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 15, 2021

Abstract

Many of the agents whom principals incentivize are impatient. We develop a prediction for how to make incentives work particularly well when agents 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 contrast, we find that increasing the frequency of payment – which should be effective if individuals are impatient over payment (rather than effort) – has no effect, suggesting limited impatience over payments. Overall, the incentive program is effective, increasing daily steps by roughly 20 percent (13 minutes of brisk walking) and improving 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, the MIT Tata Center for Technology and Design, 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, ISB, the University of 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, Varun Satish, and Emily Zhang for outstanding research assistance. We are grateful to Ned Augenblick, Abhijit Banerjee, Marianne Bertrand, Esther Duflo, Pascaline Dupas, Alex Frankel, Rick Hornbeck, Seema Jayachandran, Anett John, Supreet Kaur, Jonathan Kolstad, Ted O'Donoghue, Rohini Pande, Devin Pope, Canice Prendergast, Matthew Rabin, Heather Royer, Gautam Rao, Sheldon Ross, 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. To assess the quantitative importance of adjusting incentives for impatience, we then implement the variations in a randomized controlled trial (RCT) evaluating a promising incentives program to encourage exercise among diabetics and prediabetics in India.

When formulating our predictions, we follow the literature (e.g., Augenblick et al., 2015) and 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., 2018), 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, a first more novel variation whose efficacy increases with the discount rate over effort and a second whose efficacy increases with the discount rate over payments.

Our primary 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 higher 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.

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

One theoretical advantage of many time-bundled contracts is that they should induce extra effort from all types of people with high discount rates over effort, notably including “naïve” 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 impatient, time-inconsistent people. Commitment contracts limit people’s future options or impose penalties on their future selves in order to encourage their future selves to take 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., naïfs) 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 naïfs – to work today to access the better future opportunities.

Our second contract variation is to increase the frequency of payment, motivated by the (less novel) prediction that if individuals are impatient over payments, payment that is more frequent 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 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 equal the market interest rate for individuals with access to borrowing and saving. However, if individuals irrationally ignore financial arbitrage opportunities (Andreoni et al., 2018), or if access to credit and liquidity is limited (Carvalho et al., 2016), the discount rate over payment may approach the discount rate over consumption.

We evaluate time-bundled contracts and payment frequency using an experiment offering incentives for behavior change. Policymakers are increasingly using incentives to encourage behavior changes such as school attendance, exercise, and medication adherence. These behaviors often feature short-run costs and long-run benefits, which can cause present-biased agents to underinvest in them. One of the motivations for using incentives is precisely to mitigate this underinvestment. The use of incentives to address agent present bias makes it

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

[3]In the incentive domain, a commitment contract might pay less than the benchmark contract for low effort but no more for high effort (Kaur et al., 2015). The prediction is that time-inconsistent sophisticates might choose a commitment contract to motivate their future selves. In contrast, our prediction regarding time-bundled contracts is that they will induce more compliance from the impatient *conditional* on selection.

particularly important to understand how to make incentives work well for present-biased agents – an issue on which the evidence is thin.

Our incentives are designed to encourage walking among diabetics and prediabetics, an important policy goal. 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 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 payment discount rate.

The first goal of our experiment was to assess the quantitative importance of our theoretical predictions: we 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). The 6 pp difference represents the gap between a 3 pp positive effect among those with above-median impatience and a 3 pp negative effect among those with below-median impatience. The level of impatience is thus pivotal to whether linear or time-bundled contracts are more effective. We also calibrate a model using experimental estimates of the distribution of walking costs and find consistent results.

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 them 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 impact 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 the payments. In contrast with conventional wisdom, increasing frequency is thus not always an effective way to adjust incentives for impatience.

The second goal of our experiment was to evaluate the use of incentives to address the burden of diabetes; we find that incentives substantially increase exercise and improve the health of diabetics. Providing just 20 INR (0.33 USD) per day of compliance increases compliance by 20 pp off of a base of 30%. Average daily steps increase by 1,300, roughly a 20 percent increase or 13 minutes of brisk walking. Much of this effect also persists after the intervention ends. Our sample has high rates of diabetes and hypertension; regular exercise can prevent complications from both. Incentives moderately improve an index of health risk that includes blood sugar and body mass index. Incentives also boost mental health. 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.

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 first 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); 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., 2020), 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 naïfs (Bai et al., 2020), who make up a large share of individuals (Augenblick and Rabin, 2019). In contrast, time-bundled contracts are effective for multiple types of impatience, including partial and full naïvete.

A few other studies also examine how to make incentive contracts work well for impatient agents without using commitment but have different goals than this paper. O'Donoghue and Rabin (1999b) and Carrera et al. (2020) both examine ways to help time-inconsistent procrastinators avoid delay in completing a single task; in contrast, our objective is to maximize average effort over time.[4] Andreoni et al. (2018) customize incentives using time preference estimates with a different objective: to make agents exert the same effort on two different days. Our theoretical insights about time-bundled contracts also relate 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.[5]

Our paper is also one of the first papers to study the implications of domain-specific discounting for contract design. 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). 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. 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 higher 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.[6]

Our contribution to the literature on nonlinear contracting is empirical: we experimentally compare the efficacy of contracts with linear and nonlinear incentive structures. Other

[4] 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 avoid delay among time-inconsistent procrastinators, the optimal incentive typically involves an increasing punishment for delay over time. Carrera et al. (2020) examine whether they can help time-inconsistent procrastinators overcome startup costs by offering higher incentives upfront in a separable contract and find the approach to be ineffective empirically.

[5] Jain (2012) assumes that people discount payment and effort identically. In contrast, we allow for different discount rates over payment and effort and show that the efficacy of our time-bundled contracts for the impatient is driven by high discount rates over effort, not present-biased time preferences per se.

[6] This test complements Gardiner and Bryan (2017)'s work in the psychology literature, which finds that simultaneously increasing the frequency of feedback and the frequency of payment improves efficacy.

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 discover an interesting effect: thresholds do not work well for everyone and so create dispersion in performance. This finding complements other work examining 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 often suboptimally distort behavior and promote cheating (e.g., Jacob and Levitt, 2003). Our work helps us understand the settings in which nonlinear schemes will work best. Our result that nonlinear schemes increase variance suggests that the contracts will be more effective when the returns to compliance are not concave, and our result that (dynamic) nonlinear contracts work better for the impatient suggestive that the contracts will work better when effort discount rates are high.

Finally, we contribute 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.

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 — depend on time preferences. We first specify the agent's problem and define contract effectiveness. We then solve for compliance under a simple "base case" incentive contract which is linear, and therefore also separable, across days. We then examine two variations to the base case contract. The first variation makes the contract non-separable and, in particular, "time-bundled" (i.e., the payment for future effort increases in current effort). The second variation maintains the linear payment function from the base case and instead changes the frequency with which payments are made.

We present three main theoretical findings. First, time-bundled contracts are particularly effective when agents have high discount rates over effort. Second, time-bundled threshold contracts increase the variance of effort, which means they are more effective when the returns to effort are linear or convex than when they are concave. Third, if agents are impatient over payment, increasing the frequency of payments should increase effort.

2.1 Set-Up

Each day, an individual chooses whether to complete a binary action, and then the principal gives the individual a nonnegative payment whose amount depends on the individual's past and present actions. 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. Let m_t be the payment made to the individual on day t.

To solve for compliance, we assume that agent choices maximize the following reduced-form utility function:

$$U = \mathbb{E}\left[\sum_{t=0}^{\infty} d^{(t)} m_t - \delta^{(t)} w_t e_t\right], \tag{1}$$

where e_t is the effort cost of complying on day t, $\delta^{(t)}$ is the discount factor over effort t days in the future, and $d^{(t)}$ is the discount factor over payments received t days in the future. Both $\delta^{(t)} \leq 1$ and $d^{(t)} \leq 1$, with $\delta^{(0)} = d^{(0)} = 1$. Neither $\delta^{(t)}$ nor $d^{(t)}$ are necessarily exponential functions of t; we make no assumption on the shape of the functions except that they are weakly decreasing in t. For notational simplicity, we denote $\delta^{(1)}$ as δ and $d^{(1)}$ as d. We assume utility is linear in payments, which is likely a good approximation in our setting, as payments are small relative to overall consumption.[7]

Importantly, this specification of reduced-form utility differentiates the discount rate over payments, $d^{(t)}$, from the discount rate over effort, $\delta^{(t)}$. The specification is consistent with a standard model of utility with a single structural discount factor over consumption and effort. In that case, $\delta^{(t)}$ is the structural discount factor, while the discount rate over payments depends on the availability of borrowing and savings (Cubitt and Read, 2007). In perfect credit markets, individuals should discount future payments at the interest rate r, and so $d^{(t)} = \left(\frac{1}{1+r}\right)^t$. At the opposite extreme, with no savings or borrowing, day t payments are immediately consumed, and $d^{(t)} = \delta^{(t)}$.[8]

Time-Inconsistency and Sophistication Individuals will have time-inconsistent preferences if either $\delta^{(t)}$ or $d^{(t)}$ are non-exponential functions of t or if $d^{(t)} \neq \delta^{(t)}$. We follow O'Donoghue and Rabin (1999a) and define a sophisticate as one who is fully aware of her own preferences and discount factors (over both effort and money) and a naïf as one who "believe(s) her future selves' preferences will be identical to her current self's."

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

[8] Our approach also nests some specifications of domain-specific time preferences.

Both sophisticates and naïfs predict their future compliance decisions based on their current beliefs about their future selves' preferences. Let $w_{j,t}$ be the agent's prediction on day t about her compliance on day $j > t$. Sophisticates accurately predict how their future selves will behave while naïfs may not.

Effort Costs Let effort costs for an individual be identically (but not necessarily independently) distributed across days, with the marginal distribution of e_t given by continuous cumulative distribution function (CDF) $F(\cdot)$. Individuals know the joint distribution of effort costs in advance but do not observe the realization of e_t until the beginning of day t.

The only assumption we make about the joint distribution of costs is that $e_{t'}$ is weakly increasing in e_t, in a first order stochastic dominance sense, for all $t' > t$.[9] This assumption flexibly accommodates the range from independence across days (IID) to perfect positive correlation, just ruling out negative correlation. We allow for e_t to be negative.

Incentive Contract Structure The 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 only, and so $m_t = 0$ for all $t < T$ and $m_T \geq 0$. Define *compliance* as the expected fraction of periods complied $C = \frac{1}{T}\mathbb{E}[\sum_{t=1}^{T} w_t]$ and the expected per-day *payment* as $P = \frac{1}{T}\mathbb{E}[m_T]$.

The Principal's Objective: Contract Effectiveness We assume that the principal's objective is to maximize *effectiveness*, defined as the expected benefit to the principal from compliance less the expected payment to agents. This objective is analogous to the standard contract theory approach of maximizing output net of wage costs subject to incentive compatibility constraints.[10]

Define the benefit to the principal from compliance as $y\left(\sum_{t=1}^{T} w_t\right)$. Then we can express *effectiveness*, which is the expected benefit to the principal net of the expected payment to agents, both calculated on a per-day basis, as $\frac{1}{T}\mathbb{E}\left[y\left(\sum_{t=1}^{T} w_t\right)\right] - P$.

For this definition to be meaningful, we need to take a stand on $y(\cdot)$. For simplicity, unless otherwise noted, we assume it is linear: $y(\sum_{t=1}^{T} w_t) = \lambda \sum_{t=1}^{T} w_t$ for some $\lambda > 0$. This is reasonable in our empirical setting since the estimated marginal health benefit of days of exercise is approximately linear (Warburton et al., 2006).[11] Under the linearity

[9] i.e., $F_{e_2|e_1}(x)$ is weakly decreasing in e_1 for all x, with $F_{e_t|e_{t'}}(x)$ the conditional CDF of e_t given $e_{t'}$.

[10] This is a distinct objective from maximizing welfare, but is often used in practice. For example, in health, policymakers and insurance companies often want to maximize the total health benefits of a program relative to the program costs. We discuss the appropriateness of this objective in Section 5.4.

[11] In practice, many principals who implement incentives for exercise care about the financial savings they might achieve from participants' exercise, rather than the health benefits. For example, in our experimental setting, the government's goal was to decrease public healthcare expenditures through its public health insurance scheme. Employers and insurance companies also often fund wellness programs to try to reduce financial costs. However, if health benefits are linear in exercise, then the financial benefits may be as well.

assumption, we can also simplify our expression of effectiveness. The expected per-day benefit of compliance becomes λC. Effectiveness becomes $\lambda C - P$.

We want to compare the effectiveness of different contracts in settings where we do not know λ. Define the *cost-effectiveness* of a contract as compliance divided by expected per-day payment, C/P. Rewriting effectiveness as $\lambda C - P = C\left(\lambda - \frac{1}{(C/P)}\right)$ shows that one contract is more *effective* than another if it has strictly larger C and weakly larger C/P, or weakly larger C and strictly larger C/P (assuming the contract's effectiveness is positive). Thus, one does not need to know λ to compare the effectiveness of some contracts.

Given the assumptions above, Appendix B.1 provides the general solution to the agent's problem, which we use to solve for the compliance and effectiveness of various contracts.

2.2 Separable Linear Contracts (the Base Case)

We now solve for compliance and effectiveness under the base case contract. The contract is separable (i.e., payment for w_t depends only on w_t and not on any $w_{t'}$ for $t' \neq t$) and linear, paying m per day of compliance. Total payment is thus:

$$m_T^{\text{Base Case}} = m \sum_{t=1}^{T} w_t. \tag{2}$$

Agents comply on day t if the discounted payment outweighs the effort cost:

$$e_t < d^{(T-t)} m. \tag{3}$$

Holding all else constant, compliance is thus independent of δ.[12]

Expected payment per period P is then mC. As a result, effectiveness is $(\lambda - m)C$. Cost-effectiveness, C/P, is simply $\frac{1}{m}$ for any linear contract with positive compliance.

Observation 1. Holding all else constant, neither compliance, cost-effectiveness, nor effectiveness in the linear contract depend on $\delta^{(t)}$. This is true regardless of T or the correlation structure of costs. It is also true for any separable contract, not just for linear contracts, as neither the cost nor the benefit of compliance in a separable contract ever depend on $\delta^{(t)}$.

We will see that this result contrasts with time-bundled contracts.

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. The defining feature of time-bundled contracts is that their payment functions contain at least one dynamic complementarity (i.e.,

---

Exercise can lead to healthcare savings because it decreases expensive complications (Reiner et al., 2013).

[12] In particular, $C = \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} w_t\right] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[w_t\right] = \frac{1}{T}\sum_{t=1}^{T} F(d^{(T-t)}m)$, where the transition from the second to third term follows because of the linearity of expectations.

a period for which the payment for future compliance is increasing in current compliance). We focus our analysis on a "threshold" time-bundled contract, where there is a minimum threshold level of compliance K below which no incentive is received, and above which payment is a linear function of the number of days of compliance, $\sum_{t=1}^{T} w_t$:

$$m_T^{\text{Threshold}} = \begin{cases} m' \sum_{t=1}^{T} w_t & \text{if } \left(\sum_{t=1}^{T} w_t \geq K\right) \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

We begin by examining the effectiveness of thresholds and how it depends, first, on the discount factor over effort, δ, and second, on the shape of the benefits function. Third, we discuss other time-bundled contracts besides thresholds and clarify the connections between time-bundled contracts and commitment contracts.

Appendix B presents our formal mathematical results, which we label as propositions. In the main text, we summarize the propositions and present their testable implications, which we label as predictions.

2.3.1 Threshold Contracts and Impatience Over Effort

An important question is when thresholds are more effective than linear contracts. While many factors play a role, we focus first on the role of the discount rate over effort. Appendix B.2 compares the overall effectiveness of threshold and linear contracts formally, yielding two main take-aways. First, threshold contracts can particularly improve effectiveness when the discount rate over effort is high. For example, when δ is sufficiently small, for any linear contract there exists a threshold contract that achieves much higher cost-effectiveness with limited—if any—less compliance. Second, the discount rate over effort can be pivotal to whether threshold or linear contracts are more effective.

While the specific Appendix B.2 results require assumptions that may not hold in practice (e.g., some results require δ to be sufficiently small), they yield a testable policy implication that holds under much more general assumptions:

Prediction 1 (Threshold Effectiveness and Impatience Over Effort)**.** *Holding all else equal, threshold contracts tend to perform better relative to separable contracts, with respect to compliance and effectiveness, when the discount factor over effort, $\delta^{(t)}$, is smaller.*

Prediction 1 is based on a series of propositions, presented in Appendix B.3, showing that, under a broad range of assumptions, holding all else equal, in threshold contracts, compliance and often effectiveness are both decreasing in $\delta^{(t)}$. In contrast, with separable contracts, both compliance and effectiveness are flat in $\delta^{(t)}$ (Section 2.2). Thus, the lower is $\delta^{(t)}$, the higher are compliance and effectiveness in a threshold contract relative to a separable contract.

We illustrate the intuition behind Prediction 1 by considering a simplified case, with $T = 2$ and with effort costs that are weakly positive and IID. The threshold contract has $K = 2$: one must comply on both days of the two-day payment period to receive payment.

On day 1 of the threshold contract, the individual's motivation to comply is to have the option to be paid $2m'$ for complying on day 2. The value she places on that day 2 option is

$$\mathbb{E}\left[(d2m' - \delta e_2)w_{2,1}|w_1 = 1\right], \tag{5}$$

which is equal to the expectation of the discounted payment net of the discounted effort costs in the states of the world where the individual thinks she will comply on day 2 conditional on complying on day 1. Importantly, because the future effort cost is discounted, the value is weakly decreasing in δ, for both sophisticates and naifs: impatient people value the option more.[13]

Impatient people's greater valuation for the day 2 option is what makes the threshold contract more effective for the impatient. The impatient person's greater valuation of the option motivates her to comply more on day 1 (since she complies if $e_1 < \mathbb{E}\left[(2dm' - \delta e_2)w_{2,1}|w_1 = 1\right]$). Her greater compliance on day 1 then increases her compliance on day 2 (as she only complies on day 2 if she complied on day 1), hence increasing total compliance. Her greater total compliance increases the effectiveness of the contract.[14] Thus, the fact that time-bundled contracts link better future work opportunities with compliance today underlies their effectiveness for those with lower δ.

2.3.2 Threshold Contracts and the Shape of Benefits to Compliance

While our primary focus is on the role of the discount rate over effort in the effectiveness of thresholds, we also briefly explore the role of the shape of the benefits function, an important issue that our experiment sheds light on. Above, we assume that the benefits function, $y(\sum_{t=1}^{T} w_t)$, is linear, a reasonable assumption for our empirical setting. What if the benefits were concave instead? With concave benefits, the principal would prefer that compliance have a lower variance, all else equal. We next show that threshold contracts have a higher variance of compliance than linear contracts. This implies that thresholds are more effective

[13] $w_{2,1}$ depends on whether one is sophisticated or naïve. Since naïfs believe that their future selves have the same preferences as their present-day selves, $w_{2,1} = \mathbb{1}\{\delta e_2 < 2dm'\}$. Sophisticates correctly predict their future preferences and behavior, and so $w_{2,1} = \mathbb{1}\{e_2 < 2m'\}$. The value of the option is thus $\mathbb{E}\left[(2dm' - \delta e_2)\mathbb{1}\{e_2 < 2m'\}|e_1\right]$ for sophisticates and $\mathbb{E}\left[(2dm' - \delta e_2)\mathbb{1}\{\delta e_2 < 2dm'\}|e_1\right] = \mathbb{E}\left[\max\{2dm' - \delta e_2, 0\}|e_1\right]$ for naïfs, both of which are decreasing in δ.

[14] Effectiveness follows from compliance since an increase in compliance without a decrease in cost-effectiveness implies higher effectiveness, and the Appendix B.3 propositions show that, depending on the cost distribution, threshold cost-effectiveness tends to be flat or decreasing with $\delta^{(t)}$.

when the benefits of compliance are linear (or convex) than when they are concave.

Prediction 2 (Thresholds and Variance)**.** *Threshold contracts have higher variance in compliance than linear contracts that generate a similar level of compliance.*

Prediction 2 is based on two pieces of evidence. First, Proposition 9 in Appendix B.3 shows formally that, when $T = 2$, $d = 1$, and costs are IID and weakly positive, if a threshold and linear contract generate the same level of compliance, the threshold contract has higher variance. Second, simulation results suggest that a similar prediction would go through under much more general assumptions (e.g., larger T, positive correlation, negative costs).

Again, we illustrate the intuition behind Prediction 2 in the two-day case. In a linear contract, complying on day 1 and on day 2 are separate decisions. In contrast, in a threshold contract, complying on day 1 makes one more likely to comply on day 2. This increases the likelihood that individuals comply either on both days or zero days (as opposed to one day). Threshold contracts thus create more extreme outcomes and have higher variance.

2.3.3 Relationship between Threshold Contracts and Commitment Contracts

In this subsection, we discuss the similarities and differences between threshold contracts and commitment contracts in order to build intuition for why, unlike commitment contracts, threshold contracts are effective for impatient people who are naive. There are two main similarities. Both types of contracts are often dominated contracts from the agent's perspective ,[15] and both can increase effort of impatient, time-inconsistent agents.

An important difference between the types of contracts is that threshold contracts can generate more effort for both sophisticates and naïfs, whereas commitment contracts rely on sophistication (in particular, they rely on people choosing them, which only sophisticated people do). What underlies the difference? Appendix B.4 sheds light on this issue by investigating the full class of 2-day time-bundled contracts (i.e., contracts that pay a higher wage for day 2 compliance if the agent complies on day 1 than if she does not). In addition to thresholds, this class includes contracts where the day 2 wage is not 0 in the absence of day 1 compliance.[16] Here, we summarize the intuition, assuming $d = 1$ and $T = 2$ for simplicity.[17]

We show that the time-bundled contracts that are effective for sophisticates are effective for nearly the same reason as commitment contracts: one reason that a sophisticate

[15]Specifically, threshold contracts offer the agent weakly less money for any level of effort than in a linear contract with the same per-period payment level. Commitment contracts are dominated because, by definition, they restricts the agent's choices or imposes financial penalties compared to the status quo.

[16]e.g., a contract could pay \$5 for day 2 effort if the agent did not comply on day 1 and \$10 if she did.

[17]To tie into the commitment literature, Appendix B.4 adopts the assumption normally used in that literature that future costs are revealed on day 1 (e.g., Bai et al., 2020; O'Donoghue and Rabin, 1999a).

complies is to give her future self strong incentives to comply. While she is not imposing financial penalties for non-compliance, she is raising the price of it. Specifically, the contracts that work for sophisticates generate "commitment value": complying on day 1 is pivotal to whether the agent *actually* complies on day 2. This holds when the payment offered for day 2 compliance is greater than e_2 if (and only if) the agent complies on day 1. In this case, complying on day 1 generates a soft *"commitment"* to comply on day 2.

In contrast, the time-bundled contracts that are effective for naïfs generate "option value": complying on day 1 is pivotal to the day 1 self *wanting* her day 2 self to comply. This occurs when the payment offered for day 2 compliance is greater than δe_2, as opposed to e_2, if (and only if) the agent complies on day 1. In this case, day 1 compliance generates a valuable *"option"* to be paid for day 2 compliance. The naïf's overoptimism can make her overvalue the option *ex ante*, which means that her day 1 effort may increase more than the sophisticate's. Nonetheless, the option is normally still valuable to her *ex post*, and so she normally exercises it.[18]

We show that threshold contracts can generate both option value and commitment value, which is why they can work for both naïfs and sophisticates. In contrast, we show there are some time-bundled contracts that generate commitment value but do not generate option value and that function exactly like commitment contracts. (Thresholds never fall in this category.) In these contracts, which work for sophisticates only, day 1 compliance is analogous to taking up a commitment contract for future compliance; sophisticates incur a cost, working "below-cost" on day 1, to induce their future selves to comply.[19]

Thus, time-bundled contracts are a broad class that includes contracts that function like commitment contracts and work for sophisticates only, as well as contracts (like thresholds) that also generate option value, enabling them to work for naifs as well.

2.4 Variation 2: Payment Frequency

We now return to the base case separable linear contract from equation (2) and analyze compliance under different payment frequencies by changing the length of the payment period T. We make two intuitive predictions. See Appendix B.5 for proofs.

Prediction 3 (Frequency)**.** *If agents are impatient over the receipt of financial payments (i.e., if $d^{(t)} < 1$ for $t > 0$ and is weakly decreasing in t), then the compliance and effectiveness of the base case linear contract are weakly increasing in the payment frequency. If agents are*

[18]Naïfs could also comply in early periods expecting to be paid later and then not follow-through. Section 5.4 discusses theoretical and empirical reasons why we think this pattern is rare.

[19]That is, sophisticates comply on day 1 even when their effort cost exceeds the maximum potential financial benefit of day 1 compliance in order to induce their day 2 self to comply.

patient over the receipt of financial payments ($d^{(t)} = 1$), then payment frequency does not affect compliance or effectiveness.[20]

Prediction 4 (Payday Effects)**.** *If the discount factor over payments $d^{(t)}$ is decreasing in t, then the probability of complying in the base case linear contract increases as the payday approaches. If the discount factor over payments $d^{(t)}$ is constant in t, then the probability of complying is constant as the payday approaches.*

2.5 Empirical Tests

We design our experiment in light of the predictions above. We assess the quantitative importance of Prediction 1, that the compliance and effectiveness of time-bundled threshold contracts relative to linear contracts 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 linear and test for heterogeneity in compliance and effectiveness in the threshold relative to linear contract based on a baseline measure of impatience over effort (we address potential confounds to the impatience measure in Section 5.2.3). 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 assess Prediction 2, that thresholds have higher variance in compliance than linear contracts, we compare the variance in compliance between the threshold and linear treatment groups.

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 3) understand if agents are meaningfully impatient over payments. Finally, Prediction 4 allows us to use within-treatment variation to shed further light on the shape of the discount factor over payments, as in Kaur et al. (2015).

All of these hypothesis tests were *ex ante* or pre-specified tests except for the test of Prediction 2 regarding the variance of thresholds, which can be considered exploratory.

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

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

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."[21] After screening, we contacted eligible 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 wear a pedometer properly, 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.[22]

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 guided 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.

[21] The full list of eligibility criteria was: must be diabetic or have elevated random blood sugar ($>$ 150 if has eaten in previous two hours, $>$ 130 otherwise); 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 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.

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

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.

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).[23]

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.

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.

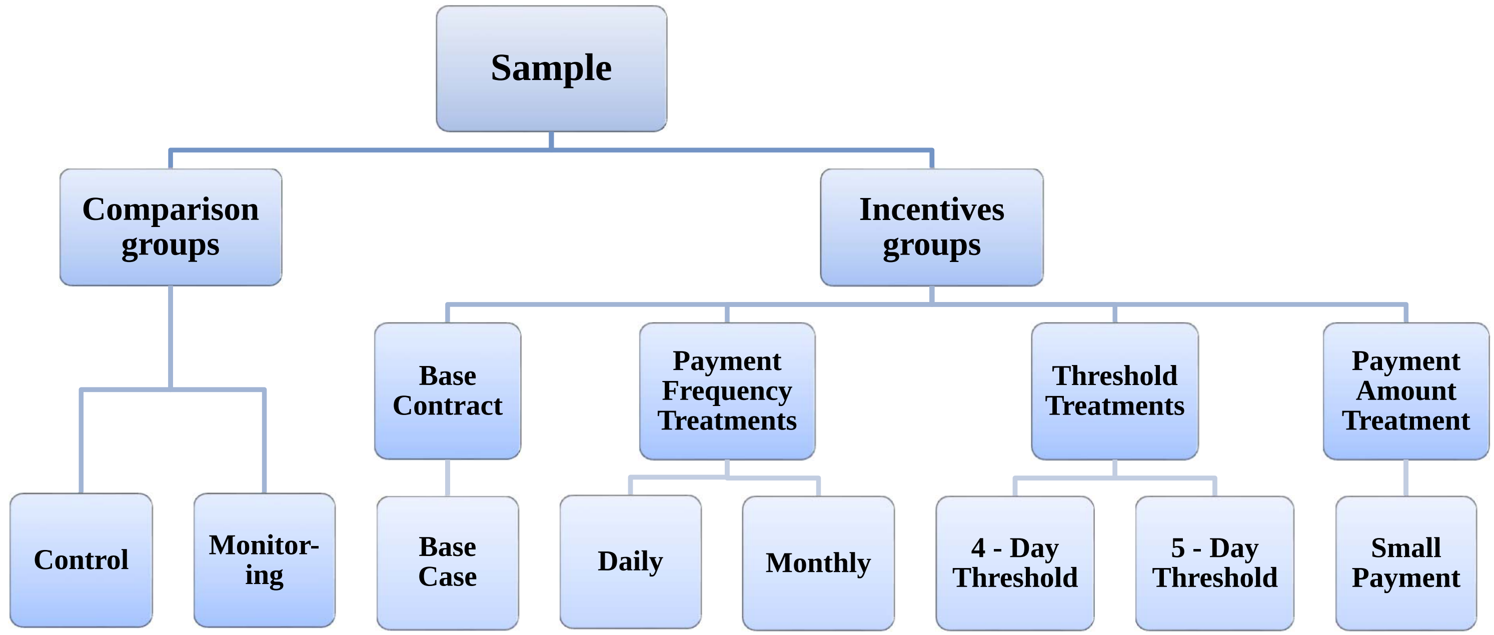
We call this the *base case* contract because all other contracts differ from it in exactly one dimension: separability, payment frequency, or payment amount. We can compare any other group to the base case group to assess the effect of changing a single contract dimension.

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

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, 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

[23] 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).



	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

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.

We follow our *ex ante* analysis plan and pool the *4-day threshold* and *5-day threshold* treatment groups for our main analyses in order to maximize statistical power.[24] We sometimes also show the results for the two thresholds separately as exploratory analyses.

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 received daily feedback on step target compliance and received a test payment of 10 INR the night before their incentive contract launched.

We allocated more of our sample to the time-bundled threshold groups than the payment frequency groups, for two reasons. First, we regard our insights about time-bundled thresholds as more novel than our insights about frequency. Second, testing our theoretical predictions regarding thresholds requires a heterogeneity analysis. In contrast, a main effects analysis of the frequency treatments is sufficient to learn about average impatience over payments and its implications for incentive design. We allocated our sample to have power to detect heterogeneous effects of the threshold treatment and main effects of the frequency treatments.

Payment Amount Our final incentive group, the *small payment group*, 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 group to learn about the distribution of walking costs and to benchmark the size of our other treatments 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 monitoring group and a pure control, to allow us to shed light on these two channels.

[24]We included the two threshold levels, with the *ex ante* intention to pool them, to reduce the risk that compliance was too high or too low (because the threshold was very easy or hard to reach) to have statistical power to to test our prediction about heterogeneity by impatience.

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.[25]

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. Since most participants did not have regular internet access to be able to sync their pedometer data, 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.[26]

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 survey. Figure A.1 shows the intervention timeline.

To ensure participants understood their contracts, surveyors also called participants several days after the intervention period had started and asked several survey questions testing

[25] 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. We control for the SMS treatment in our main regressions and test its effects in the Online Appendix.

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

their understanding of their contracts. The responses indicate that a vast majority of participants did indeed understand their assigned contract (Online Appendix Table E.1).

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 included 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.

Impatience over Effort Following the phase-in period, we conducted 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.

Our primary measure of impatience over effort 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.2, 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 tilt toward procrastination-style behaviors and so may better detect naïve time-inconsistency than other types of impatience. Our empirical heterogeneity tests using the index may thus tilt toward testing whether contracts are effective for naïfs in particular. Since naïfs and partial naïfs appear to constitute a large share of impatient individuals (Augenblick and Rabin, 2019; Bai et al., 2020), and since we consider the efficacy for naïfs to be a nice advantage of 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. Those with higher values of our impatience index have worse diets and walk less at baseline (Table A.2). Notably, the measure does not correlate with proxies for impatience over recharges (Table A.3), discussed below, suggesting that the discount rates over effort and over payments are relatively independent in our setting.

We began collecting our impatience index partway through the data collection,[27] so it is only available for the latter 54% of the sample. Luckily, that sample size is sufficient to achieve statistically signifant results. That said, to also check the robustness of our results in the full sample, we 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. Panel B of Table A.2 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 Online Appendix G. 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.[28] Further, 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

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

[28] Other suggestions of a lack of understanding include our estimates not converging for 44% of the sample and respondents failing to follow through on their chosen allocations. See Online Appendix G for details.

balances, recharge usage, and the response to a question regarding the frequency with which the person would prefer to receive payments (daily, weekly, or monthly). People who have higher balances and usage 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, close to the median for an urban household in India (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 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 7,000 steps per day in the phase-in period, comparable to average daily steps in many developed countries (Bassett et al., 2010). Panel D shows our measures of impatience over effort (Online App. Table E.2 shows summary statistics on the components of the indices). Panel E shows our proxies for impatience over mobile recharges.

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. 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

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)
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)
Prefers daily payment (=1)	0.17	0.18	0.16	0.17	0.20	0.17	0.20	0.17	0.18
	(0.38)	(0.38)	(0.37)	(0.38)	(0.40)	(0.37)	(0.40)	(0.38)	(0.39)
Prefers monthly payment (=1)	0.24	0.25	0.28	0.24	0.27	0.24	0.23	0.24	0.26
	(0.43)	(0.43)	(0.45)	(0.43)	(0.45)	(0.43)	(0.42)	(0.43)	(0.44)
F-tests for Joint Orthogonality									
P-value (relative to control)	N/A	N/A	0.76	0.30	0.67	0.44	0.31	0.37	0.49
P-value (relative to monitoring)	N/A	0.76	N/A	0.94	0.92	0.85	0.58	0.97	0.67
P-value (relative to base case)	N/A	0.44	0.85	N/A	0.47	N/A	0.79	0.97	0.46
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.

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.4). 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.[29] 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, Table C.3 shows that they do not.

Another potential concern would be if participants gave their pedometers to someone else. Reassuringly, our data suggest that this concern is limited.[30]

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 a good laboratory to explore our contract variations. Second, we explore the effect of adding a time-bundled threshold, testing our prediction that it will be more effective for those who are more impatient over effort and gauging its effectiveness and variance. Third, we assess the effect of varying payment frequency and 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

[29] 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.4). While our main specifications drop days with missing pedometer data, Table A.5 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.4), but results are robust to Lee bounds accounting specifically for that source (column 5 of Table A.5).

[30] 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).

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:

$$y_{it} = \alpha + \beta \times incentives_i + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\lambda + \varepsilon_{it}, \tag{6}$$

where y_{it} is either individual i's steps on day t during the intervention period or an indicator for individual i surpassing 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 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.

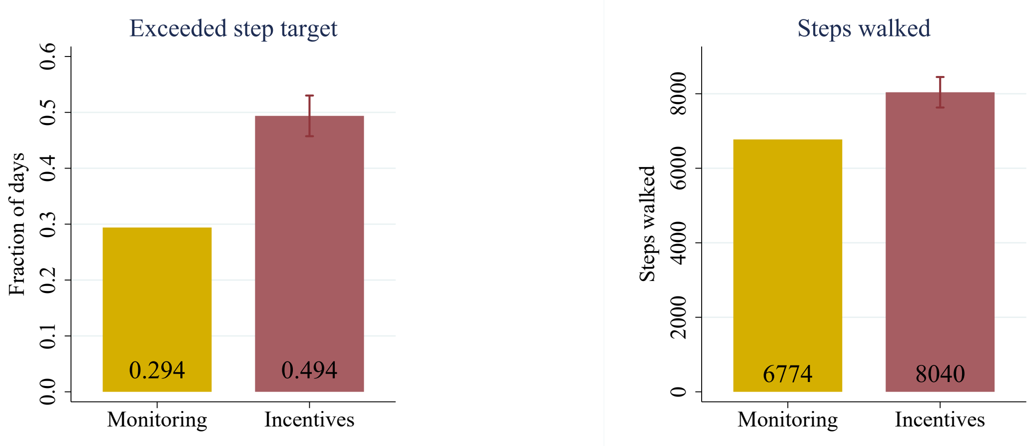
5.2 Variation 1: Time-Bundled Threshold Contracts

We now explore the impact of our contract variation designed to improve effectiveness in the face of impatience over effort: time-bundled thresholds. We first compare the effectiveness of the threshold and linear contracts in the full sample. Our theoretical analysis suggests that, under plausible conditions—such as the effort discount rates being sufficiently high—time-bundled thresholds can be more effective overall than linear contracts, making this an interesting comparison. We then test our two primary theoretical predictions. First, we test whether time-bundled thresholds increase variance (Prediction 2). Second, we test our core prediction: that time-bundled thresholds increase compliance and effectiveness more among those who are impatient over effort than among those who are not (Prediction 1).

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 (6) (Panel A) and (7) (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.4 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.

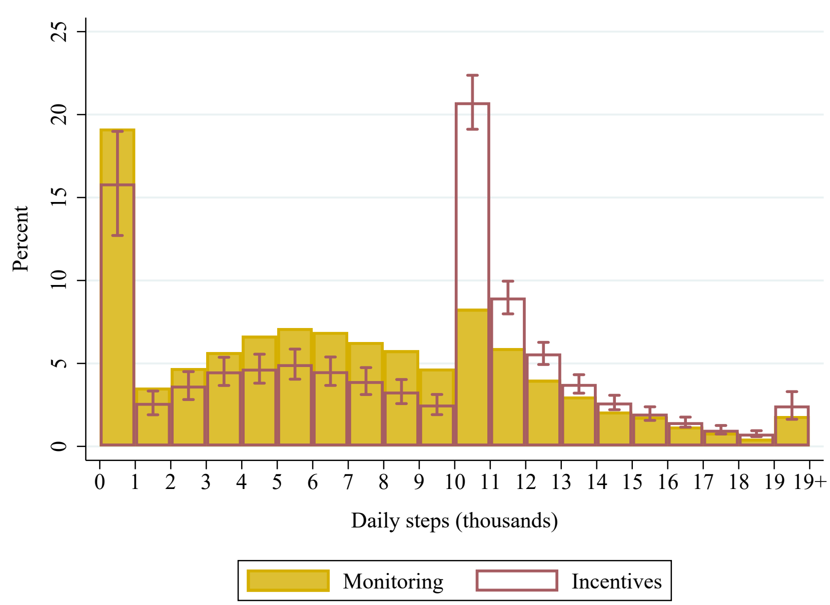


Figure 3: Incentives Shift the Distribution of Steps Walked per Day

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.

5.2.1 Average Effectiveness

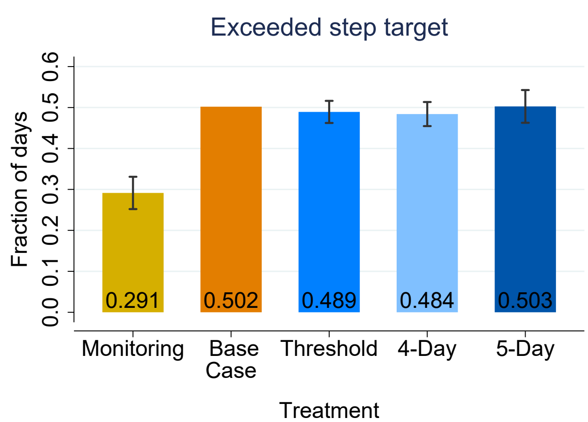
In order to establish that time-bundled threshold contracts are effective on average, we can show that they result in weakly more compliance and and weakly higher cost-effectiveness than linear contracts in the full sample, with one inequality strict.

We find that adding a time-bundled threshold does not affect average compliance relative to the base case. To test for differences across the incentive treatment groups, we estimate regressions of the following form:

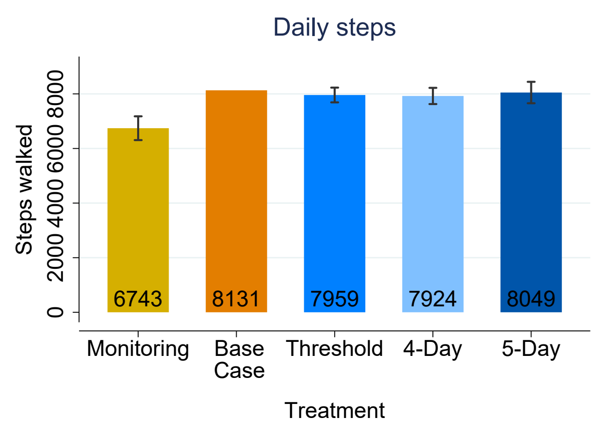
$$y_{it} = \alpha + \beta_j \times \left(incentives^j\right)_i + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\theta + \varepsilon_{it}, \tag{7}$$

where y_{it} are daily walking outcomes and $\left(incentives^j\right)_i$ is an indicator for whether individual i is enrolled in incentive treatment group $j \in$ (daily, base case, monthly, threshold, small payment). The β_j coefficients capture the average effect of each incentive treatment group relative to the monitoring group. Panel B of Table 2 displays the results.

The effect of the threshold treatment on compliance is very similar to the effect of the base case (linear) treatment on compliance, with the estimates within 1.3 pp of each other and the difference not statistically significant (p-value=0.36 from Table 2). Figure 4, Panel A displays the result graphically. It also shows the 4-day threshold group and 5-day 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, the threshold contracts are more cost-effective. Individuals in the threshold group 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, as opposed to the 100% of days that the base case group (by definition) receives payment. As a result, the cost-effectiveness of the threshold contracts are 10% and 18% higher than that of the base case contract (Table A.6, Panel B).

Because the threshold contracts have the same compliance and are more cost-effective than the base case, they are more *effective* overall. For comparison, the small payment treatment is also more cost-effective than the base case (it pays half as much per day complied), but this comes at the cost of reduced compliance (Table 2).

5.2.2 Variance

Next, we show that thresholds increase the variance of compliance in the week long payment period. 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 or seven days in the week. We also follow Iachine et al. (2010) to test for the equality of variance in week-level compliance across treatments.[31] The results confirm that the variance of compliance is significantly higher in the threshold group than the base case linear group (Table A.7).

While the increase in dispersion and in zeroes in the threshold treatment is consistent with our theoretical prediction, the increase in density at seven days in particular (instead of at the specific threshold level of four or five) is perhaps surprising. Potential explanations include that it is hard for participants to keep track of how many days they have walked or that 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

[31] We first create an individual×week-level "absolute deviation" measure equal to the absolute value of the gap between the number of days walked by an individual in a given week and the average number of days walked by her treatment group in that same week. We then regress this on a threshold group dummy in a sample that includes the threshold and base case groups only.

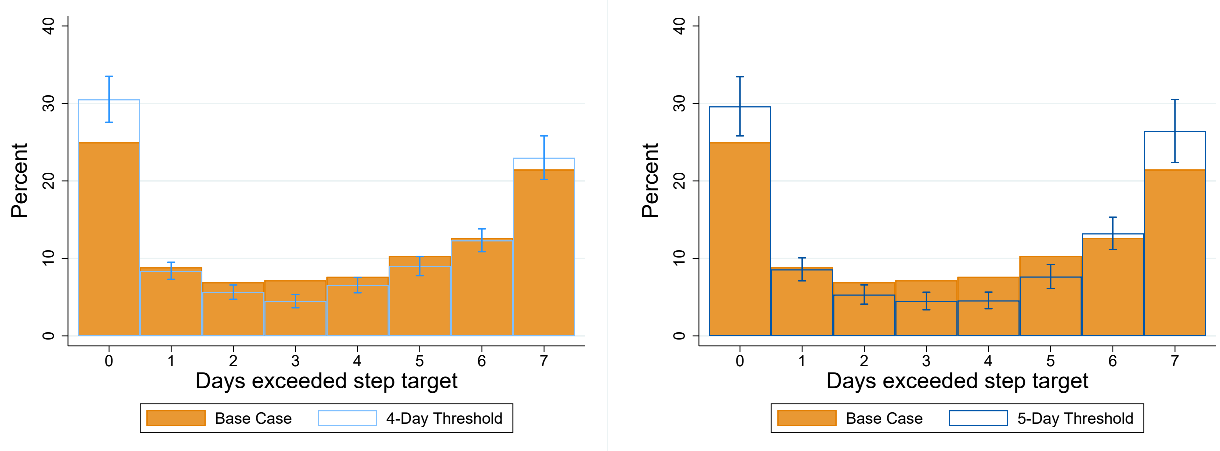


Figure 5: Threshold Contracts Increase Variance Across Weeks

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.

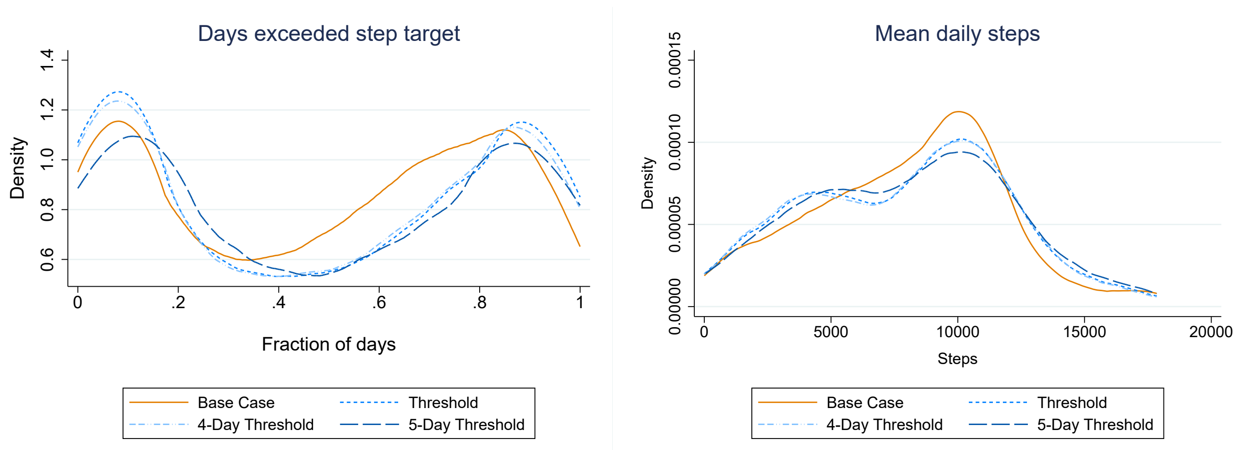


Figure 6: Threshold Contracts Increase Variance Across Individuals

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.

whom they will work well: those who are impatient over effort.

5.2.3 Heterogeneity in Threshold Effects by Impatience over Effort

We perform two exercises to assess whether, with respect to compliance and effectiveness, threshold contracts perform better relative to linear contracts when individuals are more impatient over effort (Prediction 1). First, we quantify the heterogeneity by baseline impatience in compliance in the threshold group relative to the base case (linear) group. Since Prediction 1 regards 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 condition 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 estimate, our heterogeneity estimate is the one that is relevant for policy – policymakers want to customize contract thresholds based on how their efficacy varies with observed participant impatience, irrespective of whether it is impatience itself (as opposed to the correlates of impatience) that generates the heterogeneity.

To tie our data to our theory more precisely, Appendix D also calibrates 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 effect of the time-bundled threshold by impatience:

$$y_{it} = \alpha + \beta_1 impatience_i \times thresh_i + \beta_2 thresh_i + \beta_3 impatience_i + \boldsymbol{X}'_i \pi + \boldsymbol{X}'_{it} \theta + \varepsilon_{it}, \quad (8)$$

where y_{it} is an indicator for whether individual i exceeded the 10,000-step target on day t and $thresh_i$ is an indicator for being in the threshold group. (Table A.8, Panel A shows robustness to using daily steps as the outcome, and Panel B shows the 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[32] 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 generate meaningfully more compliance among 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 compliance in the threshold group relative to the linear group 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

[32]To construct the bootstrap confidence intervals for the regressions that use the predicted impatience index, we draw bootstrap samples clustered at the individual level and, in each sample, conduct three steps: 1) run the LASSO prediction model; 2) create the predicted impatience index using that sample's LASSO coefficients, thus accounting for the error in constructing the index itself; and 3) estimate equation (8).

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

Dependent variable:	Exceeded 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]	[0.89, 5.00]	[0.04, 9.55]
Threshold	-1.3	-3.81	-1.18	-3.41**
	[-4.36, 1.76]	[-8.89,1.28]	[-3.38, 1.04]	[-5.84, -0.44]
Impatience	-2.97**	-4.68*	-2.38***	-5.3***
	[-5.36, -0.57]	[-9.46,0.10]	[-3.83, -0.78]	[-8.05, -0.98]
# 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.8, Panel B for results with the Threshold group disaggregated (unpooled). Bootstrap draws were done at the individual 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%.

the base case, the threshold generates 6 pp higher compliance for those with above-median impatience than those below the median, a large increase relative to the sample-average effect of either contract (20 pp). Recall that we only have the impatience index for the sample enrolled later in the experiment; 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 very similar (and more precise) results.

Figure 7 presents a visualization of column 4, showing that adding the threshold to the linear contract increases compliance among the more impatient while decreasing it among the less impatient. The difference between the effects is the significant 6 pp effect from column 4. Our theory (e.g., Proposition 3 in Appendix B.2) showed us that the discount rate over effort could be pivotal to whether the linear or threshold contract has higher compliance. That is the case here, which is important for policy: efforts by policymakers to individualize who receives a threshold contract based on agent impatience could substantially increase compliance.

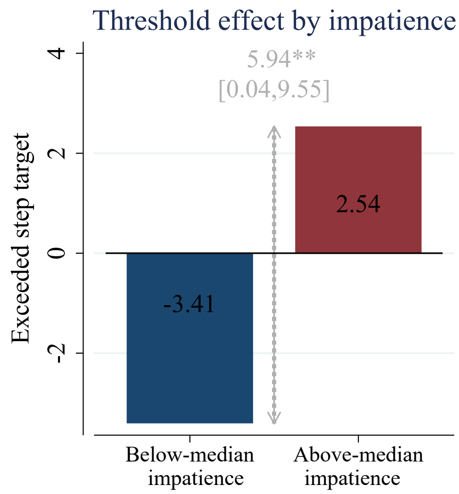


Figure 7: Time-Bundled Thresholds Increase Walking More for the Impatient

Notes: The chart plots the effect of the threshold contract relative to the base case, 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.

Prediction 1 suggested that, in addition to increasing compliance more among the impatient, threshold contracts should also increase *effectiveness* more among the impatient. Since we have already demonstrated the compliance result, demonstrating the effectiveness result requires showing that, relative to the base case, thresholds do not have lower cost-effectiveness among the impatient than the patient. Table A.6 shows that this is the case.

Table A.9 shows that the coefficient on the interaction of impatience and the threshold from Table 3 remains stable when we control for other baseline covariates and their interactions with the threshold. For example, we control for risk aversion, scheduling uncertainty, and baseline walking (a proxy for the mean of the walking cost distribution), among other covariates. The stability of the coefficient suggests 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).

5.2.4 Time-bundled Thresholds Result Summary

Our analysis creates several new findings about time-bundled threshold contracts. Consistent with our theoretical predictions, thresholds generate meaningfully greater compliance among the impatient than the patient. In the full sample, they increase effectiveness by increasing cost-effectiveness without decreasing compliance. However, they also increase variance, which means that they will be more effective in settings with linear or convex benefits of compliance than those with concave benefits. The variance in their performance across the full sample also underscores the potential policy gains from targeting the assignment of thresholds based on predictors of effectiveness and highlights the importance of our finding that impatience over effort is one such predictor.

5.3 Variation 2: Payment Frequency

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

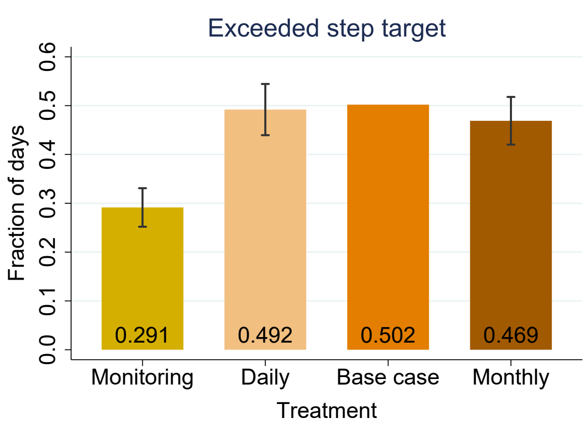
1. Between-treatment: We compare average compliance in the daily, weekly (base case), and monthly groups. We assess how payment frequency affects compliance and use Prediction 3 to shed light on the discount rate over payment.
2. Within-treatment: Within the base case and monthly groups, we examine how compliance changes as the payday approaches to shed light on the discount rate over payment using Prediction 4. 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 for making inferences about discount rates over payment using between-treatment effects.[33]

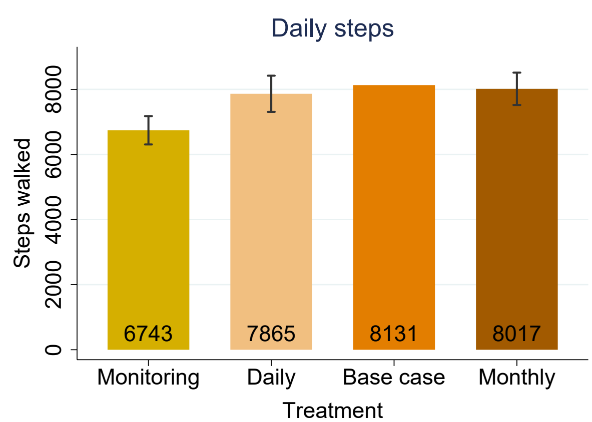
We begin with the between-treatment comparisons. Figure 8 and Panel B of Table 2 both show that the three payment frequency treatments have similar effects of walking. Compliance and steps walked are statistically indistinguishable across the three treatments. The point estimates also do not increase 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

[33] Our design mitigates some of these potential confounds, such as feedback frequency and salience (Section 3.2.1), but a couple of confounds remain. If utility were concave in payments, then the fact that higher-frequency payments break payments into smaller chunks would improve compliance and cause us to overestimate the discount rate. If instead people preferred lumpier payments since they serve as commitment devices for savings (Casaburi and Macchiavello, 2019), we would underestimate the discount rate.



(a) Probability Exceeded Step Target



(b) Average Daily Steps

Figure 8: Payment Frequency Does Not Significantly Impact Walking

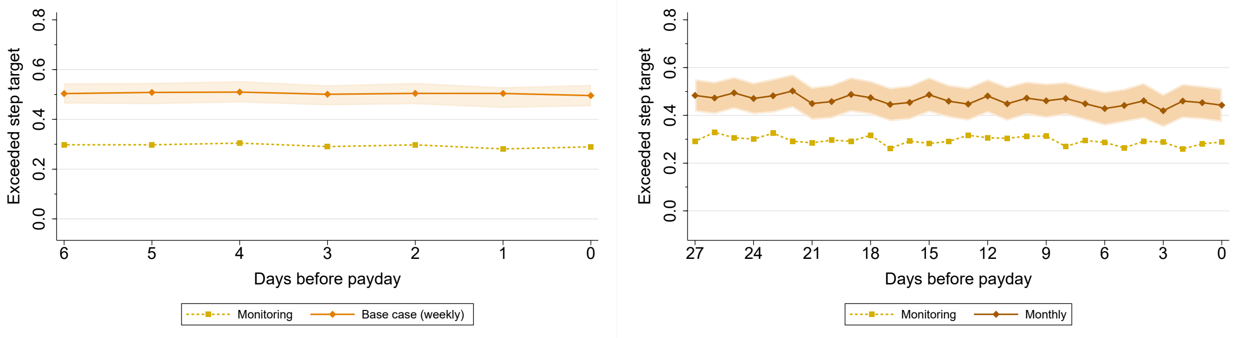
Notes: Panel A shows the average probability of exceeding the daily step target during the intervention for the three different frequency treatments (the base case treatment pays weekly). Panel B shows average daily steps during the intervention. Confidence interval bars represent tests for equality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

that it would. The lack of between-treatment frequency effects implies that the discount rate over our financial payments is small and has a relatively flat shape over the range from one day to one month. 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. We address this issue with the within-treatment analysis below.

The within-treatment analysis confirms the suggestive evidence of flat and low discount rates from the between-treatment analysis. Figure 9 shows how compliance within the base case weekly (Panel A) and monthly (Panel B) treatments changes as the payment date approaches. If agents are impatient over payments, compliance should increase as the payday approaches (Prediction 4). Yet we find that walking behavior is remarkably steady across the payment cycle. Table A.10 estimates the change in compliance as the payment date approaches within the base case and monthly groups, conditional on day-of-week fixed effects.[34] The estimates are not significantly different from zero and suggest that, if anything, compliance *decreases* as the payment date approaches.

Our confidence intervals are also tighter here. If we assume linearity of compliance in lag

[34]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.



(a) Weekly Payment Cycle

(b) Monthly Payment Cycle

Figure 9: 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.

to payment, then the confidence interval around the slope in the weekly treatment rules out the possibility that because of monetary discounting, daily payments would be more than a mere 0.3 pp more effective than weekly.

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 conflicts with Kaur et al. (2015). The reasons for the differences are an open question for future work (e.g., they may reflect different payment modes or payment amounts).

5.3.1 Payment Frequency Results Summary

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. 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

This paper evaluates ways to increase contract effectiveness (the benefit of compliance to the principal less the payout), a relevant objective in many situations. In firm and worker applications, maximizing effectiveness is often analogous to profit maximization. In public applications, policymakers are often concerned with maximizing effectiveness, perhaps because it is straightforward to explain and justify. Moving from effectiveness to welfare involves an understanding of concepts such as the social cost of public funds and the externalities of behavior, which are beyond the scope of this paper. That said, if the marginal social benefit of the incentivized behavior outweighs the marginal social cost in the "base case" version of a program – as is likely here since the estimated social benefits of walking are large relative to the private costs and incentive amounts[35] – then variations that increase compliance and effectiveness have high potential to increase social welfare.

One potential concern with our contract variations would be if they improved effectiveness and/or social welfare but did 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 naïve 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: naïfs 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. However, there are theoretical reasons to doubt that threshold contracts would actually harm any participant's welfare.[36] Two pieces of empirical evidence also suggest that our threshold contract did not reduce participants' welfare. First, at endline, we asked participants whether they were interested in continuing the program. The vast majority said that they were, with no significant difference between the threshold group and other groups and, within the threshold group, no significant difference between the more and less impatient (Table A.11). 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 behavior 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 first interpret our exercise impacts in light of the literature on related programs. We next examine how our exercise impacts changed over time, both during and after the

[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 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]First, later compliance costs must be 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 naïfs 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.

intervention. Finally, we show that the program improved cardiovascular health.

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).[37]

6.1.1 Robustness of Exercise Impacts

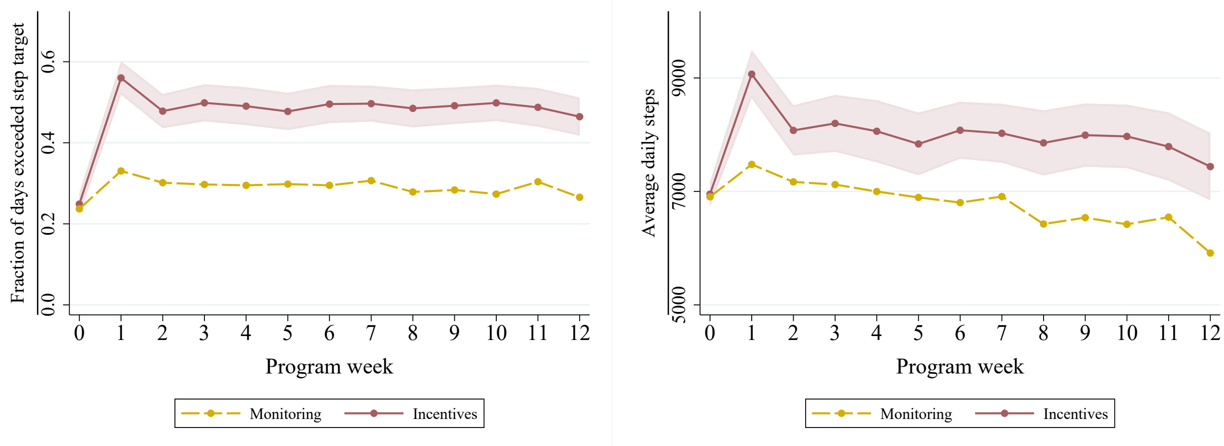
Our exercise treatment effects are robust to accounting for missing data from failure to wear pedometers. Column 3 of Table 2 reports impacts on daily steps treating days with no steps recorded as missing (which gives an unbiased estimate if participants randomly choose not to wear pedometers), and Table A.5 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 (Online App. Table E.3).

6.2 Persistence of Exercise Effects

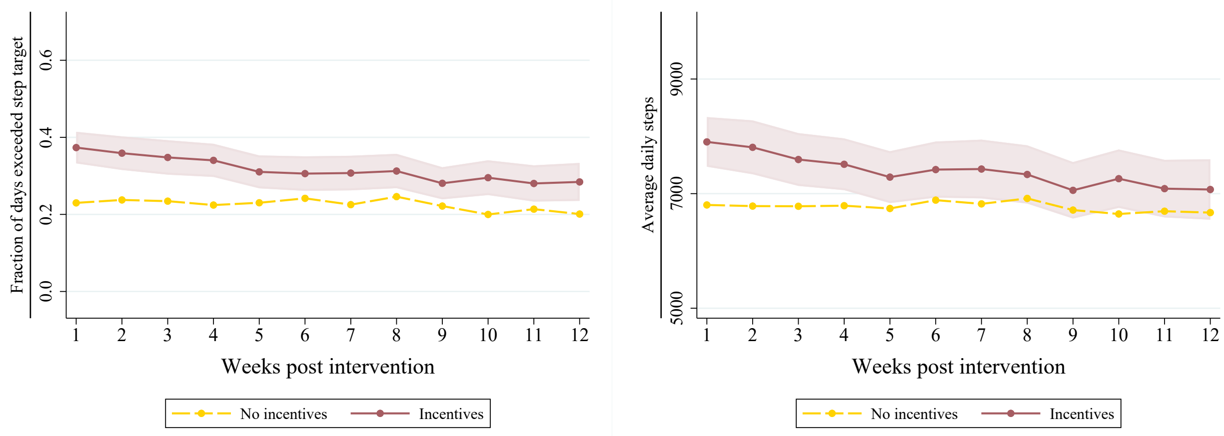
We now analyze how the exercise impacts evolve over time, both during and after the intervention. We begin with their evolution during the intervention. Panels A and B of Figure 10 show that after an initial spike at week 1, the effect of incentives on walking remains stable during the full intervention period. This suggests that policymakers could extend the program further with similar effects, an interesting finding since insurers and governments are increasingly rolling out longer-term (and even permanent) incentive programs.

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 (e.g., Royer et al., 2015; Bachireddy et al., 2019;

[37] Our estimate represents the effect of incentives relative to monitoring. Because monitoring itself may have a positive impact, the estimate is likely conservative for the overall impact of incentives on exercise. That said, a pre/post comparison shows no evidence that monitoring increases steps (Online Appendix J).



(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 10: Incentive Effects are Steady through the 12-Week Program and Persist Afterward

Notes: Panels A and C show the average probability of exceeding the step target and Panels B and D show the average daily steps walked. Panels A and B depict the intervention period and Panels C and D depict the 12 weeks subsequent to the intervention. Week 0 in Panels A and B is the phase-in period (before randomization). "No incentives" in Panels C and D represents the pooled monitoring and control groups. The shaded areas represent 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. Panels A and B are unconditional on wearing whereas Panels C and D are not, as described in footnote 38.

Patel et al., 2016; Charness and Gneezy, 2009). Panels C and D of Figure 10 depict the difference between our incentive and the pooled comparison groups for the 12 weeks after the intervention ended.[38] The incentive group walks significantly more even after incentives

[38]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 alone). While average pedometer-wearing rates declined from 87% in the intervention period to 69% post-intervention, post-intervention wearing rates are balanced across arms (Online App. Table E.4), and our results are robust to a Lee bounds exercise (Online App. Table E.5). We focus on results conditional on wearing the pedometer for greater comparability with intervention period effects but unconditional results also show persistence (Table A.12).

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: 60% and 40% as large as the intervention period effects, respectively.[39] 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:

$$y_i = \alpha + \beta_1 \times incentives_i + \beta_2 \times monitoring_i + \boldsymbol{X}_i'\gamma + \varepsilon_i, \tag{9}$$

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 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 effects for each outcome individually, we focus on the indices to infer effectiveness.

Table 4 shows that the incentive program moderately improves health. Column 1 presents the treatment effect on the health risk index. Panel A shows that 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 severity. We find stronger effects among those with more severe diabetes, although we cannot reject equality.

Table A.13 examines whether the intervention had coincident impacts on mental health or fitness. While RCTs show that exercise improves depression among the diagnosed, 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. Incentives improve the mental health index by 0.16 SD (p-value$<$ .05). In contrast, we find no effect on physical fitness, perhaps because we could only measure high-intensity fitness

[39] For reasons described in footnote 38, both the intervention period treatment effect and post-intervention period treatment effect estimates that we use to create the 60% and 40% persistence estimates come from specifications that condition on individuals wearing the pedometer. See Table A.12 notes for details.

[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).

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]	[3.52]	[0.43]	[0.041]	[0.28]
Monitoring	0.014	-0.13	1.63	1.08	0.064	0.00080
	[0.045]	[0.12]	[6.65]	[0.74]	[0.083]	[0.44]
P-value: M = I	0.14	0.55	0.22	0.13	0.14	0.63
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). 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). 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%

while our intervention motivated low-intensity exercise. Finally, we do not find impacts on diet or addictive good consumption (Online App. Table E.6).

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 naïve, thus broadening the arsenal for motivating impatient or time-inconsistent agents. The intuition behind the prediction is that agents who discount their future effort more place a higher value on future work opportunities. Time-bundled contracts link better future work opportunities with effort today, thus providing particular motivation for the impatient to exert more effort today. 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.

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 interesting topic to study is how to optimize the specific features of time-bundled contracts such as the payment period length T and threshold level K.

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 finding that time-bundled thresholds increase variance also has important policy implications. This finding suggests that policymakers and principals should use threshold contracts more in settings where the benefits of compliance are linear (or convex) than in settings where they are concave.

Our analysis of payment frequency also raises questions for future work. Increasing payment frequency is not effective in our setting because participants have limited impatience over payments. This finding suggests that, contrary to conventional wisdom, more frequent rewards are not always better, but 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 a population suffering from chronic disease. 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., M. Callen, Y. Khan, K. Jaffar, and C. Sprenger (2018). Using preference estimates to customize incentives: An application to polio vaccination drives in Pakistan. *NBER Working Paper 22019*, 1–66.

Andreoni, J., C. Gravert, M. A. Kuhn, S. Saccardo, and Y. Yang (2018). Arbitrage or narrow bracketing? On using money to measure intertemporal preferences. *NBER Working Paper Series No. 25232*.

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

Ashraf, N., D. S. Karlan, and W. Yin (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., M. Niederle, and C. Sprenger (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics 130*(3), 1067–1115.

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

Aziz, Z., P. Absetz, J. Oldroyd, N. P. Pronk, and B. Oldenburg (2015). A systematic review of real-world diabetes prevention programs: Learnings from the last 15 years. *Implementation Science 10*(1).

Bachireddy, C., A. Joung, L. K. John, F. Gino, B. Tuckfield, L. Foschini, and K. L. Milkman (2019). Effect of different financial incentive structures on promoting physical activity among adults: A randomized controlled trial. *JAMA Network Open 2*(8).

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

Bassett, D. R., H. R. Wyatt, H. Thompson, J. C. Peters, and J. O. Hill (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 A. J. Rothman (2018). Comparing types of financial incentives to promote walking: An experimental test. *Applied Psychology: Health and Well-Being 10*(2), 193–214.

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

Carrera, M., H. Royer, M. F. Stehr, J. R. Sydnor, and D. Taubinsky (2020). Who chooses commitment? evidence and welfare implications. *NBER Working Paper*.

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

Casaburi, L. and R. Macchiavello (2019). Demand and supply of infrequent payments as a commitment device: Evidence from Kenya. *American Economic Review 109*(2), 523–555.

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

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

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

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

Finkelstein, E. A., B. A. Haaland, M. Bilger, A. Sahasranaman, R. A. Sloan, E. E. K. Nang, and K. R. Evenson (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 A. D. Bryan (2017). Monetary incentive interventions can enhance psychological factors related to fruit and vegetable consumption. *Annals of Behavioral Medicine 51*, 599–609.

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

Iachine, I., H. C. Petersen, and K. O. Kyvik (2010). Robust tests for the equality of variances for clustered data. *Journal of Statistical Computation and Simulation 80*(4), 365–377.

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

Jacob, B. A. and S. D. Levitt (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.

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

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

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Kremer, M., G. Rao, and F. Schilbach (2019). Behavioral development economics. In B. D. Bernheim, S. Dellavigna, and D. Laibson (Eds.), *Handbook of Behavioral Economics - Foundations and Applications 2*, Volume 2, Chapter 5, pp. 345–458. Elsevier B.V.

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

Larkin, I. and S. Leider (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 M. Rabin (1999a). Doing it now or later. *American Economic Review 89*(1), 103–124.

O'Donoghue, T. and M. Rabin (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., D. A. Asch, R. Rosin, D. S. Small, S. L. Bellamy, J. Heuer, S. Sproat, C. Hyson, N. Haff, S. M. Lee, L. Wesby, K. Hoffer, D. Shuttleworth, D. H. Taylor, V. Hilbert, J. Zhu, L. Yang, X. Wang, and K. G. Volpp (2016, mar). Framing financial incentives to increase physical activity among overweight and obese adults. *Annals of Internal Medicine 164*(6), 385.

Qiu, S., X. Cai, U. Schumann, M. Velders, Z. Sun, and J. M. Steinacker (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., C. Niermann, D. Jekauc, and A. Woll (2013). Long-term health benefits of physical activity - a systematic review of longitudinal studies. *BMC Public Health 13*(1), 1–9.

Royer, H., M. Stehr, and J. Sydnor (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.

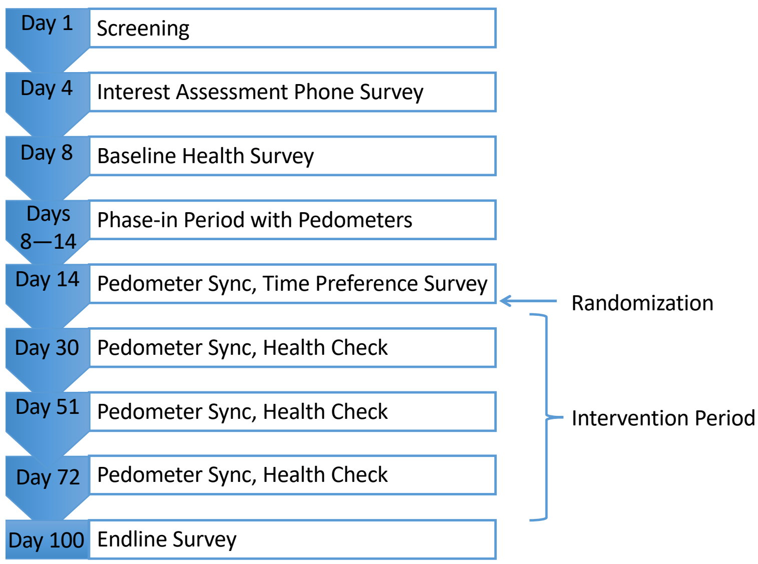
Volpp, K. G., L. K. John, A. B. Troxel, L. A. Norton, J. Fassbender, and G. Loewenstein (2008). Financial incentive based approaches for weight loss: A randomized trial. *JAMA: the Journal of the American Medical Association 300*(22), 2631–2637.

Warburton, D. E., C. W. Nicol, and S. S. Bredin (2006). Health benefits of physical activity: The evidence. *Canadian Medical Association Journal 174*(6), 801–809.

Appendices

This section contains all appendix tables and appendix figures labeled with the prefix "A" (e.g., Table A.1, Figure A.1). It also contains Appendices B, C, and D.
The Online Appendix is a separate document which contains Appendices E - L. It is 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: Measures of impatience over effort correlate in the expected direction with baseline measures of exercise, health, and behavior

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.3: No correlation between measures of impatience over effort and recharges

Covariate type:	Recharge variables				Credit constraint proxies		
Dependent variable:	Negative mobile balance	Negative yesterday's talk time	Prefers daily payment (=1)	Prefers monthly payment (=-1)	Negative wealth index	Negative monthly household income	# Individuals
Impatience Index	0.033	-0.075	-0.035	0.033	0.044*	0.037*	1760
Predicted Impatience Index	0.021	-0.012	-0.003	-0.005	-0.034*	0.025	3232

Notes: This table displays the correlations between the predicted and actual impatience indices meant to capture impatience over effort (in the rows) and baseline measures meant to proxy for the discount rates over recharges (in columns). We asked participants whether they preferred daily, weekly or monthly payments, and "Prefers Daily" ("Prefers Monthly") is an indicator that their most preferred frequency was daily (monthly). We normalize all impatience variables so that a higher value corresponds to greater impatience; hence the prediction is that coefficients should be positive if there is indeed a correlation. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.4: 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]
# 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: 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.5: Lee bounds on the impacts of incentives on exercise during the intervention

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. 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.6: Threshold treatments increase cost-effectiveness relative to the base case, with similar increases among those who are more and less impatient

		Sample defined by impatience indices			
	Full Sample	Below Median (Actual)	Above Median (Actual)	Below Median (Predicted)	Above Median (Predicted)
Treatment group	(1)	(2)	(3)	(4)	(5)
Base Case	0.050	0.050	0.050	0.050	0.050
Threshold	0.056	0.056	0.057	0.056	0.056
4-day Threshold	0.055	0.055	0.056	0.055	0.055
5-day Threshold	0.059	0.058	0.059	0.059	0.058

Notes: The table displays the cost-effectiveness of different treatment groups (in rows) and different samples (in columns). Cost-effectiveness equals average compliance divided by the average payment per day and so the units are days complied per INR. The Threshold group pools the 4-day and 5-day threshold groups.

Appendix Table A.7: Threshold Contracts Increase Dispersion Across Weeks

	Variance Measure
Threshold	0.15***
	[0.03]
Base Case mean	2.52
# Individuals	2,557
# Observations	29,189

Notes: Observations are individual × week. The sample includes the Threshold and Base Case groups. We follow Iachine et al. (2010) and define the variance measure as: $z_{ijk} = |x_{ijk} - \bar{x}_{ik}|$, where z_{ijk} is the number of days walked by individual i in treatment group j in week k. We then regress this measure on a Threshold group dummy; the null hypothesis (which we reject) is that the Threshold coefficient is 0 and that variance is equal across treatment groups. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.8: Robustness of Threshold Heterogeneity Results

Impatience measure:	Impatience index	Above median impatience index	Predicted impatience index	Above median predicted index
A. Dependent Variable = Steps				
Impatience × Threshold	289*	525	238**	521**
	[-41, 619]	[-268,1318]	[6, 420]	[6, 886]
Threshold	-143	-369	-166	-360**
	[-442, 157]	[-899, 161]	[-383, 50]	[-615, -69]
Impatience	-209	-397	-229***	-549***
	[-474, 56]	[-944, 150]	[-377, -69]	[-807, -134]
Base case mean	8,098	8,098	8,131	8,131
B. Dependent Variable = Exceeded Step Target (×100)				
Impatience × 5-day Threshold	3.52*	5.47	3.66***	7.29**
	[-0.05, 7.08]	[-2.28,13.22]	[1.23, 5.68]	[0.87, 11.16]
5-day Threshold	-1.72	-3.98	-1.71	-4.42***
	[-5.16, 1.73]	[-9.78,1.82]	[-4.02, 0.61]	[-7.11, -1.08]
Impatience × 4-day Threshold	5.00*	7.49	1.76	2.51
	[-0.94, 10.94]	[-3.93,18.91]	[-1.36, 4.93]	[-4.32, 8.48]
4-day Threshold	-0.14	-3.39	0.17	-0.84
	[-4.53, 4.26]	[-10.59,3.81]	[-3.17, 3.33]	[-4.68, 3.13]
Impatience	-2.97**	-4.68*	-2.39***	-5.32***
	[-5.36, -0.58]	[-9.45,0.10]	[-3.86, -0.78]	[-8.06, -0.99]
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. 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. Specifications in columns 1 and 2 include only participants who were enrolled after we started measuring the impatience index; columns 3 and 4 include everyone. The Threshold group pools the 4- and 5-day threshold groups. Bootstrap draws were done at the individual 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 $Threshold$ 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.9: Time preference heterogeneity robust to including other controls

Dependent variable:	Exceeded step target (×100)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
A. Predicted impatience index											
Predicted index × Threshold	3.19***	3.31***	3.19***	3.34***	3.21***	3.43***	3.28***	3.19***	3.11***	3.15***	3.17***
	[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]	[1.14,4.70]	[1.22,4.64]	[1.23,4.82]
Predicted index	-2.42***	-2.47***	-2.42***	-2.43***	-2.42***	-2.57***	-2.44***	-2.38***	-2.37***	-2.44***	-2.41***
	[-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]	[-3.49,-0.98]	[-3.56,-0.99]	[-3.58,-1.03]
Threshold	-1.26	-11.80**	-1.07	-1.27	-2.10*	-1.74	-4.34**	2.01	-1.07	-1.48	-4.26*
	[-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]	[-2.20,1.79]	[-3.63,0.50]	[-7.64,0.14]
Threshold × Covariate		0.213**	-0.470	-1.181	0.029*	0.075	1.133**	-0.901	-0.189**	0.612	0.429*
		[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]	[-2.372,-0.022]	[-3.106,5.207]	[-0.064,0.838]
Covariate		1.442**	1.131	-2.025*	-0.008	-0.286***	-0.413	1.808**	-0.821*	3.368**	4.195***
		[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]	[-1.735,0.328]	[0.165,6.090]	[3.904,4.564]
# Individuals	1959	1959	1959	1959	1959	1871	1957	1957	1959	1959	1959
# Observations	157120	157120	157120	157120	157120	150087	156973	156973	157120	157120	157120
B. Impatience index											
Impatience index × Threshold	3.97**	4.07**	3.96**	3.99***	4.13**	4.70***	3.99**	3.69**	4.26***	3.99**	4.14***
	[0.88,7.05]	[0.96,7.17]	[0.84,7.09]	[0.97,7.02]	[0.97,7.28]	[1.41,7.99]	[0.68,7.29]	[0.48,6.90]	[1.10,7.42]	[0.90,7.09]	[1.01,7.28]
Impatience index	-2.97**	-3.01**	-2.97**	-3.00**	-3.07**	-3.36**	-2.97**	-2.68**	-3.27**	-2.98**	-3.04**
	[-5.68,-0.26]	[-5.73,-0.29]	[-5.71,-0.24]	[-5.76,-0.23]	[-5.76,-0.38]	[-6.16,-0.57]	[-5.62,-0.32]	[-5.27,-0.09]	[-5.99,-0.55]	[-5.69,-0.27]	[-5.75,-0.33]
Threshold	-1.32	-10.1	-1.05	-1.31	-3.12	-1.32	-1.61	7.20	-4.38*	-1.38	-5.18
	[-4.74,2.09]	[-32.92,12.69]	[-6.31,4.21]	[-4.84,2.22]	[-6.99,0.74]	[-6.53,3.88]	[-8.03,4.80]	[-10.87,25.28]	[-8.92,0.16]	[-5.69,2.92]	[-13.17,2.82]
Threshold × Covariate		0.18	-0.64	2.36	0.063**	0.050	0.059	-2.32	2.71*	-0.043	0.55
		[-0.26,0.62]	[-8.51,7.24]	[-4.93,9.65]	[0.00,0.12]	[-0.41,0.51]	[-2.00,2.12]	[-7.27,2.62]	[-0.39,5.80]	[-10.21,10.13]	[-0.52,1.62]
Covariate		1.73	-0.88	-4.83	-0.036	-0.56***	0.29	1.96	-2.39*	2.90	4.11***
		[-0.83,4.28]	[-6.46,4.70]	[-11.43,1.78]	[-0.08,0.01]	[-0.87,-0.26]	[-1.38,1.97]	[-1.56,5.49]	[-5.09,0.31]	[-3.32,9.12]	[3.23,4.98]
Covariate used	-	Age	Female	Health risk index	Mobile balance (INR)	Yesterday's talk time (INR)	Risk aversion	Scheduling certainty	Monthly personal income (10000s INR)	Education (above median)	Baseline steps (÷1000)
# Individuals	1070	1070	1070	1070	1070	1042	1069	1069	1070	1070	1070
# Observations	85795	85795	85795	85795	85795	83519	85711	85711	85795	85795	85795
Base case mean	49.77	49.77	49.77	49.77	49.77	49.92	49.73	49.73	49.77	49.77	49.77

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. Some covariates have missing values and so their respective columns have fewer observations. Income variables are winsorized at the 5th and 95th percentiles. The unit of observation is a respondent × day. Bootstrap draws were done at the individual level, and bootstrapped 95% confidence intervals are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.10: Walking does not vary significantly across the pay cycle

Dependent variable:	Exceeded 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.11: Threshold contracts do not significantly decrease satisfaction at endline

Dependent variable:	Interest in continuing program	
Threshold	-0.0222	-0.0117
	[0.0150]	[0.0194]
Above median predicted impatience × Threshold		-0.0266
		[0.0300]
Above median predicted 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.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	0.23***	1072.8***	0.22***	1193.4***
	[0.04]	[347.77]	[0.03]	[372.70]
No incentives mean	0.3	7347.4	0.2	5687.4
% Persistence	40.0	60.4	32.6	45.0
# Individuals	903	903	903	903

Note: Table shows the average treatment effect of incentives during the post-intervention period. The omitted group is the monitoring and control groups (pooled). Each observation is a person-day; columns 1 and 2 only include days where the participant wore the pedometer (i.e., had step count > 0) and columns 3 and 4 include all days. The % Persistence row shows the treatment effect from the post-intervention period divided by the 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. The sample was enrolled in phases over time, and only those in the last phase were enrolled in the post-intervention period; we limit to that same sample for calculating the relevant intervention period treatment effect. If we calculate the % Persistence by using the full sample treatment effect, the percentages would be: 43% , 56% , 36% , 43% 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: 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 variable at baseline. Robust standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

B Theoretical Predictions Appendix

B.1 Agent Problem

Given the notation and assumptions in Section 2.1, we can express the agent's problem as follows. On day t, the agent chooses compliance, w_t, to maximize expected discounted payments net of effort costs:

$$\max_{w_t \in \{0,1\}} \mathbb{E}\left[d^{(T-t)} m_T - \sum_{j=t+1}^{T} \delta^{(j-t)} w_{j,t} e_j \middle| e_1, .., e_t, w_1, .., w_t \right] - w_t e_t, \tag{10}$$

where the expectation over future discounted payment and future discounted effort depends on the history of effort costs $(e_1, .., e_t)$ and compliance decisions $(w_1, .., w_t)$ through time t, and where $w_{j,t}$ represents the agent's prediction on day t about her compliance on day j.

Denoting $\mathbb{E}\left[d^{(T-t)} m_T - \sum_{j=t+1}^{T} \delta^{(j-t)} w_{j,t} e_j \middle| e_1, .., e_t, w_1, .., w_t \right]$ as $V_t(w_t)$, the agent will thus choose to set $w_t = 1$ (i.e., comply on day t) if the following holds:

$$\begin{gathered} V_t(0) < V_t(1) - e_t \\ \text{or} \\ e_t < V_t(1) - V_t(0). \end{gathered} \tag{11}$$

That is, on day t, the agent complies if the continuation value of complying net of the effort cost is greater than the continuation value of not complying.

B.2 The Effectiveness of Threshold and Linear Contracts

In this section, we compare the effectiveness of threshold and linear contracts under a range of effort cost assumptions, paying particular attention to how the relative effectiveness of thresholds depends on δ. For simplicity, throughout the section, we assume that $T = 2$ and that $K = 2$ and denote the threshold payment as M (i.e., $M = 2m'$). Our first proposition examines the relative performance of the contracts in the limit as δ goes to 0 under very general assumptions. It shows that, for sufficiently low δ, for any linear contract, there exists a threshold contract that achieves substantially higher cost-effectiveness with relatively little – and potentially even no – loss in compliance. In contrast, for any linear contract, one can always construct another *linear* contract with substantially higher cost-effectiveness by decreasing the payment amount, but the loss in compliance may be arbitrarily large.

Proposition 1. *Let $d = 1$ and $T = 2$. Fix all parameters other than δ, and take a linear contract that induces compliance $C > 0$.*
(a) If agents are naïve, then for sufficiently small δ, there exists a threshold contract with $K = 2$ that has at least two times higher cost-effectiveness (and $1 + \frac{1}{C}$ times higher cost-effectiveness if costs are IID) and that generates compliance $\frac{1+C}{2}$ of the linear contract.
(b) If agents are sophisticated and costs are IID, then for sufficiently small δ, there exists a threshold contract with $K = 2$ that has at least $1 + C$ times higher cost-effectiveness and that generates compliance at least $\frac{1+C}{2}$ of the linear contract.

Proof. See Online Appendix F.1. □

The potential improvements from threshold contracts demonstrated by Proposition 1 are quantitatively large. For example, when costs are IID and agents are naïve with sufficiently low δ, for a linear contract that generates $C = .5$, there exists a threshold contract that generates at least 75% as much compliance with three times higher cost-effectiveness.

Next, we show results across the full range of δ (not just for δ sufficiently low). While we make additional assumptions on the effort distribution for tractability, our next two propositions demonstrate that thresholds can be effective for those who are impatient over effort in the two limiting cases of perfectly correlated and IID effort costs.

Proposition 2. *[Perfect Correlation] Let $T = 2$. Fix all parameters other than δ, and take any linear contract that induces compliance $C > 0$. Let there be perfect correlation in costs across days ($e_1 = e_2$). Then, regardless of agent type, there exists a threshold contract that induces compliance of at least C and that has approximately $2\frac{d}{1+\delta}$ times greater cost-effectiveness than the linear contract.*

Proof. See Online Appendix F.1. □

Proposition 3. *[IID Uniform] Let $d = 1$ and $T = 2$. Fix all parameters other than δ. Let costs be independently drawn each day from a uniform[0,1] distribution.*
(a) Take any threshold contract paying $0 < M < 1$. Whether there exists a more effective linear contract depends on δ. Define $\frac{(M+1)^2}{2}$ as the "cutoff value" for naïfs and $2 - \frac{4}{(1+M)^2}$ as the "cutoff value" for sophisticates. If δ is less than the cutoff value for a given type, there does not exist a linear contract that is more effective for that type; any linear contract with at least as high cost-effectiveness will generate strictly lower compliance.[41] *In contrast, if δ is greater than the cutoff value, there always exists a linear contract that is more effective than the threshold for that type; in particular, a contract with the same cost-effectiveness and strictly higher compliance.*
(b) Take any threshold contract paying $1 \leq M < 2$.[42] *Regardless of δ, there does not exist a linear contract that is more effective.*

Proof. See Online Appendix F.1. □

B.3 Threshold Contracts and Impatience Over Effort

In this section, we first present a series of propositions that provide the theoretical underpinning for Prediction 1 from Section 2.3.1. We then present the proposition that underlies Prediction 2 from Section 2.3.1.

Propositions Underlying Prediction 1 Propositions 4 - 8 demonstrate that, holding all else equal, both compliance and effectiveness in threshold contracts are decreasing in $\delta^{(t)}$. In contrast, with *separable* contracts, holding all else equal, compliance and effectiveness are both invariant to $\delta^{(t)}$ (Section 2.2). As a result, holding all else equal, as $\delta^{(t)}$ decreases, thresholds perform better relative to separable contracts with respect to compliance and effectiveness – as stated in Prediction 1.

[41] For sophisticates, when $M < \sqrt{2} - 1$, the cutoff value is negative and so δ is never below the cutoff.
[42] Note that the principal would never pay $M > 2$ since $M = 2$ achieves 100% compliance regardless of δ.

In Propositions 4 and 5, we examine threshold contracts with $K = T$ (i.e., where one must comply on all days in order to receive payment) without making any restrictions on the cost distributions. Proposition 4 shows that, when $T = 2$, both compliance and effectiveness in the threshold contract are weakly decreasing in δ under very general conditions. Proposition 5 then shows that the result that compliance is weakly decreasing in δ in a threshold contract goes through for any T when $T = K$.

To gain tractability to examine threshold contracts with $K < T$ and threshold effectiveness when $T > 2$, we then make additional assumptions about the effort cost distribution. Proposition 6 shows that, if costs are perfectly correlated over time, both effectiveness and compliance are weakly decreasing in the threshold contract for any $T > 2$ and any $K \leq T$. Finally, to relax the perfect correlation assumption, Proposition 7 examines a simplified version of the model where costs can either be high or low, all costs are known from day 1, $K = 2$ and $T = 3$. Again, we show that compliance and effectiveness are higher when $\delta^{(t)}$ is lower.[43]

Proposition 4 ($T = 2, K = 2$, Threshold Effectiveness and Impatience Over Effort)**.** *Let $T = 2$. Fix all parameters other than δ. Take any threshold contract with $K = 2$; denote the threshold payment M. Compliance in the threshold contract is weakly decreasing in δ. In addition, as long as there is not "too much" inframarginal behavior,[44] the effectiveness of the threshold contract is weakly decreasing in δ.*

Proof. We first examine compliance and then turn to effectiveness.

Compliance Recall that the condition for complying on day 1 is to comply if $e_1 \& < V_1(1) - V_1(0)$ (equation (11)). With the threshold contract, we have that:

$$V_1(1) - V_1(0) = \mathbb{E}\left[(dM - \delta e_2)w_{2,1}|e_1, w_1 = 1\right] - \mathbb{E}\left[-\delta e_2 w_{2,1}|e_1, w_1 = 0\right] \tag{12}$$

We examine this expression separately for sophisticates and naïfs.

For sophisticates, who accurately predict their own future behavior, $w_{2,1}|^{w_1=1} = \mathbb{1}\{e_2 < M\}$ and $w_{2,1}|^{w_1=0} = \mathbb{1}\{e_2 < 0\}$. Thus:

$$\begin{aligned} V_1(1) - V_1(0) &= \mathbb{E}\left[(dM - \delta e_2)w_{2,1}|e_1, w_1 = 1\right] - \mathbb{E}\left[-\delta e_2 w_{2,1}|e_1, w_1 = 0\right] \\ &= \mathbb{E}\left[(dM - \delta e_2)\mathbb{1}\{e_2 < M\} + \delta e_2 \mathbb{1}\{e_2 < 0\}|e_1\right] \end{aligned} \tag{13}$$

We show that this is weakly decreasing in δ by showing that the argument, $(dM - \delta e_2)\mathbb{1}\{e_2 < M\} + \delta e_2 \mathbb{1}\{e_2 < 0\}$, is weakly decreasing in δ for all values of e_2. There are two cases:

[43] Note that in this case K/T is relatively high. Thresholds where K/T is very low may not always be better for impatient naïfs than patient people because they include more days where current and future effort are substitutes, which can cause naïfs to procrastinate.

[44] See equation (18) for the exact condition. When there is too much inframarginal behavior, not only will the effectiveness prediction not hold but incentives also become less likely to be a cost-effective approach. The intuition for why high levels of inframarginal behavior (combined with low $\frac{\lambda}{M}$) can flip the effectiveness prediction is as follows. If there is inframarginal behavior, then the principal effectively gets "free" compliance if people comply on day 2 only and not day 1. As we will show, lower δ increases compliance by making people more likely to comply on day 1. The benefit is extra compliance and the cost is extra payment. The cost will be particularly large if there is a lot of inframarginal behavior on day 2, because now the principal has to pay out for all of the day 2's on which day 1 compliance was induced, which the principal used to get for free.

1. $e_2 > 0$: In this case, $(dM - \delta e_2)\mathbb{1}\{e_2 < M\} + \delta e_2 \mathbb{1}\{e_2 < 0\} = (dM - \delta e_2)\mathbb{1}\{e_2 < M\}$, which is weakly decreasing in δ.
2. $e_2 \leq 0$: In this case, $(dM - \delta e_2)\mathbb{1}\{e_2 < M\} + \delta e_2 \mathbb{1}\{e_2 < 0\} = (dM - \delta e_2) + \delta e_2 = dM$, which is invariant to δ.

Thus, equation (13) is weakly decreasing in δ. That means that day 1 compliance is decreasing in δ. Hence, day 2 compliance is as well since $w_2 = 1$ if both $w_1 = 1$ and $e_2 < M$, and w_1 is weakly decreasing in δ. Thus, compliance in the threshold contract is decreasing in δ for sophisticates.

We now turn to naïfs. For naïfs, who think their day 2 selves will share their day 1 preferences, $w_{2,1}|^{w_1=1} = \mathbb{1}\{\delta e_2 < dM\}$ and $w_{2,1}|^{w_1=0} = \mathbb{1}\{\delta e_2 < 0\}$. Thus:

$$\begin{aligned} V_1(1) - V_1(0) &= \mathbb{E}\left[(dM - \delta e_2)w_{2,1}|e_1, w_1 = 1\right] - \mathbb{E}\left[-\delta e_2 w_{2,1}|e_1, w_1 = 0\right] \\ &= \mathbb{E}\left[(dM - \delta e_2)\mathbb{1}\{\delta e_2 < dM\} + \delta e_2 \mathbb{1}\{\delta e_2 < 0\}|e_1\right] \\ &= \mathbb{E}\left[\max\{dM - \delta e_2, 0\} + \delta e_2 \mathbb{1}\{e_2 < 0\}|e_1\right] \end{aligned} \tag{14}$$

Again, we show that this is decreasing in δ by showing that the argument, $\max\{dM - \delta e_2, 0\} + \delta e_2 \mathbb{1}\{e_2 < 0\}$, is weakly decreasing in δ for all values of e_2. There are two cases:

1. $e_2 > 0$: In this case, $\max\{dM - \delta e_2, 0\} + \delta e_2 \mathbb{1}\{e_2 < 0\} = \max\{dM - \delta e_2, 0\}$, which is weakly decreasing in δ.
2. $e_2 \leq 0$: In this case, for $u = -e_2 \geq 0$, we have $\max\{dM - \delta e_2, 0\} + \delta e_2 \mathbb{1}\{e_2 < 0\} = \max\{dM + \delta u, 0\} - \delta u = (dM + \delta u) - \delta u = dM$ which is invariant to δ.

Thus, equation (14) is weakly decreasing in δ. Hence day 1 compliance (and hence day 2 and total compliance) are also decreasing in δ for naïfs.

Effectiveness We first show that, if costs are positive, cost-effectiveness in the threshold is not increasing in δ. Because we already showed that compliance is decreasing in δ, this establishes that effectiveness is decreasing in δ when costs are positive. We then show sufficient conditions for threshold effectiveness to decrease in δ when costs can be negative.

To simplify notation, let e^* be the agent's cutoff value for complying in period 1, such that agents comply in period 1 if $e_1 < e^*$. From equations (13) and (14), we know that the value of e^* will depend on the agent's sophistication and, importantly, decrease in δ.

With our new notation, we can write the compliance decisions as:

$$\begin{aligned} w_1 &= \mathbb{1}\{e_1 < e^*\} \\ w_2 &= w_1 \mathbb{1}\{e_2 < M\} + (1 - w_1)\mathbb{1}\{e_2 < 0\} \\ &= w_1 \mathbb{1}\{0 < e_2 < M\} + \mathbb{1}\{e_2 < 0\} \end{aligned}$$

A Special Case: Positive Costs We first examine the restricted case where $e_1 > 0$ and $e_2 > 0$ and show that, in that case, C/P is not increasing in δ. In that case, $w_2 = w_1 w_2$. Therefore we have:

$$C/P = \frac{1}{M}\frac{\mathbb{E}[w_1+w_2]}{\mathbb{E}[w_1w_2]} = \frac{1}{M}\frac{\mathbb{E}[w_1+w_1w_2]}{\mathbb{E}[w_1w_2]} = \frac{1}{M}\left(\frac{\mathbb{E}[w_1]}{\mathbb{E}[w_1w_2]}+1\right) = \frac{1}{M}\left(\frac{\mathbb{E}[w_1]}{\mathbb{E}[w_1]\mathbb{E}[w_2|w_1=1]}+1\right)$$
$$= \frac{1}{M}\left(\frac{1}{\mathbb{E}[w_2|w_1=1]}+1\right) \tag{15}$$

Consider the first term, $\frac{1}{\mathbb{E}[w_2|w_1=1]}$. To show this is not increasing in δ, we show that $\mathbb{E}[w_2|w_1 = 1] = \mathbb{E}[\mathbb{1}\{e_2 < M\}|w_1 = 1]$ is weakly increasing in δ. Call this expression p_2^*. If costs were IID, then $p_2^* = F(M)$, which is independent of δ. To see that p_2^* is also weakly increasing in δ under our more general assumption that e_2 is weakly increasing in e_1, note that higher δ means that $w_1 = 1$ will be associated with lower values of e_1 (since e^* is decreasing in δ). This implies lower values of e_2 conditional on $w_1 = 1$, since we assume that e_2 is weakly increasing in e_1. Lower values of e_2 then mean that $p_2^* = E[w_2|w_1 = 1]$ will be weakly higher. Hence, p_2^* is weakly increasing in δ and the first term is weakly decreasing in δ. Thus, we have shown that, with positive costs, C/P is weakly decreasing in δ.

General Case Instead of using cost-effectiveness as a means to prove the result for effectiveness, we turn to the expression for effectiveness directly: $\lambda C - P$. We show the conditions under which it is weakly increasing in e^*, and hence weakly decreasing in δ.

First, we rewrite the expression for effectiveness under the threshold given what we know about C and P. (For notational simplicity, we examine $2(\lambda C - P)$ instead of $\lambda C - P$.)

$$\begin{aligned}
2(\lambda C - P) &= \lambda\mathbb{E}[w_1+w_2] - M\mathbb{E}[w_1w_2] \\
&= \lambda\left(F(e^*) + \mathbb{E}[w_1\mathbb{1}\{0<e_2<M\} + \mathbb{1}\{e_2<0\}]\right) - M\mathbb{E}[w_1\mathbb{1}\{e_2<M\}] \\
&= \lambda\left(F(e^*) + \mathbb{E}[\mathbb{1}\{e_1<e^*\}\mathbb{1}\{0<e_2<M\} + \mathbb{1}\{e_2<0\}]\right) - M\mathbb{E}[\mathbb{1}\{e_1<e^*\}\mathbb{1}\{e_2<M\}] \\
&= \lambda\left(F(e^*) + Prob(e_1<e^*, 0<e_2<M) + Prob(e_2<0)\right) - MProb(e_1<e^*, e_2<M).
\end{aligned} \tag{16}$$

We now take a derivative with respect to e^*. Let $g(e^*) = Prob(e_1 \le e^*, e_2 \in S)$, where S is some set. It is straightforward to show that $g'(e^*) = f(e^*)\, Prob(e_2 \in S|e_1 = e^*)$.[45] Thus, we have

$$\frac{d}{de^*}[2(\lambda C - P)] = \lambda[f(e^*) + f(e^*)Prob(0<e_2<M|e_1=e^*)] - Mf(e^*)Prob(e_2<M|e_1=e^*)$$

Hence, a sufficient condition for effectiveness to increase in e^* (and decrease in δ) is:

$$\lambda(1 + Prob(0<e_2<M|e_1=e^*)) \ge MProb(e_2<M|e_1=e^*) \tag{17}$$

or

$$\frac{\lambda}{M}(1 + Prob(0<e_2<M|e_1=e^*)) \ge Prob(e_2<0|e_1=e^*) + Prob(0<e_2<M|e_1=e^*)$$

[45] To show this, note that

$$\begin{aligned}
g(e^*+\epsilon) - g(e^*) &= Prob(e^* < e_1 \le e^*+\epsilon, e_2 \in S) = Prob(e^* < e_1 < e^*+\epsilon)Prob(e_2 \in S|e^* < e_1 \le e^*+\epsilon) \\
&= (F(e^*+\epsilon) - F(e^*))\, Prob(e_2 \in S|e^* < e_1 \le e^*+\epsilon).
\end{aligned}$$

Dividing by ϵ gives us: $\frac{g(e^*+\epsilon)-g(e^*)}{\epsilon} = \frac{(F(e^*+\epsilon)-F(e^*))}{\epsilon}Prob(e_2 \in S|e^* < e_1 \le e^*+\epsilon)$. Letting ϵ go to 0 and using the definition of the derivative gives that $g'(e^*) = f(e^*)\, Prob(e_2 \in S|e_1 = e^*)$.

or

$$Prob(e_2 < 0|e_1 = e^*) \leq \frac{\lambda}{M} + \left(\frac{\lambda}{M} - 1\right) Prob(0 < e_2 < M|e_1 = e^*)). \tag{18}$$

If $\lambda > M$, condition (18) will always hold. More broadly, the condition will be more likely to hold the greater λ relative to M. The condition essentially guarantees that there not be "too much" inframarginal behavior, which generally decreases the efficacy of incentives. For example, when $\lambda > M/2$, which is a reasonable condition as it guarantees that the payment to the agent for two days of compliance is less than the benefits to the principal, a sufficient condition is

$$Prob(e_2 < 0|e_1 = e^*) < Prob(e_2 > M|e_1 = e^*).$$

We have thus showed that, as long as there is not "too much" inframarginal behavior (i.e, as long as equation (18) holds), the effectiveness of a threshold contract is decreasing in δ. □

We now show that the compliance result also goes through when we increase the length of the payment period, T, maintaining the assumption that $K = T$.

Proposition 5 ($T = K$, Threshold Compliance and Impatience Over Effort)**.** *Let $T > 1$. Fix all parameters other than $\delta^{(t)}$. Take any threshold contract with threshold level $K = T$; denote the threshold payment M. Compliance in the threshold contract will be weakly decreasing in $\delta^{(t)}$ for all $t \leq T - 1$.*

Proof. See Online Appendix F.2. □

Proposition 6 (Perfect correlation, Threshold Effectiveness and Impatience over Effort)**.** *Let there be perfect correlation in costs across periods ($e_t = e_{t'} \equiv e$ for all t, t'). For simplicity, let $\delta^{(t)} < 1$ for all $t > 0$ if $\delta^{(t)} < 1$ for any t. Fix all parameters other than $\delta^{(t)}$ for some $t \leq T - 1$. Take any threshold contract with threshold level $K \leq T$. Compliance and effectiveness in the separable contract will be constant with $\delta^{(t)}$. In contrast, compliance and effectiveness in the threshold contract will be weakly decreasing in $\delta^{(t)}$. Hence, compliance and effectiveness in the threshold relative to separable contract will be decreasing in $\delta^{(t)}$.*

Proof. See Online Appendix F.2. □

To make the problem more tractable when costs are not perfectly correlated, we now consider a simplified model where $T = 3$, $K = 2$, costs take on only two values (high or low), discount factors are exponential, and agents observe all future cost realizations on day 1.

Proposition 7. *Let $T = 3$. Let the cost of effort on each day be binary, taking on either a "high value" (e_H) or a "low value" (e_L), with $e_H \geq e_L$. Let agents observe the full sequence of costs e_1, e_2, e_3 on day 1. Let $\delta^{(t)} = \delta^t$ (i.e., let the discount factor over effort be exponential) and let $d^{(t)} = 1$. Fix all parameters other than δ. Consider a threshold contract with $K = 2$, where the agent must thus comply on at least 2 days in order to receive payment. Compliance and effectiveness in the threshold contract are weakly higher for someone with a discount factor $\delta < 1$ than for someone with discount factor $\delta = 1$.*

Proof. See Online Appendix F.2. □

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

Proposition 8. *Let $T = 3$. Let costs be weakly positive and let agents observe the full sequence of costs e_1, e_2, e_3 on day 1. Let $\delta^{(t)} = \delta^t$ (i.e., let the discount factor over effort be exponential) and let $d^{(t)} = 1$. Fix all parameters other than δ. Consider a threshold contract with $K = 2$, where the agent must thus comply on at least 2 days in order to receive payment. For sophisticates, compliance and effectiveness in the threshold contract are weakly decreasing in the discount factor δ.*

Proof. See Online Appendix F.2. □

Proposition Underlying Prediction 2 We now present the result that motivates Prediction 2 from Section 2.3.1 regarding the greater variance of compliance in threshold than linear contracts.

Proposition 9 (Thresholds and Variance)**.** *Let $d = 1$ and $T = 2$. Let costs be independent across days and weakly positive. Take a threshold contract with $K = 2$ and a linear contract that generates the same level of compliance as the threshold contract, $C > 0$. The threshold contract has higher variance in compliance than the linear contract.*

Proof. See Online Appendix F.2. □

B.4 Other Types of Two-Day Time-Bundled Contracts

This section examines the full space of two-day time-bundled contracts. We begin by describing additional notation and modeling assumptions. Next, we formally define the option and commitment values described in Section 2.3.3, and characterize which time-bundled contracts can have option and commitment value. Finally, we present three formal results. First, we show that contracts must have option value to generate more effort from naïfs than from patient people. Second, we show that contracts must have commitment value to generate more effort from sophisticates than from patient people. Finally, we show that contracts can generate more effort from a sophisticated than naïve person only if day 1 compliance is analogous to taking up a commitment contract for future compliance.

Assumptions and Notation We begin from the setup presented in Section 2.1 with $T = 2$. For consistency with previous literature on contract design and impatience (e.g., Bai et al., 2020; O'Donoghue and Rabin, 1999a) we make the additional assumption that all effort costs are revealed on day 1. Finally, for notational simplicity we assume $d = 1$.[46] Thus, an impatient person is one with $\delta < 1$, and a patient person is one with $\delta = 1$.

Three parameters define a two-day contract:

1. m_1: the payment for day 1 compliance.
2. m_{2L}: the payment for day 2 compliance if the agent did *not* comply on day 1.
3. m_{2H}: the payment for day 2 compliance if the agent *did* comply on day 1.

[46] The necessary assumption is that $\delta \leq d$, which implies that agents are weakly more eager for their future self to undertake effort than from their present self. We would call an agent patient if $\delta = d$ and impatient if $\delta < d$.

For example, two-day thresholds have $m_1 = 0, m_{2L} = 0$, and $m_{2H} = M$.

As we will show, m_{2L} and m_{2H} are the two key contract parameters that interact with time preferences. Time-bundled contracts are defined by a dynamic complementarity, i.e., $m_{2H} > m_{2L}$.

Definitions of Option and Commitment Value We say a contract holds *option value* for an agent if day 1 compliance under the contract is pivotal to the agent *wanting* to comply on day 2. This occurs when the Option Value Condition is met:

$$\text{Option Value Condition: } m_{2L} \leq \delta e_2 < m_{2H}. \tag{19}$$

In this case, complying on day 1 creates an option for future paid compliance that the day 1 self wants the day 2 self to follow through on. That is, without day 1 compliance ($w_1 = 0$), the agent does not want her day 2 self to comply since $\delta e_2 \geq m_{2L}$. If instead the agent were to comply on day 1 ($w_1 = 1$), she does want her day 2 self to comply since $\delta e_2 < m_{2H}$.

In contrast, a contract holds *commitment value* for an agent if day 1 compliance is pivotal to the agent *actually* following through on day 2. This occurs when the Commitment Value Condition holds:

$$\text{Commitment Value Condition: } m_{2L} \leq e_2 < m_{2H}. \tag{20}$$

In this case, complying on day 1 commits the day 2 self to follow through: she will comply on day 2 if $w_1 = 1$ (since $e_2 < m_{2H}$) but will not comply on day 2 if $w_1 = 0$ (since $e_2 < m_{2L}$).

For patient agents, the option and commitment value conditions are equivalent. For impatient agents, however, a contract can hold option value only, commitment value only, both, or neither.

Contract Categories Columns (1) and (2) of Table B.1 divide the full space of time-bundled contracts into four categories. The categories are defined by whether the contracts meet the Option and Commitment Value conditions (equations (19) and (20)) for some $\delta < 1$, taking e_2 as given. Specifically, Option + Commitment contracts satisfy equations (19) and (20) for some $\delta < 1$, Commitment-only contracts satisfy equation (20) for some $\delta < 1$ but do not satisfy equation (19) for any $\delta < 1$, Option-only contracts satisfy equation (19) for some $\delta < 1$ but do not satisfy equation (20) for any $\delta < 1$, and Inframarginal contracts never satisfy either condition.

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

	Contract definitions		Compliance weakly decreasing in δ among:[a]	
	m_{2L}	m_{2H}	Naïfs	Sophisticates
Contract	(1)	(2)	(3)	(4)
Option + commitment	$< e_2$	$> e_2$	Y	Y
Commitment-only	e_2	$> e_2$	N	Y
Option-only	$< e_2$	$\leq e_2$	Y	N
Inframarginal	$> e_2$	$> e_2$	N	N

[a] If column says "N" for "No, compliance is not weakly decreasing in δ," it means that compliance is invariant to δ (as opposed to increasing in δ).

Results We now show three results. The first result shows that contracts with option value can induce extra effort from naïfs, and the second result shows that contracts with commitment value can induce extra effort from sophisticates (as summarized in Columns (3) and (4) of Table B.1.) Many *threshold* contracts are Option+commitment contracts, which explains why thresholds can work for both naïfs and sophisticates.[47] Our third result clarifies the relationship between our definitions of commitment value, option value and standard commitment contracts: when a contract has commitment value but not option value, then day 1 compliance is analogous to taking up a commitment contract for future compliance. As such, these contracts only work for sophisticates, not naïfs.

Proposition 10. *Fix all parameters other than δ. For naïfs, compliance is weakly decreasing in δ for Option-only and Option+ commitment contracts, but invariant to δ under Commitment-only and Inframarginal contracts.*

Proof. See Online Appendix F.3. □

Proposition 11. *Fix all parameters other than δ. For sophisticates, compliance is weakly decreasing in δ for Commitment-only and Option+ commitment contracts, but invariant to δ under Option-only and Inframarginal contracts.*

Proof. See Online Appendix F.3. □

Corollary 1. *Fix all parameters other than δ. If a sophisticate with a given $\delta < 1$ complies on day 1 of a time-bundled contract when neither a naïf with the same δ nor a patient person with $\delta = 1$ would comply, then two things must be true:*
(1) The Commitment Value condition must hold (equation (20)) but the Option Value condition must not (equation (19)), which implies the following holds:

$$\textit{Commitment Value Without Option Value Condition: } \delta e_2 < m_{2L} \leq e_2 < m_{2H}. \tag{21}$$

(2) The sophisticate must be complying on day 1 even though $e_1 \geq m_1 + m_{2H} - m_{2L}$. That is, the sophisticate complies on day 1 even though the cost of day 1 effort exceeds the maximum potential financial benefit from day 1 compliance.

Proof. See Online Appendix F.3. □

Intuition and Discussion We can see the intuition behind the results by considering the condition for complying on day 1. The agent complies on day 1 if:[48]

$$e_1 < m_1 + \begin{cases} m_{2H} - m_{2L} & \text{if the individual believes on day 1 that she will comply on day 2 regardless of her day 1 behavior} \\ m_{2H} - \delta e_2 & \text{if the individual believes on day 1 that she will comply on day 2 if and only if she complies on day 1 } \textbf{(the ``pivotal'' case)} \\ 0 & \text{if the individual believes on day 1 that she will } \textit{not} \text{ comply on day 2 regardless of her day 1 behavior} \end{cases} \tag{22}$$

[47]Since thresholds have $m_1 = m_{2H} = 0$ and $m_{2H} = M$, they are Option+commitment contracts when $M > e_2$.
[48]This comes from equation (11) in Section 2.1 which says that the agent complies on day 1 if $e_1 < V_1(1) - V_1(0)$.

Equation 22 shows that, in order for compliance to depend on δ – and, in particular, in order for compliance to be weakly decreasing in δ – we must be in the pivotal case where the impatient individual believes on day 1 that she will comply on day 2 if and only if she complies on day 1.[49]

Propositions 10 and 11 rest on the fact that, because sophisticates and naïfs have different beliefs about their future compliance, different circumstances put them in the pivotal case. Naïfs believe that their future self will follow through on their current self's preferences, and are thus in the pivotal case when their day 1 compliance is pivotal to *wanting* their day-2 self to comply (when the Option Value Condition holds). In contrast, because sophisticates understand their future preferences, they are in the pivotal case when their day 1 compliance is pivotal to their day-2 self *actually* complying (when the Commitment Value Condition holds).

Two-day threshold contracts cannot satisfy the Commitment Value Condition without also satisfying the Option Value Condition (i.e., equation (21) cannot hold for thresholds). Equation (21) requires $m_{2L} > 0$, whereas threshold contracts have $m_{2L} = 0$ by definition. Thus, by Corollary 1, no two-day threshold contract will yield higher compliance from a sophisticate than from a naïf with the same δ.

When equation (21) does hold, the naïf 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. Because of her overoptimistic beliefs, unlike the sophisticate, she is unwilling to pay a cost for commitment (Corollary 1). One way to see that sophisticates are effectively paying a cost for commitment when equation 21 holds is to note that sophisticates are willing to pay more than dollar for dollar to increase the day 2 payment from m_{2L} to m_{2H}.[50]

An interesting question is how Option + commitment contracts are able to induce more effort from impatient naïfs than from patient people while Commitment-only contracts are not. Relative to Commitment-only contracts, Option + commitment contracts "increase the stakes" by decreasing m_{2L}, which helps the naïf 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 contracts help guide naïfs to take the action in their own best interest.

B.5 Proofs of Predictions 3 and 4

Prediction 3 (Frequency)**.** *If agents are impatient over the receipt of financial payments (i.e., if $d^{(t)} < 1$ for $t > 0$ and is weakly decreasing in t), then the compliance and effectiveness of the base case linear contract are weakly increasing in the payment frequency. If agents are patient over the receipt of financial payments ($d^{(t)} = 1$), then payment frequency does not affect compliance or effectiveness.*[51]

Proof. Equation (3) implies that, in a linear contract, $C = \frac{1}{T}\sum_{t=1}^{T} F(d^{(T-t)}m)$. Compliance is thus increasing in the discount factor over payment $d^{(T-t)}$. If agents are "impatient," then $d^{(T-t)}$ is weakly decreasing in the delay to payment $T-t$. Increasing payment frequency then decreases

[49] This follows not just for day 1 compliance but for overall compliance because, on day 2, the agent will comply if $e_2 < m_2$, which is affected by δ only through δ's effect on m_2 via day 1 compliance.

[50] That is, $V_{m_{2L}}^{m_{2H}}|^{Soph\&(\delta e_2 < m_{2L} \leq e_2 < m_{2H})} = m_{2H} - \delta e_2 > m_{2H} - m_{2L}$.

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

the average delay to payment, which weakly increases compliance. If agents are patient, then the discount factor is 1 irrespective of the delay to payment and increasing payment frequency has no effect on compliance. Effectiveness follows the same pattern as compliance since cost-effectiveness is invariant to payment frequency (it is always $\frac{1}{m}$). □

Prediction 4 (Payday Effects)**.** *If the discount factor over payments $d^{(t)}$ is decreasing in t, then the probability of complying in the base case linear contract increases as the payday approaches. If the discount factor over payments $d^{(t)}$ is constant in t, then the probability of complying is constant as the payday approaches.*

Proof. Recall that, on day t, agents comply if $e_t < d^{(T-t)}m$. As the payment date approaches, the time to payment $T-t$ decreases. If $d^{(T-t)}$ is decreasing, this increases $d^{(T-t)}$ and hence increases the likelihood that $e_t < d^{(T-t)}m$. If $d^{(T-t)}$ is flat, then the likelihood that $e_t < d^{(T-t)}m$ remains constant. □

C Misreporting Steps, Confusion, and Suspensions

Procedures to Curb Misreporting Because incentive payments were determined by self-reported data and not pedometer data, we implemented a number of checks to ensure integrity of step reporting. Within each 28-day sync period, respondents who incorrectly over-reported meeting a 10k step target on over 40% of days were flagged for cheating and suspended from receiving recharges for 7 days. Those who were flagged for cheating more than once were terminated from the program. Fewer than 5% of the incentive group was suspended for cheating and only 1 participant was terminated (Table C.1)

During the intervention, 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 might 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 the prevalence of misreporting. The prevalence of "misreporting," defined as reporting steps above 10,000 when the pedometer itself records fewer than 10,000 steps, is less than 5% and, interestingly, balanced across incentive and monitoring groups (column 1 of Table C.2). The balance with the monitoring group, who had no incentives to over-report, suggests that over-reporting was mainly unintentional participant mistakes. The incentive group also appeared to put more effort into making correct step reports, with fewer divergences in either the positive or the negative directions (columns 2-4 of Table C.2).

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 [0.01]	-0.081*** [0.02]	-0.059*** [0.02]	-0.022** [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. Dates limited to each 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

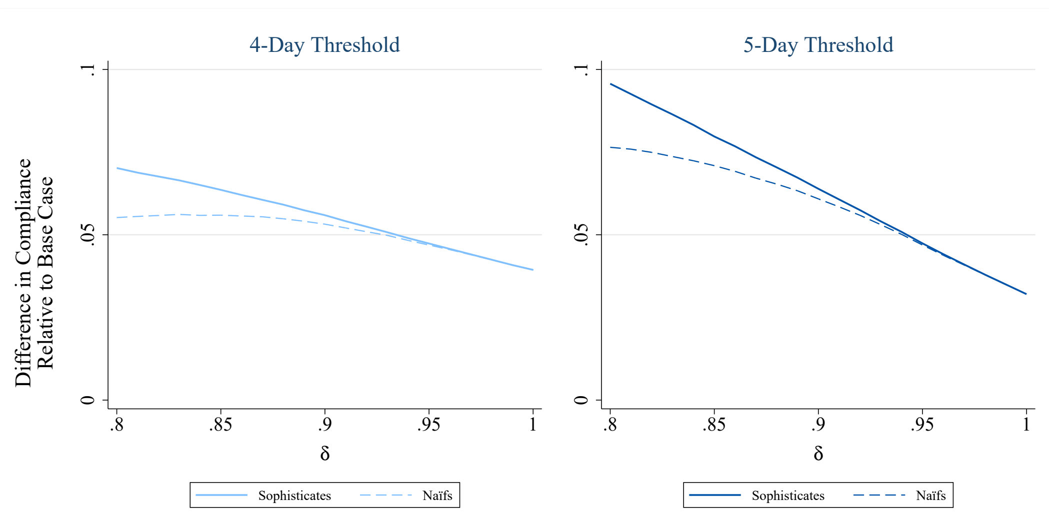
	Incentives	Monitoring	I - M	P-value: I=M
	(1)	(2)	(3)	(4)
A. Activity (by minute)				
Average daily activity	213	197	16	0.001
Average steps per minute	41	38	3	0.001
B. Time of Day				
Average start time	07:11	07:16	5	0.441
Average end time	20:49	20:50	1	0.742
C. High step counts per minute (share)				
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 × 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.

D Model Calibration for Threshold vs. Base Case

We next calibrate a model using the empirical distribution of walking costs to show that, in this setting, the performance of the threshold treatment should indeed increase meaningfully with impatience over effort. We begin with the framework from Section 2. To tractably examine contracts with seven-day payment periods and with 4- and 5-day thresholds, we simplify the model by assuming that individuals are fully patient over payments ($d = 1$), that they exponentially discount effort with discount factor δ, and that the full sequence of effort costs over the week $(e_1, \ldots, e_7)$ is known on day 1. To calibrate the average compliance in the threshold and base case (linear) contracts, we estimate the cumulative distribution function (CDF) of effort costs by fitting a uniform distribution to several moments of the CDF from the data, as described in Online Appendix D. We then use the estimated distribution to predict how relative compliance in the base case and threshold contracts would vary with δ.

The results are displayed visually in Figure D.1, with the exponential discount factor over effort δ on the x-axis and the gap between compliance in the threshold and base case linear contract on the y-axis (shown separately for the 4- and 5-day thresholds).



Appendix Figure D.1: 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.

The downward-sloping curves in the figure confirm the theoretical intuition from our model: for people who are more impatient over effort (smaller δ), there are larger compliance gains from thresholds. This is true for both naïfs and sophisticates with moderate levels of impatience.[52] 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

[52]As naïfs become more impatient ($\delta < 0.85$), the linear contract starts to gain relative to the 4-day

threshold has more bite, and where we see stronger results empirically as well (Table A.8, Panel B). For example, decreasing the effort discount rate from 1 to 0.9 increases relative compliance in the 5-day threshold contract by 3 pp among both sophisticates and naïfs.[53]

threshold, as naïfs begin to procrastinate in early periods under the threshold contract. However, even very impatient naïfs 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 days (Proposition 7 in App. B.3).

[53]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 uncertainty regarding future effort costs. However, our interest is heterogeneity by impatience, which we do not believe will change by incorporating uncertainty.