contracts, this limitation is likely minor.

These questions have two key benefits. First, they are simple for respondents to understand. Second, the psychology literature has validated that they predict real behaviors, such as poor academic performance (Kim and Seo, 2015). Reassuringly, the measures also correlate well with behavior in our sample. For example, those with higher values of our impatience index have worse diets and lower levels of baseline walking (Table A.3).

We began collecting our impatience index partway through the data collection,[23] so it is only available for the latter 54% of the sample. To check the robustness of our results in the full sample, we also create a "predicted index" using a LASSO prediction based on three similar survey questions on self-control in the lifestyle domain that were included in the baseline for all participants (e.g., "In the past week, how many times have you found yourself exercising less than you had originally planned?"). Panel B of Table A.3 lists the questions used for prediction and shows that the predicted index correlates in the expected direction with behavior measures such as the health risk index.

To measure discounting in a consistent way across multiple domains, we also adapted the convex time budget (CTB) methodology of Andreoni and Sprenger (2012) to measure time preferences over walking and mobile recharges, as described in the Online Appendix F. However, these measures are difficult to implement in the field, and we had several logistical challenges. For example, it was hard to get respondents to understand the paradigm, and likely as a result, we have an order of magnitude more law of demand violations than lab-based studies with college students.[24] Further, as described in the Online Appendix F, the impatience measures estimated using this methodology do not correlate in the expected direction with any behaviors. Thus, we judged our implementation unsuccessful and do not use these measures for analysis.

Impatience over payments Although the CTB measures were unreliable, we collected other baseline data that may proxy for impatience over mobile-recharge payments: recharge balances, recharge usage, and a measure of the marginal propensity to consume (MPC) recharges (we asked how much additional credit participants would use if they were gifted 30

[23]Challenges surfaced during our field implementation of Andreoni and Sprenger (2012) (described below).

[24]Other suggestions of a lack of understanding include our estimates not converging for roughly 44% of the sample and respondents failing to follow through on their chosen allocations.

extra INR of recharges daily over the intervention period). People who have higher balances, usage, and/or a lower MPC are less likely to be credit constrained and may have a lower discount rate over recharges.

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, 49.4 years old and has slightly more males than females. The 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 10,000 to 20,000 INR per month (Ministry of Labour and Unemployment, 2016). 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, 81% have Hba1c levels that strongly indicate diabetes, and the RBS measures show poor blood sugar control. The sample also has high rates of comorbidities: 49% have hypertension and 61% are overweight. Panel C shows that on average, participants walked just under 7,000 steps per day in the phase-in period, comparable to average daily steps in many developed countries (Bassett et al., 2010). Panels D and E show our measures of impatience over effort and our measures of mobile recharge usage (See Table A.4 for a summary of the components of the impatience index).

Baseline measures are balanced across treatment groups. Columns 2–4 of Table 1 show means for the pure control, monitoring, and incentive groups, while columns 5–9 show means separately for each incentive subgroup. To explore balance, we jointly test the equality of all characteristics in each of our three "comparison" groups (control, monitoring, and the base case incentive groups — the reference group for all incentive subgroups) with each of the treatment groups. All tests fail to reject the null that all differences are zero.

4.3 Outcomes

Our outcomes come from two datasets. The first contains time-series data of daily steps walked by each participant with a pedometer during the intervention period and (for a subset of the sample) for the 12-week period after that. We do not have daily steps for the control group during the intervention period because they did not have pedometers.

A potential issue with the daily step data is that we only observe steps taken while participants wear the pedometer. Because participants in the incentive groups are rewarded for taking 10,000 steps in a day with the pedometer, 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 incentive groups selectively make an effort to wear the pedometer when they think they will walk more but the monitoring group does not, then we will see a

Table 1: Baseline Summary Statistics in Full Sample and by Treatment Group

	Full sample	Control	Monitoring	Incentives pooled	Daily	Base case	Monthly	Threshold	Small payment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Demographics									
Age (from BL)	49.54	49.78	50.28	49.44	49.57	49.60	48.80	49.41	49.11
	(8.52)	(8.19)	(8.95)	(8.55)	(8.60)	(8.33)	(8.94)	(8.71)	(7.84)
Female (=1)	0.42	0.46	0.43	0.41	0.44	0.41	0.38	0.41	0.48
	(0.49)	(0.50)	(0.50)	(0.49)	(0.50)	(0.49)	(0.49)	(0.49)	(0.50)
Labor force participation (=1)	0.75	0.73	0.72	0.75	0.75	0.74	0.81	0.75	0.70
	(0.44)	(0.45)	(0.45)	(0.43)	(0.43)	(0.44)	(0.39)	(0.43)	(0.46)
Per capita income (INR/month)	4463	4488	4620	4447	4068	4477	4599	4461	4341
	(3638)	(4483)	(3160)	(3447)	(2765)	(3496)	(3235)	(3570)	(2615)
Household size	3.91	3.94	3.82	3.91	3.92	3.89	3.74	3.96	3.58
	(1.62)	(1.54)	(1.51)	(1.64)	(1.45)	(1.70)	(1.59)	(1.65)	(1.29)
B. Health									
Diagnosed diabetic (=1)	0.67	0.67	0.68	0.66	0.62	0.68	0.62	0.67	0.59
	(0.47)	(0.47)	(0.47)	(0.47)	(0.49)	(0.47)	(0.49)	(0.47)	(0.50)
Hba1c (mmol/mol)	8.68	8.67	8.76	8.68	8.58	8.72	8.66	8.68	8.35
	(2.33)	(2.36)	(2.40)	(2.32)	(2.36)	(2.29)	(2.44)	(2.34)	(2.14)
Random blood sugar (mmol/L)	192.42	191.32	196.07	192.51	195.58	193.26	193.30	192.23	177.38
	(89.39)	(88.73)	(86.67)	(89.87)	(91.54)	(88.25)	(98.14)	(90.42)	(77.00)
Systolic BP (mmHg)	133.35	133.33	134.06	133.34	135.25	133.27	134.18	132.84	135.62
	(19.15)	(20.34)	(17.68)	(18.99)	(21.55)	(19.07)	(19.13)	(18.35)	(21.42)
Diastolic BP (mmHg)	88.47	88.54	88.53	88.46	89.30	88.19	88.60	88.45	90.00
	(11.11)	(11.50)	(10.10)	(11.09)	(12.79)	(10.75)	(10.10)	(11.09)	(13.19)
HbA1c: Diabetic (=1)	0.82	0.82	0.81	0.82	0.77	0.84	0.79	0.81	0.77
	(0.38)	(0.38)	(0.39)	(0.38)	(0.42)	(0.36)	(0.41)	(0.39)	(0.42)
BP: Hypertensive (=1)	0.49	0.46	0.51	0.49	0.53	0.49	0.51	0.49	0.45
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Overweight (=1)	0.61	0.62	0.66	0.60	0.57	0.60	0.58	0.60	0.67
	(0.49)	(0.48)	(0.47)	(0.49)	(0.50)	(0.49)	(0.50)	(0.49)	(0.48)
BL BMI	26.42	26.52	26.47	26.40	26.41	26.47	26.39	26.30	26.99
	(4.35)	(4.34)	(3.67)	(4.39)	(5.35)	(4.53)	(4.81)	(4.07)	(4.10)
C. Walking - Phase-in									
Exceeded step target (=1)	0.25	0.25	0.24	0.25	0.25	0.23	0.27	0.26	0.27
	(0.32)	(0.31)	(0.32)	(0.32)	(0.32)	(0.30)	(0.33)	(0.33)	(0.34)
Average daily steps	6999	7066	6892	6998	7046	6810	7449	7078	7018
	(3980)	(3946)	(3697)	(4014)	(4195)	(3969)	(3857)	(4035)	(4195)
D. Impatience over effort									
Impatience index (SD's)	0.09	0.00	0.05	0.12	0.04	0.14	0.18	0.09	0.26
	(0.99)	(1.00)	(0.89)	(0.99)	(0.95)	(1.05)	(0.91)	(0.97)	(0.91)
Predicted index (SD's)	-0.05	0.00	-0.15	-0.06	-0.09	-0.02	-0.02	-0.08	-0.12
	(1.00)	(1.00)	(0.94)	(1.01)	(1.02)	(1.00)	(1.09)	(1.00)	(0.97)
E. Mobile Recharges									
Current Mobile Balance (INR)	29.26	30.80	29.48	28.98	28.61	29.69	28.55	28.45	30.05
	(49.42)	(48.79)	(48.68)	(49.88)	(38.54)	(52.08)	(63.65)	(47.96)	(36.59)
Yesterday's talk time (INR)	6.61	7.22	6.47	6.44	5.86	6.58	7.67	6.31	4.94
	(8.79)	(10.14)	(8.95)	(8.36)	(6.25)	(8.77)	(9.19)	(8.28)	(5.77)
Marginal talk time, if gifted (INR)	41.32	28.19	24.14	45.62	73.03	54.96	35.44	33.93	70.27
	(191.61)	(156.15)	(144.56)	(201.65)	(256.44)	(222.00)	(178.92)	(172.04)	(252.13)
F-tests for Joint Orthogonality									
P-value (relative to control)	N/A	N/A	0.75	0.11	0.31	0.24	0.13	0.35	0.30
P-value (relative to monitoring)	N/A	0.75	N/A	0.95	0.92	0.87	0.46	0.98	0.66
P-value (relative to base case)	N/A	0.24	0.87	N/A	0.60	N/A	0.78	0.81	0.36
Sample size									
Number of individuals	3,192	585	203	2,404	166	902	164	1,106	66
Percent of sample	100.0	18.3	6.4	75.3	5.2	28.3	5.1	34.6	2.1
Number of ind. with ped. data	2,582	–	200	2,359	163	890	163	1,079	64

Notes: Standard deviations are in parentheses. BMI is body mass index and BP is blood pressure. Overweight means BMI above 25. Hypertensive means systolic BP above 140 or diastolic BP above 90. The F-statistic tests the joint orthogonality of all characteristics to treatment assignment relative to the comparison group. The Threshold column pools both the 4-day and 5-day threshold groups.

spurious positive relationship between incentives and observed daily steps.

To minimize selective pedometer-wearing, we incentivize all monitoring and incentive 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 nonzero recorded steps) on at least 70% of days in the intervention period. The rates of pedometer-wearing are high and the difference between-treatment groups is small in magnitude (85% in monitoring versus 88% in incentives); however, the difference is statistically significant with a p-value of 0.043 (column 2 of Table A.5). To address the imbalance, we report Lee (2009) bounds accounting for missing step data due to not wearing pedometers when comparing the incentive and monitoring groups.[25] Our primary specifications do not condition on wearing the pedometer (instead setting steps and compliance to 0 on days when the pedometer was not worn), but we show that our results are robust to conditioning on wearing.

Since the pedometers record data on minute-wise (instead of day-wise) step counts for a subset of days, we can also test whether, on the days participants wore the pedometers, the incentive groups wore it for more minutes. Reassuringly, the start and end times are balanced across groups, as shown in Table C.3.

Another potential concern would be if participants gave their pedometers to someone else; we believe this concern is limited for two main reasons. First, we performed 836 unannounced audit visits with participants at their homes to verify that they were wearing their pedometers or could demonstrate where they were. In 99.6% of cases, participants were not sharing their pedometers. Second, we check whether participants' minute-wise step counts exceed what would be expected from participants of their age range and find that this is extremely rare and is balanced across incentive and monitoring groups (Table C.3).

The second outcomes dataset — the endline survey — gathered health, fitness, and lifestyle information similar to the baseline health survey. The completion rate is 97% in each one of the treatment groups (control, monitoring, and incentive; p-value for equality 0.99).

5 Empirical Results: Incentive Design

This section empirically examines the implications of impatience for incentive design. We first show that our incentive program increases compliance with the step target, making this

[25] We do not have participant pedometer data (e.g., because the pedometer broke or the sync was unsuccessful) on 6% of days. Missing pedometer data is balanced across incentive and monitoring groups (column 2, Table A.5). While our main specifications drop days with missing pedometer data, Table A.6 shows robustness to alternate specifications and Lee bounds. While missing data is balanced overall, one specific source of missing data (mid-intervention withdrawals) is imbalanced (column 5 of Table A.5), but results are robust to Lee bounds accounting specifically for that source (column 5 of Table A.6).

a good laboratory to explore our contract variations. Second, we explore the effect of adding a time-bundled threshold and test our prediction that it will be more effective for those who are more impatient over effort. Third, we analyze the effect of varying payment frequency and use the analysis to shed light on discount rates over payment. Finally, we discuss the potential welfare implications of improving contract effectiveness.

5.1 Incentives and Compliance

We first test whether providing financial incentives increases compliance with the 10,000-step target. To answer this question, we compare average compliance in the pooled incentive groups with the monitoring group, thus isolating the impact of the financial incentives alone (i.e., holding monitoring and other aspects of the full intervention constant).

We estimate regressions of the following form:

$$y_{it} = \alpha + \beta \times incentives_i + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\lambda + \varepsilon_{it}, \tag{9}$$

where y_{it} is either individual i's steps on day t during the intervention period or an indicator for whether individual i surpassed the 10,000-step target on day t; $incentives_i$ is an indicator for being in the incentive group; and $\boldsymbol{X}_i$ and $\boldsymbol{X}_{it}$ are vectors of individual- and day-level controls, respectively, described in the notes to Table 2. We cluster the standard errors ε_{it} at the individual level. The coefficient of interest, β, is the average treatment effect of incentives relative to monitoring only. Panel A of Table 2 shows the results. Figure 2 also shows the results graphically, with the 95% confidence interval depicted on the incentives bar representing a test for equality between the incentive and monitoring groups (as is the case for all the graphs in this section).

Incentives have large impacts on walking, increasing the share of days that participants reach their 10,000 step target by 20 pp (column 1 of Table 2). This effect does not simply reflect participants shifting steps from one day to another: column 2 shows that incentives increase walking by 1,266 steps per day, roughly a 20 percent increase that is equivalent to approximately 13 minutes of extra brisk walking, on average, each day. We demonstrate the robustness of this result to different specifications, including Lee bounds, in Section 6.1.1.

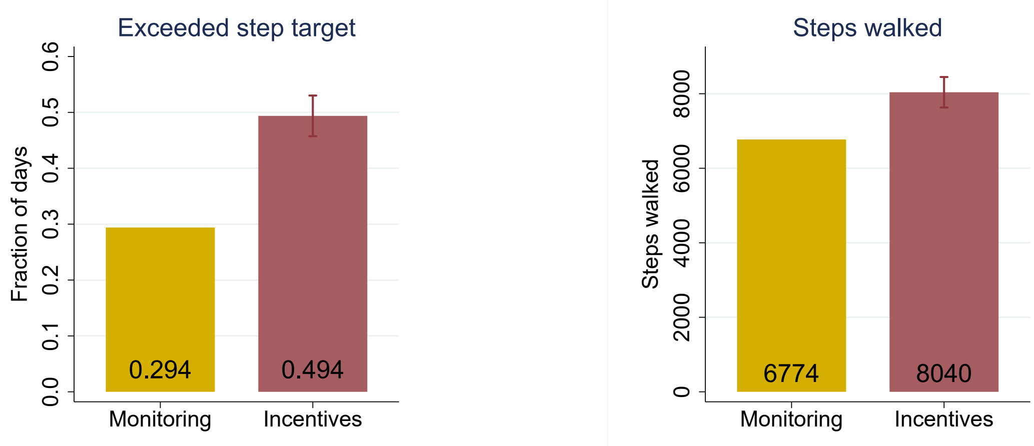
Figure 3 shows that incentives have a striking impact on the distribution of daily steps. Although there is bunching at 10,000 steps in both groups, the bunching in the incentive group is substantially more pronounced. This indicates that the incentives are motivating individuals to comply with their daily step targets.

Because we deliver incentives for walking at least 10,000 daily steps, the incentives are particularly high-powered on days when individuals would otherwise walk just under 10,000

Table 2: Impacts of Incentives on Walking

Dependent variable:	Exceeded step target	Daily steps	Daily steps (if > 0)
	(1)	(2)	(3)
A. Pooled incentives			
Incentives	0.200***	1266.0***	1161.5***
	[0.0186]	[208.7]	[188.5]
B. Unpooled incentives			
Base case	0.211***	1388.4***	1203.1***
	[0.0201]	[222.1]	[199.9]
Daily	0.201***	1122.5***	1283.1***
	[0.0303]	[331.5]	[277.9]
Monthly	0.177***	1274.2***	1179.4***
	[0.0288]	[307.4]	[271.1]
Threshold	0.198***	1216.3***	1142.6***
	[0.0199]	[220.9]	[198.5]
Small payment	0.137***	731.5*	552.9*
	[0.0383]	[386.2]	[335.0]
Monitoring mean	0.294	6,774	7,986
Controls	Yes	Yes	Yes
P-value for base case vs			
Daily	0.71	0.35	0.73
Monthly	0.18	0.65	0.91
Threshold	0.36	0.21	0.61
Small payment	0.04	0.06	0.03
# Individuals	2,559	2,559	2,557
# Observations	205,732	205,732	180,018

Notes: We report incentive effects pooled in Panel A and separately by treatment group in Panel B. The columns show coefficient estimates from regressions based on Equations 9 (Panel A) and 10 (Panel B) using daily intervention-period pedometer data. In column 1, "Exceeded step target" is an indicator variable equal to 1 if the individual exceeded their step target. Individual-level controls are a second order polynomial of age and weight, gender, height, and the average of the dependent variable during the phase-in period (before randomization). Day-level controls are month-year and day-of-week fixed effects. The sample includes the incentive and monitoring groups. The omitted category in all columns is the monitoring group. The sample size differs from Table 1 because a few participants in both the incentive and monitoring groups withdrew immediately. The likelihood of immediate withdrawal is not significantly different between-treatment groups (p-value > 0.7), see Table A.5 column 5. The Threshold group pools the 4- and 5-day Threshold groups. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.



(a) Probability Exceeded Step Target

(b) Average Daily Steps

Figure 2: Incentives Increase Average Walking

Notes: The figure displays the impact of the pooled incentive treatments on walking during the intervention period. The confidence interval represents the test of equality between the incentive and monitoring groups with the same controls as Table 2. Panel A shows the average probability of exceeding the daily step target; Panel B shows average daily steps walked.

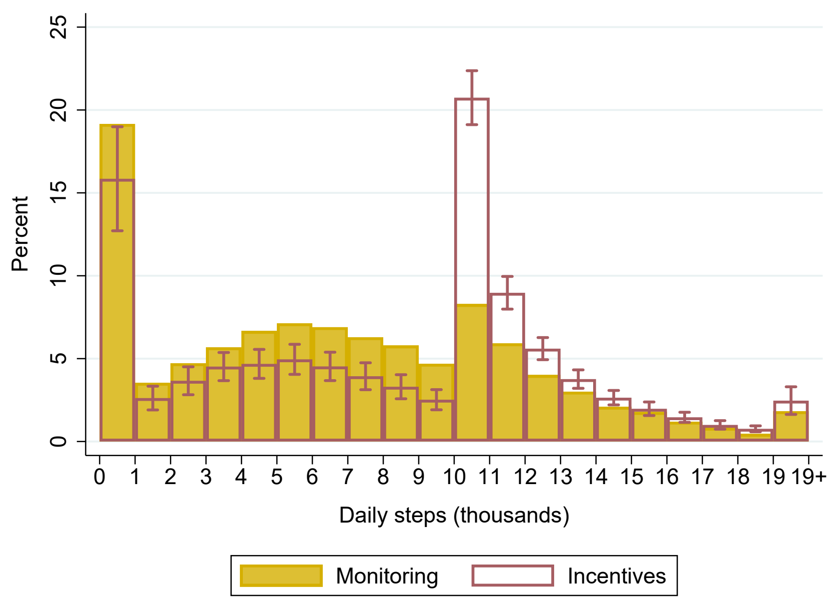


Figure 3: Incentives Shift the Distribution of Steps Walked per Day

Notes: The figure displays the impact of the pooled incentive groups relative to the monitoring group during the intervention period. The confidence intervals represent tests of equality between the incentive and monitoring groups with the same controls as Table 2.

steps and low-powered on days when they would otherwise walk far fewer steps or would reach the target no matter what. Therefore, our incentive program could in theory only improve compliance on days when individuals would have otherwise walked nearly 10,000 steps. Figure 3 provides evidence that this is not the case. Incentives shift the entire distribution of daily steps rather than simply pushing marginal participants over the step target: there is less mass everywhere below the step target and more mass everywhere above.

Having established that our incentives increase compliance, we next use the experiment to explore the effectiveness of incentive contract variations designed to improve performance in the face of impatience over effort and over financial payments, respectively.

5.2 Time-Bundled Threshold Contracts

In this section, we first analyze the effects of time-bundled thresholds in the full sample and then explore the heterogeneity in their effects by impatience over effort. Our primary prediction is about heterogeneity: relative to linear contracts, time-bundled thresholds should increase compliance among those who are impatient over effort relative to those who are not.

The impact of threshold contracts on average compliance, in contrast, is theoretically ambiguous. That said, many simple models yield predictions about the impact of threshold contracts on cost-effectiveness and dispersion. While we do not derive these predictions formally, we lay out the intuition here. First, threshold contracts often decrease the cost of achieving a given level of compliance. The reason is that threshold contracts pay the same as linear contracts when people meet their threshold and less (nothing) when people do not. Second, threshold contracts often increase dispersion. To see this, imagine people who would walk two days less than the threshold in a linear contract in a given week. For some of these people, adding the threshold would cause them to increase their compliance to reach the threshold; for others, they would give up on the incentives and decrease walking. Dispersion at the person-week level (the unit for the contract) would thus go up.

Evidence on both the cost-effectiveness and dispersion impacts of thresholds is scant. We begin by exploring these effects, as well as the sample-average effect on compliance, before testing our prediction about the heterogeneity in the impact of thresholds by impatience.

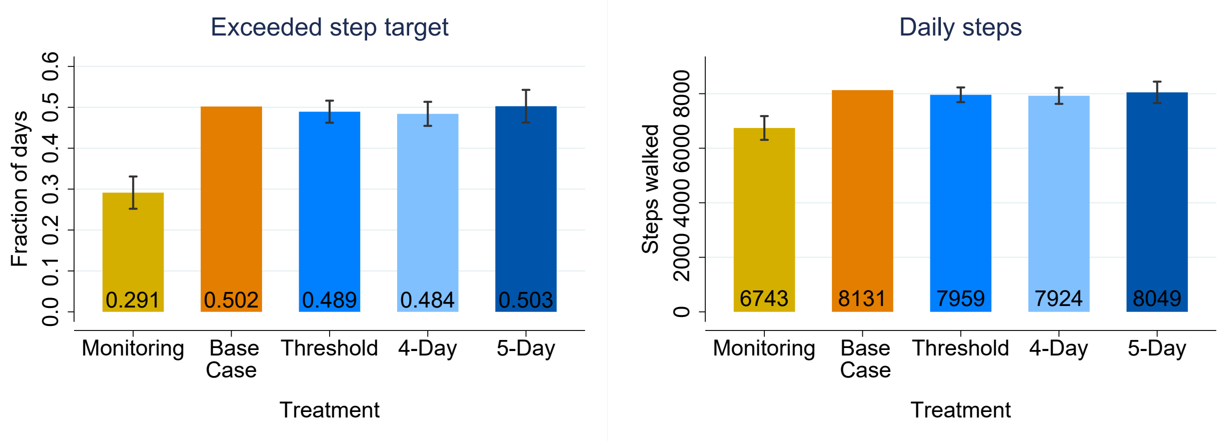
5.2.1 Average Effectiveness, Cost-Effectiveness, and Dispersion

Panel B of Table 2 evaluates the average effects of all of our incentive contract variations relative to the monitoring group, estimating regressions of the following form:

$$y_{it} = \alpha + \beta_j \times \left(incentives^j\right)_i + \boldsymbol{X}'_i\gamma + \boldsymbol{X}'_{it}\theta + \varepsilon_{it}, \tag{10}$$

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, threshold, small payment). Recall that all other treatments vary from the base case contract on exactly one dimension (time-bundling, payment frequency, or payment amount); the bottom rows of the table thus show the p-values for the significance of the difference between each incentive treatment group and the base case group.

We find that adding a time-bundled threshold does not affect the average level of exercise. Figure 4 and Table 2, Panel B, row 4 show that the threshold treatment groups meet the daily step target roughly as frequently as the base case (linear) group does. Step-target compliance is within 1.3 pp of compliance in the base case group, with the differences not statistically significant. Figure 4 also shows the two threshold groups separately; neither has meaningfully different compliance than the base case.



(a) Probability Exceeded Step Target

(b) Average Daily Steps

Figure 4: Adding a Time-Bundled Threshold Does Not Significantly Affect Average Walking

Notes: The figure compares the time-bundled threshold treatments with the base case (linear) incentive treatment. Panel A shows the average probability of exceeding the daily step target during the intervention period; Panel B shows average daily steps walked during the intervention period. The confidence intervals represent tests of equality between the base case incentive group and each other treatment group, with the same controls as Table 2. The Threshold group pools the 4- and 5-day threshold groups.

However, consistent with the intuition laid out above, the threshold contracts generate exercise more cost-effectively. Individuals in the threshold groups only receive payment for exceeding the step target if they do so on at least four or five days in a given week; when they comply on fewer days, they are not rewarded. We find that the 4-day and 5-day threshold groups are paid on only 90% and 85% of the days they achieve the step target, respectively. They are thus paid an average of 18 INR and 17 INR per day of compliance, less than the 20 INR paid (by definition) to the base case group. Importantly, these cost savings of 10%

and 15% are achieved while generating the same amount of walking among participants. For comparison, the incentive paid per day walked is also lower in the small payment group (10 INR per day walked), but this comes at the cost of reduced steps overall (Table 2). Because threshold contracts do not reduce overall compliance but pay out for only a subset of compliance, they are more cost-effective than the base case contract.

We now examine whether thresholds increase the dispersion of walking at the week level and find that they do. Figure 5 shows histograms of the number of days the step target was met per week in the threshold and base case groups. The threshold contracts have a large bimodal effect, causing significantly more individuals to achieve their step target zero days in the week or seven days in the week. The increase in dispersion and in zeroes is consistent with the intuition laid out earlier. The increase in density at seven days in particular (instead of at the specific threshold level of four or five) is perhaps more surprising. Potential explanations include that it is hard for participants to keep track of how many days they have walked or it is easier to schedule walking every day in a given week than on a subset of days.

Thresholds do not just increase dispersion across weeks but also across individuals. Figure 6 plots the density of each individual's probability of exceeding her step target, and mean daily steps, over the intervention. The threshold treatments have thicker tails, with more people walking at the high and low ends. A Brown Forsythe test for equal variance finds that the pooled threshold treatments significantly increase the variance of average steps across the population (p-value < 0.001). Thus, although thresholds do not work well for everyone, they work very well for some people.

The bimodal effects of thresholds highlight the importance of understanding for whom they work best. We next test our theoretical prediction about one type of individual for whom they will work well: those who are impatient over effort.

5.2.2 Heterogeneity in Time-Bundled Threshold Effects by Time Preferences

We perform two exercises to assess the quantitative advantage of time-bundled contracts for those with higher impatience over effort. First, we quantify the heterogeneity by baseline impatience in compliance with the threshold contracts relative to the base case (linear) contract. Since Prediction 1 is a prediction about heterogeneity in the threshold effect *holding all else constant*, this heterogeneity analysis will only be a direct test of the theory if impatience is not correlated with other variables that influence the effectiveness of the threshold. To shed light on whether this assumption holds here, we control for many covariates interacted with the threshold and show that the estimated relationship is robust. Moreover, even if there were omitted variables affecting the estimates, the heterogeneity we estimate is still relevant for policy: policymakers want to customize contract thresholds based on the

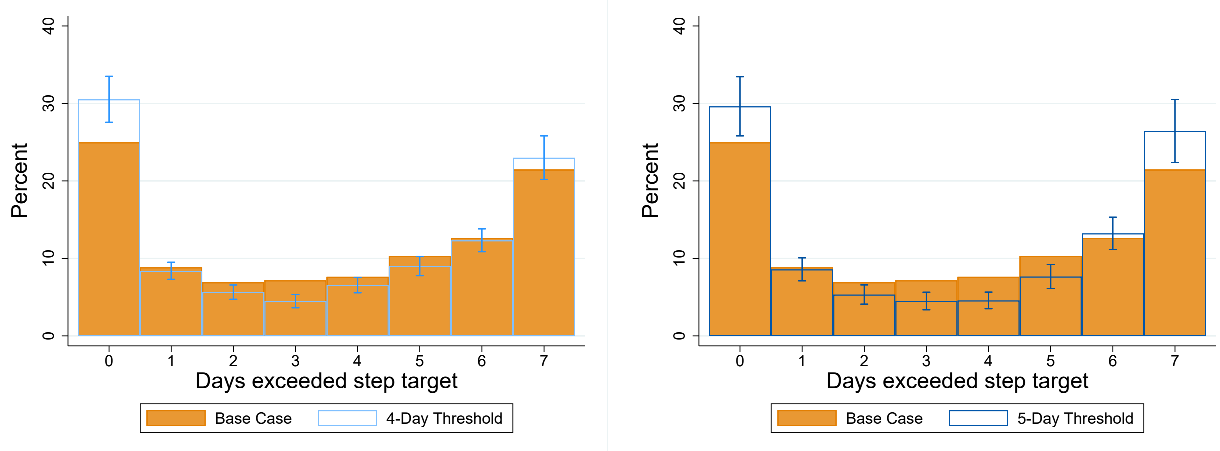


Figure 5: Threshold Contracts Increase Dispersion Across Weeks

Notes: This figure shows the distribution of the number of days walked each week during the intervention period. Data are at the respondent-week level. Confidence intervals represent a test of equality between the base case and 4- or 5-day treatment from a regression with the same controls as Table 2.

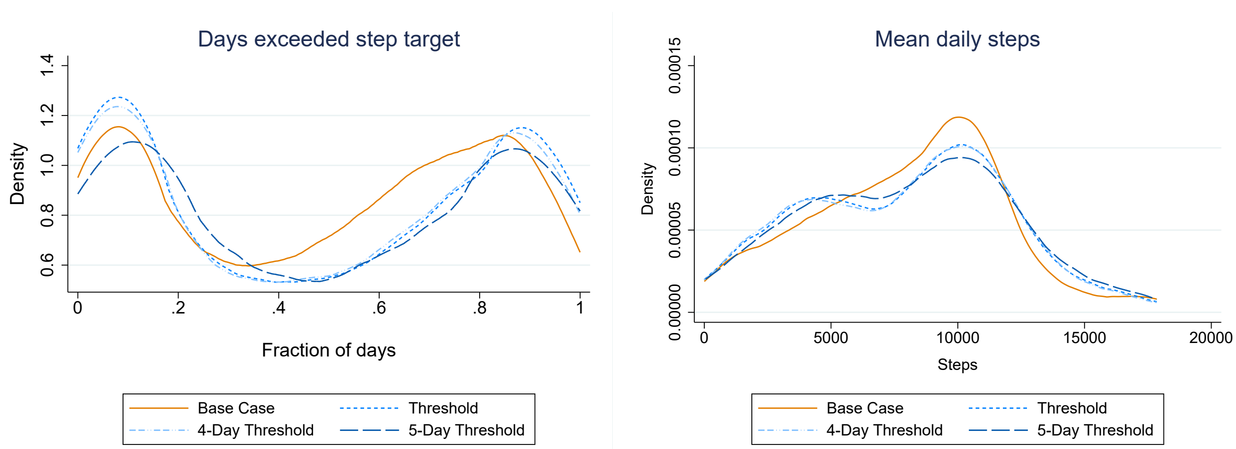


Figure 6: Threshold Contracts Increase Dispersion Across Individuals

Notes: This figure shows the distribution of the fraction of days walked and average steps for participants in the threshold contract groups over the intervention period compared with the base case (linear) contract. The Threshold group pools the 4-day and 5-day threshold groups.

heterogeneity in their efficacy with respect to observed participant impatience, irrespective of whether it is impatience itself that generates the heterogeneity.

To more precisely tie our data to our theory, we also calibrate a model to determine whether the gap in predicted compliance between the threshold and linear contracts varies with the discount rate over effort. All analyses yield consistent results.

Heterogeneity by Baseline Impatience We use a regression of the following form to test for heterogeneity in the effects of time-bundled thresholds by impatience:

$$y_{it} = \alpha + \beta_1 impatience_i \times thresh_i + \beta_2 thresh_i + \beta_3 impatience_i + \boldsymbol{X}_i'\pi + \boldsymbol{X}_{it}'\theta + \varepsilon_{it}, \quad (11)$$

where y_{it} is an indicator for whether individual i surpassed the 10,000-step target on day t and $thresh_i$ is an indicator for being in the threshold group (see Table A.7, Panel A for impacts on daily steps and Panel B for results with the 4-day and 5-day threshold groups unpooled). Measures of individual impatience are denoted by $impatience_i$; because some measures are estimated, we present bootstrap confidence intervals in the table[26] as well as Gaussian standard errors and p-values in table notes when available.

We restrict the sample to only the base case and threshold groups, so the only difference between groups is whether their contract has a time-bundled threshold. The key coefficient of interest is β_1, which captures how the effect of the threshold (relative to the base case) varies with impatience. Our prediction is that $\beta_1 > 0$.

Table 3 shows that, consistent with our prediction, thresholds work meaningfully better for those with higher impatience over effort. Column 1 uses the impatience index (i.e., our standardized index of questions on impatience and self-control from the psychology literature) as the measure of impatience. Having a one standard deviation higher value of the impatience index increases the average performance of the threshold contracts relative to linear contracts by 4 pp (statistically significant at the 5% level). To aid in interpretation, column 2 uses a dummy for having an above-median value of the impatience measure. Relative to the base case, the threshold works 6 pp better for those with above-median impatience than those below the median. Recall that we only have the impatience index for the sample enrolled later in the intervention; to improve power and to use the full sample, columns 3 and 4 use the predicted impatience index, which is available for the full sample, as the impatience measure. We find similar (and more precise) results. Relative to the base case, the threshold works 6 pp better for those with above-median impatience than those with below-median, a large increase relative to the sample-average effect of either contract (20 pp).

Figure 7 presents a visualization of column 4, showing that adding the threshold to the

[26] To construct the bootstrap confidence intervals, we draw repeated bootstrap samples clustered at the individual level. In each sample, we first re-run the LASSO prediction model. Second, we re-create the predicted index in that sample given the LASSO coefficients from that sample, thus accounting for the error in constructing the index itself. Finally, we estimate equation 11 in that sample.

Table 3: Time-Bundled Thresholds Increase Walking More for the Impatient

Dependent variable:	Met step target (×100)			
Impatience measure:	Impatience index	Above median impatience index	Predicted impatience index	Above median predicted index
Sample:	Late	Late	Full	Full
	(1)	(2)	(3)	(4)
Impatience × Threshold	3.8**	5.97*	3.12***	5.94**
	[0.57, 7.03]	[-0.86,12.81]	[1.26, 4.71]	[0.96, 8.90]
Threshold	-1.3	-3.81	-1.18	-3.41**
	[-4.36, 1.76]	[-8.89,1.28]	[-3.02, 0.64]	[-5.38, -0.89]
Impatience	-2.97**	-4.68*	-2.38***	-5.3***
	[-5.36, -0.57]	[-9.46,0.10]	[-3.55, -0.99]	[-7.46, -1.40]
# Individuals	1,075	1,075	1,969	1,969
# Observations	86,215	86,215	157,946	157,946
Base case mean	50.4	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; its units in columns 1 and 3 are standard deviations. The sample includes the base case and time-bundled threshold incentive groups only. The "Late" sample includes only participants who were enrolled after we started measuring the impatience index; the Full sample includes everyone. The Threshold group pools the 4- and 5-day threshold groups. See Table A.7, Panel B for results with the Threshold group disaggregated (unpooled). Bootstrap draws were done at the person level, and bootstrapped 95% confidence intervals are in brackets. The Gaussian standard errors and p-values for the column 1 $Impatience \times Threshold$ coefficient are 1.9 and 0.046, respectively; for column 2 the corresponding values are 3.78 and 0.114 . Controls are the same as Table 2. Significance levels: * 10%, ** 5%, *** 1%.

linear contract increases compliance among the more impatient while decreasing it among the less. The difference between the effects is the significant 6 pp effect from column 4. Although we had no prediction for whether the threshold would have a positive or negative effect in each group (just that the effect would be *more* positive among the more impatient), it is important for policy that in our case, the effects are positive for one group and are negative for the other. This means that efforts by policymakers to individualize who receives a threshold contract based on agent impatience could substantially increase effectiveness.

Table A.8 presents estimates of Equation 11, controlling for other baseline covariates and their interactions with the threshold treatments. For example, we control for risk aversion,

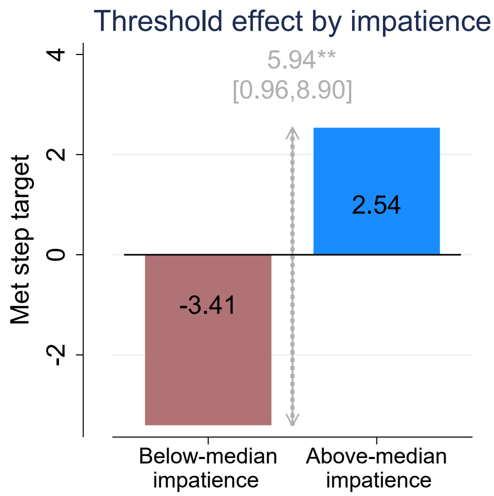


Figure 7: Time-Bundled Thresholds Increase Walking More for the Impatient

Notes: The chart plots the average probability of meeting the step target in the threshold contracts relative to the base case (linear) contract, estimated separately for those with below-median predicted impatience (left bar) versus above-median predicted impatience (right bar). The height of the vertical arrow shows the difference between the treatment effects, with the 95% confidence interval in brackets. All estimates come from Table 3 column 4.

scheduling uncertainty, and baseline walking (a proxy for the mean of the walking cost distribution), among other covariates. The coefficient on the interaction of impatience and the threshold remains stable, suggesting that it is likely impatience itself (and not its correlates) driving the estimated relationship. Another potential confound that was difficult to measure at baseline (and hence which we do not control for) is the individual-level propensity for habit formation. However, reassuringly, the propensity to form habits does not appear to be correlated with impatience in our setting, as impatience does not predict the persistence of incentive effects after payments stop (results available upon request).

Model Calibration We next 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 contracts with seven-day payment periods and with 4- and 5-day thresholds. We simplify the model slightly by assuming that individuals are fully patient over payments ($d_m = 1$) and exponentially discount exercise effort at rate (δ).

To calibrate the average compliance in the threshold and base case (linear) contracts, we need to estimate the cumulative distribution function (CDF) of walking costs. We do this by fitting a uniform distribution to several moments of the CDF from the data, as described in Appendix D. We then use the estimated distribution to predict how relative compliance in the base case and threshold contracts would vary with the discount rate over effort.

The results are displayed visually in Figure 8, with the exponential discount factor over walking δ on the x-axis and the gap between performance in the threshold and base case on the y-axis (shown separately for the 4- and 5-day thresholds). The downward-sloping curves in the figure confirm the theoretical intuition from our model: for people who are more impatient over walking (smaller δ), there are larger compliance gains from thresholds. This is true for both naifs and sophisticates for moderate levels of impatience.[27] In addition, the increase in performance of the threshold contract as impatience increases is quantitatively important, especially for the 5-day threshold contract, where the threshold has more bite. For example, decreasing the effort discount rate from 1 to 0.9 increases relative compliance in the 5-day threshold contract by roughly 3 pp among both sophisticates and naifs.[28]

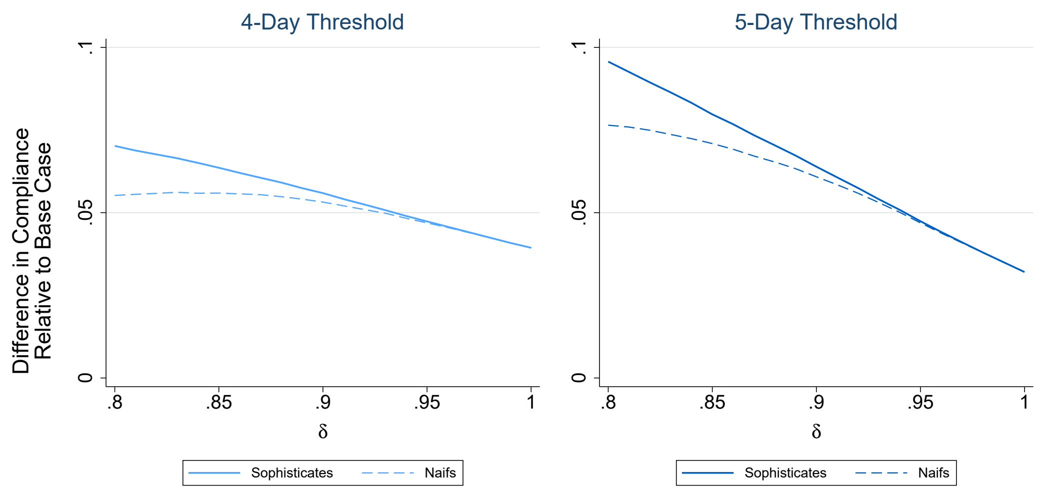


Figure 8: Threshold Relatively More Effective for More Impatient in Calibrated Model

Notes: The figure shows the difference between compliance in each Threshold contract relative to the Base Case as predicted by a walking model with uniform walking costs calibrated to our data. We assume exponential discounting over effort, with δ^t the discount factor over effort t periods in advance.

5.2.3 Time-bundled Thresholds Result Summary

Our analysis creates several new findings about time-bundled threshold contracts. Consistent with our theoretical predictions, thresholds generate meaningfully more compliance

[27]As naifs become more impatient ($\delta < 0.85$), the linear contract starts to gain relative to the 4-day threshold: as naifs become very impatient, they procrastinate in early periods under the threshold contract. However, even very impatient naifs still do better with the threshold than completely patient people ($\delta = 1$), which is our theoretical prediction when the threshold level is less than the number of periods (App. B.4).

[28]The calibration overestimates the average effect of the threshold, which in practice we found to be zero. This is likely because our model does not incorporate risk aversion over uncertain walking costs, which would decrease the average performance of the threshold. However, our main interest is heterogeneity by impatience, which we do not believe will change by incorporating uncertainty and risk aversion.

among the impatient than the patient. In the full sample, they have advantages and disadvantages, improving cost-effectiveness but increasing dispersion. The variance in their performance across the full sample underscores the potential policy gains from targeting the assignment of thresholds based on predictors of efficacy and highlights the importance of our finding that impatience over effort is one such predictor.

5.3 Payment Frequency

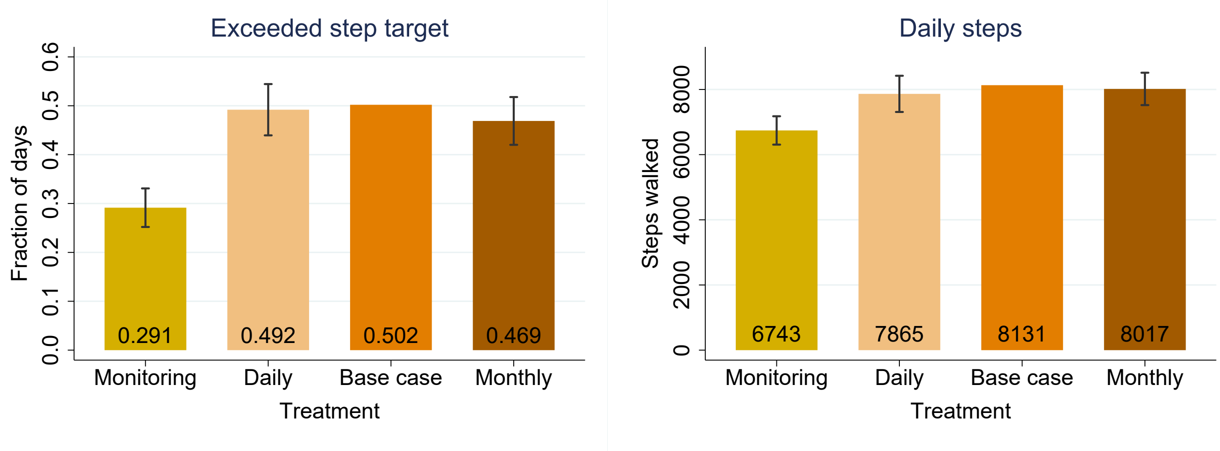
We conduct two primary analyses to better understand the roles of payment frequency and the discount rate over financial payments in incentive design:

1. Between-treatment: We compare average compliance in the daily, weekly (base case), and monthly groups. We assess how payment frequency affects compliance and use Prediction 2 to shed light on the level and shape of discount rates over payment.
2. Within-treatment: Within the base case and monthly groups, we examine how compliance changes as the payday approaches to shed light on the shape of discount rates over payment using Prediction 3. Kaur et al. (2015) uses similar variation to study discounting.

The approaches are complementary. The between-treatment approach answers the policy question of whether payment frequency matters, while the within-treatment approach has more statistical power and rules out potential confounds to the between-treatment effects.

We begin with the between-treatment comparisons. Figure 9 and Panel B of Table 2 both show that the three payment frequency treatments have similar effects of walking. The impacts on both compliance and steps walked are statistically indistinguishable. The point estimates also do not increase monotonically with frequency, as would be expected if differences reflected discounting instead of statistical noise. We thus do not find evidence that increasing payment frequency in the range from daily to monthly affects compliance – a perhaps surprising finding given the conventional wisdom that it would.

The lack of between-treatment frequency effects implies that the discount rate over financial payments is small and has a relatively flat shape over the range from one day to one month. If we interpret these findings from the lens of a quasi-hyperbolic ("beta-delta") model, the lack of difference between weekly and monthly frequency implies that the long-run discount factor, delta, may be relatively close to 1, while the lack of difference between the daily and weekly frequencies implies either that the present bias parameter beta is also relatively close to 1 (i.e., that there is limited present bias), or that the "beta window" (i.e., the "present" that is not discounted) is shorter than a day. These findings add evidence from a novel field-based test to extensive lab-based studies that trace out the payment dis-



(a) Probability Exceeded Step Target

(b) Average Daily Steps

Figure 9: Payment Frequency Does Not Significantly Impact Walking

Notes: Panel A shows the average probability of exceeding the daily step target during the intervention for the three different frequency treatments (the base case treatment pays weekly). Panel B shows average daily steps during the intervention. Confidence interval bars represent tests for equality between each group and the base case incentive group and are from regressions with the same controls as Table 2.

count rate over time and find mixed evidence on its shape.[29] Field evidence is particularly important because, in the lab, narrow bracketing may cause discount rates measured using payments to reflect discount rates over *consumption* instead (Andreoni et al., 2018b).

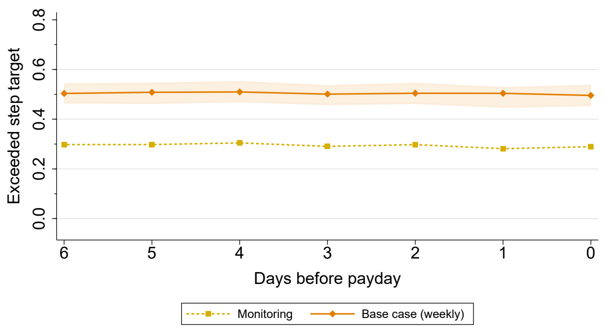
One important caveat to these results is that the between-treatment effects are somewhat imprecise, and we have limited power to reject large discount rates.[30] We address this issue with the within-treatment analysis below.

Using our between-treatment effects to make inferences about discount rates over payment requires that no other confounds drive the response to payment frequency. At the design stage, we attempted to mitigate potential confounds as well as possible.[31] Many of the remaining confounds would improve the effectiveness of higher-frequency payments. For example, if utility were concave in the payment amount, then the fact that higher-frequency payments also break payments of a given size into smaller chunks would improve compliance. Since these confounds would lead us to overestimate the discount rate over payments, they cannot be driving our finding that payment discount rates are small. That said, there do

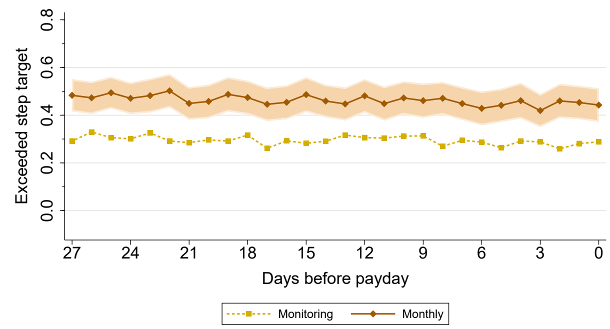
[29] While Andreoni and Sprenger (2012) find no evidence of present bias over payments and Balakrishnan et al. (2020) find that individuals are only impatient over payments in the very immediate future, Janssens et al. (2017) find present bias over payments even when the "present" period is actually "tomorrow."

[30] We cannot rule out that daily payments have an effect 4 pp higher than the base case or that monthly payments have an effect 8 pp lower.

[31] For example, higher-frequency payments could improve compliance by providing more frequent feedback; to address this, we hold text messages constant across all treatment groups.



(a) Weekly Payment Cycle



(b) Monthly Payment Cycle

Figure 10: The Probability of Exceeding the Step Target Is Stable over the Payment Cycle

Notes: The 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, day-of-week fixed effects, day-of-week relative to survey day-of-week fixed effects, and the controls in Table 2. The shaded area represents a collection of confidence intervals from tests of equality within each daily period between the incentive and monitoring groups from regressions with the same controls as Table 2.

exist potential confounds whose effects run in the opposite direction.[32] The within-treatment analysis is less subject to these confounds, as variables like payment size are fixed within treatment. That analysis, thus, allows us to more cleanly assess discount rates over payment.

The within-treatment analysis confirms the suggestive evidence of flat and low discount rates from the between-treatment analysis. Figure 10 shows how compliance within the base case weekly (Panel A) and monthly (Panel B) treatments changes as the payment date approaches. Prediction 3 shows that, if agents are impatient over payments, compliance increases as the payday approaches. Yet we find that walking behavior is remarkably steady across the payment cycle. Table A.9 estimates the change in compliance as the payment date approaches within the base case and monthly groups, conditional on day-of-week fixed effects.[33] The estimates are not significantly different from zero and suggest that, if anything, compliance *decreases* as the payment date approaches. For each day closer to the payday, compliance is 0.11 p.p. lower in the weekly group and 0.08 p.p. lower in the monthly group.

Our confidence intervals are also tighter here. If we assume linearity of compliance in lag to payment, then the confidence interval around the slope in the weekly treatment rules out

[32] For example, Casaburi and Macchiavello (2019) suggest that people might prefer lumpier payments since they can serve as commitment devices for savings.

[33] Intervention launch visits were made seven days per week, allowing us to control for day-of-week and payday day-of-week when estimating payment cycle effects. To address the concern that launch survey dates were endogenous to participants' schedules, we randomly varied the delay between the survey date and the contract launch (and hence the payday). We then control for fixed effects of day-of-week relative to the launch survey date, thereby isolating variation in the payment cycle within a given number of days from the survey day-of-week.

the possibility that because of monetary discounting, daily payments would be more than a mere 0.3 pp more effective than weekly. We also calculate an implied payment discount rate following Kaur et al. (2015). We use our small payment and base case groups to estimate the elasticity of walking to payment and combine that estimate with the slope of walking as payday approaches from column 1 of Table A.9. The implied discount rate is negative and even the maximum estimate implied by the top of our confidence interval is only 0.6%.

Although these results are consistent with recent lab-based work (e.g., Augenblick et al. 2015) in showing limited discounting over payments, the absence of payday spikes and low implied daily discount rate over payments conflicts with Kaur et al. (2015), who back out an implied daily discount rate of 4%. The reasons for the differences are an open question for future work (e.g., they may reflect different countries or payment amounts).

Evidence from self-reported contract demand further substantiates our finding that there is limited impatience over payments.[34] When we asked participants upfront whether they would prefer payments at a daily, weekly, or monthly frequency, the modal answer was weekly, preferred by 58% of participants. Daily payments were the least common choice, preferred by only 17% of participants.

5.3.1 Summary of Results

Our analysis suggests two main findings. First, changing the payment frequency between monthly and daily does not have meaningful effects on average compliance in our setting, suggesting that in contrast with conventional wisdom increasing payment frequency is not always an effective policy to improve compliance. Second, on average, the model of discounting over payments that best describes our participants is one of patience over mobile recharges.

5.4 Effectiveness and Welfare

Our focus in this paper is on maximizing contract "effectiveness" (the compliance achieved for a given payout). Improving contract effectiveness is an appropriate objective in many situations. In firm and worker applications, maximizing effectiveness is often analogous to profit maximization. In public applications with a social cost of public funds, if the incentivized behavior has a positive marginal social benefit – as is likely in our setting since the estimated social benefits to walking are large relative to the private costs – then maximizing effectiveness should increase social welfare.[35] Policymakers themselves are often concerned

[34] In addition, we do not find larger frequency treatment effects among those who appear more impatient according to our (imperfect) proxies for impatience over mobile recharges. Although we did not power our frequency groups to examine heterogeneity, for completeness, the Online Appendix shows these results.

[35] Exercise generates health benefits and financial savings by reducing the incidence of expensive complications (Reiner et al., 2013). Baseline exercise is likely inefficiently low due to both internalities and

with maximizing effectiveness, perhaps because it is straightforward to explain and justify.

One potential concern with our approach would be if the contract variations we examine improve effectiveness and/or social welfare but do not cause a Pareto improvement, instead decreasing the welfare of some individuals relative to a no-incentives benchmark. This concern is potentially relevant for the threshold contract, and is particularly vivid in light of evidence that pre-commitment contracts can decrease welfare among partially naive individuals who pay upfront for commitment but fail to follow through (e.g., Bai et al., 2020).

Are there similar concerns with offering threshold contracts, even though individuals do not pay upfront for them? In fact, there is a potentially analogous issue: naifs may comply in early periods of a threshold contract (a form of paying upfront) but fail to receive compensation because they do not follow through in later periods. Our theory suggests this concern is small for two reasons. First, later compliance costs must be much larger than earlier costs for lack of follow through to be an issue: as the compliance approaches the threshold, the incentives for marginal compliance become more and more high powered. Second, even if naifs do comply upfront but fail to follow through, this could still increase private welfare if they undercomply without incentives due to internalities like present bias.

Two pieces of empirical evidence also suggest that the program did not reduce individual participants' welfare. First, at endline, we asked participants whether they were interested in continuing the program. The vast majority said that they were interested, with no significant difference between the threshold groups and other groups and, within the threshold group, no significant difference between the more and less impatient (Table A.10).[36] Second, impatient people are no more likely (and in fact are less likely) than patient people to comply and *not* be paid for it under threshold contracts (results available upon request).

6 Empirical Results: Program Evaluation

The impacts of an incentive program on health and healthy behaviors are of policy interest, especially among a population like ours that has a high risk of complications from noncommunicable disease. This section delves into the impact of incentives on exercise patterns and health. We summarize the literature on exercise interventions for diabetics and on pedometer-based incentive programs and interpret our estimates in light of the existing evidence. We next examine how our exercise impacts changed over time, both during and after the intervention. Finally, we show that the program improved cardiovascular health.

externalities, with the externalities stemming from the fact that in many places, including India, health insurance schemes mean that individuals do not bear the full cost of their own health care.

[36] While we would ideally separate impatient *naifs* in particular, our survey questions are particularly well positioned to detect naive impatience, so this result is still reassuring that the program did not harm naifs.

6.1 Exercise Effects

Interventions previously shown to improve exercise among diabetics and prediabetics have required highly trained staff to engage in frequent and personally-tailored interactions with participants (Aziz et al., 2015; Qiu et al., 2014), and hence have had limited scalability. Since evidence conclusively shows that exercise has important health benefits for diabetics (Qiu et al., 2014), developing scalable approaches to generate exercise among diabetics is a crucial policy priority. Although scalable, low-intensity programs – and pedometer-based incentives in particular – have successfully generated exercise among non-diabetic populations, whether such approaches can also be effective among diabetics is an open question.

Encouragingly, our estimates suggest that low-intensity pedometer-based incentives can be very successful among diabetics. Our treatment effect on daily steps (1,266 from Column 2 of Table 2) is at the high end of effect sizes found in other populations, which range from only 1.5 steps in Bachireddy et al. (2019) to 1,050 steps in Finkelstein et al. (2016). Our treatment effect represents the effect of incentives relative to the monitoring group. Because monitoring itself may have an independent positive impact, our estimate is likely a lower bound on the overall impact of incentives on exercise.[37]

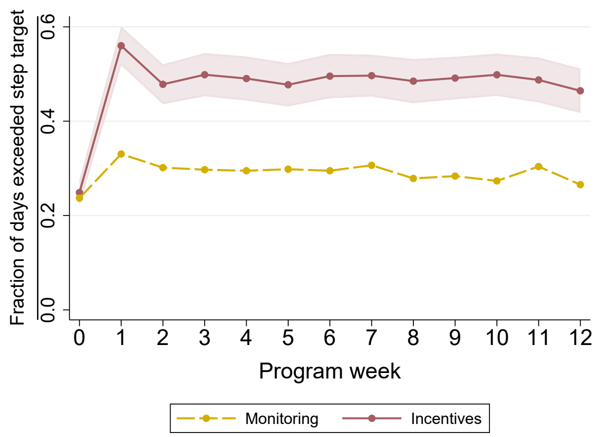
6.1.1 Robustness of Exercise Impacts

Our exercise treatment effects are robust to accounting for missing data from failure to wear pedometers. Column 3 of Table 2 reports impacts on daily steps treating days with no steps recorded as missing (which gives an unbiased estimate if participants randomly choose not to wear pedometers), and Table A.6 reports Lee bounds which account for the non-random patterns of missing data; both strategies find similar effects. The estimates are also robust to excluding the control variables from the regression (Table H.7 in the Online Appendix).

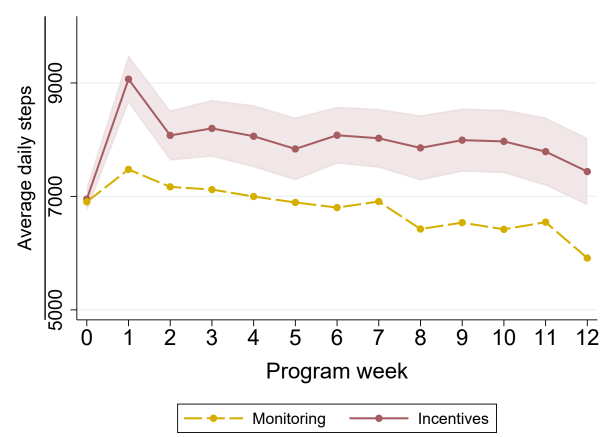
6.2 Persistence of Exercise Effects

We now analyze how the exercise impacts evolve over time, both during and after the intervention. We begin with their evolution during the intervention. Because insurers and governments are increasingly rolling out longer-term (and even permanent) incentive programs, it is important to understand whether one can sustain incentive effects throughout the intervention. Panels A and B of Figure 11 show that after an initial spike at week 1, the effect of incentives on walking remains stable during the full intervention period, suggesting that policymakers could extend this intervention further with similar effects.

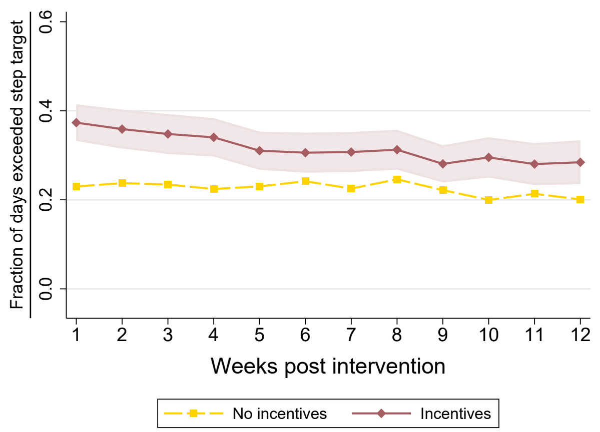
[37] A pre/post comparison shows no evidence that monitoring increases steps (see Online Appendix).



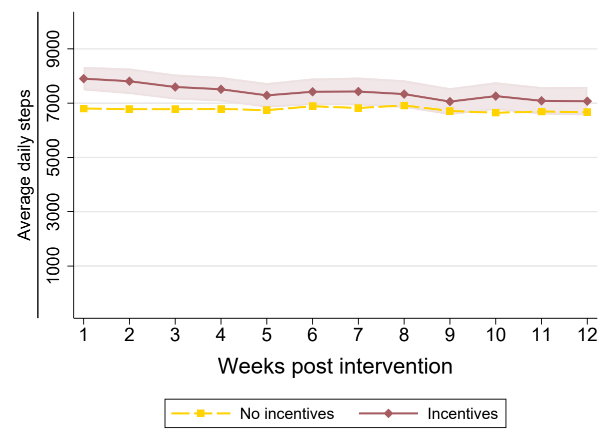
(a) Step-Target Compliance During Intervention



(b) Daily Steps Walked During Intervention



(c) Step-Target Compliance Post Intervention



(d) Daily Steps Walked Post Intervention

Figure 11: Incentive Effects are Steady through the 12-Week Program and Persist Afterward

Notes: For the pooled incentive and monitoring groups in each weekly period, Panel A shows the average probability of exceeding the step target and Panel B shows the average daily steps walked over the intervention period. Week 0 is the phase-in period (before randomization). Panels C and D show the same dependent variables as Panels A and B, respectively, over the 12 weeks subsequent to the intervention period for the pooled incentive and pooled comparison (monitoring and control) groups. The shaded area represents a collection of confidence intervals from tests of equality within each weekly period between the incentive and comparison groups from regressions with the same controls as Table 2. Intervention period graphs are unconditional on wearing whereas post-intervention period graphs are not, as described in footnote 39.

Do the effects of incentives also persist after the payments stop? Studies of similar exercise programs find mixed results regarding whether the effects persist both throughout the intervention and after incentives end.[38] Panels C and D of Figure 11 depict the difference

[38]For example, the treatment effects in Royer et al. (2015) and Bachireddy et al. (2019) fade out *during the intervention period* whereas they persist in Patel et al. (2016) and Finkelstein et al. (2016). Similarly, while Charness and Gneezy (2009) find that the effects of a roughly 4-week program incentivizing gym visits persist 7 weeks *after incentives end*, Royer et al. (2015) find the effect of a similar intervention is indistinguishable from zero 8 weeks after incentives are removed.

between our incentive and the pooled comparison groups for the 12 weeks after the intervention ended.[39] The incentive group walks significantly more even after incentives end, with impacts persisting until the last week of measurement. Table A.12 shows that the post-intervention treatment effects on steps and compliance are statistically significant and large: 55.5% and 43.1% as large as the intervention period effects, respectively. Our short-run incentive program may thus induce habit formation, enabling long-term impacts.

6.3 Health and Lifestyle Effects

We now assess the impacts of our programs on health outcomes. Our experiment was powered to detect the difference between incentive groups (pooled) and the pure control group. Table 4 reports results from regressions of the following form:

$$y_i = \alpha + \beta_1 \times incentives_i + \beta_2 \times monitoring_i + \boldsymbol{X}_i'\gamma + \varepsilon_i, \tag{12}$$

where y_i is a health outcome at endline for individual i; $incentives_i$ is an indicator for being in the incentive 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 intent-to-treat (ITT) effects on our primary outcome of health as well as on two secondary outcomes, anaerobic fitness and mental health, and two potential confounders, diet and addictive substance use. To maximize power and avoid multiple testing concerns, we create a single index of all variables in each category by taking the average of each variable, standardized by the mean and standard deviation in the control group.[40] While we report ITT estimates for each outcome individually, we focus on the indices to infer effectiveness.

Table 4 suggests that the incentive program caused moderate improvements in health.[41] Column 1 presents the treatment effect on the "Health Risk Index," which averages the five health risk factors displayed in the table. Panel A shows that, across the full population, incentives improve the index by 0.05 SDs, significant at the 10% level. Since we hypothesized ex ante that health outcomes among those with more severe diabetes might be more responsive to exercise, Panel B also examines the health impacts separately by baseline diabetes

[39] We pool comparison groups for power. The results are similar when we compare incentives with control alone (the post-intervention monitoring group is too small to analyze separately). While average pedometer-wearing rates declined from 87% in the intervention period to 69% post-intervention, wearing rates in the post-intervention period are balanced across arms (Table A.11) and our results are robust to a Lee bounds exercise (Table H.6 in the Online Appendix). We focus on results conditional on wearing the pedometer for greater comparability with intervention period effects; unconditional results also show persistence (Table A.12).

[40] For individuals who have nonmissing responses to at least one index component, we impute missing components as the sample mean following Kling et al. (2007).

[41] Each physical health outcome is trimmed using World Health Organization guidelines to trim biologically implausible health outcome measurements (i.e., z-scores < -4 or > 4).

Table 4: Incentives Moderately Improve Health

A. Sample-Average Impacts	Health risk index	HbA1c	Random blood sugar	Mean arterial BP	Body mass index	Waist circumference
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	-0.045*	-0.072	-5.67*	0.081	-0.049	-0.18
	[0.025]	[0.070]	[3.42]	[0.42]	[0.042]	[0.27]
Monitoring	0.014	-0.13	1.63	1.08	0.064	0.00080
	[0.044]	[0.12]	[6.07]	[0.75]	[0.074]	[0.48]
P-value: M = I	0.13	0.57	0.18	0.14	0.09	0.67
B. Heterogeneity by Hba1c						
Incentives × Above Median Hba1c	-0.074**	-0.15	-12.1**	-0.18	0.060	-0.18
	[0.036]	[0.10]	[4.79]	[0.61]	[0.061]	[0.39]
Incentives × Below Median Hba1c	-0.024	-0.031	-2.81	0.31	-0.14**	-0.18
	[0.035]	[0.097]	[4.56]	[0.59]	[0.058]	[0.37]
Control mean	0.00	8.44	193.83	103.02	26.45	94.44
# Individuals	3,063	3,061	3,062	3,051	3,053	3,054
P-value: I × Above Median Hba1c = I × Below Median Hba1c	0.32	0.40	0.16	0.57	0.02	1.00

Notes: The omitted category is the pure control group. Controls are the same as Table 2, along with second order polynomials of the dependent variable at baseline. The Health Risk Index is the simple average of the variables in columns 2-6, standardized with the control group mean and standard deviation. Hba1c is the average plasma glucose concentration (%), RBS is the blood glucose level (mg/dL), and mean arterial BP is the mean arterial blood pressure (mm Hg). In Panel B we control for both the main effects of above-median HbA1c and below-median HbA1c and their interactions with a monitoring group dummy. Thus, the interaction terms represent the total effects of incentives for those with above- or below-median Hba1c. Robust standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%

severity. We find somewhat stronger effects on health risk among those with more severe diabetes at baseline, although we cannot reject equality.

Table H.5 examines whether the intervention had coincident impacts on mental health or fitness. While RCTs strongly indicate that exercise improves depression among the diagnosed (Kvam et al., 2016), there is scant evidence on its mental health effects among people without a depression diagnosis. We measure mental health using seven questions from RAND's 36-Item Short Form Survey. The incentive program significantly improves mental health. In contrast, we find no effect on physical fitness, perhaps because our surveys could only measure high-intensity fitness while our intervention motivated low-intensity exercise. Finally, we do not find significant impacts on diet or addictive good consumption, as shown in the Online Appendix.

7 Conclusion

This paper investigates incentive design for impatient agents. Starting from a model where agents discount consumption and financial payments differently, we formulate incentive contract variations that interact with impatience in each domain. First, relative to linear contracts, we show that compliance with time-bundled contracts is increasing in agents' discount rates over effort. One useful feature of this prediction is that it holds regardless of whether agents are time-consistent or time-inconsistent, sophisticated or naive, thus broadening the arsenal for motivating impatient or time-inconsistent agents. The intuition behind the prediction is that the time-bundled contracts enable the principal to purchase future effort from participants instead of current effort, which is advantageous when participants discount their future effort and are willing to effectively sell it "at a discount." Our second prediction is that higher-frequency payments induce more effort if agents discount future financial payments. To assess the quantitative importance of these predictions, we implement an RCT to incentivize walking among 3,200 diabetics and prediabetics in India.

Our empirical findings regarding time-bundling are promising for policy and open up new research directions. We find that time-bundled contracts are an effective way to motivate the impatient, inducing more effort than linear contracts for those with above-median impatience. However, they induce less effort than linear contracts for those with below-median impatience. Their heterogeneous efficacy increases dispersion, highlighting the potential promise of trying to target the contracts only to those who are more impatient. One question for future research is whether such targeting could be done effectively at scale. Another potential topic to study is how to optimize the specific features of time-bundled contracts such as the payment period length T and threshold level C. Higher C and T should increase the advantage of the contracts for the impatient by increasing the number of periods of future costs that are "bundled" with present costs. However, higher C and higher T may also decrease the overall performance of the contracts, especially if there is substantial uncertainty about future costs. Future work can illuminate these trade-offs.

Our insight that impatience increases the value of time-bundling for the principal in principal-agent relationships could have broad applicability. Dynamic incentives are widespread, and we find that high discount rates over effort may be a potential explanation. A common dynamic incentive is a labor contract where an individual could be fired if she does not exert enough effort today, so effort today increases her future payoff to effort. While standard models show one reason such contracts enhance effort is simply the high stakes of job loss, our work suggests that these contracts have extra bite if the agent discounts her future effort.

Our analysis of payment frequency also raises new questions for future work. We find that increasing payment frequency is not effective in our setting because participants appear to have limited impatience over payments. Our finding suggests that, contrary to conventional wisdom, more frequent rewards are not always better, but also leaves open an important question: under what circumstances are agents impatient over payment and under what circumstances are they patient?

Finally, we find that an incentive program for walking improves health and leads to a large and persistent increase in walking among the study population. Our study thus provides some of the first evidence of a scalable, low-touch intervention with the potential to decrease the large and growing burden of chronic disease worldwide.

References

Andreoni, J. et al. (2018a). Using preference estimates to customize incentives: An application to polio vaccination drives in pakistan. *NBER Working Paper*, 22019:1–66.

Andreoni, J. et al. (2018b). Arbitrage or narrow bracketing? on using money to measure intertemporal preferences. *NBER Working Paper Series*, No. 25232.

Andreoni, J. and Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review*, 102(7):1–28.

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 inconsistency in real effort tasks. *The Quarterly Journal of Economics*, 130(3):1067–1115.

Augenblick, N. and Rabin, M. (2019). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, 86:941–975.

Aziz, Z. et al. (2015). A systematic review of real-world diabetes prevention programs: Learnings from the last 15 years. *Implementation Science*, 10(1).

Bachireddy, C. et al. (2019). Effect of different financial incentive structures on promoting physical activity among adults: A randomized controlled trial. *JAMA Network Open*, 2(8).

Bai, L., Handel, B. R., Miguel, E., and Rao, G. (2020). Self-control and demand for preventive health: Evidence from hypertension in india. *Review of Economics and Statistics*, Forthcoming.

Balakrishnan, U., Haushofer, J., and Jakiela, P. (2020). How soon is now? evidence of present bias from convex time budget experiments. *Experimental Economics*, 23:294–321.

Bassett, D. R. et al. (2010). Pedometer-measured physical activity and health behaviors in u.s. adults. *Medicine and Science in Sports and Exercise*, 42(10):1819–1825.

Burns, R. J. and Rothman, A. J. (2018). Comparing types of financial incentives to promote walking: An experimental test. *Applied Psychology: Health and Well-Being*, 10(2):193–214.

Carrera, M., Royer, H., Stehr, M., and Sydnor, J. (2020a). The structure of health incentives: Evidence from a field experiment. *Management Science*, 66(5):1783–2290.

Carrera, M. et al. (2020b). Who chooses commitment? evidence and welfare implications. *NBER Working Paper*.

Carvalho, L. S., Meier, S., and Wang, S. W. (2016). Poverty and economic decision-making: Evidence from changes in financial resources at payday. *American Economic Review*, 106(2):260–284.

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.

Chassang, S. (2013). Calibrated incentive contracts. *Econometrica*, 81(5):1935–1971.

Cubitt, R. P. and Read, D. (2007). Can intertemporal choice experiments elicit time preferences for consumption? *Experimental Economics*, 10:369–389.

Cutler, D. M. and Everett, W. (2010). Thinking outside the pillbox - medication adherence as a priority for health care reform. *The New England Journal of Medicine*, 362(17):1553–1555.

Edmans, A., Gabaix, X., Sadzik, T., and Sannikov, Y. (2012). Dynamic ceo compensation. *Journal of Finance*, 67(5):1603–1647.

Finkelstein, E. A. et al. (2016). Effectiveness of activity trackers with and without incentives to increase physical activity (trippa): A randomised controlled trial. *The Lancet Diabetes and Endocrinology*, 4(12):983–995.

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.

Howells, L., Musaddaq, B., McKay, A. J., and Majeed, A. (2016). Clinical impact of lifestyle interventions for the prevention of diabetes: An overview of systematic reviews. *BMJ Open*, 6(12):1–17.

International Diabetes Federation (2019). *Idf Diabetes Atlas*. International Diabetes Federation, Brussels, Belgium, 9 edition.

Jacob, B. A. and Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, 118(3):843–877.

Jain, S. (2012). Self-control and incentives: An analysis of multiperiod quota plans. *Marketing Science*, 31(5):855–869.

Janssens, W., Kramer, B., and Swart, L. (2017). Be patient when measuring hyperbolic discounting: Stationarity, time consistency and time invariance in a field experiment. *Journal of Development Economics*, 126(December 2016):77–90.

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. In Bernheim, B. D., Dellavigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics - Foundations and Applications 2*, volume 2, chapter 5, pages 345–458. Elsevier B.V.

Kvam, S., Lykkedrang, C., Hilde, I., and Hovland, A. (2016). Exercise as a treatment for depression: A meta-analysis. *Journal of Affective Disorders*, 202:67–86.

Laibson, D. (2015). Why don't present-biased agents make commitments? *American Economic Review*, 105(5):267–272.

Larkin, I. and Leider, S. (2012). Incentive schemes, sorting, and behavioral biases of employees: Experimental evidence. *American Economic Journal: Microeconomics*, 4(2):184–214.

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

Lazear, E. P. (1981). Agency, earnings profiles, productivity, and hours restrictions. *American Economic Review*, 71(4):606–620.

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

Long, J. A. (2012). "buddy system" of peer mentors may help control diabetes. *LDI Issue Brief*, 17(6):1–4.

Ministry of Labour and Unemployment (2016). Report on fifth annual employment - unemployment survey (2015-16). Technical report, Labour Bureau, Government of India, Chandigarh.

O'Donoghue, T. and Rabin, M. (1999a). Doing it now or later. *American Economic Review*, 89(1):103–124.

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

Oyer, P. (1998). Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality. *The Quarterly Journal of Economics*, 113(1):149–185.

Patel, M. S. et al. (2016). Framing financial incentives to increase physical activity among overweight and obese adults. *Annals of Internal Medicine*, 164(6):385.

Qiu, S. et al. (2014). Impact of walking on glycemic control and other cardiovascular risk factors in type 2 diabetes: A meta-analysis. *PLoS ONE*, 9(10).

Reiner, M., Niermann, C., Jekauc, D., and Woll, A. (2013). Long-term health benefits of physical activity - a systematic review of longitudinal studies. *BMC Public Health*, 13(1):1–9.

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.

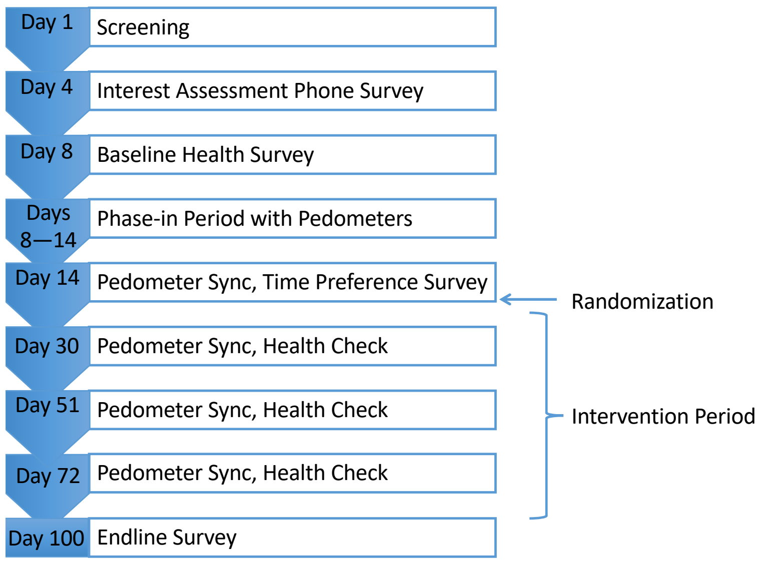
Tuckman, B. W. (1991). The development and concurrent validity of the procrastination scale. *Educational and Psychological Measurement*, 51:473–480.

Volpp, K. G. et al. (2008). Financial incentive based approaches for weight loss: A randomized trial. *JAMA: the Journal of the American Medical Association*, 300(22):2631–2637.

Appendices

This section contains all appendix tables, appendix figures, and Appendices B - D. The Online Appendix is a separate document available at: faculty.chicagobooth.edu/-/media/faculty/rebecca-dizon-ross/research/incentivedesignapp.pdf

Appendix Figure A.1: Experimental Timeline for Sample Participant



Notes: This figure shows an experimental timeline for a participant. Visits were scheduled according to the participants' availability. We introduced variation into the timing of incentive delivery by delaying the start of the intervention period by one day for randomly selected participants. The intervention period was exactly 12 weeks for all participants.

Appendix Table A.1: Enrollment statistics

Total screened: 57,599

Total eligible: 7,781

Stage:	# Individuals	% of total eligible
	(1)	(2)
Successfully contacted	6,965	90%
Interested in enrolling	5,552	71%
Completed baseline survey	3,438	44%
Successfully enrolled	3,192	41%

Appendix Table A.2: Participants displayed high levels of understanding of their assigned contracts

Contract Type	Questions	% Correct	
		First call	Any Call
Base Case $n = 902$	How many recharges would you receive on (*payment day of week*) if you walked 10,000 steps on exactly 1 day over the period (*payment day of week*) to (*payment day of week* -1) ?	0.96	
	How many times over the course of this week would you receive recharges if you walked 10,000 steps on exactly 5 days over the period (*payment day of week*) to (*payment day of week* -1) ?	0.96	
4-day Threshold $n = 794$	What is the minimum number of days that you need to walk to get a recharge?	0.88	
	How many recharges would you receive at the end of this week if you walked 10,000 steps on exactly 1 day this week?	0.87	
	How many recharges would you receive at the end of this week if you walked 10,000 steps on exactly 4 days this week?	0.88	
	How many recharges would you receive at the end of this week if you walked 10,000 steps on exactly 6 days this week?	0.88	
5-day Threshold $n = 312$	What is the minimum number of days that you need to walk to get a recharge?	0.89	
	How many recharges would you receive at the end of this week if you walked 10,000 steps on exactly 1 day this week?	0.93	
	How many recharges would you receive at the end of this week if you walked 10,000 steps on exactly 5 days this week?	0.94	
	How many recharges would you receive at the end of this week if you walked 10,000 steps on exactly 6 days this week?	0.93	
Daily $n = 166$	How many times over the course of this week would you receive recharges if you walked 10,000 steps on exactly 1 day ?	0.97	
	How many times over the course of this week would you receive recharges if you walked 10,000 steps on exactly 5 days ?	0.99	
Monthly $n = 164$	How many recharges would you receive on (*payment day of week*) if you walked 10,000 steps on exactly 1 day over this week ?	0.98	
	How many recharges would you receive on (*payment day of week*) if you walked 10,000 steps on exactly 5 days in this week ?	0.99	
Monitoring $n = 205$	How do you report your steps to us?	0.99	
	How large is the fitbit wearing bonus?	0.41	

Notes: Each participant in the monthly, base case, and threshold groups was always paid on the same day of the week, which is labeled *"payment day of week"*. This was the day they launched in the program (e.g., if a base case participant started in the program on a Wednesday, they would be paid every Wednesday for their exercise between the previous Wednesday to Tuesday). Although monthly group participants were paid at the end of the month, their payment day of the week would still correspond with this day.

Appendix Table A.3: Impatience measures correlate in the expected direction with baseline measures of behavior and health

Covariate type:	Exercise		Baseline Indices			
Dependent variable:	Daily steps	Daily exercise (min)	Negative health risk index	Negative vices index	Healthy diet index	# Individuals
A. Impatience Index Measures						
Impatience index	-0.080***	-0.070***	-0.017	-0.052	-0.185***	1,760
1. I'm always saying: I'll do it tomorrow	-0.059	-0.100***	-0.012	-0.031	-0.150***	1,760
2. I usually accomplish all the things I plan to do in a day	-0.054	-0.053	-0.012	-0.043*	-0.151***	1,760
3. I postpone starting on things I dislike to do	-0.041*	0.006	0.004	-0.053	0.047	1,760
4. I'm on time for appointments	-0.053	0.002	-0.021	0.010	-0.097***	1,760
5. I often start things at the last minute and find it difficult to complete them on time	-0.041*	-0.066***	-0.009	-0.043*	-0.209***	1,760
B. Predicted index measures						
Predicted index	0.001	-0.038	-0.061***	0.020	0.005	3,232
1. In the past week, how many times have you found yourself exercising less than you had planned?	0.016	-0.009	-0.060***	0.010	0.027	3,232
2. In the past 24 hours, how many times have you found yourself eating foods you had planned to avoid?	-0.001	0.053***	-0.059***	0.015	0.033*	3,232
3. Do you worry that if you kept a higher balance on your phone, you would spend more on talk time?	-0.027	-0.063***	-0.018	0.025	-0.038	3,232

Notes: This table displays the correlations between our impatience measures and a number of baseline health and behavior measures. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize health and behavior measures so that higher values correspond to healthier behavior; hence we expect all correlations to be negative. Panel A displays the impatience index along with the five questions from which it is generated. Panel B shows the predicted index along with the three questions from which it is generated. 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%.

Appendix Table A.4: Baseline impatience summary statistics in full sample and by treatment group.

	Full Sample	Control	Monitoring	Incentives pooled	Daily	Base case	Monthly	Threshold	Small payment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Impatience index									
1. I am continually saying: I'll do it tomorrow	2.22	2.07	2.14	2.26	2.13	2.28	2.40	2.23	2.51
	(1.46)	(1.40)	(1.30)	(1.48)	(1.50)	(1.47)	(1.38)	(1.49)	(1.68)
2. I usually accomplish all the things I plan to do in a day	0.64	0.57	0.67	0.66	0.59	0.66	0.60	0.67	0.68
	(0.97)	(0.89)	(0.98)	(0.98)	(0.97)	(0.99)	(0.87)	(0.99)	(1.03)
3. I postpone starting on things I dislike to do	3.97	3.76	4.17	4.00	4.07	4.04	3.99	3.98	3.68
	(1.31)	(1.37)	(1.11)	(1.30)	(1.27)	(1.27)	(1.28)	(1.31)	(1.56)
4. I'm on time for appointments	0.47	0.49	0.37	0.47	0.48	0.50	0.52	0.44	0.51
	(0.86)	(0.88)	(0.73)	(0.86)	(0.82)	(0.93)	(0.90)	(0.81)	(0.84)
5. I often start things at the last minute and find it difficult to complete them on time	2.51	2.53	2.30	2.52	2.29	2.51	2.72	2.49	3.19
	(1.54)	(1.53)	(1.47)	(1.55)	(1.48)	(1.54)	(1.58)	(1.55)	(1.58)
# Individuals	1,740	316	111	1,313	86	487	93	610	37
B. Predicted index									
1. In the past week, how many times have you found yourself exercising less than you had planned?	0.53	0.56	0.45	0.52	0.51	0.58	0.55	0.48	0.47
	(1.01)	(1.02)	(0.93)	(1.02)	(1.04)	(1.05)	(1.08)	(0.98)	(1.01)
2. In the past 24 hours, how many times have you found yourself eating foods you had planned to avoid?	0.21	0.17	0.25	0.21	0.21	0.23	0.25	0.20	0.15
	(0.54)	(0.49)	(0.56)	(0.55)	(0.61)	(0.58)	(0.70)	(0.49)	(0.44)
3. Do you worry that if you kept a higher balance on your phone, you would spend more on talk time? (=1)	0.13	0.15	0.10	0.13	0.11	0.12	0.13	0.14	0.12
	(0.34)	(0.35)	(0.31)	(0.34)	(0.32)	(0.33)	(0.34)	(0.34)	(0.33)
# Individuals	3,192	585	203	2,404	166	902	164	1,106	66

Notes: Standard deviations in parentheses. Components of the impatience index range from 1 to 5, 1 being very false and 5 being very true. Responses to question 2 from the predicted index range from 0 to 3. The Threshold column pools 4- and 5-day threshold groups.

Appendix Table A.5: Missing pedometer data during the intervention period

Dep. Variable:	No Steps data	**Reason** no steps data		**Reason** no data from fitbit			
		Did not wear fitbit	No data from fitbit	Lost data entire period	Immediate withdrawal	Mid-intervention withdrawal	Other reasons
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Incentives	-0.0151	-0.0309**	0.0152	-0.00178	0.00542	0.0164**	-0.00480
	[0.0176]	[0.0144]	[0.0124]	[0.00506]	[0.00724]	[0.00691]	[0.00594]
Controls	X	X	X	X	X	X	X
# Individuals	2,607	2,559	2,607	2,607	2,607	2,607	2,607
# Observations	218,988	205,732	218,988	218,988	218,988	218,988	218,988
Monitoring mean	0.19	0.15	0.05	0.00	0.01	0.01	0.02

Notes: This table reports balance of missing data by treatment status. Each observation is an individual×day. There are two reasons why data can be missing: people did not wear their pedometers (column 2) or we do not have data from the person's pedometer (column 3). Columns 2 + 3 = Column 1 except that column 2 conditions on there not being missing data for consistency with our main step analyses whereas columns 1 and 3 do not (column 2 results similar without this restriction). Columns 4-7 summarize reasons for why steps data might have been missing, and sum up to column 3. Some people have no data during the entire intervention period (columns 4 and 5) because their pedometers broke and all intervention data was lost (4), or because they withdrew immediately after being assigned a treatment group (5). Others only have missing data for part of the intervention period, either because they withdrew midway through the period (6) or had a broken fitbit or a failed sync (7). "Did not wear fitbit" takes value 1 when steps = 0 for that day. Controls are the same as Table 2. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.6: Lee bounds on the impacts of incentives on exercise during the intervention period

Definition of missing:	No steps data	No data from fitbit	Did not wear fitbit	Lost data entire period	Withdrew immedi-ately	Mid-period withdrawal	Other reasons
A. Daily steps							
Regression estimate	1269	1338	1269	1338	1338	1338	1338
(conditional on nonmissing data)	[245]	[261]	[245]	[261]	[261]	[261]	[261]
Lee lower bound	1053	1230	882	1315	1297	1226	1303
	[62]	[44]	[53]	[43]	[43]	[43]	[43]
Lee upper bound	1426	1572	1571	1351	1430	1581	1358
	[55]	[48]	[51]	[42]	[44]	[44]	[42]
B. Met 10k step target							
Regression estimate	0.223	0.205	0.223	0.205	0.205	0.205	0.205
(conditional on nonmissing data)	[0.024]	[0.022]	[0.024]	[0.022]	[0.022]	[0.022]	[0.022]
Lee lower bound	0.215	0.200	0.208	0.204	0.203	0.200	0.204
	[0.005]	[0.004]	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]
Lee upper bound	0.232	0.216	0.242	0.206	0.209	0.217	0.206
	[0.005]	[0.004]	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]
# Individuals	2,557	2,559	2,557	2,559	2,559	2,559	2,559
# Observations	180,018	205,732	180,018	205,732	205,732	205,732	205,732

Notes: This table reports regression estimates and Lee bounds accounting for different types of missing pedometer data. The regression estimates and Lee bounds condition on data not being missing, using different definitions of missing data in each column. All estimates are of the effect of incentives pooled relative to the monitoring group. Note that regression estimates are not comparable to those reported in Table 2 because each column conditions on the "type of missing" indicator in the first row being equal to 0 and does not include controls.

Appendix Table A.7: Robustness of Threshold Heterogeneity Results

	Impatience index	Above median impatience index	Predicted impatience index	Above median predicted index
	Late	Late	Full	Full
	(1)	(2)	(3)	(4)
A. Dependent Variable = Steps				
Impatience × Threshold	289*	525	238**	521**
	[-41, 619]	[-268,1318]	[49, 392]	[64, 832]
Threshold	-143	-369	-166*	-360***
	[-442, 157]	[-899, 161]	[-348, 18]	[-589, -118]
Impatience	-209	-397	-229***	-549***
	[-474, 56]	[-944, 150]	[-350, -90]	[-749, -183]
Base case mean	8,098	8,098	8,131	8,131
B. Dependent Variable = Met Step Target (×100)				
Impatience × 5-day Threshold	3.52*	5.47	3.66***	7.29**
	[-0.05, 7.08]	[-2.28,13.22]	[1.58, 5.43]	[1.74, 10.36]
5-day Threshold	-1.72	-3.98	-1.71*	-4.42***
	[-5.16, 1.73]	[-9.78,1.82]	[-3.59, 0.25]	[-6.71, -1.64]
Impatience × 4-day Threshold	5.00*	7.49	1.76	2.51
	[-0.94, 10.94]	[-3.93,18.91]	[-0.87, 4.36]	[-3.39, 7.16]
4-day Threshold	-0.14	-3.39	0.17	-0.84
	[-4.53, 4.26]	[-10.59,3.81]	[-2.56, 2.81]	[-3.98, 2.44]
Impatience	-2.97**	-4.68*	-2.39***	-5.32***
	[-5.36, -0.58]	[-9.45,0.10]	[-3.56, -1.00]	[-7.48, -1.39]
Base case mean	50.4	50.4	50.2	50.2
# Individuals	1,075	1,075	1,969	1,969
# Observations	86,215	86,215	157,946	157,946

Notes: Panel A shows that the threshold heterogeneity reported in Table 3 is robust to using daily steps as the outcome. Each Panel reports results from a different outcome variable estimated on the same sample. Panel B shows heterogeneity in the 4-day and 5-day threshold treatments by impatience with threshold groups disaggregated (unpooled). The impatience measure changes across columns; its units in columns 1 and 3 are standard deviations. The sample includes the base case and time-bundled threshold incentive groups only. The "Late" sample includes only participants who were enrolled after we started measuring the impatience index; the Full sample includes everyone. The Threshold group pools the 4- and 5-day threshold groups. Bootstrap draws were done at the person level, and bootstrapped 95% confidence intervals are in brackets. For Panel A: The Gaussian standard errors and p-values for the column 1 $Impatience \times Threshold$ coefficient are 1.9 and 0.046, respectively; for column 2 the corresponding values are 3.78 and 0.114 . For Panel B: The Gaussian standard errors and p-values for the column 1 $Impatience \times$ 5-*day Threshold* coefficient are 2.07 and 0.090, respectively; for column 2 the corresponding values are 4.11 and 0.183 . The Gaussian standard errors and p-values for the column 1 $Impatience \times$ 4-*day* coefficient are 3.09 and 0.106, respectively; for column 2 the corresponding values are 5.66 and 0.186 . Controls are the same as Table 2. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.8: 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	3.12***	3.24***	3.12***	3.29***	3.14***	3.11***	3.21***	3.11***	2.96***	3.07***	3.14***
	[1.26,4.71]	[1.42,4.84]	[1.24,4.70]	[1.43,4.88]	[1.28,4.73]	[1.26,4.65]	[1.36,4.79]	[1.28,4.71]	[1.09,4.54]	[1.22,4.64]	[1.27,4.70]
Predicted index	-2.38***	-2.42***	-2.38***	-2.39***	-2.37***	-2.3***	-2.39***	-2.33***	-2.29***	-2.39***	-2.47***
	[-3.55,-0.99]	[-3.58,-1.03]	[-3.55,-0.97]	[-3.53,-1.01]	[-3.54,-0.98]	[-3.49,-0.91]	[-3.56,-1.01]	[-3.50,-0.95]	[-3.42,-0.91]	[-3.56,-0.99]	[-3.61,-1.05]
Threshold	-1.18	-11.7**	-.946	-1.19	-2.02*	-1.24	-4.42**	1.97	-.187	-1.5	-.941
	[-3.02,0.64]	[-22.15,-1.16]	[-3.25,1.36]	[-3.00,0.66]	[-4.04,0.07]	[-3.02,0.60]	[-7.88,-1.20]	[-5.31,9.50]	[-2.23,1.79]	[-3.63,0.50]	[-4.04,1.21]
Threshold × Covariate		0.213**	-0.569	-1.560	0.029*	0.009	1.189**	-0.868	-0.002*	1.067	-0.062
		[0.008,0.422]	[-3.992,2.840]	[-4.622,1.665]	[-0.006,0.063]	[-0.070,0.080]	[0.190,2.221]	[-2.816,1.114]	[-0.005,0.000]	[-3.106,5.207]	[-0.235,0.240]
Covariate		1.550**	1.123	-2.127*	-0.008	-0.250***	-0.428	1.771**	-0.001*	3.132**	-3.339***
		[0.359,2.580]	[-1.893,4.594]	[-4.653,0.419]	[-0.033,0.021]	[-0.364,-0.129]	[-1.112,0.263]	[0.279,3.252]	[-0.003,0.000]	[0.165,6.090]	[-8.436,-2.391]
# Individuals	1,969	1,969	1,969	1,969	1,969	1,969	1,969	1,969	1,969	1,969	1,969
# Observations	168,672	168,672	168,672	168,672	168,672	168,672	168,672	168,672	168,672	168,672	168,672
B. Impatience index											
Impatience index × Threshold	3.80**	3.90**	3.79**	3.84**	3.96**	3.80**	3.81**	3.95**	3.52**	3.84**	3.74**
	[0.57,7.03]	[0.84,6.97]	[0.58,7.01]	[0.61,7.07]	[0.73,7.18]	[0.79,6.80]	[0.63,6.99]	[0.72,7.18]	[0.31,6.73]	[0.65,7.02]	[0.51,6.97]
Impatience index	-2.97**	-3.01**	-2.97**	-3.00**	-3.06**	-2.76**	-2.96**	-3.10**	-2.67**	-2.98**	-2.92**
	[-5.36,-0.57]	[-5.35,-0.67]	[-5.33,-0.61]	[-5.41,-0.58]	[-5.42,-0.70]	[-5.13,-0.39]	[-5.24,-0.68]	[-5.52,-0.67]	[-5.19,-0.15]	[-5.40,-0.55]	[-5.21,-0.63]
Threshold	-1.30	-10.3	-0.97	-1.30	-3.10	-2.07	-1.47	-0.34	7.22	-1.47	-2.24
	[-4.36,1.76]	[-34.1,13.4]	[-5.05,3.10]	[-4.31,1.70]	[-6.79,0.59]	[-5.18,1.03]	[-8.27,5.32]	[-4.14,3.46]	[-4.95,19.4]	[-5.09,2.15]	[-8.48,4.00]
Threshold × Covariate		0.18	-0.76	1.71	0.063	0.19*	0.019	-0.0025	-2.32	0.39	0.12
		[-0.28,0.65]	[-8.14,6.62]	[-3.29,6.70]	[-0.016,0.14]	[-0.0065,0.38]	[-2.33,2.36]	[-0.0078,0.0029]	[-5.68,1.03]	[-7.96,8.74]	[-0.58,0.81]
Covariate		1.79*	-0.76	-4.85*	-0.036	-0.63***	0.29	-0.0011	1.95	2.90	-2.96
		[-0.27,3.84]	[-5.34,3.82]	[-9.95,0.25]	[-0.088,0.016]	[-0.92,-0.34]	[-1.34,1.93]	[-0.0053,0.0031]	[-0.90,4.81]	[-4.12,9.92]	[-8.85,2.94]
Covariate used	-	Age	Female	Health risk index	Mobile balance (INR)	Yesterday's talk time (INR)	Risk aversion	Scheduling certainty	Daily personal income	Education (above median)	Baseline steps (over 1000)
# Individuals	1,075	1,075	1,075	1,075	1,075	1,075	1,075	1,075	1,075	1,075	1,075
# Observations	92,148	92,148	92,148	92,148	92,148	92,148	92,148	92,148	92,148	92,148	92,148
Base case mean	50.4	50.4	50.4	50.4	50.4	50.4	50.4	50.4	50.4	50.4	50.4

Notes: The sample is restricted to the base case (linear) group and the 2 threshold groups, 4-day threshold and 5-day threshold, pooled together here as "Threshold." All columns control for the baseline value of the dependent variable and the same controls as Table 2. Panel A uses the predicted index as the measure for impatience while Panel B uses the impatience index; the units for both impatience measures are standard deviations. The unit of observation is a respondent × day. Bootstrap draws were done at the person level, and bootstrapped 95% confidence intervals are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.9: Walking Does Not Vary Significantly across the Pay Cycle

.

Dependent variable:	Met step target (×100)				
Payment frequency:	Weekly		Monthly		
	(1)	(2)	(3)	(4)	(5)
Days before payday	0.11		0.08		
	[0.09]		[0.05]		
Payday		-0.63		0.12	
		[0.55]		[1.02]	
Payweek					-0.12
					[1.02]
# Individuals	890	890	163	163	163
# Observations	71,672	71,672	13,333	13,333	13,333
Sample mean	50.2	50.2	49.3	49.3	49.3

Notes: The columns show the effect of days until payday 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 base case (weekly) treatment group, and the sample in columns 3 and 4 is restricted to the monthly treatment group. Regressions control for payday day-of-week fixed effects, day-of-week fixed effects, day-of-week relative to launch survey day-of-week fixed effects, a day-of-contract-period time trend, and the controls in Table 2. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.10: Threshold contracts do not significantly decrease satisfaction at endline

Dependent variable:	Interest in continuing program	
Impatience measure:	Above median predicted index	
	(1)	(2)
Threshold	-0.0222	-0.0117
	[0.0150]	[0.0194]
Impatience × Threshold		-0.0266
		[0.0300]
Impatience		0.0541**
		[0.0211]
Base case (omitted) mean	0.880	0.880
# Individuals	2607	2607

Notes: This table shows predictors of satisfaction with the walking program. We ask respondents at endline if they are interested in continuing the program for an extra 3 months. The impatience measure is a dummy for being above-median on the predicted impatience index. Controls are the same as Table 2, as well as the main effect for impatience and treatment indicators (both main effects and interactions with impatience) for being in the daily, monthly, small payment, or monitoring treatments. The omitted group is the base case (weekly) group. The Threshold group pools the 4- and 5-day threshold groups. Standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.11: Missing pedometer data during the post-intervention period

Dep. Variable:	No Steps data	**Reason** no steps data		**Reason** no data from fitbit			
		Did not wear fitbit	No data from fitbit	Lost data entire period	Immediate withdrawal	Mid-intervention withdrawal	Other reasons
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Incentives	-0.00616	0.000425	-0.0109	0.00187	0.00214	-0.00872*	-0.00622
	[0.0232]	[0.0211]	[0.0202]	[0.00384]	[0.0191]	[0.00497]	[0.00596]
Controls	X	X	X	X	X	X	X
# Individuals	1,254	1,122	1,254	1,254	1,254	1,254	1,254
# Observations	105,336	91,756	105,336	105,336	105,336	105,336	105,336
Monitoring mean	0.40	0.31	0.13	0.00	0.10	0.01	0.02

Notes: This table reports the reasons that we do not have step data by treatment status in the post-interevntion period. Each observation is an individual × day. Controls are the same as Table 2. Column 1 reports fitbit data missing for any reason, which can include mid-intervention withdrawal, fitbit sync issues, and not wearing the fitbit, amongst others. Columns 2 + 3 = Column 1, except that column 2 conditions on there being no missing data for consistency with our main step analyses whereas columns 1 and 3 do not. Columns 4-7 summarize reasons for why steps data might have been missing, and the variables in columns 4-7 sum up to the variable in column 3. Some people have no data during the entire measurement period, as summarized in columns 4 and 5. The omitted category is the pooled control and monitoring groups. "Did not wear fitbit" takes value 1 when fitbit data is non-missing and fitbit steps = 0. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.12: The Effects of Incentives Persist After the Intervention Ends

	Conditional on wearing fitbit		Unconditional on wearing fitbit	
Dependent variable:	Compliance	Daily Steps	Compliance	Daily Steps
	(1)	(2)	(3)	(4)
Incentives	0.093***	648.3***	0.071***	537.2**
	[0.02]	[195.82]	[0.01]	[220.90]
No incentives mean	0.3	7347.4	0.2	5687.4
% Persistence	43.4	55.9	35.7	43.0
# Individuals	1,122	1,122	1,122	1,122

Note: Table shows the average treatment effect of incentives during the post-intervention period. The omitted group is the monitoring and control groups (pooled). We considered a participant to have worn the pedometer if their step count > 0. Each observation is a person-day; columns 1 and 2 only include days where the participant wore the pedometer and columns 3 and 4 include all days. The % Persistence row shows the treatment effect from the post-intervention period divided by the corresponding treatment effect from the intervention period, where the intervention period treatment effect comes from a specification using the same dependent variable and pedometer-wearing condition. Controls are the same as Table 2. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.13: Predicted and Actual Impatience Indicies Do Not Correlate with Other Proxies for Impatience over Recharges

Covariate type:	Recharge variables			Credit constraint proxies			
Dependent variable:	Negative mobile balance	Negative yesterday's talk time	Marginal talk time, if gifted	Negative wealth index	Negative monthly household income	Negative monthly personal income	# Individuals
Impatience Index	0.033	-0.075	0.082***	0.044*	0.061***	0.009	1760
Predicted Impatience Index	0.021	-0.012	0.050	-0.034*	0.006	-0.017	3232

Notes: This table displays the correlations between predicted and actual impatience indices and baseline measures that should be related to credit constraints and discount rates over recharges. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize the proxies so that higher values correspond to higher expected discount rates; hence the prediction is that coefficients should be positive. All CTB parameters have been winsorized at the top and bottom 1 percentile to remove outliers. Significance levels: * 10%, ** 5%, *** 1%.

B Theoretical Predictions Appendix

B.1 Proof of Prediction 1

Prediction 1. *Holding all else equal, average compliance in the time-bundled threshold contract relative to the separable contract is weakly decreasing in the discount factor over effort, $\delta(k)$.*

Proof. We first use Equation 5 to rewrite compliance under a separable linear contract for the case where $T = 2$:

$$\sum_{t=1}^{2} w_t|^{\text{Separable}} = \mathbb{1}\{e_1 < d_m(1)m\} + \mathbb{1}\{e_2 < m\}. \tag{13}$$

We now solve for compliance under the time-bundled threshold. On day 1, the agent considers if it is worth the effort to comply on *both* days in order to be paid. She compares the present discounted cost of both days' effort, $e_1 + \delta(1)e_2$, with the value of the payment, $d_m(1)2m$. She wants to comply on both days if the costs are low enough:

$$e_1 + \delta(1)e_2 < d_m(1)2m. \tag{14}$$

Equation 14 is more likely to hold if agents discount effort more. The cost of both days' effort from the day 1 perspective is decreasing in $\delta(1)$ since day 2 effort is discounted. The role of $\delta(1)$ for desired compliance in the threshold contract contrasts with the separable contract, where $\delta(1)$ does not affect compliance: agents only compare present effort costs with future payment (Equation 13).

Equation 14 tells us whether the individual's day 1 self wants to comply in both days, but if individuals are time-inconsistent, then it is not sufficient to determine compliance. We consider sophisticates' compliance first. Sophisticates know that their day 2 selves may not share their day 1 selves' preferences, and may not comply on day 2 even if Equation 14 holds. Sophisticates will comply only if Equation 14 holds *and* they know they will follow through on day 2, which happens if $e_2 < 2m$. The compliance of a sophisticate under the threshold is thus

$$\sum_{t=1}^{2} w_t|^{\text{Threshold, Sophisticate}} = \mathbb{1}\{e_1 + \delta(1)e_2 < d_m(1)2m\} \times \mathbb{1}\{e_2 < 2m\} \times 2. \tag{15}$$

Inspection of Equations 13 and 15 shows that Prediction 1 holds for sophisticates since compliance in Equation 15 is decreasing in $\delta(1)$, whereas compliance in Equation 13 is invariant to $\delta(1)$.

We next consider naifs. Unlike sophisticates, naifs assume their day 2 selves will follow through as their day 1 selves desire. They thus comply on day 1 if Equation 14 holds and comply on day 2 if they complied on day 1 and $e_2 < 2m$:

$$\sum_{t=1}^{2} w_t|^{\text{Threshold, Naif}} = \mathbb{1}\{e_1 + \delta(1)e_2 < d_m(1)2m\} \times (1 + \mathbb{1}\{e_2 < 2m\}). \tag{16}$$

Inspection of Equations 13 and 16 proves Prediction 1 for naifs: naif compliance under the time-bundled threshold, but not the separable contract, is decreasing in $\delta(1)$.

The relative effectiveness of the threshold contract is therefore increasing in $\delta(1)$ for naifs and sophisticates. Agents who discount future effort more have a lower total discounted cost of reaching the threshold and thus have higher compliance for a given payment level. □

B.2 Other Types of Time-Bundled Contracts

This section examines the full space of two-day time-bundled contracts. We first define notation to describe the key contract parameters. Second, we characterize which types of contracts will generate higher compliance among an impatient person than a fully patient person, formally characterizing the concepts of option and commitment described in Section 2.3.2. Third, we present proofs describing the types of two-day time-bundled contracts for which Prediction 1 holds, separately for sophisticates and for naifs.

B.2.1 Notation

The two main contract parameters that interact with time preferences are:

1. m_{2L}: the payment for day 2 compliance if the agent did *not* comply on day 1.
2. m_{2H}: the payment for day 2 compliance if the agent *did* comply on day 1.

Since we examine time-bundled contracts in this section, we assume $m_{2H} > m_{2L}$ (the dynamic complementarity). To simplify notation, we set $d_m(1) = 1$ for the remainder of the subsection. All results in this section (as throughout the paper) assume that agents only comply if the payment is strictly larger than the cost; they do not comply when the cost and payment are equal.

Since behavior under a given contract depends on the specific cost realization (and, in particular, the value of e_2), we categorize contracts for a given value of e_2.

B.2.2 Contracts that Yield Higher Compliance Among an Impatient Person with a Given $\delta(1) < 1$ than a Fully Patient Person

Prediction 1 holds among agents who think that their day 2 compliance is pivotal to their day 1 compliance (we show this more rigorously in Section B.2.3). In this subsection, we describe the contracts that meet this condition for naive and sophisticated agents for a given $\delta(1)$ and e_2, using the notions of "option" and "commitment". That is, we characterize the contracts

For an impatient naif who has discount factor over effort $\delta(1) < 1$, the condition holds when day 1 compliance creates option value, wherein day 1 compliance is pivotal to whether the day 1 naif *wants* to comply on day 2. Option contracts satisfy:

$$\text{Option: } m_{2L} \leq \delta(1)e_2 < m_{2H} \tag{17}$$

For sophisticates with the same discount factor $\delta(1)$, the condition holds when day 1 compliance creates commitment, wherein day 1 compliance is pivotal to whether the agent will *actually* follow through on day 2. Commitment contracts satisfy:

$$\text{Commitment: } m_{2L} \leq e_2 < m_{2H} \tag{18}$$

Some contracts satisfy both equations 17 and 18 and are thus effective for both types (i.e., generate more effort from both types than from a patient person with $\delta(1) = 1$):

$$\text{Option + commitment: } m_{2L} \leq \delta(1)e_2 \text{ and } e_2 < m_{2H} \tag{19}$$

Appendix Table B.1: Two-day time-bundled contracts

	Contract definitions		**Comparative statics**	
	Payment for day 2 compliance:		Sign of slope[a] of compliance w.r.t. $\delta(1)$	
	If did not comply on day 1 (m_{2L})	If complied on day 1 (m_{2H})	Naif	Sophisticate
Contract	(1)	(2)	(3)	(4)
Option + commitment	$< e_2$	$> e_2$	≤ 0	≤ 0
Commitment-only	e_2	$> e_2$	0	≤ 0
Option-only	$< e_2$	$\leq e_2$	≤ 0	0
Inframarginal	$> e_2$	$> e_2$	0	0

[a] Compliance takes on the values 0 and 1. We denote the slope as ≤ 0 when compliance as a function of $\delta(1)$ is a step function with a derivative that is always either 0 or $-\infty$.

In contrast, contracts with commitment but not option work for sophisticates but not naifs:

$$\text{Commitment only: } \delta(1)e_2 < m_{2L} \leq e_2 < m_{2H} \tag{20}$$

And contracts with option but not commitment work for naifs but not sophisticates:

$$\text{Option only: } m_{2L} \leq \delta(1)e_2 < m_{2H} \leq e_2 \tag{21}$$

B.2.3 Predictions and Proofs for General $\delta(1)$

We define four contract types, also summarized in the first two columns of Table B.1

1. Option + commitment: $m_{2L} < e_2 < m_{2H}$
2. Commitment-only: $m_{2L} = e_2 < m_{2H}$
3. Option-only: $m_{2L} < m_{2H} \leq e_2$
4. Inframarginal: $e_2 < m_{2L} < m_{2H}$

Note that contract types 1, 2, and 3 satisfy equations 19, 20, and 21, respectively, for some $\delta(1) < 1$, with contract type 2 in fact satisfying equation 20 for all $\delta(1) < 1$.

We define some additional notation for the proofs. Let m_1 be the payment for day 1 compliance, and the realized day 2 payment be $m_2 = (1 - w_1)m_{2L} + w_1 m_{2H}$. Finally, let an individual's day-1 willingness to pay (WTP) in day 2 dollars to have the principal increase her day 2 payment for compliance from payment level m to m' be $WTP_m^{m'}$.

We now demonstrate the following predictions, summarized in Table B.1 columns (3)-(4).

Prediction 4. *For naifs, for a given cost realization, compliance is weakly decreasing in $\delta(1)$ for Option-only and Option + commitment contracts, but invariant to $\delta(1)$ under Commitment-only and Inframarginal contracts.*

Proof. Time-bundled contracts with a two-day payment period reward walking on day 1 with m_1 and an increase in the day-2 payment from m_{2L} to m_{2H}. Individuals thus comply on day 1 if

$$e_1 < m_1 + WTP_{m_{2L}}^{m_{2H}}. \tag{22}$$

As a result, day 1 (and hence total)[42] compliance is decreasing in $\delta(1)$ if and only if $WTP_{m_{2L}}^{m_{2H}}$ decreases in $\delta(1)$.

We now show that, for naifs, $WTP_{m_{2L}}^{m_{2H}}$ only decreases in $\delta(1)$ for Option-only and Option + commitment contracts. Because naifs think their day 2 selves will comply when $m > \delta(1)e_2$, their WTP to increase the day 2 payment from 0 to $m > 0$ is

$$WTP_0^m|^{\text{Naif}} = \max(m - \delta(1)e_2, 0). \tag{23}$$

Since $WTP_{m_{2L}}^{m_{2H}} = WTP_0^{m_{2H}} - WTP_0^{m_{2L}}$, then for any $m_{2H} > m_{2L}$:

$$WTP_{m_{2L}}^{m_{2H}}|^{\text{Naif}} = \begin{cases} m_{2H} - m_{2L} & \text{if } \delta(1)e_2 \leq m_{2L} < m_{2H} \\ m_{2H} - \delta(1)e_2 & \text{if } m_{2L} < \delta(1)e_2 < m_{2H} \\ 0 & \text{if } m_{2L} < m_{2H} \leq \delta(1)e_2 \end{cases} \tag{24}$$

Equation 24 shows that $WTP_{m_{2L}}^{m_{2H}}|^{\text{Naif}}$ is weakly decreasing in $\delta(1)$ for all parameter values. It is strictly increasing in $\delta(1)$ when $m_{2L} < \delta(1)e_2 < m_{2H}$, which is the option case: from the day 1 perspective, walking on day 2 is not worth a payment of m_{2L} but is worth a payment of m_{2H}. As a result, increasing the payment from m_{2L} to m_{2H} creates a lucrative option that the naif believes she will exploit on day 2. The more she discounts effort, the more valuable the option is. Since there exists a range of $\delta(1) \leq 1$ for which $m_{2L} < \delta(1)e_2 < m_{2H}$ for all Option-only and Option + commitment contracts, Equation 24 implies that, for those contracts, there exists a range of e_1, m_1, and $\delta(1) \leq 1$ for which $WTP_{m_{2L}}^{m_{2H}}$ (and hence compliance) is strictly decreasing in $\delta(1)$.

To see that compliance is not decreasing in $\delta(1)$ under Commitment-only or Inframarginal contracts, note that Equation 24 shows that for those contracts $WTP_{m_{2L}}^{m_{2H}} = m_{2H} - m_{2L}$, which is not decreasing in $\delta(1)$. The intuition is that, in these contracts, the naif believes on day 1 that she will comply on day 2 regardless of her day 1 action. She thus values the increase in m_2 from m_{2L} to m_{2H} at exactly its cash value, which is invariant to $\delta(1)$. □

Note that, under Commitment-only contracts, the naif is mistaken about her day 2 compliance: she thinks she will comply on day 2 regardless of her day 1 action, but in reality will only comply on day 2 if she complies on day 1. If naifs were not mistaken, then Prediction 1 would also hold for Commitment-only contracts for naifs.

How are Option + commitment contracts able to induce more effort from impatient naifs than from patient people (i.e., have Prediction 1 hold) while Commitment-only contracts are not? Relative to Commitment-only contracts, Option + commitment contracts "increase the stakes" by decreasing m_{2L}, which helps the naif realize that she would only comply on day 2 if she complies on day 1. By worsening the consequences of noncompliance on day 1, the Option + commitment

[42]Total compliance follows from day 1 compliance since day 2 compliance is increasing in day 1 compliance (because $m_{2H} > m_{2L}$) and is not otherwise affected by $\delta(1)$.

contracts help guide naifs to take the action in their own best interest (complying on day 1 when it is truly in their best day 1 interest to do so) by preventing them from overestimating their day 2 compliance were they not to comply in day 1.

Another interesting pattern for naifs is that, although Option-only contracts induce more compliance from impatient naifs on day 1 than from patient people, they do not induce extra compliance on day 2: no agent (regardless of $\delta(1)$) will comply on day 2 of an Option-only contract. In contrast, if an Option + commitment contract induces additional compliance from an impatient naif on day 1, it will also induce extra compliance on day 2 because agents who comply on day 1 of Option + commitment contracts always follow-through on day 2.

Prediction 5. *For sophisticates, for a given cost realization, compliance is weakly decreasing in $\delta(1)$ for Commitment-only and Option + commitment contracts, but invariant to $\delta(1)$ under Option-only and Inframarginal contracts.*

Proof. Sophisticates know they will not comply on day 2 for a payment less than e_2, yielding the following WTP for a day 2 payment $m > 0$:

$$WTP_0^m|^{\text{Soph}} = (m - \delta(1)e_2) \times 1\{e_2 < m\}. \tag{25}$$

WTP to increase m_2 from m_{2L} to m_{2H} is thus

$$WTP_{m_{2L}}^{m_{2H}}|^{\text{Soph}} = \begin{cases} m_{2H} - m_{2L} & \text{if } e_2 < m_{2L} < m_{2H} \text{ (Inframarginal)} \\ m_{2H} - \delta(1)e_2 & \text{if } m_{2L} \leq e_2 < m_{2H} \text{ (Option + Comm. or Comm. only)} \\ 0 & \text{if } m_{2L} < m_{2H} \leq e_2 \text{ (Option-only)} \end{cases} \tag{26}$$

The critical cutoff for day 2 payment over which it is differentially valuable for more impatient sophisticates is thus e_2, not $\delta(1)e_2$ as it was for naifs. Even though the day 2 contract appears to be a lucrative option as soon as m_2 surpasses $\delta(1)e_2$, sophisticates know their day 2 selves will not follow through on the option if $m_2 < e_2$. Thus Option-only contracts are not more effective for low-$\delta(1)$ sophisticates. In contrast, in contracts with commitment, sophisticates know that day 1 compliance is pivotal to day 2 compliance, and the more they discount future effort, the higher the net benefits of compliance.

To complete the proof, note that compliance with Inframarginal contracts is invariant to $\delta(1)$ for the same reason as for naifs. □

Interestingly, under Commitment-only and Option + Commitment contracts, sophisticates are willing to pay more than dollar for dollar to increase the day 2 payment from m_{2L} to m_{2H}.[43] In this way, these time-bundled contracts operate like standard commitment contracts.

B.3 Predictions with Uncertainty

We now show that, under reasonable distributional assumptions, Prediction 1 holds in the case where future effort costs are unknown (i.e., the agent knows the distribution of her future effort costs but not the realizations). For simplicity, we examine the same two-day model as Section 2 and consider the case where effort costs are weakly positive and $d_m = 1$. Assume the agent's day 2 costs are distributed according to the distribution function $F(\cdot)$. We examine sophisticates first.

[43] That is, if $\delta(1) < 1$, $WTP_{m_{2L}}^{m_{2H}}|^{m_{2H} > e_2, m_{2L} = e_2} = m_{2H} - \delta(1)e_2 > m_{2H} - m_{2L}$.

Sophisticates We solve backwards. Day 2 compliance can be expressed as

$$w_2 = \mathbb{1}\{w_1 = 1 \ \& \ e_2 < 2m\}, \tag{27}$$

and conditional on day 1 compliance, day 2 compliance occurs with probability $F(2m)$.

The sophisticate anticipates her day 2 behavior, and thus complies on day 1 if the expected payments are greater than the discounted sum of expected costs:

$$e_1 + \delta(1)F(2m)E[e_2|e_2 < 2m] < F(2m)2m \tag{28}$$

The lower is $\delta(1)$, the more likely is this equation to hold.

Equations 27 and 28 show that Prediction 1 holds for sophisticates with cost uncertainty. Overall compliance is decreasing in $\delta(1)$ since day 1 compliance decreases with $\delta(1)$ (equation 28) and day 2 compliance increases with day 1 compliance (equation 27).

Naifs The solution for naifs is more complicated because naifs' beliefs about their own future compliance depend on $\delta(1)$. As such, we need to put some structure on the distribution of e_2. We show the result here assuming that e_2 is distributed uniformly from 0 to k.[44]

We solve forwards. On day 1, the naif thinks she will comply on day 2 if:

$$w_2 = \mathbb{1}\{w_1 = 1 \ \& \ \delta(1)e_2 < 2m\}. \tag{29}$$

She then complies on day 1 if the expected payments are greater than the discounted sum of expected costs:

$$e_1 + \delta(1)F\left(\frac{2m}{\delta(1)}\right)E[e_2|\delta(1)e_2 < 2m] < F\left(\frac{2m}{\delta(1)}\right)2m \tag{30}$$

With uniform costs, this expression simplifies to the following condition for day 1 compliance:

$$e_1 < \begin{cases} 2m - \delta(1)\frac{k}{2} & \text{if } \delta(1) < \frac{2m}{k} \\ \frac{(2m)^2}{\delta(1)2k} & \text{if } \delta(1) \geq \frac{2m}{k} \end{cases} \tag{31}$$

The lower is $\delta(1)$, the more likely equation 31 is to hold, as both of the "cutoffs" for compliance ($2m - \delta(1)\frac{k}{2}$ and $\frac{(2m)^2}{\delta(1)2k}$) are decreasing in $\delta(1)$ (the cutoffs are equal when $\delta(1) = \frac{2m}{k}$). Thus day 1 compliance is decreasing in $\delta(1)$ for naifs. Like sophisticates, naifs then comply on day 2 if equation 27 holds.

Hence, overall compliance for naifs is decreasing in $\delta(1)$. Day 1 compliance is decreasing in $\delta(1)$ (equation 31), and day 2 compliance is increasing in day 1 compliance (equation 27).

B.4 Predictions with Thresholds Less than 100%

In the main text, we examine dynamic thresholds where the threshold level of compliance is set at 100% (i.e., where someone must comply on 100% of days to receive payment, so the threshold level C is equal to the payment period length T). In Online Appendix E, we show that Prediction 1 still holds under certain cost assumptions when the threshold is less than

[44] The result goes through with many other distribution functions. The key condition is that the PDF of e_2 cannot be too much greater between $2m$ and $\frac{2m}{\delta(e)}$ than it is right below $2m$.

100% (i.e., when $C < T$). We outline the predictions here and show the proofs in Online Appendix E.

We consider a three-day model ($T = 3$) where all payments are made on day 3. People discount effort t days in the future with an exponential discount factor of δ^t, $\delta \leq 1$. Given the short time frame, we assume the discount factor over payments is one. In the threshold contract, people are paid m per day complied if they comply on at least two days (i.e., $C = 2$). In the nonthreshold contract, people are paid m per day complied regardless of how many days they comply.

We then prove the following predictions. To build intuition, we begin with the simplified case where the per-day cost of walking is constant across days and show the following:

Prediction 6. *When the cost of walking is constant across days, compliance in the threshold contract relative to the nonthreshold contract is weakly decreasing in the discount factor δ.*

We then examine more realistic cases where costs do not have to be constant across days.

Prediction 7. *Assume the cost of effort in each day is binary, taking on either a "high value" (e_H) or a "low value" (e_L), with $e_H \geq e_L$. Compliance in the threshold contract relative to the nonthreshold is weakly higher for someone with a discount factor $\delta < 1$ than for someone with discount factor $\delta = 1$.*

Prediction 7A. *For sophisticates, regardless of the cost distribution, compliance in the threshold contract relative to the nonthreshold contract is weakly decreasing in the discount factor δ.*

See Online Appendix E for proofs and discussion.

C Misreporting Steps, Confusion, and Suspensions

Procedures to Curb Misreporting Because incentive payments for walking were determined by self-reported data and not pedometer data, we implemented a number of checks during the intervention to ensure integrity of step reporting. Within each 28 day sync period, respondents who were found to have incorrectly over-reported meeting a 10k step target on over 40% of days were flagged for cheating and contacted by a member of our field team. Those who were flagged were suspended from receiving recharges for 7 days. Those who were flagged for cheating more than one time were terminated from the program. As shown in Table C.1, fewer than 5% of the incentive group was suspended for cheating and only 1 participant was terminated.

During the intervention period, we also attempted to flag participants who appeared to be confused about how to read their pedometers or report properly. Our pedometers record daily steps until midnight, and because respondents typically reported their daily steps via our IT system before midnight, we expected that even if people report correctly, reported steps may be slightly under pedometer steps. We tagged those whose reported steps were either more than 10% higher than their pedometer steps or more than 15% lower than their pedometer steps as “confused.” Those who were flagged as simply “confused” received tutorials from the surveyors on how to use the step-reporting system.

Rates of Misreporting and Confusion Although our analysis only uses pedometer data (not reported data), so misreporting would not bias our conclusions, it is still interesting to examine whether misreporting was prevalent in practice. We find that the prevalence of “misreporting” behavior, defined as reporting walking at least 10,000 steps when the pedometer itself records fewer than 10,000 steps, is less than 5% and, interestingly, relatively balanced across incentive and monitoring groups. See column 1 of Table C.2. The balance with the monitoring group, who had no incentives to over-report, suggests that much of what looks like intentional misreporting was simply participant mistakes or confusion. We also find that the incentive group appeared to put more effort into making sure their step reports were correct, with less examples of divergences in either the positive or negative directions (columns 2-4 of Table C.2).

D Calibrating the CDF of Walking Costs

Here, we describe how we estimate the cumulative distribution function (CDF) of walking costs using the step data for our calibration exercise in Section 5.2.2.

To estimate values of the CDF, we take advantage of the fact that payments are not discounted on payday, and so the probability of walking on paydays is equal to the probability that walking costs are less than the payment amount – or the CDF of the payment amount. Therefore, average payday walking in the small payment treatment uncovers $F(10)$, and average payday walking in the base case and daily groups uncover $F(20)$. Average walking in the monitoring treatment uncovers $F(0)$. We also use two additional moments: the probability of walking for the 4-day (5-day) threshold group when one had already walked three days (four days), and it is the last day of the contract period uncovers $F(80)$ ($F(100)$). With these five data points in hand, we fit a linear (i.e., uniform) walking cost CDF using a linear regression of $F(x)$ on (x).

Appendix Table C.1: Summary statistics on audits and suspensions

	Count		Share	
	Incentives	Monitoring	Incentives	Monitoring
	(1)	(2)	(3)	(4)
Shared fitbit ever*	3	0	0.004	0.000
Suspended for cheating	100	N/A	0.042	N/A
Terminated for cheating	1	N/A	0.000	N/A
Total:	2,404	203	0.92	0.08

*Notes: We randomly audited around 1,000 individuals from both the incentive and monitoring groups to look for evidence of pedometer sharing. The first row in columns (3) and (4) is conditional on being audited.

Appendix Table C.2: Misreporting, confusion and cheating by contract group

Variable type:	Reporting	Confusion		
Dependent variable:	Incorrectly reported over 10k steps	Over-reported or under-reported	Over-reported by at least 10%	Under-reported by at least 15%
	(1)	(2)	(3)	(4)
Incentives	0.0079	-0.081***	-0.059***	-0.022**
	[0.01]	[0.02]	[0.02]	[0.01]
Monitoring mean	0.049	0.272	0.167	0.104
# Individuals	2,542	2,542	2,542	2,542
#Observations	173,131	173,131	173,131	173,131

Notes: Each observation is a respondent × day. Column 2 shows whether a respondent over-reported by at least 10% or under-reported by at least 15%. The omitted group is the monitoring group. Analysis is restricted to dates falling within an individual's contract period. Controls include baseline steps as well as all other variables included in Table 2 to maintain consistency with other step analyses. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table C.3: Summaries from the minute-level pedometer data

	Incentives	Monitoring	I - M	P-value: I=M
	(1)	(2)	(3)	(4)
A. Activity (by minute)				
Average daily activity	213	197	16	0.001
Average steps per minute	41	38	3	0.001
B. Time of Day				
Average start time	07:11	07:16	5	0.441
Average end time	20:49	20:50	1	0.742
C. High step counts per minute (share)				
Steps > 242	0	0	0	-
Steps > 150	1.3e-06	0	1.3e-06	-
# Individuals:	2,368	201		

Notes: This table presents various statistics at the respondent $\times$ minute level. High step count 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 intervention period.

Because the final two CDF values are estimated on a selected sample, they are only valid for estimating the population-level cost distribution if costs are i.i.d. across people. Therefore, while the calibration exercise is useful for exposition, our model may not perfectly fit observed behavior.

Online Appendix for:

Incentivizing Behavioral Change:
The Role of Time Preferences

Shilpa Aggarwal
Indian School of Business

Rebecca Dizon-Ross
University of Chicago

Ariel Zucker
UC Berkeley

Table of Contents

E Theoretical Predictions: Additional Proofs

F CTB Time Preference Measurement

G Heterogeneity in Frequency Effects by (Proxies for) Impatience over Payment

H Program Evaluation: Appendices and Additional Analysis

I Monitoring Treatment Impacts on Walking

J Lifestyle Modification Guidelines

E Theoretical Predictions: Additional Proofs

We now provide the proofs for Predictions 6 - 7A, which show that Prediction 1 still holds under certain cost assumptions when the threshold is less than 100%.

We consider a three-period model where all payments are made in period 3. People discount effort t periods in the future with an exponential discount factor of δ^t, $\delta \leq 1$. Given the short time frame, we assume the discount factor over payments is one. In the threshold contract, people are paid m per period complied if they comply in at least two periods. In the nonthreshold contract, people are paid m per period complied regardless of how many periods they comply with.

We define some useful notation:

- X_t is the "walking stock" coming into period t (i.e., sum from period 1 to period $t-1$ of whether the person complied $X_t = \sum_{i=1}^{t-1} w_i$).
- $w_t(X_t)$ is a dummy for whether the person complies in period t as a function of the walking stock coming into period t.

E.1 Constant Costs

We first build intuition by examining the simplified case where the per-period cost of walking is constant across periods: $e_t = e$ for all t.

Prediction 6. *When the cost of walking is constant across periods, compliance in the threshold contract relative to the nonthreshold contract is weakly decreasing in the discount factor δ.*

Proof. Compliance in the nonthreshold contract does not depend on δ. Thus the prediction holds if we can show that compliance in the threshold contract is weakly decreasing in δ. We examine the problem separately for three different cost cases: $e \geq 2m$, $m \leq e < 2m$, and $e < m$. The first and third cases are not interesting. For the case $e \geq 2m$, it will never be worth it to walk in any period, and so $\sum_{t=1}^{3} w_t = 0$ for all δ. On the other hand, if $e < m$, since the cost of walking is low relative to the incentive level, then it is worth it for the participant to walk in all periods regardless of δ, and so $\sum_{t=1}^{3} w_t = 3$ for all δ.

We now show the proof for the interesting case where $m \leq e < 2m$. In this case, walking in the nonthreshold contract would be zero when $\delta = 1$, so we are exploring whether there are any δ for which the threshold contract could generate walking when the nonthreshold contract could not.

We begin by solving for sophisticates, and then for naifs. For sophisticates, we solve backwards.

Sophisticates in Period 3: $m \leq e < 2m$. Behavior will depend on the walking stock X_3. There are three cases:

1. "In the money" ($X_3 = 2$). The person walks if $e < m$. Thus they never walk:

$$w_3(2) = 0. \tag{32}$$

2. "On the cusp" ($X_3 = 1$). The person walks if $e < 2m$. Thus they always walk:

$$w_3(1) = 1. \tag{33}$$

3. "Out of the money" ($X_3 = 0$). The person walks if $e < 0$. Thus they never walk:

$$w_3(0) = 0 \tag{34}$$

Sophisticates in Period 2: $m \leq e < 2m$.

1. On the cusp ($X_2 = 1$). The person walks if

$$-e + 2m + w_3(2)(m - \delta e) \geq w_3(1)(-\delta e + 2m)$$
$$\Rightarrow \delta e \geq e.$$

   If $\delta = 1$, then this means they walk in period 2 and do not walk in period 3; if $\delta < 1$, then they do not walk in period 2 but do walk in period 3.

2. On track ($X_2 = 0$). The person walks if $e + \delta e < 2m$. Note that this is the same "cost-bundling" equation that causes 100% threshold contracts to work better for those with lower δ. The equation means that they walk in both periods 2 and 3 if $\delta < \frac{2m}{e} - 1$.

Sophisticates in Period 1: $m \leq e < 2m$**.** The sophisticate knows that she will never walk in all three periods since $w_3(2) = 0$. She also knows that she will always achieve the threshold if she walks in period 1 or 2, since $w_3(1) = 1$. If $\delta < 1$, then to achieve two periods of walking, she would rather walk in periods 2 and 3 than walk in periods 1 and 3 or periods 1 and 2, from today's perspective. So when $\delta < \frac{2m}{e} - 1$, she will wait until period 2 and then walk. Thus, we just have to check if there are any scenarios where it is worth it for her to walk in period 1 and period 3 when she would not walk in periods 2 and 3. From period 1's perspective, it is worth it to walk in periods 1 and 3 if $-e + \delta^2 e \leq 2m$, which means $\delta < \sqrt{\frac{2m}{e} - 1}$. Since this is higher than the threshold for walking in period 2 ($\delta < \frac{2m}{e} - 1$), she will thus walk in period 1 if $\frac{2m}{e} - 1 < \delta < \sqrt{\frac{2m}{e} - 1}$.

Sophisticates Summary: $m \leq e < 2m$**.** We summarize walking as a function of δ in Table E.1, which shows that total walking weakly decreases in δ as a result of the same cost-bundling logic that drives the efficacy of the 100% threshold contract for the impatient.

Appendix Table E.1: Walking by Sophisticate in 3-Day Contract with 2-Day Threshold, by δ

	Walks on day:			
Discount factor over walking (δ)	1	2	3	Total days walked
$\sqrt{\frac{2m}{e} - 1} < \delta \leq 1$	N	N	N	0
$\frac{2m}{e} - 1 < \delta \leq \sqrt{\frac{2m}{e} - 1}$	Y	N	Y	2
$\delta < \frac{2m}{e} - 1$	N	Y	Y	2

Notes: Per-period cost e is assumed to be weakly positive.

Naifs We now examine the case where $m \leq e < 2m$ for naifs. By definition, naifs always think they will follow through on the plan that is best from today's perspective; as a result, we can "solve forward," examining what is best for the naif to do in any period assuming she will follow-through on it.

When $m \leq e < 2m$, it is never optimal for the naif to walk in all three periods (since the payment for walking in period 3 conditional on walking in periods 1 and 2 is just $m \leq e$). Thus, the naif walks in a maximum of two periods. If $\delta = 1$, then the naif does not care which periods she walks in but will never walk since the condition for walking twice would be that $e + e < 2m$, which can never hold given the costs.

If $\delta < 1$, then the naif will always want to postpone walking to periods 2 and 3. In period 2, she will walk if $e + \delta e < 2m$, which will hold for any $\delta < \frac{2m}{e} - 1$.

Thus, as shown in Table E.2, our prediction holds due to the same cost-bundling intuition we saw before.

One interesting point is that for $\frac{2m}{e} - 1 < \delta \leq \sqrt{\frac{2m}{e} - 1}$, the naif does not walk even though it would have been in her best interest in period 1 to do so. This is because she thinks she will walk in period 2 instead. However, our main prediction of interest still holds for naifs: the threshold still works better for the naif the lower δ is. □

Appendix Table E.2: Walking by Naif in 3-Day Contract with 2-Day Threshold, by δ

	Walks on day:			
Discount factor over walking (δ)	1	2	3	Total days walked
$\frac{2m}{e} - 1 < \delta \leq 1$	N	N	N	0
$\delta < \frac{2m}{e} - 1$	N	Y	Y	2

Notes: Per-period cost e is assumed to be weakly positive.

E.1.1 Non-Constant Costs

We now allow for costs to vary across periods. We assume for simplicity that costs are all weakly positive (the results are the same when we allow costs to be negative, but it complicates the notation). We show the prediction under a reasonable case where effort costs are binary.

Prediction 7. *Assume the cost of effort in each period is binary, taking on either a "high value" (e_H) or a "low value" (e_L), with $e_H \geq e_L$. Compliance in the threshold contract relative to the nonthreshold is weakly higher for someone with a discount factor $\delta < 1$ than for someone with discount factor $\delta = 1$.*

Proof. Since compliance in the non threshold contract is invariant to δ, the prediction holds if we can show that compliance in the threshold contract is weakly higher for $\delta < 1$ than for $\delta = 1$. We first consider different values of e_H and e_M. First, if $e_H < m$, then $\sum_{t=1}^{3} w_t = 3$ for all δ and so the prediction trivially goes through. Second, if $e_L \geq m$, then $\sum_{t=1}^{3} w_t = 0$ for $\delta = 1$. However, some people with $\delta < 1$ may walk in at least one period due to the standard cost-bundling effect (e.g., if they have costs of e_L every period and if $e_L + \delta e_L < 2m$, then they would walk twice). Thus the prediction goes through in that case as well. We thus have proved the prediction in the cases where $e_H < m$ and $e_L \geq m$ and so we next consider the cases where $e_H \geq m$ and $e_L < m$.

To prove the prediction, we examine all 8 potential sequences of costs and prove it separately for each case. Note that we only consider the cases where $e_H \geq m$ and $e_L < m$.

1. Cases 1 and 2: e_L, e_L, e_L and e_H, e_H, e_H
   Since in these cases, costs are constant across periods, the prediction goes through by the proofs for the constant cost prediction (Prediction 6).
2. Case 3: e_H, e_H, e_L: Again neither sophisticates nor naifs walk in period 1 but both walk in period 2 and period 3 if $e_H + \delta e_L < 2m$ (note that by the assumptions above, since $e_L < m$, they will always follow-through so there is no follow-through constraint). Thus total compliance is decreasing in δ.
3. Case 4: e_H, e_L, e_H. Again nobody walks in period 1. Sophisticates walk in periods 2 and 3 if $e_L + \delta e_H < 2m$ and $e_H < 2m$. Naifs walk in period 2 if $e_L + \delta e_H < 2m$ and in period 3 if they've walked in period 2 and $e_H < 2m$. Again total compliance is decreasing in δ.

4. Case 5: e_L, e_H, e_H. Sophisticates walk in period 1 if $e_L + \delta^2 e_H < 2m$ and they know they will follow through ($e_H < 2m$). Naifs walk in period 1 if $e_L + \delta^2 e_H < 2m$. Neither type walks in period 2 since $e_H \geq m$. Both types walk in period 3 if they walked in period 1 and $e_H < 2m$. Again total compliance is thus decreasing in δ.

5. Cases 6, 7, and 8: e_L, e_H, e_L; e_L, e_L, e_H; and e_H, e_L, e_L. All people, regardless of δ, walk in the two periods where the cost is e_L, since $e_L + e_L < 2m$. Nobody walks in the period where the cost is e_H since they know they will walk in the other periods and $e_H \geq m$. Thus the prediction (trivially)holds.

□

For sophisticates, we can also show a stronger result. In simulations with most realistic cost distributions, this stronger result goes through for naifs as well.

Prediction 7A. *For sophisticates, regardless of the cost distribution, compliance in the threshold contract relative to the nonthreshold contract is weakly decreasing in the discount factor δ.*

Proof. We work backward. In period 3, behavior will depend on the walking stock X_3:

$$\begin{aligned} w_3(2) &= 1\{e_3 < m\} \\ w_3(1) &= 1\{e_3 < 2m\} \\ w_3(0) &= 1\{e_3 < 0\}. \end{aligned}$$

We show that the prediction holds by showing that it holds under all potential cases for e_3.

Case 1: $m \leq e_3 < 2m$ In this case, walking in period 3 is

$$\begin{aligned} w_3(2) &= 0 \\ w_3(1) &= 1 \\ w_3(0) &= 0. \end{aligned}$$

Note that this implies the person will never walk three times. Walking in period 2 is

$$\begin{aligned} w_2(1) &= 1\{e_2 \leq \delta e_3\} \\ w_2(0) &= 1\{e_2 + \delta e_3 < 2m\}. \end{aligned}$$

In period 1, consider two cases:

1. $e_2 + \delta e_3 < 2m$: she knows she will walk at least twice, and the only question is whether to walk now or later. If $e_1 < min\{\delta e_2, \delta^2 e_3\}$, then she will walk in period 1; if not, then she will wait and walk in periods 2 and 3. Either way, she walks twice.
2. $e_2 + \delta e_3 \geq 2m$: she knows she will not walk later, so she will walk if $e_1 + min\{\delta e_2, \delta^2 e_3\} < 2m$.

Thus we can see that when $m \leq e_3 < 2m$, overall compliance is as follows:

$$\text{Days walked} = \begin{cases} 2 & \text{if } e_2 + \delta e_3 \leq 2m \text{ OR } e_1 + \delta min\{e_2, \delta e_3\} \leq 2m \\ 0 & \text{otherwise.} \end{cases}$$

Thus, compliance is obviously decreasing in δ.

Case 2: $e_3 \geq 2m$ In this case, the person will never walk in period 3 regardless of the walking stock. Thus, overall compliance is as follows:

$$\text{Days walked} = \begin{cases} 2 & \text{if } e_1 + \delta e_2 < 2m \text{ AND } e_2 < 2m \\ 0 & \text{otherwise.} \end{cases}$$

This is again decreasing in δ.

Case 3: $e_3 < m$ In this case, walking in period 3 is

$$\begin{aligned} w_3(2) &= 1 \\ w_3(1) &= 1 \\ w_3(0) &= 0. \end{aligned}$$

There are two cases to consider for e_2:

1. $e_2 < m$: in this case (for $\delta \leq 1$), discount rates do not matter since the person will walk regardless in periods 2 and 3. Then they walk in period 1 if $e_1 < m$.
2. $e_2 \geq m$: in this case, the person will not walk in period 2 with walking stock 1. Thus, the maximum the person will ever walk is two periods (the first or the second and then the third).

$$\text{Days walked} = \begin{cases} 2 & \text{if } (e_1 + \delta^2 e_3 < 2m \;\&\; e_3 < 2m) \text{ or } (e_2 + \delta e_3 < 2m \;\&\; e_3 < 2m) \\ 0 & \text{otherwise.} \end{cases}$$

Thus days walked is again weakly decreasing in δ.

Thus, the proposition (that compliance is weakly decreasing in δ) is proved since we have shown it holds for all potential values of e_3. □

F CTB Time Preference Measurement

We adapted the convex time budget (CTB) methodology of Andreoni and Sprenger (2012) to try to measure time preferences in two domains, walking and mobile recharges. Here we discuss why we believe our measurement was not a reliable measure of time preferences in this setting.

First, the methodology is complicated. It was difficult to explain to our sample, who had limited familiarity with screens, sliders, or complicated exercises. Due to survey length constraints, we also included fewer questions (and gave less practice) than previous laboratory studies.

A number of patterns suggest that participant understanding was limited. First, law of demand violations are far more common than in previous studies.[45] As shown in Table F.3, 57% of the sample violated the law of demand at least once. For reference, participants in the Augenblick et al. (2015) had 16 opportunities to violate monotonicity, while ours had just 2. If understanding were similar in both contexts one would expect a higher share of the Augenblick et al. (2015) sample to ever violate the law of demand, but the share in their sample was only 16%.

Second, in the effort task, there was low follow-through on the incentivized activity: fewer than 50% of participants selected to complete the step task did so despite large rewards (500 INR) for completion. While this partly reflects a logistical glitch (we failed to give respondents intended

[45] We can only examine law of demand violations in the effort domain because we did not include exchange rate variation in the recharge domain, so cannot estimate the demand curve.

Appendix Table F.3: Law of demand violations in effort allocations

	# of violators	% of sample
	(1)	(2)
Violates 0/7	1,337	41.4
Violates 7/14	1,515	46.9
Violates at least once	1,830	56.7
Violates both	1,022	31.6
Total:	3,232	100

Notes: Violators allocate more steps to the future date when we increase the interest rate from 1 to 1.25. We varied the exchange rate for two questions: today versus 7 days from now, and 7 versus 14 days; rows 1 and 2 show violations for these two questions separately and row 3 and 4 show percentages of people who violated at least once or both.

reminder calls the day before their activity), the lack of follow-through may also indicate a lack of respondent understanding. Regardless, the poor follow-through is problematic methodologically: identification requires that, when participants make their allocation decisions, they think they will follow-through with certainty, which seems unrealistic given how few followed through in practice.

Third, respondents on average allocated more steps to today than the future even when the interest rate was 1:1. Although they could be future-biased, the following other potential explanations are concerning for interpretation: respondents were confused; they saw steps as consumption instead of a cost (violating the first order conditions underlying estimation); or uncertainty over future walking costs and schedules led participants to want to finish steps sooner, which would confound discount rate estimates with risk aversion and uncertainty.

Fourth, day-specific-shocks appear to be important in practice. 19% of respondents' allocations of steps to the sooner date are neither monotonically weakly increasing nor monotonically weakly decreasing across questions which feature the same sooner date (today) but a monotonically decreasing later date (questions 2-6). These allocations cannot be rationalized with a discount rate that is either weakly decreasing *or* increasing with lag length without day-specific utility shocks. The same holds for 24% of respondents in the recharge domain. These types of shocks would also confound estimation.

Fifth, the CTB parameter estimates themselves are not robust and are inconsistent with typical priors. First, we do not have estimates for a large, endogenous share of the sample. The estimates do not converge (i.e., we are unable to estimate discount rate parameters) for 38 to 44% of the sample in the recharge domain, and 23 to 44% of the sample in the steps domain. Moreover, many of the participants with estimates that converge in the effort domain have an estimated $\alpha < 1$, which violates the first order conditions for estimation and is often associated with non-sensible δ and β estimates. When we exclude these estimates, we are left with estimates for only 34 to 38% of the sample in the effort domain. Second, we have a high rate of negative estimated discount rates: 26% for steps and 30% for recharges. This is more that the usual rate of negative individual-level estimates.

Finally, the CTB estimates do not correlate with any of the behaviors one would expect them to. The CTB estimates in the steps/effort domain do not correlate with exercise and health, and the estimates in the recharge domain do not correlate consistently with our proxies for impatience over recharges (e.g., balances). For all of these reasons, we do not think our CTB estimates are a reliable measure of discount rates in this setting and do not use them for analysis.

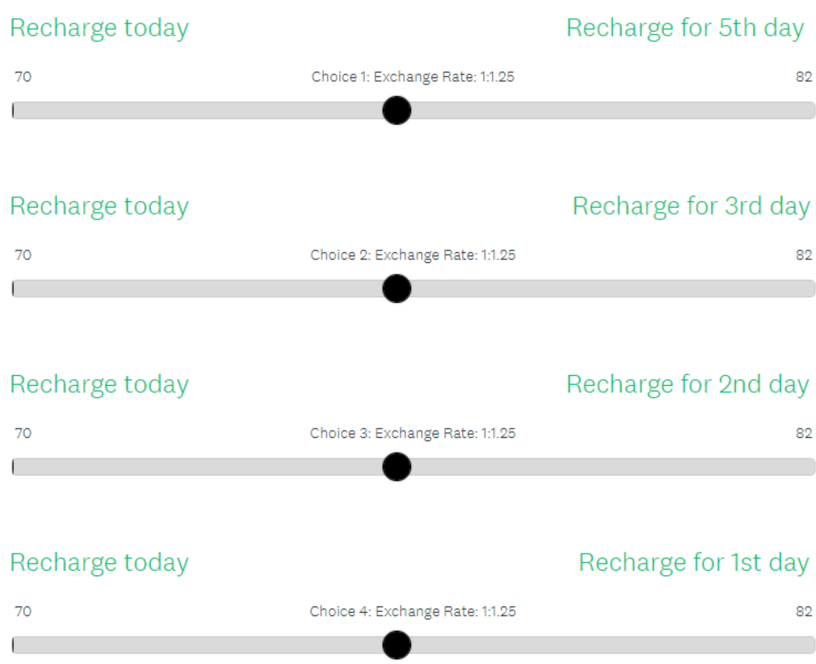
To try to measure time preferences in two domains, walking and mobile recharges, we adapted the convex time budget (CTB) methodology of Andreoni and Sprenger (2012). Below we discuss the methodology used and then show that the measures do not correlate with the behaviors that we would expect.

F.1 Estimation Methodology

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

$$c_t + \frac{1}{r} c_{t+k} = m \tag{35}$$

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



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

We asked participants to make six allocations in the recharge domain, and eight allocations in the step domain, as summarized in Table F.4. We assume a time-separable and good-separable CRRA utility function with quasi-hyperbolic discounting. In the domain of recharges, individuals will then seek to maximize utility,

$$U\left(c_t, c_{t+k}\right) = \frac{1}{\alpha}\left(c_t - \omega\right)^{\alpha} + \beta\delta^k \frac{1}{\alpha}\left(c_{t+k} - \omega\right)^{\alpha} \tag{36}$$

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

$$C\left(c_t, c_{t+k}\right) = \frac{1}{\alpha}\left(c_t + \omega\right)^{\alpha} + \beta\delta^k \frac{1}{\alpha}\left(c_{t+k} + \omega\right)^{\alpha} \tag{37}$$

The variation in consumption choices as the budget constraint varies identify the time preference parameters – in particular, the daily discount factor δ and the present-bias parameter β – as well as the concavity or convexity of preferences α. Due to budget and time constraints, we had to keep the module short and so did not implement interest rate variation for the recharge tradeoffs, only for the step tradeoffs. Thus α is identified for the effort estimation only, not the recharge one; for the recharge estimation, we calibrate α using the estimate of α from Augenblick et al. (2015) in the financial payment domain.

We recover individual-level structural estimates of time preference and concavity parameters from the allocations, (c_t, c_{t+k}), using a two-limit Tobit specification of the intertemporal Euler condition following Augenblick et al. (2015).

$$\log\left(\frac{c_t + \omega}{c_{t+k} + \omega}\right) = \frac{\log(\beta)}{\alpha - 1} 1_{t=0} + \frac{\log(\delta)}{\alpha - 1} k - \frac{1}{\alpha - 1}\log\left(r\right) \tag{38}$$

Details on the estimation strategy can be found in the Online Appendix of Augenblick et al. (2015). Because our predictions concern overall impatience, not whether an individual is time-consistent, on the time preference side we want one single summary measure capturing impatience. To do so, we estimate two different variants. In one, we set $\beta = 1$ for everyone at the estimation stage and simply estimate δ at the individual level. In the second, we estimate the equation as above, allowing both β and δ to vary at the individual level, and use $\beta \times \delta$ as our measure of individual-level impatience. In both estimation procedures, we allow α to vary at the individual-level in the steps domain, since we considered individual-level convexity of the step function to be an important potential confound.[46] However, the results we describe next are similar if we do not allow α to vary at the individual-level for steps.

Our CTB environment builds on a number of features from previous studies. First, the choices are made after the one-week phase-in period in which all participants have pedometers and report their daily steps, ensuring that participants are familiar with the costs of walking. This allows for meaningful allocations of steps between sooner and later dates. Second, the responses are designed to be incentive compatible; all respondents were informed that we would implement their choice from a randomly selected survey question. We set the probabilities such that for most respondents the randomly selected survey question was a multiple price list of lotteries over money (which measures risk preferences), but for a few a CTB allocation was selected. Because the allocations might have interfered with any walking

[46]Indeed, when we estimate impatience (e.g., δ) but do not allow α to vary, that estimated δ correlates as strongly with α as it does with the δ estimated allowing α to vary, suggesting that convexity is an important confound indeed.

Appendix Table F.4: CTB allocation parameters

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

Notes: This table summarizes the parameters of the six CTB allocations made over recharges, and the eight CTB allocations made over steps.

program offered, we excluded the 40 respondents who were randomly selected to receive one of their allocations from the experimental sample.[47] To try to ensure that participants complete the allocated steps, we offer a large cash completion bonus of 500 INR in the step domain if the allocation is selected to be implemented, and the steps are completed as allocated, with the bonus to be delivered 15 days from the date of the survey (which is 1 day after the latest "later" day used).

We also take a number of precautions to avoid various potential confounds, including confounds reflecting fixed costs or benefits of taking an action, or confounds due to the time of day of measurement.[48] However, we were not able to fully address one potential

[47] This means we have CTB data from a total of 3,232 people: the 3,192 in the experimental sample plus the 40 selected to receive "real-stakes" allocations. For completeness, we summarize in this section the CTB data for all 3,232 but the results are the same if we restrict to the experimental sample.

[48] To avoid confounds related to fixed costs or benefits, such as the effort of wearing a pedometer or the psychological benefit of receiving a free recharge, we include minimum allocations on both sooner and later days in each domain. The minimum allocations were chosen to be high enough that any fixed costs would be included (e.g. one could not easily achieve the minimums by simply shaking the pedometer) but low enough to avoid corner solutions. In the step domain, this required a modification of the CTB methodology: individual-specific minimum allocations. Our step allocations also featured individual-specific total step budgets m, which were chosen to be large enough that achieving them would require some effort beyond simply wearing the pedometer but small enough that participants would certainly achieve them in exchange for the completion bonus. Specifically, minimum steps on each day are calculated as $\frac{X}{10}$, and the total step budget m is $X + 2\frac{X}{10}$, respectively, where $X \in \{3000, 4000, 5000\}$ is the element closest to the participant's average daily walking during the phase-in period. That is, minimum steps are one of 300, 400, or 500 on each day, and the total step budget is one of 3,600, 4,800, or 6,000. To avoid confounding impatience with the time of day that the baseline time-preference survey was administered (which could influence the desirability of walking and/or recharges delivered in the next 24 hours), as well as to capture heterogeneity in time preferences including any present-bias for very short beta-windows, we required that all walking on any date be conducted within a 2 hour period, which was chosen to start at the time immediately after the time-preference survey would end (e.g., if the survey ended at 4pm, the time period for any day's walking would be 5-7pm). The short window could potentially bias our overall measures of impatience downwards, as uncertainty about future schedules in a short time window could lead participants to want to get their walking done early when they had more certainty over their schedule. However, our primary purpose was to capture heterogeneity in time-preferences, and we considered the potential loss in validity of aggregate time preference estimates to be worth the ability to capture heterogeneity in time preferences in the time frames

Appendix Table F.5: Summary Statistics For CTB Parameters

	Full sample		$\alpha < 1$	
Parameters estimated:	$\beta\delta$	δ	$\beta\delta$	δ
	(1)	(2)	(3)	(4)
A. Effort				
Beta	2.066	–	1.573	–
Delta	0.883	0.997	1.015	0.999
Alpha	0.244	0.723	1.673	1.576
% of sample:	77.2	56.3	33.8	37.9
# Individuals:	2,494	1,821	1,092	1,225
B. Recharges				
Beta	1.004	–	–	–
Delta	0.997	0.997	–	–
% of sample:	55.9	62.2	–	–
# Individuals:	1,808	2,010	–	–

Notes: This table displays means and convergence rates of individual-level CTB parameters in both the effort and recharge domains. Columns 1 - 2 display average values for the parameters from the full sample of individuals with parameters that converged. In the effort domain, in columns 3 - 4, we ignore all individuals whose estimated *alpha* was below 1, as handled similarly in Andreoni and Sprenger (2012), as that is inconsistent with the first order conditions. We winsorize all parameters at the top and bottom 1 percentiles. We allow α to vary at the individual level in the effort domain, and in the recharge domain, we calibrate α to be 0.975, which is the estimated value in Augenblick et al. (2015). Delta is estimated by allowing δ to vary at the individual level and setting β to 1. Beta-delta is estimating by allowing both δ and β to vary. We derive these two parameters from an estimation that allows δ and β to vary at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

confound to our estimates of time-preferences across individuals: variation across people in the cost of walking over time, or in the benefit of receiving a recharge over time. For example, an individual with a particularly busy week after the time-preference survey, and therefore relatively high costs to steps in the near-term relative to the distant future, will appear to be particularly impatient over steps in our data (he will wish to put off walking). An individual with a relatively free week just after the time-preference survey will instead appear particularly forward-looking (he will not wish to put off walking). The same concerns can also arise with recharges.

F.2 CTB Estimates

Table F.5 displays the summary statistics as well as the convergence statistics diussed in more detail in the upfront text.

Tables F.6 and F.7 show that the estimated CTB parameters do not correlate in the

near to the present.

expected direction with measured behaviors. In particular, Table F.6 shows that the CTB estimates in the steps/effort domain do not correlate with exercise and health,[49] and Table F.7 shows that the estimates in the recharge domain do not correlate with recharge balances, usage, or credit constraint proxies. The CTB measures do correlate at the 10% level with our measure of marginal propensity to consume recharges, but the correlations go in opposite directions for the two CTB measures (δ from an estimation setting $\beta = 1$ vs. $\beta\delta$ estimated allowing both parameters to vary) so is likely noise.

[49] Table F.6 shows the correlations when we exclude the effort estimates from participants with estimated $\alpha < 1$, but the results are similar when we include all estimates together.

Appendix Table F.6: CTB Estimates of Discount Factors Over Steps Do Not Correlate with Measured Behaviors

Covariate type:	Exercise		Baseline indices			
Dependent variable:	Daily steps	Daily exercise (min)	Health index	Negative vices index	Healthy diet index	# Individuals
Delta	-0.019	0.009	-0.040	0.010	0.027	1,225
Beta-delta	0.016	0.018	0.014	0.010	0.027	1,092

Notes: This table displays the correlations between CTB parameters in the effort domain and a few baseline health covariates. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize health outcomes so that higher values correspond to healthier outcomes. All CTB parameters have been winsorized at the top and bottom 1 percentile to remove outliers. Delta is measured from an estimation that allows δ and α to vary at the individual level, while excluding β. Beta-delta is a measure of beta times the average delta over one week. We estimate the two parameters by allowing β, δ, and α to vary at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.7: CTB Estimates of Discount Factors Over Recharges Do Not Correlate with Other Proxies for Impatience over Recharges

Covariate type:	Recharge variables			Credit constraint proxies			
Dependent variable:	Negative mobile balance	Negative yesterday's talk time	Marginal talk time, if gifted	Wealth index	Monthly household income	Monthly personal income	# Individuals
Delta	0.026	0.008	-0.042*	-0.016	0.002	0.008	1,892
Beta-delta	-0.012	-0.004	0.044*	-0.001	0.018	-0.012	1,701

Notes: This table displays the correlations between CTB parameters in the recharge domain and baseline measures that should be related to credit constraints and discount rates over recharges. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize the proxies so that higher values correspond to higher expected discount rates; hence the prediction is that coefficients should be positive. All CTB parameters have been winsorized at the top and bottom 1 percentile to remove outliers. We use two main estimation specifications, and to identify parameters, we calibrate α to be 0.975, the value of α estimated in Augenblick et al. (2015). Delta is estimated by allowing δ to vary at the individual level and excluding β. Beta-delta is a measure of the average delta over one week multiplied by beta. We derive these two parameters from an estimation that allows δ and β to vary at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

G Heterogeneity in Frequency Effects by (Proxies for) Impatience over Payment

This appendix explores the possibility that immediate incentive delivery is a driver of incentive effectiveness among the subset of participants who are more impatient over payment. If so, we expect a positive interaction between more immediate incentive delivery and our (imperfect) proxies for 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:

$$\begin{aligned} y_{it} = &\alpha + \beta_0 Impatience_i + \beta_1 daily_i + \beta_2 monthly_i \\ &+ \beta_3 Impatience_i \times daily_i + \beta_4 Impatience_i \times monthly_i + \boldsymbol{X}_i' \gamma + \varepsilon_{it}, \end{aligned} \tag{39}$$

where y_{it} is a daily walking outcome; $Impatience_i$ is one of our three proxies for impatience in the recharge domain (the negative of someone's baseline balances, the negative of someone's baseline daily recharge usage, or a measure of constraints), all normalized so that higher values proxy for higher levels of impatience; and $daily_i$ and $monthly_i$ are indicators for being assigned to the daily and monthly treatments, respectively. β_1 and β_2 represent the effects of daily and monthly relative to the base case weekly payment (respectively). The coefficients of interest are β_3 and β_4, showing whether the effects of daily or monthly relative to Weekly are differentially large for those who are more impatient. If impatience over recharges is a mechanism through which more immediate incentive delivery increases effectiveness, then we expect the daily treatment to be more effective ($\beta_3 > 0$) and the monthly treatment to be less effective ($\beta_4 < 0$) for more impatient individuals.

We report our results in Table J.3. We find no evidence that suggests that sooner payments work better for those we would expect to be more impatient, with no clear pattern across measures and the one result significant at the 5% level going the wrong way.

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 incentive 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 payments. We estimate the interaction between individual payday effects and our baseline proxies for impatience over recharges using regressions of the following form:

$$\begin{aligned} y_{it} = &\alpha + \beta_0 \left(Impatience\,Measure\right)_i + \beta_1 \left(Payday\right)_{it} \\ &+ \beta_2 \left(Payday\right)_{it} \times \left(Impatience\,Measure\right)_i + \boldsymbol{X}_i' \gamma + \varepsilon_{it}, \end{aligned} \tag{40}$$

where y_i, $(Impatience\,Measure)$, and $\boldsymbol{X}_i$ are defined as in equation 39; and $(Payday)_{it}$ is an indicator for whether day t is a payday for individual i. To test whether more impatient individuals respond more to more immediate payment, we test whether $\beta_2 > 0$.

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

Dependent variable:	Met step target		
Impatience measure:	Negative mobile balance	Negative yesterday's usage	Marginal talk time, if gifted
	(1)	(2)	(3)
Daily × Impatience	-0.055**	-0.046	-0.0093
	[0.026]	[0.033]	[0.016]
Monthly × Impatience	-0.034	-0.055**	-0.022
	[0.021]	[0.026]	[0.014]
Daily	-0.011	0.0034	-0.0060
	[0.026]	[0.028]	[0.028]
Monthly	-0.031	-0.030	-0.038
	[0.025]	[0.025]	[0.026]
Impatience	-0.0052	-0.043*	0.011
	[0.012]	[0.026]	[0.010]
Base case mean	0.50	0.50	0.51
# Individuals	2,558	2,450	2,388

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. We use three mobile recharge variables collected at baseline as proxies for impatience over recharges: mobile balance, yesterday's talk time in INR, and unconstrained recharge usage if we were to gift individuals recharges. Variables are normalized by the standard deviations of the control group. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize the proxies so that higher values correspond to higher expected discount rates. 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 are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

Our results are shown in Table G.2. We find no evidence that even those individuals who are most impatient over payments react to more immediate reward delivery over the payment cycle.

Appendix Table G.2: Payday effects are not bigger for those who are more impatient

Dependent variable:	Met step target		
Impatience measure:	Negative mobile balance	Negative yesterday's usage	Marginal talk time, if gifted
	(1)	(2)	(3)
Impatience × Payday	0.0058	0.0075	0.0021
	[0.0050]	[0.0067]	[0.0034]
Payday	0.034	0.029	0.042*
	[0.025]	[0.025]	[0.025]
Impatience	-0.0048	0.017	-0.013**
	[0.0087]	[0.011]	[0.0064]
Base case mean	0.50	0.50	0.51
# Base case	890	845	826
# Monthly	163	160	155
# Individuals	1,053	1,005	981

Notes: This table shows heterogeneity in the "payday" effects for those in the base case incentive and the monthly incentive groups, by proxies for impatience over recharges. Our proxies include baseline measures for mobile balance, yesterday's usage, and unconstrained usage, which is a self-reported estimate of usage in INR if recharges were gifted. Impatience measures are normalized by the standard deviations of the control group, and such that higher values correspond to greater impatience. We normalize the proxies so that higher values correspond to higher expected discount rates. Payday effects are defined as the difference in a daily exercise behavior on paydays compared to all other days. Standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

H Program Evaluation: Appendices and Additional Analysis

Appendix Table H.1: Impacts of incentive program and monitoring on diet and addictive consumption.

A. Healthy diet	Healthy diet index	Wheat meals	Meals with vegeta-bles	Servings of fruit	Negative of rice meals	Negative of junkfood pieces	Negative of spoons sugar in coffee	Negative of sweets yester-day)	Avoid un-healthy food
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Incentives	0.046	0.023	0.058*	0.040	0.026	-0.026	-0.024	-0.025	0.0045
	[0.044]	[0.029]	[0.030]	[0.035]	[0.033]	[0.066]	[0.047]	[0.038]	[0.018]
Monitoring	0.018	0.011	0.076	0.063	-0.010	0.13	-0.030	-0.039	-0.040
	[0.084]	[0.053]	[0.054]	[0.061]	[0.060]	[0.10]	[0.081]	[0.082]	[0.033]
Control mean	0.00	0.49	0.58	0.53	-2.34	-0.91	-1.12	-0.35	0.83
P-value: M = I	0.71	0.81	0.70	0.68	0.50	0.08	0.94	0.86	0.13
# Individuals	3,068	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	-0.013	0.034	-0.034	-0.061
	[0.038]	[0.038]	[0.027]	[0.095]
Monitoring	-0.00021	0.019	-0.014	-0.027
	[0.060]	[0.069]	[0.037]	[0.14]
Control mean	0.00	0.13	0.11	1.02
P-value: M = I	0.81	0.81	0.46	0.79
# Individuals	3,068	3,068	3,068	3,068

Notes: 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. For the two indices, controls are the same as Table 2, along with second order polynomials of all questions underlying the indices at baseline. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

H.1 SMS Treatment Impacts

We here present the effects of the SMS treatment, which we included in the experiment to appease our government partners who were interested in its efficacy. We estimate regressions of the following form, using the same outcome variables as in Table 2 and Section 6:

$$y_i = \alpha + \beta_1 \times incentives_i + \beta_2 \times monitoring_i + \beta_3 \times SMS_i + \boldsymbol{X}_i'\gamma + \varepsilon_i \tag{41}$$

where y_i is a health or lifestyle outcome at endline for individual i; $incentives_i$ is an indicator for being in the incentive group; $monitoring_i$ is an indicator for being in the monitoring group; SMS is an indicator for being in the SMS treatment; and $\boldsymbol{X}_i$ is a vector of controls, shown in the table notes. β_1 is the ITT effect of incentives relative to the pure control group, β_2 is the ITT effect of the monitoring relative to the pure control group, and β_3 is the ITT effect of being in the SMS reminders group. Note that, when the outcome variable represents exercise measured using pedometer data, we omit the $monitoring_i$ dummy and the reference group becomes the monitoring group (since that data is unavailable for the control group).

Table H.2 contains the walking impacts of the SMS treatment and Table H.3 contains the health impacts. We do not see any significant effects on either. While the coefficients for exercise are near 0, the coefficient for health impacts is non-trivial in magnitude and, if anything, positive; recall that here positive coefficients are associated with *worse* health. Although this coefficient is not significant and likely reflects statistical noise, to probe further, in Table H.4, we estimate a model including all of the interaction effects between the SMS treatment and the monitoring or incentive treatments. Although these estimates should be interpreted as suggestive since (a) we did not plan to run this specification ex ante, and (b) the interaction effects are only marginally significant, it appears that one reason for the SMS Treatment's marginally negative average effect may be that the SMS treatment does not interact well with incentives.

Appendix Table H.2: Impacts of SMS treatments and 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	0.200***	1266.0***	1161.5***
	[0.0186]	[208.7]	[188.5]
SMS	-0.000795	123.1	46.96
	[0.0209]	[208.2]	[182.4]
B. Unpooled incentives			
Base case	0.211***	1388.4***	1203.1***
	[0.0201]	[222.1]	[199.9]
Daily	0.201***	1122.5***	1283.1***
	[0.0303]	[331.5]	[277.9]
Monthly	0.177***	1274.2***	1179.4***
	[0.0288]	[307.4]	[271.1]
Threshold	0.198***	1216.3***	1142.6***
	[0.0199]	[220.9]	[198.5]
Small payment	0.137***	731.5*	552.9*
	[0.0383]	[386.2]	[335.0]
SMS	-0.00241	112.6	34.79
	[0.0209]	[208.0]	[182.4]
Monitoring mean			
Controls	Yes	Yes	Yes
P-value for Base case vs			
Daily	0.71	0.35	0.73
Monthly	0.18	0.65	0.91
Threshold	0.36	0.21	0.61
Small payment	0.04	0.06	0.03
# 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 sample includes the incentive and monitoring groups. Controls are the same as Table 2. The omitted category in all columns is the monitoring group. Standard errors, in brackets, are clustered at the individual level. The Threshold group pools the 4- and 5-day hreshold groups. A number of individuals in the incentive and monitoring groups immediately withdrew from the contract period, but there was no statistically significant difference in likeliness to withdraw by group (p-val > 0.7).

Appendix Table H.3: Impacts of incentives, monitoring, and SMS treatments on health

	Health risk index	HbA1c	Random blood sugar	Mean arterial BP	Body mass index	Waist circum-ference
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	-0.045*	-0.072	-5.67*	0.081	-0.049	-0.18
	[0.025]	[0.070]	[3.42]	[0.42]	[0.042]	[0.27]
Monitoring	0.014	-0.13	1.63	1.08	0.064	0.00080
	[0.044]	[0.12]	[6.07]	[0.75]	[0.074]	[0.48]
SMS treatment	0.043	0.12	-4.05	0.66	0.099*	0.43
	[0.032]	[0.090]	[4.41]	[0.55]	[0.054]	[0.35]
Control mean	0.00	8.44	193.83	103.02	26.45	94.44
# Control	561	560	561	560	559	559
P-value: M = I	0.13	0.57	0.18	0.14	0.09	0.67
P-value: SMS = I	0.03	0.10	0.77	0.40	0.03	0.16
P-value: SMS = M	0.59	0.10	0.45	0.65	0.70	0.47
# Individuals	3,063	3,061	3,062	3,051	3,053	3,054

Notes: Standard errors in brackets. For the Health Risk Index, controls are the same as Table 2, along with second order polynomials of all health variables underlying the index at baseline. Controls for all other outcomes are the same as Table 2. The Health Risk Index is an index created by the average of endline Hba1c, RBS, MAP, BMI, and waist circumference standardized by their average and standard deviation in the control group. Hba1c is the average plasma glucose concentration (%), RBS is the blood glucose level (mg/dL), MAP is the mean arterial blood pressure (mm Hg), and BMI is the body mass index. The omitted category in all columns is the pure control group. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table H.4: Impacts of incentives, monitoring, SMS treatments, and their interactions on 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	-0.061**	-0.13*	-7.72**	0.12	-0.056	-0.20
	[0.026]	[0.074]	[3.61]	[0.45]	[0.044]	[0.29]
Monitoring	0.018	-0.16	2.27	1.04	0.084	0.068
	[0.047]	[0.13]	[6.39]	[0.79]	[0.078]	[0.50]
SMS treatment	-0.068	-0.34*	-18.5*	0.93	0.064	0.32
	[0.073]	[0.20]	[10.0]	[1.24]	[0.12]	[0.79]
Incentives × SMS	0.15*	0.60***	20.0*	-0.40	0.065	0.20
	[0.082]	[0.23]	[11.2]	[1.39]	[0.14]	[0.88]
Monitoring × SMS	-0.040	0.21	-7.47	0.39	-0.20	-0.66
	[0.15]	[0.41]	[20.2]	[2.46]	[0.24]	[1.57]
Control mean	0.00	8.44	193.83	103.02	26.45	94.44
# Control	561	560	561	560	559	559
P-value: M = I	0.06	0.84	0.08	0.19	0.05	0.55
P-value: SMS = I	0.92	0.29	0.26	0.49	0.31	0.49
P-value: SMS = M	0.28	0.41	0.06	0.94	0.88	0.77
# Individuals	3,063	3,061	3,062	3,051	3,053	3,054

Notes: Standard errors in brackets. All definitions and controls follow Table 2. The omitted category in all columns is the pure control group. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table H.5: Impact of incentives on fitness and mental health

A. 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	0.097**	0.090**	0.027	0.058	0.065	0.016	0.089**	0.054*
	[0.045]	[0.045]	[0.045]	[0.047]	[0.047]	[0.044]	[0.042]	[0.033]
Monitoring	0.16**	0.075	0.12	0.095	0.037	0.12*	0.17**	0.051
	[0.073]	[0.075]	[0.077]	[0.083]	[0.081]	[0.075]	[0.066]	[0.053]
Control mean	0.00	3.06	3.48	3.35	3.30	3.86	4.40	4.71
P-value: M = I	0.33	0.82	0.15	0.61	0.70	0.11	0.15	0.95
# Individuals	3,068	3,068	3,068	3,068	3,068	3,068	3,068	3,068

B. Fitness	Fitness time trial index	Seconds to walk 4m	Seconds for 5 sit-stands
	(1)	(2)	(3)
Incentives	0.013	0.033	-0.10
	[0.044]	[0.041]	[0.12]
Monitoring	0.056	0.071	-0.082
	[0.074]	[0.072]	[0.19]
Control mean	0.00	3.88	13.18
P-value: M = I	0.52	0.54	0.90
# Individuals	2,890	2,825	2,793

Notes: The Mental health index averages the values of seven questions adapted from RAND's 36-Item Short Form Survey (SF-36). A large value of the Fitness time trial index indicates low fitness. The omitted category is the pure control group. Controls are the same as Table 2, along with second order polynomials of the dependent variabe at baseline. Robust standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table H.6: Lee bounds on the impacts of incentives on exercise: Post-intervention period

Definition of missing:	No steps data	No data from fitbit	Did not wear fitbit	Lost data entire period	Withdrew immedi-ately	Mid-period withdrawal	Other reasons
A. Daily steps							
Regression estimate	765	471	765	471	471	471	471
(conditional on nonmissing data)	[238]	[246]	[238]	[246]	[246]	[246]	[246]
Lee lower bound	731	366	689	459	448	304	459
	[58]	[61]	[55]	[38]	[40]	[40]	[38]
Lee upper bound	840	503	934	515	554	522	515
	[99]	[41]	[86]	[39]	[57]	[39]	[39]
B. Met 10k step target							
Regression estimate	0.102	0.068	0.102	0.068	0.068	0.068	0.068
(conditional on nonmissing data)	[0.020]	[0.016]	[0.020]	[0.016]	[0.016]	[0.016]	[0.016]
Lee lower bound	0.101	0.063	0.099	0.067	0.067	0.060	0.067
	[0.004]	[0.004]	[0.004]	[0.003]	[0.003]	[0.003]	[0.003]
Lee upper bound	0.105	0.069	0.110	0.070	0.072	0.070	0.070
	[0.006]	[0.003]	[0.006]	[0.003]	[0.004]	[0.003]	[0.003]
# Individuals	1,122	1,122	1,122	1,122	1,122	1,122	1,122
# Observations	62,858	91,756	62,858	91,756	91,756	91,756	91,756

Notes: Table reports regression estimates and Lee bounds accounting for different types of missing pedometer data in the post-intervention period. The regression estimates condition on data not being missing, using different definitions of missing data in each column, and then the Lee bounds are estimated again allowing the definition of missing data to vary by column. Panel A reports results using average daily steps as the dependent variable, and Panel B reports results using proportion of days met 10k step target as the dependent variable. The omitted category is the monitoring group. The number of observations is reported for the Lee bounds regressions. Note that regression estimates reported in columns 1 - 3 are not comparable to those reported in Table 2 because each column conditions on the “type of missing” indicator in the first row being equal to 0 and does not include controls. 132 people have no data during the period. The most common reason for this was immediate withdrawal.

Appendix Table H.7: Impacts of Incentives on Walking, Without Baseline Controls.

Dependent variable:	Compliance	Daily steps	Daily steps (if > 0)
	(1)	(2)	(3)
A. Pooled incentives			
Incentives	0.21***	1337.6***	1271.4***
	[0.022]	[261.1]	[246.1]
B. Unpooled incentives			
Base case	0.21***	1356.6***	1208.8***
	[0.024]	[277.0]	[258.6]
Daily	0.21***	1202.7***	1363.9***
	[0.034]	[389.5]	[346.0]
Monthly	0.20***	1568.6***	1482.3***
	[0.035]	[393.8]	[365.4]
Threshold	0.21***	1337.9***	1315.2***
	[0.024]	[277.1]	[259.3]
Small payment	0.15***	820.5	658.5
	[0.049]	[524.0]	[477.9]
# Individuals	2,559	2,559	2,557
Observations	205,732	205,732	180,018

Notes: This table replicates the Table 2 estimates without including the baseline controls. The Threshold group pools the 4- and 5-day threshold groups. Significance levels: * 10%, ** 5%, *** 1%.

I Monitoring treatment impacts on walking

The health results suggest that the monitoring treatment 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? We now present an analysis of the effects of monitoring on exercise. Because we do not have pedometer walking data from the control group, we use a before-after design. We find that monitoring alone has limited impact on overall steps. Monitoring does however change the distribution of steps, increasing the share of days on which participants met the 10,000 step target but decreasing the steps taken on other days for a null effect on total exercise.

Our before-after design compares 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 affects 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 11, panels A and B. Walking increases immediately during the intervention period for the monitoring group, although the effects decay over time.

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

$$\begin{aligned} y_{it} = {} & \alpha + \beta_1 Intervention\,Period_{it} + \beta_2 incentives_i + \beta_3 (Intervention\,Period_{it} \times incentives_i) \\ & + \boldsymbol{X}_i' \gamma + \boldsymbol{\mu}_m + \varepsilon_{it}, \end{aligned} \tag{42}$$

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

Table I.1 presents the results. Column 2 shows that 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 36% of the estimated impact of incentives. In contrast, the estimated effect on steps is very small in magnitude, varies across specifications, and is in fact sometimes negative (columns 4-6).

Thus, the monitoring treatment, if anything, appears to do more to make walking consistent across days than it does to increase total steps.

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

	Achieved 10K steps			Daily steps		
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	0.012	0.013	0.012	66.7	66.4	48.9
	[0.024]	[0.024]	[0.014]	[268.1]	[266.9]	[112.3]
Intervention period	0.057***	0.073***	0.064***	-130.4	108.0	-18.5
	[0.020]	[0.020]	[0.020]	[237.8]	[240.8]	[234.1]
Intervention period X Incentives	0.19***	0.19***	0.19***	1270.9***	1258.9***	1212.7***
	[0.021]	[0.021]	[0.021]	[248.6]	[249.2]	[243.4]
Monitoring phase-in mean	0.24	0.24	0.24	6904.80	6904.80	6904.80
Year-month FEs	No	Yes	Yes	No	Yes	Yes
Individual controls	No	No	Yes	No	No	Yes
# Monitoring	203	203	203	203	203	203
# Incentives	2,401	2,401	2,401	2,401	2,401	2,401
# Individuals	2,604	2,604	2,604	2,604	2,604	2,604
Observations	221,214	221,214	221,214	221,214	221,214	221,214

Notes: This table shows coefficient estimates from regressions of the form specified in Equation 42. The outcomes are from daily panel data from the pedometers. Standard errors, in brackets, are clustered at the individual level. Individual controls are the same as Table 2. The omitted category in all columns is the monitoring group in the phase-in period. The coefficient in the second row, on $Intervention\,Period_{it}$, corresponds to the pre-post estimate of the monitoring effect. Significance levels: * 10%, ** 5%, *** 1%.

J Lifestyle Modification

ABDUL LATIF JAMEEL
Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

IFMR LEAD
ABDUL LATIF JAMEEL POVERTY ACTION LAB

Lifestyle Modification Guidelines

The Ministry of Health and Family Welfare, Government of Tamil Nadu recommends that all hypertensives, diabetics *and* people at risk for those diseases follow the life style modifications below. These modifications will help to maintain blood pressure, blood sugar, and body weight at healthy and controlled levels. This is essential for the prevention and management of diabetes and hypertension.

PHYSICAL ACTIVITY

- Do moderate physical activity, such as brisk walking for 30-40 minutes per day at least 4 times every week, but preferably daily. Brisk walking is walking at a pace where you find speaking difficult, but not impossible.
- Do some sort of physical exercise every day.
- Try to reach a total of 10,000 steps, or about 8km of walking, per day.
- In order to increase your level of physical activity, perform all your household chores such as cleaning, washing clothes, gardening etc. yourself.

MAINTAIN AN IDEAL WEIGHT FOR HEIGHT

- Avoid being overweight for your height.
- Being overweight increases the risks of diabetes, hypertension, heart attacks, stroke, and paralysis.

HEALTHY DIET

- Reduce your consumption of sugar, salt, cooking oil, and dairy fat.
- If you are a non-vegetarian, replace fatty meats with skinless chicken and fish.
- Increase your consumption of fruit and green leafy vegetables. As a general rule, you must have at least 500 grams of fruits and vegetables in a day.
    - However, avoid vegetables like potatoes and tubers.
    - In addition, some diabetics should avoid fruits. If you are diabetic you should check with a medical professional about your fruit intake.
- Stop smoking and avoid alcohol.

DECREASE YOUR STRESS LEVELS

- Physical activities from walking to dancing, listening to and playing music, meditation, and yoga can all reduce stress.

MAINTAIN NORMAL BLOOD PRESSURE AND BLOOD SUGAR

- Uncontrolled blood pressure can lead to cardiovascular diseases such as hypertension, stroke, and heart attack.
- Uncontrolled blood sugar can lead to diabetes.
- Maintaining healthy and controlled blood sugar levels is the best way to prevent complications of diabetes such as foot and nerve damage, loss of eyesight, and kidney failure.
- If you are at risk for NCDs, check your blood sugar and blood pressure at least once a year at an NCD screening location. The nearest location is: Coimbatore Government Hospital [address]

Appendix Figure J.1: Lifestyle Modification Advice Delivered at Baseline

Appendix Table J.2: Replication of Table G.1

Dependent variable:	Exceeded step target		
Impatience measure:	Daily Payment (=1)	Monthly Payment (=-1)	Marginal talk time, if gifted
	(1)	(2)	(3)
Daily × Impatience	0.0070	0.013	-0.0093
	[0.024]	[0.026]	[0.016]
Monthly × Impatience	0.015	0.0049	-0.022
	[0.023]	[0.026]	[0.014]
Daily	-0.013	-0.0083	-0.0060
	[0.027]	[0.027]	[0.028]
Monthly	-0.035	-0.033	-0.038
	[0.025]	[0.025]	[0.026]
Impatience	0.010	0.014	0.011
	[0.017]	[0.015]	[0.010]
Base case mean	0.50	0.50	0.51
# Individuals	2,559	2,559	2,388

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. We use three mobile recharge variables collected at baseline as proxies for impatience over recharges: mobile balance, yesterday's talk time in INR, and unconstrained recharge usage if we were to gift individuals recharges. Variables are normalized by the standard deviations of the control group. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize the proxies so that higher values correspond to higher expected discount rates. 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 are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table J.3: Replication of Table G.2

Dependent variable:	Exceeded step target		
Impatience measure:	Daily Payment (=1)	Monthly Payment (=-1)	Marginal talk time, if gifted
	(1)	(2)	(3)
Impatience × Payday	-0.00029	-0.0048	0.0021
	[0.0052]	[0.0053]	[0.0034]
Payday	0.036	0.035	0.042*
	[0.025]	[0.025]	[0.025]
Impatience	0.022**	0.011	-0.013**
	[0.0094]	[0.0095]	[0.0064]
Base case mean	0.50	0.50	0.51
# Base case	890	890	826
# Monthly	163	163	155
# Individuals	1,053	1,053	981

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. We use three mobile recharge variables collected at baseline as proxies for impatience over recharges: mobile balance, yesterday's talk time in INR, and unconstrained recharge usage if we were to gift individuals recharges. Variables are normalized by the standard deviations of the control group. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize the proxies so that higher values correspond to higher expected discount rates. 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 are in brackets. Significance levels: * 10%, ** 5%, *** 1%.