# Designing Incentives for Impatient People: An RCT Promoting Exercise to Manage Diabetes

Shilpa Aggarwal	Rebecca Dizon-Ross	Ariel Zucker *
Indian School of Business	University of Chicago	UC Santa Cruz

September 6, 2024

#### Abstract

Many people are impatient. We develop a prediction for how to make incentives work particularly well when people are impatient over effort: implement *time-bundled* contracts that make the payment for future effort increase in current effort. We test and find empirical support for this prediction using a randomized evaluation of an incentive program for exercise (walking) among diabetics in India. On average, a time-bundled contract generates as much effort as a time-separable linear contract, yet at a 15% lower cost. Moreover, time-bundled contracts perform roughly 30% better among individuals with above-median impatience over effort than those with below-median. Pooled across contracts, incentives increase daily steps by roughly 20% and improve blood sugar control relative to a control group.

<sup>\*</sup>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, Santa Cruz, arzucker@ucsc.edu. A previous version of this working paper was released as NBER Working Paper No. 27079 under the title "Incentivizing Behavioral Change: The Role of Time Preferences." 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). This RCT was registered as AEARCTR-0001818. 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, Ruoyu Chen, Katherine Daehler, Yashna Nandan, Varun Satish, and Emily Zhang for outstanding research assistance. We are grateful to Ned Augenblick, Abhijit Banerjee, Lori Beaman, Marianne Bertrand, Arun Chandrasekhar, Esther Duflo, Pascaline Dupas, Alex Frankel, Rick Hornbeck, Seema Jayachandran, Anett John, Emir Kamenica, Supreet Kaur, Jonathan Kolstad, Ted O'Donoghue, Kate Orkin, Rohini Pande, Devin Pope, Canice Prendergast, Matthew Rabin, Heather Royer, Gautam Rao, Sheldon Ross, Frank Schilbach, and Justin Sydnor for helpful conversations and feedback and to numerous seminar and conference participants for insightful discussions. All errors are our own.

### 1 Introduction

Policymakers are increasingly using incentives to encourage behaviors that have immediate costs but yield benefits in the future, such as saving, exercising, and studying (e.g., Gertler et al., 2019; Carrera et al., 2020; Fryer, 2011). A key motivation for these incentives is to offset underinvestment due to impatience or high discounting of the future, a common trait (e.g., Mahajan et al., 2020; Augenblick and Rabin, 2019; O'Donoghue and Rabin, 1999a). Given this motivation, it is critical that incentive contracts perform well when people are impatient.

This paper proposes and validates a novel strategy for increasing the performance of incentives in the face of impatience: implement *time-bundled* contracts in which the payment for future effort increases with current effort. Notably, this approach is designed to be effective in the face of impatience over effort, an important consideration given empirical findings that discount rates can be domain-specific and higher over effort than payment (e.g., Augenblick et al., 2015). We use a randomized controlled trial (RCT) to compare time-bundled contracts to a more standard time-separable contract (in which the payment for current effort depends only on current effort). We find that time-bundled contracts significantly improve contract performance for individuals who are impatient over effort and hence are an effective way to adapt incentives for impatience. In contrast, we test a more traditional strategy (more frequent payments) that should be effective if people are impatient over payment and find no evidence of its effectiveness. Our RCT, which randomizes these contracts among participants in an incentive program for exercise among diabetics and prediabetics, also shows that the program could be a powerful tool in the global fight against chronic disease.

We begin by showing theoretically that, relative to time-separable contracts, time-bundled contracts are more effective when individuals discount their future effort costs more. To illustrate the intuition, imagine you need a worker to perform two days of work. Consider first a time-bundled *threshold* contract that pays the worker a lump sum on day two if and only if she worked on both days. For the contract to induce two days of work, the total payment must exceed the worker's present discounted cost of effort.<sup>1</sup> 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 linearly on day two for each day of work, a larger minumum payment of \$20 is required to induce two days of work: \$10 per day. Time-bundled contracts exploit the fact that, when individuals have high effort discount rates, it is "cheaper" to buy their future (discounted) effort than their current effort.

Time-bundled threshold contracts should induce extra effort from all types of people with high discount rates over effort: time-consistent or time-inconsistent and, among time-inconsistent, sophisticated or naive (i.e., aware or unaware) about their own present bias. The contracts'

<sup>&</sup>lt;sup>1</sup>This example assumes zero short-run discounting of payments for simplicity.

potential for effectiveness among naifs is particularly valuable, since naive time inconsistency is common (Mahajan et al., 2020) but naifs are difficult to motivate (Bai et al., 2021). Timebundled thresholds work for naifs because they leverage current discount rates, which even naifs understand. That is, even naifs discount their future effort and will sell it at a discount today.

The second (and less novel) strategy we consider is to increase the frequency of payment, which should be effective if individuals are impatient over *payments*. Scholars have long theorized that because people are impatient, "the more frequent the reward, the better" (Cutler and Everett, 2010). Indeed, DellaVigna and Pope (2018) describes more frequent payment as the main way to adjust incentives for present bias. However, they also acknowledge that increasing payment frequency should only be effective if people heavily discount payments, which even those who heavily discount effort may not do (Augenblick et al., 2015).

An exercise incentive program for diabetics and prediabetics provides an apt context to assess the effectiveness of time-bundled contracts and increased payment frequency in tailoring incentives for impatience. The healthy behaviors (like exercise) that help prevent and manage diabetes and other lifestyle diseases feature short-run utility costs (such as effort costs) but only long-run benefits, making people with high discount rates over utility less likely to invest in them (Hunter et al., 2018). Indeed, evidence suggests that people with early-onset diabetes or high BMI have higher discount rates over utility than the general population (e.g., Courtemanche et al., 2014; Cobb-Clark et al., 2022), making them well-suited for assessing incentive strategies for mitigating impatience.

Our incentive program monitored participants' walking for 3 months using pedometers and provided financial incentives in the form of mobile phone credits for achieving a daily step target of 10,000 steps. Among participants randomly selected to receive incentives, we randomly varied the contract. In the "base case" contract, payment was a time-separable (in particular, linear) function of the number of days the participant complied with the step target, with payments made weekly. To evaluate time-bundling, we randomized some participants to receive timebundled threshold contracts (which we also refer to more simply as threshold contracts). These contracts only rewarded compliance with the step target if the step target was met a minimum number of days that week. We used two threshold levels: four and five days. Both contracts paid at the end of the week, like the Base Case. To explore payment frequency, we then randomized two additional linear contracts that paid daily and monthly.

Our primary empirical contribution is to validate time-bundled contracts as a strategy for tailoring incentives for impatience over effort, and we present three main empirical findings. First, on average across the full sample, the time-bundled threshold contracts perform better than the time-separable linear contract—they achieve the same sample-average level of compliance, but do so at a lower cost to the principal. For example, the 5-day threshold contract pays out nearly 15% less in incentives than the linear contract for the same level of compliance,

because it does not pay out for every day of compliance like the linear contract does. This improves performance from the perspective of a policymaker who wants to maximize the benefits of compliance net of the incentive costs.<sup>2</sup>

The second finding is that high levels of impatience over effort in our sample are an important mechanism driving the effectiveness of time-bundled threshold contracts, as the contracts are significantly more effective for those with higher impatience over effort. Specifically, heterogeneity analysis using a measure of impatience over effort shows that, relative to the time-separable contract, the time-bundled threshold contracts increase compliance with the step target by 6 percentage points (pp) more for those with above-median impatience than for those with belowmedian impatience. This difference is large, equivalent to roughly 30% of the sample-average effect of either contract (20 pp). The 6 pp estimate represents the difference between a 3 pp positive effect among those with above-median impatience and a 3 pp negative effect among those with below-median impatience. In addition to their effects on compliance, the thresholds also improve cost-effectiveness (i.e., decrease the payout per day of compliance) among both less and more impatience, while having an ambiguous effect for those with lower impatience. Although our analysis exploits non-random variation in impatience across the population, we provide evidence suggesting that confounding factors do not drive our results.

The better relative performance of the time-bundled threshold among effort-impatient people suggests that a policymaker can improve incentive performance by customizing thresholds based on impatience over effort. Customized incentives could be leveraged for a range of policy goals, such as savings, preventive health, and school attendance (e.g., Beaman et al., 2014; Breza and Chandrasekhar, 2019; Barrera-Osorio et al., 2011). Incentives could be customized at the population level by using threshold contracts for populations that are particularly impatient, such as those with chronic diseases. Customization could also occur at the individual level (Andreoni et al., 2023). Individual-level customization can be challenging to implement since impatience is often not observable; however, we provide evidence suggesting that such personalization would be feasible, for example by showing that a principal could use more easily observed characteristics to proxy for impatience.

Finally, we find that increasing the frequency of payment has no impact in our setting. While the finding is somewhat imprecise, we present multiple other pieces of evidence indicating that participants have low discount rates over the contract payments (mobile phone credits).<sup>3</sup> The low discount rates over payment and lack of impact of high-frequency payments in our setting

 $<sup>^{2}</sup>$ We discuss other objectives (e.g., welfare maximization) later in the paper. This statement assumes linear benefits from compliance, which are likely in many contexts, including exercise (e.g., Banach et al., 2023).

<sup>&</sup>lt;sup>3</sup>While it is possible that people would be more impatient over payments delivered with a different modality, limited impatience over payments is not rare (e.g., Augenblick et al., 2015; Tanaka et al., 2010).

make it important to identify other methods to adjust incentives for impatience and highlight the significance of our finding that time-bundled contracts are one such method.

A second contribution of our evaluation is to demonstrate that incentives for exercise are a useful tool that could help decrease the burden of chronic disease in India and beyond. Chronic lifestyle diseases such as diabetes represent a severe threat to health and development in low and middle income countries (LMICs). The cost of diabetes alone is estimated to be 1.8% of GDP annually in LMICs (Bommer et al., 2017), with 12% of adults estimated to have the disease (International Diabetes Federation, 2019). Although there is widespread agreement that the key to addressing the burden is to promote lifestyle changes such as better exercise (World Health Organization, 2009), the existing evidence-based interventions promoting such changes in this population are prohibitively expensive (Howells et al., 2016). Governments are thus eager for scalable interventions to promote lifestyle change among diabetics. Our RCT was funded by the Government of Tamil Nadu, one of the most populous states in India, who sought an intervention to scale up across their state to address their exploding diabetes epidemic.

Pooling across our incentive contracts, we show that, relative to a control group, our relatively low-cost incentives program substantially increases exercise and moderately improves health among our diabetic population. Average daily steps increase by roughly 20 percent during the intervention. Roughly 50% of the treatment effect on steps persists even after the intervention and payments end. Incentives also improve blood sugar control. Given the dearth of evidence-based lifestyle-change interventions for diabetics, our scalable program offers a promising strategy to combat diabetes in LMICs.

### **1.1** Contributions to the Literature

This paper's primary contribution is to theoretically investigate and empirically validate time-bundled contracts as a novel strategy for motivating a wide range of people with high discount rates over effort. In doing so, we connect a classic literature on dynamic incentives (e.g., Lazear, 1981; Lambert, 1983) with a newer literature on domain-specific time preferences and high discount rates over effort (e.g., Augenblick et al., 2015), providing the first examination of how domain-specific discounting affects the design of dynamic incentive contracts. We describe how our work contributes to the time-preferences and dynamic incentives literatures in turn.

**Time Preferences** Our finding that time-bundled contracts effectively adapt incentives for impatience adds to a small literature proposing incentive designs to motivate impatient agents. O'Donoghue and Rabin (1999b) theoretically explores incentive schemes to motivate time-inconsistent procrastinators to more quickly complete a task which requires effort only in a single period.<sup>4</sup> In contrast, our solution of time-bundled contracts hinges on the agent incurring effort in multiple periods. Carrera et al. (2020) examines whether larger time-separable

<sup>&</sup>lt;sup>4</sup>They show that optimal single-period task incentives feature an increasing punishment for delay over time.

incentives upfront can help time-inconsistent people overcome startup costs, but finds no empirical evidence that they do. DellaVigna and Pope (2018) shows that decreasing the lag until payment does not significantly increase effort — an approach targeting impatience over *payment*, while time-bundled contracts target impatience over *effort*.

Finally, researchers have also motivated impatient agents with commitment devices (e.g., Royer et al., 2015; Schilbach, 2019). Commitment is a useful tool, but it is not a panacea. Take-up of commitment devices is typically modest (Laibson, 2015), undermining their use as a broad policy solution. Moreover, unlike time-bundled thresholds, commitment devices are only effective for sophisticated time-inconsistents, and can even be harmful for naifs (e.g., Bai et al., 2021; John, 2020). Our work broadens the arsenal by considering an approach that can succeed in settings with naivete or without commitment demand.

**Dynamic Incentives** We contribute to the literature on dynamic incentives, specifically on contracts that "defer compensation" to the future (as time-bundled contracts do).<sup>5</sup> This literature has been primarily theoretical, exploring reasons that time-bundled, deferred compensation contracts may be better or worse than time-separable (e.g., Lazear, 1979; Rogerson, 1985).

Our first contribution is empirical: despite the extensive theoretical literature, we conduct, to our knowledge, the first rigorous empirical comparison of the two types of contracts.<sup>6</sup> Our comparison explores their distributional impacts and the mechanisms driving their relative performance. This comparison is valuable because, while both contract types are popular, their relative performance is theoretically ambiguous.

Our second contribution is to introduce and test a new theoretical channel for the effectiveness of time-bundled contracts: agent discounting of *effort*. Previous dynamic contracting papers use the same discount rate for effort and payments. In such models, time bundling can be effective when barriers, such as unobservable effort, prevent the principal from compensating agents for their exact effort cost at the end of each period (e.g., Lazear, 1979, 1981).<sup>7</sup> Building on evidence that people discount utility and payments differently (e.g., Chapman and Elstein, 1995; Augenblick et al., 2015), we introduce domain-specific discounting and show that high agent discount rates over effort give principals an additional reason to prefer time-bundled contracts, even without traditional barriers to direct end-of-period compensation, thus broadening

<sup>&</sup>lt;sup>5</sup>The literature refers to time-bundling as deferring compensation or backloading. Both terms have various meanings, including making payment functions non-separable over time and changing payment timing. We introduce the term time-bundling to clarify that our focus is on non-separability, not payment timing.

<sup>&</sup>lt;sup>6</sup>In a lab experiment where college students acted as workers and as firms without exerting any real effort, Huck et al. (2011) compares deferred compensation contracts to one another but not to separable contracts.

<sup>&</sup>lt;sup>7</sup>This is relevant for us as effort costs vary by period, making it hard to pay the exact effort cost each period. Under this theory, time-bundling's effectiveness arises because pay in early periods only motivates early effort, while later time-bundled pay can motivate effort in all periods. Another theory is that time-bundling manages marginal utility under concavity (Rogerson, 1985), which is less relevant here due to the small incentive amounts.

the applicability of these contracts beyond the commonly cited scenarios.<sup>8</sup>

The paper proceeds as follows. Section 2 presents our theoretical predictions. Sections 3 and 4 discuss the study setting and design. Sections 5 and 6 present our results on incentive design and our program evaluation of incentives, respectively. Section 7 concludes.

### 2 Theoretical Predictions

This section examines the effectiveness of time-bundled contracts and shows that, under a broad range of assumptions, they are particularly effective when individuals have high discount rates over effort. We first specify the individual's problem and define the principal's goal: contract effectiveness. We then solve for effectiveness under a simple *base case* incentive contract which is linear across days, and therefore time-separable. Next, we examine the impact of a time-bundled contract, where the payment for future effort increases in current effort, focusing on a time-bundled *threshold* contract that pays only if a threshold level of effort is reached.

We present two key results applicable to various types of impatience, including timeinconsistent sophistication, time-inconsistent naivete, and time consistency. The first result is that the effectiveness of the time-bundled threshold contract relative to the linear contract is increasing in the discount rate over effort. While it is possible to find specific parameter values that are exceptions, this result holds in many typical and empirically relevant cases. Our second result is that the most effective time-bundled threshold contract is more effective than the most effective linear contract if the discount rate over effort is sufficiently high, and less effective if the discount rate over effort is low. While this result strengthens the first by speaking to the overall effectiveness of threshold and linear contracts rather than just heterogeneity, it requires more specific conditions such as assumptions about the effort cost distribution. Finally, we briefly explore high-frequency payments as a strategy to adjust incentives for impatience over payment rather than effort, demonstrating their effectiveness when payment discounting is high.

### 2.1 Set-Up

Each day, an individual chooses whether to complete a binary action. Define  $w_t$  as an indicator for whether the individual *complies* (i.e., completes the action) on day t. Define  $m_t$  as the payment made by the principal to the individual on day t;  $m_t$  depends on the individual's compliance decisions through time t.

To solve for compliance, we assume that individual choices on day t maximize the following reduced-form utility function:

<sup>&</sup>lt;sup>8</sup>Our exploration of impatience and incentives relates to Jain (2012), which assumes identical discount rates for effort and money, showing that with quasi-hyperbolic discounting and barriers to immediate payment, firms can increase profits by offering two-period quotas. In contrast, our model with domain-specific discounting offers stronger and more general results: time-bundled contracts are more effective for agents with high effort discount rates, even without barriers to direct end-of-period compensation or quasi-hyperbolic discounting. Moreover, while Jain (2012) is purely theoretical, we empirically test and confirm our insights.

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

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 (for notational simplicity, we denote  $\delta^{(1)}$  as  $\delta$  and  $d^{(1)}$  as d). 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 assume only that they are weakly decreasing in t. We assume utility is linear in payments, which is likely a good approximation in our setting, as payments are small relative to overall consumption.

Importantly, this reduced-form utility function differentiates the discount factor over payments,  $d^{(t)}$ , from the discount factor over effort,  $\delta^{(t)}$ . The specification is consistent with a standard model of utility with a single structural discount factor over consumption and effort (e.g., Augenblick et al., 2015). In that case,  $\delta^{(t)}$  is the structural discount factor, while  $d^{(t)}$ depends on the availability of borrowing and savings. For example, in perfect credit markets, individuals should discount future payments at the interest rate r, and so  $d^{(t)} = \left(\frac{1}{1+r}\right)^t$ .

Time-Inconsistency and Sophistication Individuals will have time-inconsistent preferences either if  $\delta^{(t)}$  or  $d^{(t)}$  are non-exponential functions of t, or if  $d^{(t)} \neq \delta^{(t)}$ . Among timeinconsistent agents, we follow O'Donoghue and Rabin (1999a) in distinguishing sophisticates, who are aware of their discount factors (over both effort and money), from naifs, who "believe [their] future selves' preferences will be identical to [their] current self's." Thus, sophisticates accurately predict how their future selves will behave, while naifs may not.<sup>9</sup>

**Effort Costs** Let  $e_t$  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 day t.  $e_t$  can be negative, as agents may comply without payment.

**Incentive Contract Structure and Compliance** 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.

Define *compliance*, the expected fraction of days on which the individual complies, as  $C = \frac{1}{T} \mathbb{E}[\sum_{t=1}^{T} w_t]$  and the expected per-day *payment* as  $P = \frac{1}{T} \mathbb{E}[m_T]$ .

The Principal's Objective: Effectiveness We assume that the principal aims to maximize *effectiveness*, defined as the expected per-day benefit to the principal from compliance less the expected payment to agents P. Maximizing effectiveness is analogous to the standard contract theory approach of maximizing output net of wage payments subject to incentive compatibility

<sup>&</sup>lt;sup>9</sup>With domain-specific discounting, naivete can stem from misunderstanding how the future self will either (a) value current effort relative to money, or (b) discount effort or money further in the future.

constraints.<sup>10</sup> For the definition to be operable, we need to take a stand on the expected benefit function. We assume the expected benefit is linear in compliance, equal to  $\lambda C$  for some  $\lambda > 0$ . This simplifying assumption is reasonable in our empirical setting since the estimated marginal health benefit of days of exercise is approximately linear (Warburton et al., 2006; Banach et al., 2023). With linear benefits, effectiveness becomes  $\lambda C - P$ .

We want to compare the effectiveness of different contracts even when we do not know  $\lambda$ . To do so, define *cost-effectiveness* as compliance divided by expected per-day payment, C/P. One can then easily show that one contract is more *effective* than another if it has strictly larger compliance and weakly larger cost-effectiveness, or weakly larger compliance and strictly larger cost-effectiveness.<sup>11</sup>

### 2.2 Time-Separable Linear Contracts (the Base Case)

We now solve for compliance and effectiveness under the base case contract. The contract is linear, paying m per day of compliance. Total payment is therefore:

$$m_T^{\text{Base Case}} = m \sum_{t=1}^T w_t.$$
<sup>(2)</sup>

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

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

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

**Observation 1.** In a time-separable contract, holding all else constant, neither compliance, cost-effectiveness, nor effectiveness depend on  $\delta^{(t)}$ .<sup>12</sup>

We will see that this observation does not hold for time-bundled contracts.

### 2.3 Time-Bundled Contracts and Impatience over Effort

We now examine the effect, relative to the Base Case, of making the contract time-bundled while maintaining the same payment period length. We pay particular attention to the relationship between the effectiveness of time-bundled contracts and the discount factor over effort. We first use a simple case to show the intuition for why time-bundled contracts are more effective when the discount factor is small. We then present key testable implications, which we label predictions. Appendix B presents our formal mathematical results, which we label propositions.

<sup>12</sup>In the linear case, compliance is  $\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} w_t\right] = \frac{1}{T}\sum_{t=1}^{T} F(d^{(T-t)}m)$ , which is not directly related to  $\delta^{(t)}$ .

<sup>&</sup>lt;sup>10</sup>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 5.5.

<sup>&</sup>lt;sup>11</sup>This is true assuming effectiveness is positive. To see this, rewrite effectiveness as  $C\left(\lambda - \frac{1}{(C/P)}\right)$ .

Time-bundled contracts contain at least one period in which the payment for future compliance is increasing in current compliance. We focus on a *threshold* time-bundled contract, where there is a minimum threshold level of compliance K; however, our testable predictions hold for other types of time-bundled contracts as well. In a threshold contract, if compliance is below K, no incentive is received. If compliance is above K, payment is a linear function of the number of days of compliance, with a rate of m' per day. Total payment in the threshold contract is thus:

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

Intuition: Threshold Effectiveness Decreases in  $\delta$  We first consider a simplified case with d = 1, T = 2, K = 2, and effort costs that are weakly positive and known from day 1. On day 1 of the threshold contract, the individual complies if both the discounted (by d = 1) payment outweighs the present discounted cost of effort on days 1 and 2, and she expects to comply on day 2 if she complies on day 1. Day 1 compliance is hence:

$$w_1 = \begin{cases} 1 & \text{if } (e_1 + \delta e_2 < 2m') \text{ and (believe } w_2 \text{ will be 1 if } w_1 = 1) \\ 0 & \text{otherwise.} \end{cases}$$
(5)

Importantly, because future effort costs are discounted, the present discounted effort cost is lower — and day 1 compliance higher — the lower is  $\delta$ . This holds for both sophisticates and naifs, as outlined in more detail in Appendix B.1.

Impatient people's higher compliance on day 1 underlies the threshold's greater effectiveness for them. On day 2, individuals comply if  $w_1 = 1$  and the payment exceeds their effort costs:

$$w_2 = \begin{cases} 1 & \text{if } e_2 < 2m' \text{ and } w_1 = 1\\ 0 & \text{otherwise.} \end{cases}$$
(6)

Impatient people's higher day 1 compliance thus leads to higher day 2 compliance as well. Their greater total compliance makes the contract more effective.<sup>13</sup>

**Testable Predictions** We next return to the full model with d unrestricted,  $T \ge 2$  and  $K \le T$ , and extend the intuition from the 2-period model to make two testable predictions in more general models. We first ask how the effectiveness of threshold contracts depends on impatience over effort. Appendix B.2 presents a series of propositions investigating this comparative static, which we summarize in the following prediction:

<sup>&</sup>lt;sup>13</sup>Effectiveness follows from compliance since an increase in compliance without a decrease in cost-effectiveness implies higher effectiveness, and the Appendix B.2 propositions show that, depending on the cost distribution, threshold cost-effectiveness tends to be flat or decreasing with  $\delta^{(t)}$ .

**Prediction 1** (Comparative Static in  $\delta^{(t)}$  of Time-Bundled Threshold versus Time-Separable Effectiveness). Holding all else equal, under many conditions, compliance and effectiveness in time-bundled threshold contracts relative to time-separable contracts decrease in the discount factor over effort,  $\delta^{(t)}$ .

Holding all else equal, compliance and effectiveness in time-bundled threshold contracts decrease in  $\delta^{(t)}$  under a broad range of assumptions. In contrast, in time-separable contracts, both compliance and effectiveness are flat in  $\delta^{(t)}$  (Observation 1). Thus, the lower  $\delta^{(t)}$  is, the higher compliance and effectiveness are in a threshold relative to time-separable contract.

While Prediction 1 speaks to the heterogeneity in the performance of threshold relative to separable contracts by  $\delta^{(t)}$ , it is also important from a policy perspective to understand which type of contract performs better for any given level of  $\delta^{(t)}$ . Making some additional assumptions for tractability, Appendix B.3 presents a series of propositions comparing both optimized threshold and separable (in particular, linear) contracts, and threshold and linear contracts offering the same payment per day (as in our experiment).<sup>14</sup> These are summarized in the following prediction:

**Prediction 2** (The Level of Time-Bundled Threshold versus Time-Separable Linear Effectiveness by  $\delta$ , T = 2). Holding all else equal, under many conditions:

(a) When δ is sufficiently low, threshold contracts are more effective than linear contracts that offer the same payment amount per day. When δ is sufficiently high, the reverse is true.
(b) When δ is sufficiently low, the most effective contract is a threshold contract. When δ is sufficiently high, the most effective contract is linear.

While the assumptions underlying Prediction 2 are more restrictive than those underlying Prediction 1, the prediction suggests that principals will often prefer threshold to linear contracts when individuals are sufficiently impatient over effort.

### 2.4 Payment Frequency and Impatience over Payment

We now briefly explore a strategy for improving the performance of incentives if people are impatient over payment rather than effort: increasing payment frequency. We return to the base case separable linear contract from equation (2) and analyze compliance under different payment frequencies (i.e., different T's). Appendix B.4 contains the proof.

**Prediction 3** (Frequency). If agents are impatient over 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 financial payments ( $d^{(t)} = 1$ ), then payment frequency does not affect compliance or effectiveness.

<sup>&</sup>lt;sup>14</sup>In many empirical applications, constructing the optimal contract is not feasible as it requires knowledge of both the discount rate and the distribution of costs.

#### 2.5 Empirical Tests

Our theoretical analysis informed the design of our experiment. Among participants who receive incentives in our experiment, we randomly vary whether the contract is linear or is a threshold contract offering the same payment per day as the linear (m' = m). To assess the empirical relevance of Prediction 2 — that, under certain assumptions, the threshold contract has higher effectiveness than the linear when discount factors over effort are low — we compare the effectiveness of the two contracts in the full sample. To assess our more general Prediction 1 and investigate whether impatience is a contributor to the effectiveness of thresholds, we test for heterogeneity in the effect of the threshold relative to the linear contract based on a measure of impatience over effort. Finally, to shed light on the role of payment frequency and the discount rate over payments (per Prediction 3), we randomize the frequency of payments.<sup>15</sup>

### 3 Experimental Design

### 3.1 Sample Selection and Pre-Intervention Period

We conducted our experiment in the South Indian city of Coimbatore, Tamil Nadu. India is facing a diabetes epidemic, and prevalence is highest in urban areas of southern states (Anjana et al., 2011). We selected our sample through public screening camps held in various locations including government hospitals, markets, religious institutions, and parks, in order to recruit a diverse socioeconomic group. 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.<sup>16</sup> After screening, eligible individuals were invited by phone to participate in a program encouraging walking.

Surveyors visited the participants at their homes or workplaces for a pre-intervention visit to conduct a baseline health survey, deliver lifestyle modification advice, and enroll them in a one-week phase-in period to familiarize them with our procedures and collect baseline walking data. Surveyors gave participants pedometers for the duration of the program; we gathered step data by syncing the pedometers with a central database. Because syncing requires an internet

 $<sup>^{15}</sup>$ Before launching our experiment, we drafted a pre-analysis plan that guided our design and power calculations. While we did not polish it for public posting at the time, we have since posted the draft (last modified before we launched endline data collection) in our AEA registry to demonstrate that our key subsample and heterogeneity analyses were conceived *ex-ante*.

<sup>&</sup>lt;sup>16</sup>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.

connection, which most participants did not have, pedometer step data were not available in real time. Thus, we also asked participants to report their daily step count to an automated calling system which called every evening and prompted them to enter the step count recorded on their pedometer. During the pre-intervention visit, surveyors demonstrated how to wear a pedometer, report steps, and check text messages from our reporting system. Surveyors asked participants to wear the pedometer and report their steps each day of the phase-in period.

At the end of the phase-in period, surveyors visited respondents to sync the data from the pedometers and conduct a baseline time-preference survey. After all baseline data were collected, surveyors described to participants their randomly assigned treatment group by guiding them through a contract describing the intervention period.<sup>17</sup> We exclude from the sample all participants who withdrew or were found ineligible prior to receiving their contracts, leaving a final experimental sample of 3,192 individuals. The sample represents 41% of the screened, eligible population. We began screening in October 2016 and enrolled participants on a rolling basis from February–November 2017, with endline data collection launched in May 2017.

### 3.2 Experimental Design and Contract Launch

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

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

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

Within the incentive group, we randomized participants to receive one of six incentive contracts for walking, as shown in Figure 1.

### 3.2.1 Incentive Sub-Groups

All incentive sub-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).<sup>18</sup> After reporting steps, participants immediately received text-message confirmations of their step

<sup>&</sup>lt;sup>17</sup>All participants who completed the baseline survey were randomly assigned to treatment prior to this visit. The randomization was stratified by baseline HbA1c (a measure of blood sugar control) and a simple one-question proxy for impatience using a randomization list generated in Stata.

<sup>&</sup>lt;sup>18</sup>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).



Figure 1: Experimental Design

report, payment earned, and the payment date. We also sent participants weekly text messages summarizing their walking behavior and total payments earned.

Each of the six incentive subgroups received a different incentive contract with three dimensions of variation: time-separability, payment frequency, and payment amount.

**The Base Case** Participants in *Base Case* received a time-separable, linear contract paying 20 INR (0.3 USD) 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 it differs from all other contracts in exactly one dimension: time-separability, payment frequency, or payment amount. We can compare any other group to the Base Case to assess the effect of changing a single contract dimension.

**Threshold Contracts** Participants in the threshold groups received contracts that differ from the base case contract only in time-separability: the threshold contracts use time-bundled threshold payment functions. Participants in the 4-Day Threshold received 20 INR for each day of compliance only if they met the target at least four days in the weeklong 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, participants in the 5-Day Threshold received 20 INR 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 sub-groups to walk at least four or five days per week when initially explaining the contracts.<sup>19</sup> To maximize statistical power, we pool the 4- and 5-Day Threshold for our main analyses. We show results for the two threshold groups separately in some exploratory analyses.<sup>20</sup>

**Payment Frequency** The contracts for two groups, *Daily* and *Monthly*, differ from the base case contract only in the payment frequency. In Daily, recharges were delivered at 1:00 am the same night participants reported their steps. In Monthly, recharges were delivered every four weeks for all days of compliance in the previous four weeks.

Higher payment frequency could increase both the salience of compliance and trust in the payment system. To hold these factors constant, all incentive sub-groups received daily feedback on their compliance and a test payment of 10 INR the night before their contract launched.

**Payment Amount** Participants in our final incentive group, *Small Payment*, received contracts that differ from the base case only by the amount of incentive paid. This group received 10 INR, instead of 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.

We allocated more of our sample to the 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, we need a heterogeneity analysis to test Prediction 1 about thresholds, but only a main effects analysis to test Prediction 3 about payment frequency.

### 3.2.2 Comparison Groups

The incentive program could affect behavior because it provides financial incentives or simply because it monitors walking behavior. We include two control groups in our experiment, a monitoring group and a pure control, to allow us to isolate the effects of financial incentives on steps while also testing whether the full program impacts health relative to pure control.

<sup>&</sup>lt;sup>19</sup>For the threshold groups, the target days-per-week was the same as their assigned threshold level. For the other groups, it was randomized between 4 and 5 in the same proportion as in the threshold groups.

 $<sup>^{20}</sup>$ We pre-specified in our AEA registry that we would pool the threshold groups (see the power calculations section). 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 test our prediction about heterogeneity by impatience.

**Monitoring** Monitoring 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 daily step report confirmation texts and weekly text message summaries, as in the incentive groups. Finally, during the upfront explanation of the contract, surveyors delivered 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** Control participants 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.<sup>21</sup>

### 3.2.3 Contract Understanding

To ensure participants understood their contracts, a few days after each participant was assigned their contract, a surveyor called them to ask several questions testing their understanding of their contract. If participants got an answer wrong, the surveyor would explain the correct response. The responses indicate that a vast majority of participants did indeed understand their assigned contract (Online Appendix Table F.1).

#### 3.3 The Intervention Period and After

During the 12-week intervention period, participants received incentives, which were based on both their assigned contracts and their reported steps. 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 pedometer data, not the reported data.<sup>22</sup>

At these visits, we also conducted short surveys to collect biometric 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.

Finally, to assess the persistence of our treatment effects on exercise, we gave pedometers to the final 1,254 participants enrolled in our experiment (including Control participants) for 12 weeks after the intervention period had ended. We hereafter refer to this period as the postintervention period. Participants no longer reported steps daily or received incentive payments,

<sup>&</sup>lt;sup>21</sup>To accommodate a request from our government partners, we tested one additional intervention: weekly text message reminders to engage in healthy behaviors (the "SMS treatment"). Ten percent of the sample, cross-randomized with all treatments, received the messages, which we control for in our regressions.

<sup>&</sup>lt;sup>22</sup>Online Appendix G contains detailed statistics on misreporting. Misreporting rates are similar across monitoring and incentive groups, suggesting misreports were primarily accidental.

but surveyors still returned every four weeks to sync their pedometers.

### 4 Data and Outcomes

This section first describes our measures of baseline information — including health, walking and time-preferences — and presents summary statistics. Next, it describes our two sources of outcomes data: pedometer data and a health survey.

### 4.1 Baseline Data: Demographics, Health and Walking

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 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 (BP), a measure of hypertension; and a short mental health assessment. During the phase-in period (between the baseline health survey and randomization), we collected one week of baseline pedometer data.

### 4.2 Time Preferences Data

Impatience over Effort Following the phase-in period, we conducted a baseline timepreference survey to measure impatience over effort. 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." Notably, our sample was elderly and had limited education, and had difficulty with the screen-based convex time budget (CTB) measure of Andreoni and Sprenger (2012a); although we implemented a full CTB module, the data were of such poor quality that we do not use them for analysis.<sup>23</sup> Our heterogeneity analyses instead leverage four other measures of impatience over effort collected during the time-preference survey, with relatively consistent results.

Although impatience measures tend to be noisy (Kremer et al., 2019), and ours may be particularly so, measurement error will bias our heterogeneity results toward zero. Thus, true heterogeneity across time preferences is likely even larger than what we measure.

Impatience Index and Predicted Impatience Index: Our preferred measure of impatience over effort is an index of responses to simple survey questions from the psychology literature on impatience and procrastination. The questions, listed in Panel A of Table A.1, are a subset of the Tuckman (1991) and Lay (1986) scales chosen *ex ante* by our field team as translating

 $<sup>^{23}</sup>$ Respondents did not understand the CTB method well, and we have an order of magnitude more law-ofdemand violations than lab-based studies with college students. Moreover, as described in Online Appendix J.3, our CTB estimates do not converge for 44% of the sample, they do not correlate in the expected direction with any behaviors we measure, and respondents did not follow through with their chosen allocations. These issues make the CTB estimates unusable for analysis.

well to our setting. Each question asks respondents to respond on a Likert scale of agreement with a statement such as "I'm continually saying 'I'll do it tomorrow'." We construct the index (hereafter: the impatience index) by averaging the standardized question responses.

The Tuckman and Lay scales are validated predictors of real behaviors such as poor academic performance (Kim and Seo, 2015). The impatience index also predicts behavior in our sample: those with higher index values walk less and have worse diets at baseline (Table A.1).

We further validate our impatience index by showing that it predicts an incentivized measure of effort impatience in Appendix C. After the completion of our experiment, we elicited incentivized choices from a separate sample of similar participants (n=71) regarding the number of effort tasks they wanted to complete on different days for different piece rates following the methodology of Augenblick (2018) (we were unaware of the Augenblick (2018) methodology when we conducted our experiment in 2016.) Reassuringly, Appendix C.1.3 shows that those with higher impatience index also make more effort-impatient choices, choosing relatively more tasks in the future than the present.<sup>24</sup>

We began collecting the impatience index partway through the experiment,<sup>25</sup> so it is only available for the latter 55% of the sample. The available sample yields sufficient power to conduct heterogeneity analyses. That said, to check the robustness of our results in the full sample, we fit a "predicted index" using a LASSO model with three survey questions on self-control asked of all participants. Panel B of Table A.1 lists the questions and shows that the predicted index correlates in the expected direction with measures of impatient behavior.

*Demand for Commitment:* Our third impatience measure relies on participants' demand for contracts that are financially-dominated but increase incentives for future effort, a common (but coarse) indicator of sophisticated present bias in the literature (e.g., Ashraf et al. 2006; Kaur et al. 2015).<sup>26</sup> We presented participants with two choices: each between the base case contract and one of the financially-dominated contracts, either the 4-day or 5-day threshold. Our measure is the simple average of the two indicators for choosing the threshold contracts (Online Appendix Table F.2 shows that our findings are robust to different ways of aggregating).

<sup>&</sup>lt;sup>24</sup>Specifically, Figure C.1(a) shows the gap between tasks chosen for the future versus the present is more than twice as large for those with above- than below-median impatience index. Table C.1 also shows that a structurally estimated effort discount factor is large and statistically indistinguishable from 1 among people with below-median impatience, but significantly smaller than 1 among those with above-median impatience.

<sup>&</sup>lt;sup>25</sup>We initially planned to only use measures estimated from the full CTB. We added the impatience index after challenges surfaced in our data collection which made the full CTB estimates unusable for analysis.

<sup>&</sup>lt;sup>26</sup>While Carrera et al. (2022) shows that this measure is biased upwards due to measurement error caused by participants misunderstanding their utility under the contracts, the authors argue it is still "useful as one imperfect measure of awareness [sophistication] of time-inconsistency." While we prefer our primary measure because it is can detect naive impatience and because it is less coarse, we believe the demand for commitment measure is useful both to show robustness to an incentive-compatible measure and because it provides a proxy of sophistication that we can use to distentangle behavior between impatient sophisticates and impatient naives.

The elicitation was incentive-compatible, as we assigned a small fraction of the sample to their selected contract for each choice.<sup>27</sup> To ensure understanding, we provided visual-aid based explanations of payment in both contracts, emphasizing the dominated payment schedule in the threshold contracts, followed by quizzes to test understanding. Demand for commitment is relatively high in our setting, with 51% and 46% of the sample preferring the 4-day and 5-day thresholds to the linear contract, respectively.

Simple CTB Questions: Our final measure uses two simplified questions that follow the CTB paradigm of selecting intertemporal effort allocations. However, instead of allowing participants to allocate steps from a continuous convex walking budget, these questions required respondents to select a preferred allocation between just two discrete points from such a budget.<sup>28</sup> For example, one question asked the respondent whether they would rather walk (A) 30 minutes today and 60 minutes one week from today, or (B) 60 minutes today and 20 minutes one week from today, both in exchange for the same large payment. Our impatience measure is the average of the indicators for choosing the option with more walking later (e.g., option (A) above), but our findings are robust to different ways of combining the answers (Online Appendix Table F.2). Table A.1 shows this measure correlates in the expected direction with baseline exercise: people who are more impatient according to this measure exercise less.

**Impatience Over Payments** Our theory predicts that impatience over effort affects the performance of time-bundled thresholds, and so we focus on measuring impatience over effort for heterogeneity analysis. However, we also collected several measures of impatience over payments to better understand our setting and for use in robustness checks.

We collected three proxies for impatience over recharges at baseline: a real-stakes measure of demand for more frequent payment, recharge balances, and recharge usage.<sup>29</sup>

We complement these proxies with more direct measures of impatience over payments that we collected after randomization. We use these measures to shed light on the overall levels of impatience over payments in our sample (Appendix C.2) and to show evidence that impatience over effort and over payments vary independently (Appendix C.3).

The first more direct measure uses a series of seven real-stakes  $Simple \ CTB$  questions in the recharge domain that we collected at the second fitbit sync visit for a subset of partici-

<sup>&</sup>lt;sup>27</sup>This subsample is excluded from all analyses. To make this and related preference elicitations incentivecompatible, we used the strategy method: participants were presented with 11 different choices and informed that one of their choices would be randomly selected for implementation.

<sup>&</sup>lt;sup>28</sup>These questions were asked as a "warm-up" for the (unsuccessful) full CTB module. The Simple CTB seems to have performed better than the full version, as described in more detail in Online Appendix J.2.

<sup>&</sup>lt;sup>29</sup>Demand for more frequent payment is an incentive-compatible measure gathered by asking participants' preferences among the daily, base case, and monthly contracts. Higher balances and/or usage indicate a person's recharge purchases are less constrained, and thus, their discount rate over recharges is more likely to be low.

pants.<sup>30</sup> (Since our focus is impatience over effort, the term "Simple CTB" without further specification hereafter refers to the Simple CTB over effort measure.) Participants were asked to choose between four allocations of recharges between an earlier and a later date (selected from a discretized CTB budget set), and were told upfront that a randomly selected fraction of them would receive their choice from a randomly selected question. The second measure harnesses *Paycycle Effects* — the degree to which participants' compliance increases as the payday approaches — following the methodology of Kaur et al. (2015).

Impatience Over Effort versus Payments Our theoretical predictions rely on there being a distinction between discount rates over effort ( $\delta$ ) and payment (d), as we take comparative statics with respect to one holding the other constant. Appendix C provides two pieces of evidence that  $\delta$  and d are distinct. First, Section C.2 shows that, in our setting, populationlevel estimates of  $\delta$  and d are significantly different, with  $\delta < d$ . Second, Section C.3 shows that, at the individual level, there is no correlation between our measures of impatience over effort and our measures of impatience over payment.

### 4.3 Summary Statistics

The first column of Table 1 displays the baseline characteristics of our sample. The sample is, on average, 50 years old. The average monthly household income is approximately 16,000 INR (about 240 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). Panels D and E show our measures of impatience over effort and impatience over payment.

Baseline measures are balanced across treatment groups. Columns 3 and 6 show means for the control and base case groups, while columns 4–5 and 7–10 show differences in means relative to Control and Base Case, respectively. To explore balance, we jointly test the equality of all characteristics relative to Control for Incentives and Monitoring or the Base Case for all other incentives sub-groups. All tests fail to reject the null that all differences are zero. Covariates are also balanced in the subsample with post-intervention period data.

<sup>&</sup>lt;sup>30</sup>We began collecting these data after problems with the full CTB surfaced. Rather than further bog down the lengthy time-preference survey, we chose to add these simpler questions to a later encounter with participants.

	Groups				Incentives sub-groups					
	Full S	ample	Control	Incentives	Monitoring	Base Case	Threshold	Daily	Monthly	Small Payment
	Mean	SD	Mean	Coef	Coef	Mean	Coef	Coef	Coef	Coef
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A Domographies	( )					( )	,			. ,
A. Demographics	49.56	8.51	49.78	-0.33	0.50	49.60	-0.20	-0.03	-0.80	-0.50
1150	10.00	0.01	10.10	(0.39)	(0.46)	10.00	(0.61)	(0.96)	(0.26)	(0.64)
Female $(=1)$	0.42	0.49	0.46	-0.04	-0.02	0.41	-0.01	0.03	-0.04	0.07
				(0.06)	(0.57)		(0.79)	(0.55)	(0.38)	(0.27)
Labor force participation $(=1)$	0.74	0.44	0.73	0.02	-0.01	0.74	0.01	0.01	0.07	-0.04
	4405	9641	4400	(0.23)	(0.84)	4 4 777	(0.62)	(0.76)	(0.06)	(0.42)
Per capita income (iNK/month)	4400	5041	4488	(0.82)	(0.71)	4477	(0.92)	(0.17)	(0.69)	(0.78)
Household size	3.91	1.62	3.94	-0.03	-0.11	3.89	0.07	0.03	-0.14	-0.31
				(0.69)	(0.37)		(0.34)	(0.81)	(0.31)	(0.14)
B. Health										
Diagnosed diabetic $(=1)$	0.67	0.47	0.67	0.00	0.01	0.68	-0.01	-0.06	-0.06	-0.09
				(0.90)	(0.73)		(0.62)	(0.14)	(0.11)	(0.14)
Blood sugar index	0.01	0.92	0.01	0.00	0.04	0.02	-0.01	-0.02	-0.01	-0.17
Hb 1 1 a (mmal/mal)	0 60	9.99	° 67	(0.91)	(0.59)	° 79	(0.78)	(0.82)	(0.87)	(0.14)
HDATC (IIIII01/III01)	0.00	2.33	0.07	(0.92)	(0.64)	0.12	(0.71)	(0.45)	(0.76)	-0.38
Random blood sugar (mmol/L)	192.52	89.44	191.32	1.18	4.75	193.26	-1.03	2.32	0.05	-15.88
0 ( , , ,				(0.77)	(0.51)		(0.80)	(0.76)	(0.99)	(0.15)
Systolic BP (mmHg)	133.38	19.16	133.33	0.01	0.73	133.27	-0.44	1.98	0.90	2.35
				(0.99)	(0.65)		(0.60)	(0.23)	(0.58)	(0.34)
Diastolic BP (mmHg)	88.48	11.10	88.54	-0.08	-0.01	88.19	0.26	1.11	0.41	1.81
HbA1a: Diabatia (-1)	0.82	0.38	0.82	(0.88)	(0.99)	0.84	(0.00)	(0.24)	(0.65)	(0.20)
IDAIC. Diabetic $(-1)$	0.82	0.30	0.82	(0.88)	(0.76)	0.04	(0.03)	(0.02)	(0.10)	(0.13)
BP: Hypertensive $(=1)$	0.49	0.50	0.46	0.04	0.05	0.49	0.00	0.04	0.02	-0.03
				(0.13)	(0.21)		(0.98)	(0.33)	(0.69)	(0.59)
Overweight $(=1)$	0.61	0.49	0.62	-0.02	0.04	0.60	0.00	-0.04	-0.03	0.06
				(0.33)	(0.36)		(0.98)	(0.36)	(0.51)	(0.32)
BMI	26.42	4.34	26.52	-0.12	-0.05	26.47	-0.17	-0.06	-0.08	(0.52)
C Walking phase in				(0.50)	(0.83)		(0.51)	(0.00)	(0.04)	(0.57)
Exceeded step target $(-1)$	0.25	0.32	0.25	0.00	-0.01	0.23	0.03	0.02	0.04	0.04
Exceeded step target (-1)	0.25	0.52	0.25	(0.90)	(0.58)	0.25	(0.03)	(0.46)	(0.11)	(0.27)
Average daily steps	7004	3981	7066	-68	-174	6810	268	236	639	208
				(0.71)	(0.58)		(0.14)	(0.49)	(0.06)	(0.68)
D. Impatience over effort										
Impatience index (SD's)	0.09	0.99	0.00	0.12	0.05	0.14	-0.05	-0.10	0.05	0.12
	0.05	1 00	0.00	(0.06)	(0.64)	0.00	(0.46)	(0.40)	(0.70)	(0.50)
Predicted index (SD's)	-0.05	1.00	0.00	-0.06 (0.23)	-0.15	-0.02	-0.06	-0.07	(0.00)	-0.09
E Mabila machanaca				(0.23)	(0.05)		(0.17)	(0.42)	(0.97)	(0.40)
Current mobile balance (INB)	20.34	49.59	30.80	1.83	1 33	20.60	1.95	1.00	1 14	0.36
Current mobile balance (hvit)	23.04	43.03	30.00	(0.42)	(0.74)	23.03	(0.58)	(0.80)	(0.80)	(0.96)
Yesterday's talk time (INR)	6.58	8.76	7.22	-0.78	-0.75	6.58	-0.26	-0.72	1.09	-1.64
				(0.06)	(0.35)		(0.50)	(0.33)	(0.15)	(0.15)
Prefers daily $(=1)$	0.17	0.38	0.18	0.00	-0.02	0.17	0.01	0.03	0.04	0.02
	0.04	0.49	0.05	(0.79)	(0.48)	0.04	(0.58)	(0.29)	(0.26)	(0.73)
Prefers monthly $(=-1)$	0.24	0.43	0.25	-0.01 (0.58)	(0.02)	0.24	(0.00)	(0.03)	-0.01 (0.75)	(0.02)
F-tests for joint orthogonality				(0.00)	(0.43)		(0.30)	(0.00)	(0.10)	(0.11)
n-value				0.11	0.85		0.89	0.56	0 79	0.44
Sample size	•	•		0.11	0.00		0.00	0.00	0.10	11.0
Number of individuals	3 109		585	2 101	203	009	1 106	166	164	66
Percent of sample	100.0		18.3	75.3	6.4	28.3	34.6	5.2	5.1	2.1
Number of ind, with ped, data	2.559		0	2.359	200	890	1.079	163	163	64

Table 1: Baseline Summary Statistics and Balance Across Treatment Groups

Notes: "Mean" columns show group means; "Coef" columns show coefficients from regressing the variable on a treatment indicator among the treatment group and its main comparison group (for Incentives and Monitoring, the main comparison group is Control; for Threshold, Daily, Monthly, and Small Payment, it is Base Case). p-values are in parentheses. BMI is body mass index, BP is blood pressure, overweight is BMI>25, hypertensive is systolic BP>140 or diastolic BP>90. Current mobile balance is the available phone credit on the respondent's phone, and yesterday's talk time is the monetary equivalence of minutes used the day before the baseline survey. Threshold pools 4- and 5-day threshold groups. In Incentives and Monitoring, the total number of individuals is larger than the number with pedometer data as some participants withdrew immediately. The F-tests are from separate regressions for each treatment group.

### 4.4 Outcomes: Exercise

We measure exercise using a time-series dataset of daily steps walked by each participant with a pedometer during the intervention period and (for a subset of the sample) 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. All analyses use *pedometer* steps as the outcome; however, payments to participants were based on *reported* steps.<sup>31</sup>

### 4.4.1 Data Quality Controls

A potential issue with the daily step data is that we only observe steps taken while participants wear the pedometer. Because participants in the incentive groups are rewarded for taking 10,000 steps in a day with the pedometer, they have an additional incentive to wear the pedometer. This could lead to a potential selection issue if the incentive group participants wear their pedometers more than the monitoring group.

To minimize selective pedometer-wearing in the intervention period, we incentivized participants to wear their pedometers. We offered a cash bonus of 200 INR ( $\approx 3$  USD) if participants wore their pedometer (i.e., had positive steps) on at least 80% of days. As a result, pedometer wearing rates are high, and the difference between treatments is small: 85% in Monitoring versus 88% in Incentives. However, the difference is statistically significant (Table A.2, column 2). To address the imbalance, we show robustness to Lee (2009) bounds accounting for missing step data due to not wearing pedometers.<sup>32</sup> Our primary specifications do not condition on wearing the pedometer (instead we set 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.

We also assess whether the incentive group wore their pedometers for more minutes per day, conditional on wearing. To do so, we use data recorded by pedometers on the times that the participant put it on and took it off. Reassuringly, these times are balanced across arms (Online Appendix Table F.4, Panel B).

To encourage participants to wear their pedometers in the post-intervention period, we provided all participants with a small incentive for wearing their pedometers on a sufficiently high fraction of days. While average pedometer-wearing rates declined somewhat to 69% (from 87% in the intervention period), the rates are balanced across arms.

Another concern is that participants might give their pedometers to someone else. Our data suggest that this concern is limited. First, we performed 835 unannounced audit visits

<sup>&</sup>lt;sup>31</sup>Although incentives were delivered for reported steps, we cross-checked reports with actual pedometer data after every pedometer sync. Anyone who was overreporting was initially warned, then suspended, and eventually terminated from the program if the behavior continued. Online Appendix G provides more detail.

 $<sup>^{32}</sup>$ 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 are balanced across Incentives and Monitoring (column 3, Table A.2). While our main specifications drop days with missing pedometer data, Online Appendix Table F.3 shows robustness to alternate specifications and Lee bounds. While missing data are balanced overall, one specific source of missing data (mid-intervention withdrawals) is imbalanced (column 6 of Table A.2). Results are robust to Lee bounds accounting specifically for that source (column 5 of Online App Table F.3).

to participants' homes. In 99.6% of visits, participants were not sharing their pedometers. Second, we check if participants' minute-wise step counts exceed age-based expectations. This is very rare, and balanced across Incentives and Monitoring (Online App. Table F.4).

### 4.5 Outcomes: Health

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 treatment group (Control, Monitoring, and Incentive; *p*-value for equality 0.99).

Our primary health outcome is blood sugar, the main clinical marker of diabetes. Our preferred outcome variable for blood sugar is a standardized index of two measures: HbA1c (longer-term blood sugar control) and RBS (short-term blood sugar control). While our AEA registry prespecified HbA1c as our blood sugar measure, accurately measuring HbA1c in the field proved challenging.<sup>33</sup> Thus, we also measure RBS, another blood sugar measure strongly associated with diabetes severity (Bowen et al., 2015)<sup>34</sup> that is easier to accurately measure. Our measures of RBS and HbA1c have independent predictive power for blood sugar control,<sup>35</sup> so our preferred measure incorporates both, but we also present the measures separately.

Since exercise is also associated with improvements in hypertension and cardiovascular health, we measured blood pressure, BMI, and waist circumference as secondary health outcomes. We combine these three measures with the two blood sugar measures to construct a standardized "health risk index".

We also gathered information on two secondary health outcomes: mental health and anaerobic fitness. We measure mental health using seven questions from RAND's 36-Item Short Form Survey. Anaerobic fitness is measured via two fitness tests (time to complete five stands from a seated position, and time to walk four meters). Following Kling et al. (2007), we impute missing components of all indices as the mean within an individual's group (Control, Monitoring, or Incentive) for individuals who have at least one nonmissing index component.

### 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 setting to explore our contract variations. Second, we show that adding a time-bundled threshold increases effectiveness. Third, we show that the threshold is particularly effective for those with higher impatience over effort, in line with our theoretical prediction that impa-

<sup>&</sup>lt;sup>33</sup>The only available measurement tool (the SD A1cCare analyzer from SD Biosensor) was temperaturesensitive and error prone, and in a validation subsample it did not align with gold-standard lab measurements.

<sup>&</sup>lt;sup>34</sup>RBS is sensitive to recent activity such as eating and is therefore problematic as a clinical measure, but yields a good measure of average glycemic control in a sample.

<sup>&</sup>lt;sup>35</sup>Online Appendix Table F.5 shows that baseline RBS strongly predicts endline HbA1c in Control even conditional on baseline HbA1c, and vice versa.

tience over effort is a mechanism for its effectiveness. Finally, we find that higher-frequency payments do not increase effectiveness, suggesting limited impatience over payments.

### 5.1 Incentives Increase Exercise

We first test whether providing financial incentives increases steps and compliance with the 10,000-step target during the intervention period. To do so, we compare outcomes in the pooled incentive groups with the monitoring group, thus isolating the impact of the financial incentives alone. We estimate regressions of the following form:

$$y_{it} = \alpha + \beta Incentives_i + \mathbf{X}'_i \gamma + \mathbf{X}'_{it} \lambda + \varepsilon_{it}, \tag{7}$$

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<sub>i</sub>* is an indicator for being in the incentive group; and  $X_i$  and  $X_{it}$  are vectors of individual- and day-level controls, respectively, described in the notes to Table 2. We exclude the control group, for whom we have no pedometer data. We cluster the standard errors at the individual level. The coefficient of interest,  $\beta$ , is the average treatment effect of Incentives relative to Monitoring only. Panel A of Table 2 shows the results.

Incentives have large impacts on walking, increasing the share of days that participants reach their 10,000-step target by 20 pp or roughly 70 percent (column 1 of Table 2). This effect does not simply reflect participants shifting steps from one day to another: column 2 of Table 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 each day. This treatment effect is at the high end of effect sizes for pedometer incentives (found in non-diabetic populations in developed countries), which range from only 1.5 steps in Bachireddy et al. (2019) to 1,050 steps in Finkelstein et al. (2016).

The estimated effects of incentives on exercise 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 Online Appendix Table F.3 reports Lee bounds which account for the non-random patterns of missing data, with similar results.

### 5.2 Time-bundled Threshold Contracts Increase Average Effectiveness

We begin our analysis of time-bundled thresholds by comparing the average performance of the threshold and the (linear) base case contracts. Prediction 2 suggests that when the effort discount rate is high, as it appears to be in our sample (Appendix C.2), time-bundled threshold contracts tend to be more effective overall than linear contracts. In order to establish that the thresholds are effective on average, we can show that they result in weakly more compliance and weakly higher cost-effectiveness than the base case contract in the full sample, with one inequality strict, as described in Section 2. We thus examine compliance

Dependent variable:	Exceeded step target	Daily steps	Daily steps $(if > 0)$	Earned payment when target met
	(1)	(2)	(3)	(4)
A. Pooled incentives				
Incentives	$0.200^{***}$ [0.0186]	$1266.0^{***}$ $[208.7]$	$1161.5^{***}$ [188.5]	$0.952^{***}$ [0.00305]
B. Unpooled incentives				
Base Case	$0.211^{***}$ [0.0201]	$1388.4^{***} \\ [222.1]$	$1203.1^{***}$ [199.9]	$1.006^{***}$ [0.00262]
Threshold	$0.198^{***}$ [0.0199]	$1216.3^{***} \\ [220.9]$	$1142.6^{***} \\ [198.5]$	$0.892^{***}$ [0.00546]
Daily	$0.201^{***}$ [0.0303]	$\frac{1122.5^{***}}{[331.5]}$	$\frac{1283.1^{***}}{[277.9]}$	$1.003^{***}$ [0.00362]
Monthly	$0.177^{***}$ [0.0288]	$1274.2^{***} \\ [307.4]$	$1179.4^{***} \\ [271.1]$	$1.002^{***}$ [0.00325]
Small Payment	$0.137^{***}$ [0.0383]	$731.5^{*}$ [386.2]	$552.9^{*}$ [335.0]	$1.000^{***}$ [0.00479]
p-value for Base Case vs				
Daily	0.710	0.350	0.730	0.360
Monthly	0.180	0.650	0.910	0.190
Threshold	0.360	0.210	0.610	< 0.001
Small Payment	0.040	0.060	0.030	0.180
Monitoring mean	0.294	6,774	$7,\!\overline{986}$	0
# Individuals	2,559	2,559	2,557	2,394
# Observations	205,732	205,732	180,018	99,406

 Table 2: Incentives Increase Average Walking

Notes: This table shows the treatment effect of incentives on walking (relative to Monitoring). Incentive groups are pooled in Panel A and considered separately in Panel B. The columns show estimates of coefficients from equation (7) (Panel A) and (8) (Panel B) using intervention-period pedometer data at the individual-day level. "Exceeded step target" is an indicator for whether the individual exceeded their step target and "Earned payment when target met" is an indicator for receiving payment on a given day, conditional on meeting the step target. All regressions control for gender, an indicator for being in the cross-randomized text message group, average daily steps in the phase-in period (before randomization), second order polynomials of age, weight, and height, and year-month and day-of-week fixed effects. Online Appendix Table F.6 shows robustness to excluding controls, adding stratum fixed effects, or selecting controls by double-Lasso. The sample includes the incentive and monitoring groups. The omitted category is Monitoring. Threshold pools the 4- and 5-day threshold groups. Standard errors, in brackets, are clustered at the individual level. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.

and cost-effectiveness in turn.

**Compliance** We find that adding a time-bundled threshold does not change average compliance relative to the Base Case. Specifically, to test for differences across the incentive treatment groups, we estimate regressions of the following form:

$$y_{it} = \alpha + \sum_{j} \beta_{j} \times \left(incentives^{j}\right)_{i} + \mathbf{X}_{i}^{\prime} \gamma + \mathbf{X}_{it}^{\prime} \theta + \varepsilon_{it}, \tag{8}$$

where  $y_{it}$  are daily walking outcomes and  $(incentives^j)_i$  is an indicator for whether individual i is enrolled in incentive treatment group  $j \in (\text{Daily, Base Case, Monthly, Threshold, Small Payment})$ . The  $\beta_j$  coefficients capture the average effect of each incentive treatment group relative to Monitoring. Panel B of Table 2 displays the results.

The effect of the threshold contract on compliance is very similar to the effect of the base case contract, with the estimates within 1.3 pp of each other and the difference not statistically significant (p-value=0.360). Figure 2(a) displays the result graphically. It also shows the 4- and 5-Day Threshold separately—neither has meaningfully different compliance than Base Case.



Figure 2: Thresholds Do Not Affect Average Walking But Increase Cost-Effectiveness

Notes: The figure compares the base case treatment with all other incentive treatments during the intervention period. Panel (a) shows the average probability of exceeding the daily step target, with the dashed line representing the Monitoring mean. Panel (b) shows the fraction of days on which the participants received payments, conditional on meeting the step target (the Monitoring mean here is 0). In both Panels, the confidence intervals represent tests of equality between Base Case and each treatment group using the same controls as in Table 2. Data are at the individual-day level. Threshold pools the 4- and 5-day threshold groups.

**Cost-Effectiveness and Overall Effectiveness** While compliance is similar, the threshold contracts are more cost-effective than the base case contract. Individuals in the threshold groups only receive payment for exceeding the step target if they do so on at least four or five days in a given week; when they comply on fewer days, they are not rewarded. As shown in Figure 2(b), 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. This difference, which is mathematically equivalent to a difference in cost-effectiveness since the groups receive the same daily payment rate, is significant at the 1% level, as shown in Table 2 column  $4.^{36}$  As a result, the cost-effectiveness of the threshold contracts are 11% and 17% higher than that of the base case contract (Table A.3).

Because the threshold contracts have the same compliance and are more cost-effective than Base Case, they are more effective overall. For comparison, the small payment treatment is also more cost-effective than Base Case (it pays half as much per day complied), but this comes at the cost of reduced compliance, as shown in Panel B of Table 2. The fact that the threshold contracts achieve the same compliance as the base case contract for lower cost implies that a budget-neutral threshold (i.e., a threshold contract with the same average cost as Base Case) would have higher compliance than Base Case.

**Distributional Impacts and Effectiveness in Other Settings** Equal compliance and higher cost-effectiveness only necessarily imply higher effectiveness if the benefits of compliance are linear. While the estimated health benefits of compliance are approximately linear in our setting (Warburton et al., 2006; Banach et al., 2023), many settings have nonlinear benefits. In those settings, effectiveness depends not just on average compliance but also on its distribution and variance.

Theory suggests that thresholds could increase the variance of compliance by decreasing intermediate effort just below the threshold (Grant and Green, 2013). This could decrease the effectiveness of thresholds for principals who particularly value compliance improvements among those with low average compliance (i.e., principals with concave benefits to compliance). Appendix D assesses the effect of thresholds on the distribution of compliance (e.g., Figure D.4). While the thresholds have some impact, the magnitude of the impact is small, implying that thresholds would be preferred even by many principals with concave benefit functions, provided their benefit functions are not too concave.

### 5.3 Mechanisms: Effort Impatience Contributes to Threshold Effectiveness

Our theory indicates that high discount rates over future effort may be an important contributor to the effectiveness of threshold contracts. This section presents empirical evidence supporting this theoretical prediction, as we show that the threshold is more effective for more impatient individuals. Specifically, relative to the base case contract, the threshold

$$\frac{C}{P} = \frac{C}{C \times \text{daily payment rate} \times \text{fraction of days on which earned payment when step target met}}$$

and note that the C cancels out, and all but Small Payment have the same *daily payment rate* of 20 INR.

 $<sup>^{36}</sup>$ For all groups but Small Payment, we can test for equal cost-effectiveness by testing for equality in the Figure 2(b) outcome of *fraction of days on which earned payment when step target met*. To see the equivalence, express cost-effectiveness as

contract generates significantly more compliance from more impatient individuals without any loss in cost-effectiveness. Since Predictions 1 and 2 regard heterogeneity in the threshold effect *holding all else constant*, this heterogeneity analysis is a direct test of the theory only if impatience is not correlated with other variables that influence the effectiveness of the threshold. We show that the estimated heterogeneity is robust to controlling for many covariates interacted with Threshold, suggesting that this condition holds. We also use machine learning to demonstrate the importance of impatience for predicting the Threshold effect.

**Compliance** We use a regression of the following form to test for heterogeneity in the effect of the time-bundled threshold on compliance by impatience:

$$y_{it} = \alpha + \beta_1 Impatience_i \times Thresh_i + \beta_2 Thresh_i + \beta_3 Impatience_i + \mathbf{X}'_i \pi + \mathbf{X}'_{it} \theta + \varepsilon_{it}, \quad (9)$$

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 a threshold group. Measures of individual impatience are denoted by  $Impatience_i$ . Because some of the measures are estimated, we present bootstrap confidence intervals in the table as well as Gaussian standard errors and *p*-values in table notes when available.

We restrict the sample to 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  $\beta_1$ , which captures how the effect of Threshold (relative to Base Case) varies with impatience. Our prediction is that  $\beta_1 > 0$ .

Table 3 shows that, consistent with the theory, relative to the Base Case, thresholds generate meaningfully more compliance among those with higher impatience over effort. Column 1 uses the impatience index as the measure of impatience. Having a one standard deviation higher value of the impatience index increases compliance in Threshold relative to Base Case by 4 pp (statistically significant at the 5% level). Column 2 uses a dummy for having an above-median value of the impatience index. While this estimate leverages less of the variation available in the data and hence has lower power, it is easier to interpret. Relative to Base Case, Threshold generates 6.5 pp higher compliance for those with abovemedian impatience than those with below-median (p-value<0.05). This represents a large increase, equal to over 30% of 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 verify the results in the full sample, columns 3 and 4 use the predicted impatience index, which is available for the full sample. We find very similar (and more precise) results, with p-values <0.01 and <0.05 in columns 3 and 4, respectively.

The point estimates in columns 2 and 4 imply that, relative to the linear contract, the threshold contract increases compliance among the more impatient (by 2-3 pp), while decreasing it among the less patient (by 3-4 pp).

Columns 5 and 6 of Table 3 show that these heterogeneity results are robust to using our

Dependent variable:		E	xceeded step	target ( $\times 100$	)	
	(1)	(2)	(3)	(4)	(5)	(6)
Impatience $\times$ Threshold	$3.80^{**}$ [0.22, 7.38]	$6.50^{**}$ [0.39,12.60]	$3.12^{***}$ [1.00, 5.17]	$5.94^{**}$ [0.28, 10.01]	$6.06^{**}$ [0.42, 11.71]	$4.70^{*}$ [-0.70, 10.11]
Threshold	-1.30 [-5.24, 2.64]	-4.13* [-8.80,0.54]	-1.18 [-3.36, 0.93]	-3.41** [-5.92, -0.47]	-4.29** [-8.55, -0.02]	-3.78** [-7.51, -0.05]
Impatience	-2.97** [-5.54, -0.39]	-5.03* [-10.55,0.49]	-2.38*** [-3.88, -0.85]	-5.3** [-8.13, -0.69]	-2.37 [-7.74, 3.00]	-2.67 [-6.05, 0.71]
Impatience measure:	Impatience index	Above- median impatience index	Predicted impatience index	Above- median predicted index	Chose commitment	Simple CTB
Sample:	Late	Late	Full	Full	Full	Full
Base Case mean	50.4	50.4	50.2	50.2	49.9	50.2
# Individuals	1,075	1,075	$1,\!969$	$1,\!969$	1,798	$1,\!967$
# Observations	86,215	86,215	$157,\!946$	157,946	144,099	157,799

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

Notes: This table shows heterogeneity by impatience over effort in the effect of the threshold contracts relative to the linear Base Case. The sample includes the base case and threshold groups only. The impatience measure changes across columns; its units in columns 1 and 3 are standard deviations. "Chose commitment" is the average of indicators for preferring the 4- and 5-day threshold to the base case contract. "Simple CTB" is an average of two indicators for impatient walking choices. See Online Appendix Table F.2 for robustness to different ways of combining the Chose commitment and Simple CTB choices. Column 5 has fewer observations because the questions used to construct the "Chose commitment" measure had a "no preference" response option which we treat as missing. Online Appendix Table F.8 shows that the results are similar when we instead assume "no preference" responses indicate a preference for either option. The "Late" sample includes only participants who were enrolled after we started measuring the impatience index; the Full sample includes everyone. Threshold pools the 4- and 5-day threshold groups. See Online Appendix Table F.9 Panel B for results with the threshold groups disaggregated. Data are at the individual-day level. Bootstrapped 95% confidence intervals are in brackets. Bootstrap draws were clustered at the individual level. For regressions using the predicted impatience index we conduct three steps in each bootstrap sample to construct the 95% confidence interval: 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 (9). 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, 3.77 and 0.076; for column 5, 3 and 0.044; for column 6, 2.83 and 0.097. Controls are the same as in Table 2. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.

alternative measures of impatience: incentivized demand for commitment, and the simple CTB measure (*p*-values <0.05 and <0.10, respectively). Although the impatience index is our preferred measure, given our *ex ante* intention that it be primary, we find the robustness across multiple types of measures reassuring.

The heterogeneity based on the demand for commitment measure suggests that the threshold contracts work well among sophisticated impatient people in particular. To shed light on whether the thresholds also work well for impatient naifs, we re-estimate equation (9) but exclude from the sample the participants who demanded commitment. Since this restriction should exclude most sophisticated impatient people for whom the threshold will increase compliance, the *Impatience*  $\times$  *Threshold* coefficient should be primarily identified by naifs.

The results, shown in Table A.4, suggest that threshold contracts are also relatively more effective among naifs. The *Impatience*  $\times$  *Threshold* remains positive and relatively large across all impatience measures, and maintains at least 10% significance except when the sample size is small (columns 1 and 2). Thus, consistent with our theory, both impatient naifs and impatient sophisticates appear to have higher compliance with threshold contracts, and it is only patient people who appear to do more poorly in the threshold groups (as evidenced by the negative main effect of Threshold in Table A.4).

**Cost-Effectiveness and Effectiveness** Prediction 1 suggested that, in addition to increasing compliance more among people who are impatient over effort, threshold contracts should also increase *effectiveness* more. Since we have already established the compliance result, to demonstrate the effectiveness result it is sufficient to show that, relative to Base Case, the threshold contracts do not decrease cost-effectiveness more among the impatient than the patient. Table A.5 shows that this is true.<sup>37</sup> Paired with the compliance result, this implies that the threshold increases effectiveness more for those with higher impatience over effort than lower impatience over effort.

In addition to caring about the threshold's relative effects among more and less impatient people, a policymaker may also care whether the threshold has higher effectiveness than the linear contract for each group in our specific context. For those with above-median impatience, the threshold increases both compliance and cost-effectiveness and is thus more effective overall than the linear contract. This important finding is consistent with Prediction 2 and implies that principals could increase effectiveness by using thresholds for impatient populations. For those with below-median impatience, the answer is more ambiguous. Relative to Base Case, Threshold decreases compliance but increases cost-effectiveness. Whether a principal would prefer it for this population thus depends on the principal's specific value of compliance ( $\lambda$  from Section 2).

**Robustness of the Compliance Heterogeneity by Impatience** Impatience over effort is correlated with other factors, such as baseline exercise levels, that may also independently influence the performance of thresholds. For example, if impatient people are more likely to also have counterfactual walking that is right below the threshold level (as opposed to above

<sup>&</sup>lt;sup>37</sup>We do see that the threshold is slightly less cost-effective when the impatience measure is demand for commitment. While the significance of this coefficient at the 10% level (out of six specifications) may be due to chance, demand for commitment (which is equivalent to a preference for the threshold contract over the base case) requires both (sophisticated) impatience and a perception that the participant will regularly receive payment under the threshold contract (e.g., because of low walking costs). This second factor may underlie the heterogeneity in cost-effectiveness.

or far below), that could independently cause them to respond more to the threshold. To shed light on whether this type of factor plays a role in the heterogeneity we see, Figure 3 examines the robustness of the Table 3 estimates to controlling for other baseline covariates and their interactions with Threshold, such as the mean of baseline steps (a proxy for the mean of the walking cost distribution), the standard deviation of baseline steps (a proxy for the variance of the walking cost distribution), and fixed effects for the number of days the individual walked at least 10,000 steps in the baseline period (a proxy for how close to the threshold the person's counterfactual walking is). We also control for risk aversion and "scheduling uncertainty" (the stated frequency with which unexpected events arise) and our proxies for impatience over payment (and all of their interactions with Threshold), which could all influence the performance of threshold contracts, among other controls.

The coefficients on the interaction of impatience and Threshold remain stable as we add these additional controls. Panels A and B of Figure 3 show this for the actual and predicted impatience index, respectively, and Online Appendix Figure F.1 shows this for our other impatience measures. The coefficient stability suggests that it is impatience itself (and not its correlates) driving the estimated relationships.

A potential confound that was difficult to measure at baseline (and hence which we do not control for) is the propensity for habit formation. However, we measure the propensity for forming habits at endline by assessing how much of the treatment effect of incentives persists after payments stop. Table F.10 in the Online Appendix reassuringly shows that impatience does not predict persistence, suggesting that the propensity to form habits is not correlated with impatience in our setting.

Even if omitted variables were affecting our Table 3 heterogeneity estimates, the estimates are still 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.

Machine learning also confirms the importance of impatience for driving the treatment effect of thresholds. First, using a causal forest (Athey et al., 2019), we find that the impatience index has a particularly strong signal in predicting the threshold effect, as it exhibits the highest variable "importance" among all covariates (Figure A.2). Second, Figure A.3 shows that the Chernozhukov et al. (2018) approach, which addresses multiple hypothesis testing and overfitting concerns, also identifies impatience as a statistically significant predictor of threshold effectiveness. While it is not necessary for us to use the Chernozhukov et al. (2018) approach (since we planned *ex ante* to analyze heterogeneity based on impatience), we find it reassuring that this robust method identifies impatience as a significant predictor.

### 5.3.1 Policy Implications of Time-Bundled Thresholds Results

We find that, in the full sample, time-bundled thresholds increase effectiveness by increasing cost-effectiveness without decreasing compliance. Moreover, consistent with theory, we



Figure 3: Threshold Heterogeneity by Impatience is Robust to a Variety of Controls

Notes: Panel A displays robustness of the *Threshold*×*Impatience* coefficient from column 1 of Table 3 (actual impatience index) to including various additional controls, interacted with *Threshold*, in the regression. As a reference, the "No additional controls" row just displays the *Threshold*  $\times$  *Impatience* coefficient and 95% confidence interval from column 1 of Table 3. The next 17 rows show estimates of the  $Threshold \times Impatience$ coefficient with two additional controls: the main effect of the covariate listed and the covariate interacted with *Threshold*. The final "All controls" row shows estimates of the *Threshold*  $\times$  *Impatience* coefficient from a regression where we control simultaneously for all covariates included in the previous 17 rows (both main effects and interactions with *Threshold*). Panel B is analogous but based on column 3 of Table 3 (predicted impatience index). Baseline steps (mean) and baseline steps (sd) represent the mean and standard deviation of the baseline steps distribution. Baseline step target compliance (FEs) are fixed effects for the number of days the individual walked at least 10,000 steps in the baseline period. Scheduling uncertainty represents the individual's stated frequency of facing unexpected events that would prevent them from walking for 30 minutes in a given day. Risk aversion is an incentivized measure from a multiple price list. The health risk index is an index created by taking the average of endline HbA1c, RBS, mean arterial BP, waist circumference, and BMI, standardized by the control group mean and standard deviation. The blood sugar index is constructed by taking the mean of endline HbA1c and RBS standardized by their average and standard deviation in the control group. Income is winsorized at the 5th and 95th percentiles. Mobile balance and Yesterday's talk time are as in Table 1. Data are at the respondent-day level and include the threshold and base case groups only. All confidence intervals are constructed via bootstrap, with bootstrap draws as in Table 3. See Online Appendix Figure F.1 for a version with our alternate impatience measures.

provide evidence that one of the mechanisms for the effectiveness of thresholds is impatience over future effort. Specifically, we show that time-bundled thresholds generate meaningfully greater compliance and effectiveness among the impatient than the patient.

Our findings suggest that time-bundled thresholds are a useful policy tool for adapting incentives to address impatience over effort. Policymakers could tailor time-bundled thresholds at the population level, using them for groups known for greater impatience, such as those with chronic disease (Wainwright et al., 2022) or younger people (Read and Read, 2004).

Alternatively, policymakers could personalize the assignment of time-bundled thresholds within a population. One approach would be to personalize contracts by measuring impatience, which Andreoni et al. (2023) finds is feasible.<sup>38</sup> Although participants might misreport their discount rates, evidence suggests that this type of misreporting is limited in the context of incentives for behavior change.<sup>39</sup> Another approach would be to personalize on predictors of impatience that are harder to manipulate. Appendix E demonstrates that a prediction of impatience based on such characteristics (e.g., gender and BMI) predicts heterogeneity in the Threshold effect.

However, given the inherent noise in any impatience measure, the effectiveness of personalization depends on both the quality of the measure and the overall level of impatience in the population. In populations with high overall impatience, like ours, the benefits of personalization may be limited. Online Appendix H explores whether personalizing using any of our impatience measures improves effectiveness relative to applying the threshold universally. We find that the answer is ambiguous: personalization increases compliance but decreases cost-effectiveness. A promising approach to improve the performance of personalization could involve using multiple impatience measures to reduce exclusion errors. For example, assigning the threshold to those who rate highly on either of two measures proves more effective than relying on a single measure alone.

### 5.4 Payment Frequency Does Not Meaningfully Change Effectiveness

We now explore the roles of payment frequency and the discount rate over financial payments in incentive design. To do so, we compare average compliance in the daily, base case (weekly), and monthly groups. Columns 1-3 of Panel B of Table 2 and Figure 2(a) show that the three payment frequency treatments have similar effects on 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. The lack of between-treatment frequency

<sup>&</sup>lt;sup>38</sup>Andreoni et al. (2023) customized the parameters of a contract for 2-day vaccination drives to equalize worker effort across both days, using discount rates measured in a simple effort allocation experiment. They succeeded: customized contracts resulted in more equal effort than randomized contracts.

<sup>&</sup>lt;sup>39</sup>Dizon-Ross and Zucker (2023) explores an incentive program for steps in the same setting as this study. While that paper's goal is to evaluate the use of mechanism design to personalize step targets, and it does not implement any time-bundled contracts, it does show that participants do not manipulate their observable characteristics to avoid assignment to a financially-dominated contract with a higher step target.



Figure 4: 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 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 same controls as 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 in Table 2. Panel (a) includes the monitoring and base case groups; Panel (b) includes the monitoring and monthly groups. Data are at the individual-day level.

effects implies that the discount rate over our financial payments is small. However, our precision here is somewhat low. To gain precision, we also examine how compliance changes as the payday approaches in the base case and monthly groups. If people are impatient over payments, compliance should increase as the payday approaches (as shown in both Kaur et al. 2015 and Prediction 4 in Appendix B.4). Yet, Figure 4 shows that walking behavior is remarkably steady across the payment cycle. The estimates here are more precise, allowing us to rule out even small effects of decreasing the lag until payment on compliance.<sup>40</sup>

The limited effect of increased payment frequency theoretically hinges on the discount factor over our contract payments, which Appendix C.2 shows are close to 1. While this estimate is specific to our sample and payment modality, limited impatience over payments is not rare (e.g., Augenblick et al., 2015; Andreoni and Sprenger, 2012a).<sup>41</sup> Thus, we expect that lackluster payment frequency effects may be common. Indeed, DellaVigna and Pope (2018) also finds limited impact of randomizing the payment lag among US participants on mTurk. Note that the effects of payment frequency are relevant for considering time-bundled contracts as well: if higher-frequency payments were more effective, it could present

<sup>&</sup>lt;sup>40</sup>Specifically, Online Appendix Table F.12 shows estimates of the change in compliance as the payment date approaches within the base case and monthly groups, conditional on day-of-week fixed effects. The estimates are precise and near zero, allowing us to rule out even small effects of more immediate payment. For example, if we assume linearity of compliance in lag to payment, then the confidence interval around the slope in the base case treatment rules out the possibility that, because of monetary discounting, daily payments would generate a mere 0.3 pp more compliance than Base Case.

<sup>&</sup>lt;sup>41</sup>Cross-country preference surveys suggest that low impatience over payments is more common in higherincome countries but varies, with some lower-income countries exhibiting low impatience (Falk et al., 2023). Indeed, several studies in developing countries estimate small discount rates over payment (e.g., Mbiti and Weil (2013) in Kenya, and Tanaka et al. (2010) in Vietnam), although some find larger (Kaur et al., 2015).

challenges for time-bundled contracts, which require a delay between effort and payment.

### 5.5 Effectiveness and Welfare

This paper evaluates ways to increase contract effectiveness, 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 objects such as the social cost of public funds 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 appears to be true in our case<sup>42</sup>— then variations that increase compliance and effectiveness will likely increase social welfare.

One potential concern with our time-bundled threshold contract would be if it improved effectiveness or welfare but was not Pareto improving, instead decreasing some individuals' welfare relative to a no-incentives benchmark. This concern is particularly vivid in light of evidence that commitment contracts can decrease welfare among partially naive individuals who pay upfront for commitment but fail to follow through (e.g., Bai et al., 2021).

Even though individuals do not pay upfront for threshold contracts, there is a potentially analogous issue. Naifs may comply on the early days of a threshold contract (a form of paying upfront) but fail to receive compensation because they do not follow through on the later days. However, there are theoretical reasons to doubt that this would happen much in practice.<sup>43</sup> Empirical evidence also suggests that our threshold contract did not reduce participants' welfare. At endline, most participants expressed interest in continuing the program, with no significant difference between the threshold and other groups, nor between more and less impatient participants in the threshold groups. Additionally, there is no heterogeneity by impatience in the likelihood of complying but not being paid under the threshold.<sup>44</sup>

### 6 Empirical Results: Program Evaluation

The impacts of incentive programs on health and behavior are of policy interest, especially among populations at high risk of complications from chronic disease. This section examines the exercise impacts over time and provides evidence that the program improved health.

<sup>&</sup>lt;sup>42</sup>The estimated social benefits of walking include both large private health benefits (e.g., Warburton et al., 2006; Banach et al., 2023) and positive fiscal externalities such as reduced public healthcare costs (e.g., Sangarapillai et al., 2021), and are estimated to be large relative to the social costs of these interventions.

<sup>&</sup>lt;sup>43</sup>Specifically, later compliance costs must be much larger than earlier costs for lack of follow through to be an issue: as the compliance approaches the threshold, the incentives for marginal compliance become more and more high powered. See Appendix B.1 and especially footnote 52 for more detail.

<sup>&</sup>lt;sup>44</sup>Our effectiveness results are specific to time-bundled dynamic contracts and do not hold for all dynamically non-separable contracts. For example, Carrera et al. (2022) empirically analyzes a threshold-style contract that differs from ours in (a) not paying "on the margin" for compliance exceeding the threshold and (b) using a low threshold level (K/T). Both features imply that, in that contract, current effort often *decreases* the payment for future effort and hence make the contract not time-bundled.

### 6.1 The Impacts of Incentives Persist During and After the Intervention Period

Chronic disease management requires lasting lifestyle changes, underscoring the need for programs that yield sustained improvements in exercise. We show the treatment effects of our incentives intervention on exercise over time, first during the intervention period and then after the intervention ends. Figure A.4 estimates equation (7) separately by week of the intervention for walking outcomes. After an initial spike at week 1, the effect of incentives on walking remains stable for the full intervention period. This suggests that longer-term (and even permanent) programs have the potential to promote sustained exercise improvements, an encouraging finding as insurers and governments are increasingly rolling out such programs.

Do the effects of incentives persist after the payments stop? Studies of similar exercise programs find mixed results (e.g., Royer et al., 2015; Charness and Gneezy, 2009). To examine persistence, we estimate equation (7) using the pedometer data from the 12 weeks after the intervention ended.<sup>45</sup>

Table A.6 shows that the incentive group continues to walk significantly more then the comparison groups after incentives end. The treatment effect on steps is statistically significant and large: around 10% of the comparison group mean (columns 2 and 3), or roughly 60% of the size of the treatment effect of incentives on steps during the intervention period (which was 15% of the comparison group mean).<sup>46</sup> Figure A.5 shows that effects persist until the end of the 12-week post-intervention period. Our short-run incentive program may thus induce habit formation, resulting in long-term impacts.

### 6.2 Incentives Moderately Improve Health

We now examine whether the incentives program measurably improves health. We powered our RCT to detect the difference in health outcomes (which are relatively noisy) between the pooled incentive groups and Control. (While we did not power it to compare Incentives with Monitoring, we include this comparison alongside our comparison with Control for completeness.) Table 4 reports results from regressions of the following form:

$$y_i = \alpha + \beta_1 Incentives_i + \beta_2 Monitoring_i + \mathbf{X}'_i \gamma + \varepsilon_i, \tag{10}$$

where  $y_i$  is an endline health outcome for individual *i* and  $X_i$  is a vector of controls (shown in the table notes).  $\beta_1$  represents the overall effect of the incentive program.

The results suggest that the program moderately improves blood sugar and cardiovascular health. Column 1 presents the treatment effect on our preferred blood sugar measure, the

 $<sup>^{45}</sup>$ While we have pedometer data from Control during this period, sample size is limited: we collected post-intervention period data from only a third of participants. We thus pool Control and Monitoring, so the *Incentives* coefficient represents the effect of incentives relative to this pooled comparison group. Results are similar when we compare Incentives with Control alone; with only 70 people post-intervention, Monitoring is too small to analyze alone.

<sup>&</sup>lt;sup>46</sup>Since we compare the effect of Incentives relative to Control in the post-intervention period with the effect of Incentives relative to *Monitoring* in the intervention period, we will overestimate persistence if Monitoring alone increases steps. However, Online Appendix K suggests that monitoring does not affect steps.

index incorporating both HbA1c and RBS. Incentives improve the index by 0.05 standard deviations, significant at the 10% level. Columns 2 and 3 display HbA1c and RBS separately. Column 4 shows that incentives improve our health risk index by 0.05 standard deviations, significant at the 10% level.

Dependent variable:	Blood sugar index	HbA1c	Random blood sugar	Health risk index	Blood sugar index	HbA1c	Random blood sugar	Health risk index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Incentives	$-0.05^{*}$ $[0.03]$	-0.07 $[0.07]$	$-6.1^{*}$ [3.5]	$-0.05^{*}$ $[0.03]$	-0.10** [0.05]	-0.1 [0.1]	$-12.6^{**}$ [5.9]	-0.08** [0.04]
Monitoring	-0.02 [0.05]	-0.1 $[0.1]$	1.8 $[6.6]$	0.02 [0.04]	-0.06 [0.08]	-0.3 $[0.2]$	1.5 [10.5]	-0.05 $[0.07]$
p-value: I = M	0.491	0.534	0.188	0.119	0.574	0.294	0.133	0.556
Sample	Full	Full	Full	Full	Above- median blood sugar	Above- median blood sugar	Above- median blood sugar	Above- median blood sugar
Control mean # Individuals	$0\\3,067$	8.44 3,066	$193.83 \\ 3,067$	$\begin{array}{c} 0\\ 3,068\end{array}$	$.64 \\ 1,530$	$10.09 \\ 1,529$	$248.26 \\ 1,530$	.45 1,531

Table 4: Incentives Moderately Improve Blood Sugar and Cardiovascular Health

Notes: p-value I = M is Incentives = Monitoring. Columns 1-4 report results estimated in the full sample while columns 5-8 report results estimated in the sample with above-median blood sugar index. Observations are at the individual-level. HbA1c is the average plasma glucose concentration (%). Random blood sugar (RBS) is the blood glucose level (mg/dL). The blood sugar index is constructed by taking the mean of endline HbA1c and RBS standardized by their average and standard deviation in the control group. The health risk index is an index created by taking the average of endline HbA1c, RBS, mean arterial BP, waist circumference, and BMI, standardized by the control group mean and standard deviation. See Online Appendix Table F.13 for treatment effects on the index components not shown here. We follow World Health Organization guidelines to trim biologically implausible physical health outcomes and index components (i.e., z-scores < -4 or > 4). All specifications control for the baseline value of the dependent variable (or index components for indices), the baseline value of the dependent variable (or index components, and their second-order polynomials, as well as endline completion date, hour of endline completion, and dummy for late completion. Online Appendix Table F.14 shows that the estimates are similar but less precise when we omit the control variables, add stratum fixed effects, or use controls selected by double-Lasso. Robust standard errors are in brackets. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.

Since health outcomes among those with more severe diabetes are likely to be more responsive to exercise, we separately assess health impacts among those with higher blood sugar in columns 5-8 of Table  $4.^{47}$  As expected, the estimated health improvements are larger among those with above-median blood sugar. Incentives decrease the blood sugar index by 0.10 standard deviations and decrease RBS by 13 mg/DL, both significant at the 5% level.

Online Appendix Table F.15 examines whether the intervention had coincident impacts on mental health or fitness. Incentives improve the mental health index by 0.10 SD. In contrast,

<sup>&</sup>lt;sup>47</sup>While this subsample comparison was not formally pre-registered, our registry mentions that we stratified randomization by baseline HbA1c, a step we took to maximize statistical power for this subsample comparison. Our Analysis Plan, discussed in footnote 15, also outlines this analysis (see the first bullet of Hypothesis 1).

we find no effects on physical fitness, perhaps because we measure higher-intensity fitness while our intervention motivated lower-intensity exercise. Finally, we do not find impacts on diet or addictive good consumption (Online Appendix Table F.16).

### 6.3 Incentives and Chronic Disease: Results Summary and Discussion

Overall, these results show that incentives for exercise are a scalable, effective intervention to decrease the burden of diabetes in resource-poor settings. Exercise has important long-run health benefits for diabetics (e.g., Qiu et al., 2014), and our incentives substantially increase exercise during and after the intervention.

We also find clinically meaningful treatment effects on blood sugar: the estimated program impact of lowering RBS by 6 mg/dl would bring someone near the diabetes threshold a quarter of the way to healthy blood sugar levels.<sup>48</sup> In addition, an exploratory analysis shows that the treatment effects on RBS grow over the intervention period (Figure A.6), and hence might continue to amplify after the program given that exercise habits persist.

Our findings contribute to the health literature by providing the first experimental evidence of the impact of a pedometer-based intervention on blood sugar control. This is particularly significant given the scalability of such interventions in resource-poor settings. Previous interventions shown to improve health outcomes among diabetics require highly trained staff for frequent, personalized interactions (Aziz et al., 2015; Qiu et al., 2014).<sup>49</sup> In contrast, our intervention is scalable, low-cost, and induces lasting behavior change, with the potential to generate health savings that exceed program costs.<sup>50</sup> As a result, programs like ours could be essential tools in mitigating the global impacts of chronic disease.

### 7 Conclusion

This paper provide new insights into how to adjust incentives for impatience. We show both theoretically and empirically that, relative to time-separable contracts, the performance of time-bundled contracts is significantly higher among participants who are more impatient over effort. One useful feature of this prediction is that it holds regardless of whether agents are time-consistent or time-inconsistent, sophisticated or naive, thus broadening the arsenal for motivating impatient individuals. The intuition behind the prediction is that timebundled contracts enable the principal to purchase future effort from participants instead of current effort, which is advantageous when participants discount their future effort and are willing to effectively sell it "at a discount." The success of time-bundled contracts in

 $<sup>^{48}\</sup>mathrm{For}$  RBS measured in the morning, values less than 100 mg/dl are normal, 100-125 mg/dl indicate prediabetes, while above 126 mg/dl indicate diabetes.

<sup>&</sup>lt;sup>49</sup>Other incentive interventions for diabetics have targeted non-exercise outcomes and have found limited success and face similar scalability concerns (e.g., Finkelstein et al., 2017; Desai et al., 2020).

<sup>&</sup>lt;sup>50</sup>The per-person incentive program cost is 1,700 INR (26 USD), which is only 7% of the estimated annual direct cost of care for a diabetic in Tamil Nadu, or 28% of the direct cost of care during the 3-month intervention period (Tharkar et al., 2010). Interventions generating similar short-run levels of exercise among diabetics in other contexts have produced cost savings of this order of magnitude (Nguyen et al., 2008).

adjusting incentives for impatience is particularly striking when compared to the failure of higher-frequency payments in our sample. More frequent payments only work if individuals are impatient over *payment*, which may not be the case even for those with high discount rates over utility. In contrast, time-bundled contracts succeed by leveraging impatience over effort.

We explore time-bundled contracts using an experiment evaluating incentives for behavior change. This is a particularly apt setting for exploring the relationship between incentives and impatience, as a key rationale for incentivizing behavior change (e.g., savings, preventive health behaviors) is to mitigate underinvestment due to present bias and impatience. Adapting these types of incentives for impatience may thus be particularly impactful.

Our particular empirical setting also allows us to make a second contribution: we show that an incentive program for walking increases exercise and health in a diabetic population. In doing so, we provide some of the first evidence of a scalable, low-cost intervention with the potential to decrease the large and growing burden of chronic disease worldwide.

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 they do not exert enough effort today, so effort today increases their future payoff to effort. While standard models show one reason such contracts enhance effort is the high stakes of job loss in the presence of imperfect information, our work suggests that these contracts have extra bite if the agent discounts their future effort.

Our empirical findings regarding time-bundling are promising for policy and open up new research directions. One question for future research is how to optimize the specific features of time-bundled contracts such as the payment period length and threshold level. Future research can also probe external validity, exploring whether time-bundled contracts are indeed more effective than time-separable contracts in other populations with high discount rates of effort. Future work could also go further in exploring how to personalize time-bundled contracts at scale at the individual level, evaluating the options we explore in this paper (e.g., targeting via observables). Together, the answers to these questions will allow policymakers to effectively employ time-bundled contracts to motivate impatient people.

### References

- Andreoni, J., M. Callen, K. Hussain, M. Y. Khan, and C. Sprenger (2023). Using preference estimates to customize incentives: an application to polio vaccination drives in Pakistan. *Journal of the European Economic Association 21*(4), 1428–1477.
- Andreoni, J. and C. Sprenger (2012a). Estimating time preferences from convex budgets. American Economic Review 102(7), 1–28.
- Andreoni, J. and C. Sprenger (2012b). Risk preferences are not time preferences. American Economic Review 102(7), 3357–3376.
- Anjana, R. M., R. Pradeepa, M. Deepa, M. Datta, V. Sudha, and et al. (2011, dec). Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural india: Phase i results of the indian council of medical research-india diabetes (icmr-indiab) study. *Diabetologia* 54(12), 3022–3027.
- 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.
- Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. Annals of Statistics 47(2), 1179–1203.
- Augenblick, N. (2018). Short-term time discounting of unpleasant tasks. Unpublished manuscript.
- 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(3), 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, et al. (2019). Effect of different financial incentive structures on promoting physical activity among adults: A randomized controlled trial. JAMA Network Open 2(8).
- Bai, L., B. Handel, E. Miguel, and G. Rao (2021). Self-control and demand for preventive health: Evidence from hypertension in india. *Review of Economics and Statistics* 103(5), 835–856.
- Banach, M., J. Lewek, S. Surma, P. E. Penson, and A. S. et al. (2023, 8). The association between daily step count and all-cause and cardiovascular mortality: a meta-analysis. *European Journal of Preventive Cardiology*.
- Barrera-Osorio, F., M. Bertrand, L. L. Linden, and F. Perez-Calle (2011). Improving the design of conditional cash transfer programs : Evidence from a randomized evaluation in colombia organ donations. *American Economic Journal: Applied Economics* 3(April), 167–195.
- 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.
- Beaman, L., D. Karlan, and B. Thuysbaert (2014). Saving for a (not so) rainy day: A randomized evaluation of savings groups in mali. Technical report, National Bureau of Economic Research.
- Bommer, C., E. Heesemann, V. Sagalova, J. Manne-Goehler, R. Atun, et al. (2017). The global economic burden of diabetes in adults aged 20–79 years: A cost-of-illness study. *The Lancet Diabetes and Endocrinology* 5(6), 423–430.
- Bowen, M. E., L. Xuan, I. Lingvay, and E. A. Halm (2015). Random blood glucose: a robust risk factor for type 2 diabetes. *The Journal of Clinical Endocrinology & Metabolism 100*(4), 1503–1510.
- Breza, E. and A. G. Chandrasekhar (2019). Social networks, reputation, and commitment: evidence from a savings monitors experiment. *Econometrica* 87(1), 175–216.

- 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. Stehr, J. Sydnor, and D. Taubinsky (2022). Who chooses commitment? evidence and welfare implications. *The Review of Economic Studies* 89(3), 1205–1244.
- Chapman, G. B. and A. S. Elstein (1995). Valuing the future: Temporal discounting of health and money. Medical decision making 15, 373–386.
- Charness, G. and U. Gneezy (2009). Incentives to exercise. *Econometrica* 77(3), 909–931.
- Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val (2018). Generic machine learning inference on heterogenous treatment effects in randomized experiments. *NBER Working Paper*.
- Cobb-Clark, D. A., S. C. Dahmann, D. A. Kamhöfer, and H. Schildberg-Hörisch (2022). The predictive power of self-control for life outcomes. *Journal of Economic Behavior Organization 197*, 725–744.
- Courtemanche, C., G. Heutel, and P. McAlvanah (2014, 06). Impatience, Incentives and Obesity. The Economic Journal 125(582), 1–31.
- 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.
- DellaVigna, S. and D. Pope (2018). What motivates effort? Evidence and expert forecasts. *The Review* of Economic Studies 85(2), 1029–1069.
- Desai, J. R., G. Vazquez-Benitez, G. Taylor, S. Johnson, J. Anderson, et al. (2020). The effects of financial incentives on diabetes prevention program attendance and weight loss among low-income patients: the we can prevent diabetes cluster-randomized controlled trial. *BMC public health* 20(1), 1–11.
- Dizon-Ross, R. and A. Zucker (2023). Mechanism design for personalized policy: A field experiment incentivizing exercise. Technical report, Working paper, University of Chicago.
- Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science* 69(4), 1935–1950.
- Finkelstein, E. A., B. A. Haaland, M. Bilger, A. Sahasranaman, R. A. Sloan, et al. (2016). Effectiveness of activity trackers with and without incentives to increase physical activity (trippa): A randomised controlled trial. *The Lancet Diabetes and Endocrinology* 4(12), 983–995.
- Finkelstein, E. A., K.-W. Tham, B. A. Haaland, and A. Sahasranaman (2017). Applying economic incentives to increase effectiveness of an outpatient weight loss program (trio)-a randomized controlled trial. Social Science & Medicine 185, 63–70.
- Fryer, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. The Quarterly Journal of Economics 126(4), 1755–1798.
- Garon, J.-D., A. Masse, and P.-C. Michaud (2015). Health club attendance, expectations and self-control. Journal of Economic Behavior Organization 119, 364–374.
- Gertler, P., S. Higgins, A. Scott, and E. Seira (2019). Increasing financial inclusion and attracting deposits through prize-linked savings. *Unpublished manuscript*.
- Grant, D. and W. B. Green (2013). Grades as incentives. *Empirical Economics* 44(3), 1563–1592.
- Halevy, Y. (2008). Strotz meets allais: Diminishing impatience and the certainty effect. American Economic Review 98, 1145–1162.
- 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.
- Huck, S., A. J. Seltzer, and B. Wallace (2011). Deferred compensation in multiperiod labor contracts: An experimental test of lazear's model. *American Economic Review* 101(2), 819–843.
- Hunter, R. F., J. Tang, G. Hutchinson, S. Chilton, D. Holmes, and F. Kee (2018). Association be-

tween time preference, present-bias and physical activity: implications for designing behavior change interventions. *BMC Public Health* 18, 1–12.

- International Diabetes Federation (2019). *Idf Diabetes Atlas* (9 ed.). Brussels, Belgium: International Diabetes Federation.
- Jain, S. (2012). Self-control and incentives: An analysis of multiperiod quota plans. Marketing Science 31(5), 855–869.
- John, A. (2020). When commitment fails: evidence from a field experiment. *Management Science* 66(2), 503–529.
- John, A. and K. Orkin (2022). Can simple psychological interventions increase preventive health investment? Journal of the European Economic Association 20(3), 1001–1047.
- 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.
- Lambert, R. A. (1983). Long-term contracts and moral hazard. The Bell Journal of Economics and Management Science 14(2), 441–452.
- Lay, C. H. (1986). At last, my research article on procrastination. Journal of Research in Personality 20, 474–495.
- Lazear, E. P. (1979). Why is there mandatory retirement? Journal of Political Economy 87(6), 1261– 1284.
- 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.
- Mahajan, A., C. Michel, and A. Tarozzi (2020). Identification of time-inconsistent models: The case of insecticide treated nets. Technical report, National Bureau of Economic Research.
- Mbiti, I. and D. N. Weil (2013). The home economics of e-money: velocity, cash management, and discount rates of m-pesa users. *American Economic Review* 103(3), 369–374.
- Ministry of Labour and Unemployment (2016). Report on fifth annual employment unemployment survey (2015-16). Technical report, Labour Bureau, Government of India, Chandigarh.
- Nguyen, H. Q., M. L. Maciejewski, S. Gao, E. Lin, B. Williams, and J. P. LoGerfo (2008). Health care use and costs associated with use of a health club membership benefit in older adults with diabetes. *Diabetes Care* 31(8), 1562–1567.
- 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.
- 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*  $ONE \ 9(10)$ .

- Read, D. and N. L. Read (2004). Time discounting over the lifespan. Organizational behavior and human decision processes 94(1), 22–32.
- Rogerson, W. P. (1985). Repeated moral hazard. *Econometrica: Journal of the Econometric Society*, 69–76.
- 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.
- Sangarapillai, T., M. Hajizadeh, S. S. Daskalopoulou, and K. Dasgupta (2021, 8). Cost-comparison analysis of a physician-delivered step-count prescription strategy. CJC Open 3, 1043–1050.
- Schilbach, F. (2019). Alcohol and self-control: A field experiment in india. American economic review 109(4), 1290–1322.
- Tanaka, T., C. F. Camerer, and Q. Nguyen (2010). Risk and time preferences: Experimental and household survey data from Vietnam. American Economic Review 100(1), 557–571.
- Tharkar, S., A. Devarajan, S. Kumpatla, and V. Viswanathan (2010, sep). The socioeconomics of diabetes from a developing country: A population based cost of illness study. *Diabetes Research and Clinical Practice* 89(3), 334–340.
- Tibshirani, J., S. Athey, R. Friedberg, V. Hadad, D. Hirshberg, et al. (2023). grf: Generalized Random Forests. R package version 2.3.0.
- Tuckman, B. W. (1991). The development and concurrent validity of the procrastination scale. Educational and Psychological Measurement 51, 473–480.
- Wainwright, K., P. Romanowich, and M. A. Crabtree (2022). Associations between impulsivity and self-care adherence in individuals diagnosed with type 2 or prediabetes. *PloS one* 17(3), e0263961.
- 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.
- Welz, M., A. Alfons, M. Demirer, and V. Chernozhukov (2022). GenericML: Generic Machine Learning Inference. R package version 0.2.2.
- World Health Organization (2009). Global health risks, mortality and burden of disease attributable to selected major risks.

## 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 - E. The Online Appendix contains Appendices F - K and 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 in 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 Figure A.2: Causal Forest Selects Impatience Index as Important Predictor of Threshold Effect

Notes: This figure displays the importance of each predictor included in the causal forest prediction of the Threshold treatment effect on average compliance at the individual level. Variable importance is a weighted sum of the number of splits on the variable of the causal forest at each depth. We limit the sample to those we collected the impatience index from. Predictors include the actual impatience index and controls shown in Panel A of Figure 3, except that this analysis uses continuous versions of the baseline compliance and education variables (because the importance analysis more naturally handles continuous variables). Missing values of predictor variables are imputed with the treatment-group mean; we also include an indicator for whether each variable is missing (each of which the analysis assigned importance values of 0). We implement the Causal Forest using the GRF package in R (Tibshirani et al., 2023). Results for the other impatience measures are shown in Online Appendix Figure F.2.



(a) Average Impatience Increases with Predicted (b) Impatience Significantly Higher in Highest vs. Threshold Treatment Effect Lowest Quartiles of Predicted Threshold Effect

Appendix Figure A.3: Classification Analysis Shows Impatience Varies With Predicted Threshold Effect

Notes: The figure displays heterogeneity in the actual impatience index across quartiles of the predicted treatment effects of Threshold, following the method in Chernozhukov et al. (2018) Panel (a) shows the value of the actual impatience index in each quartile of predicted Threshold treatment effects with 95% confidence intervals, showing that the average level of the impatience index is increasing in the predicted Threshold effect. Panel (b) shows the difference in the actual treatment effect between the most-affected quartile and the other three quartiles with 95% confidence intervals, showing that the difference in the impatience index between the top and bottom quartiles of predicted Threshold effect is statistically significant. We conduct the classification analysis using the GenericML package in R (Welz et al., 2022), which selects the method of best fit among lasso, random forest, and support vector machine; support vector machine is selected in our case. Predictors include the controls shown in Figure 3 and indicators for whether each variable is missing. Qualitatively similar results for the other impatience measures are shown in Online Appendix Figure F.3.



Appendix Figure A.4: Incentive Effects are Steady through the 12-Week Program

Notes: Panel (a) shows the average probability of exceeding the step target and Panel (b) shows the average daily steps walked, both during the intervention period. Week 0 is the phase-in period (before randomization). The shaded areas represent a collection of confidence intervals from tests of equality within each weekly period between the incentive and monitoring groups from regressions with the same controls as in Table 2. Data are at the individual-week level. Both graphs are unconditional on wearing the pedometer. Graphs look similar when condition on wearing the pedometer except that, in both groups, there is less downward trend over time.



Appendix Figure A.5: Incentive Effects Persist After the 12-Week Program

Notes: Panel (a) shows the average probability of exceeding the step target and Panel (b) shows the average daily steps walked, both in the 12 weeks following the intervention. "No incentives" represents the pooled monitoring and control groups; the Panels look very similar when we compare with the control group only. The shaded areas represent a collection of confidence intervals from tests of equality within each weekly period between the incentive and no incentive groups from regressions with the same controls as in Table 2. All graphs are unconditional on wearing the pedometer. Data are at the individual-week level. Graphs look similar when condition on wearing the pedometer except that, in both groups, there is less downward trend over time.



Appendix Figure A.6: Blood Sugar Treatment Effects Grow Over Time

Notes: Figures show how the impact of incentives on random blood sugar (RBS) evolves over time by presenting the treatment effect of incentives on RBS separately for each time RBS was measured. Panel A shows the full sample and Panel B restricts to those with above-median baseline values of the blood sugar index. Survey week 0 was the baseline survey measurement; survey week 12 was the endline survey measurement; and survey weeks 3, 6, and 9 were the measurements at the pedometer sync visits held every three weeks during the intervention period. Observations are at the individual level. The "No incentives" group represents the pooled monitoring and control groups. As in our other graphs of trends over time, we pool the two comparison groups (control and monitoring) for power. Results are similar but slightly less precise if we compare incentives with control alone. For each survey, we regress random blood sugar on the incentives dummy and control for the same controls as in the random blood sugar specification in Table 4. The shaded areas represent a collection of 95% confidence intervals from those regressions. The *p*-values for the significance of the increase over time are .05 and .02 for the Panels A and B, respectively.

Appendix Table A.1: Measures of Effort	Impatience	Correlate with	Baseline	Exercise,	Health,	and I	Behavior
--	------------	----------------	----------	-----------	---------	-------	----------

			Ce	orrelation wi	th		
		Baseline	exercise	Ba	aseline indio	ces	
	Mean	Daily steps	Daily exercise (min)	Negative health risk index	Negative vices index	Healthy diet index	# Indi- viduals
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Impatience index measures							
Impatience index	0.092	-0.080***	-0.070***	-0.016	-0.052	-0.181***	1,740
1. I'm always saying: I'll do it tomorrow	2.217	-0.059	-0.101***	-0.010	-0.031	$-0.147^{***}$	1,740
2. I usually accomplish all the things I plan to do in a day	0.643	-0.054	-0.052	-0.012	$-0.043^{*}$	$-0.149^{***}$	1,740
3. I postpone starting on things I dislike to do	3.967	$-0.042^{*}$	0.004	0.004	-0.052	0.050	1,740
4. I'm on time for appointments	0.468	-0.054	0.006	-0.021	0.008	-0.097***	1,740
5. I often start things at the last minute and find it difficult to complete them on time	2.506	-0.039	-0.069***	-0.009	-0.043*	-0.207***	1,740
B. Predicted index measures							
Predicted index	-0.052	0.000	-0.036	$-0.064^{***}$	0.021	0.004	$3,\!192$
1. In the past week, how many times have you found yourself exercising less than you had planned?	0.526	0.015	-0.006	-0.064***	0.007	0.026	3,192
2. In the past 24 hours, how many times have you found yourself eating foods you had planned to avoid?	0.208	-0.001	0.050***	-0.058***	0.015	$0.034^{*}$	3,192
3. Do you worry that if you kept a higher balance on your phone, you would spend more on talk time?	0.131	-0.027	-0.062***	-0.018	0.031*	-0.038	3,192
C. Other impatience measures							
Chose commitment	0.485	0.045	-0.005	-0.027	0.011	0.015	$2,\!871$
Simple CTB	0.532	-0.120***	-0.028	-0.003	-0.018	-0.020	$3,\!190$

Notes: This table displays the correlations between impatience measures and baseline behavior and health. Each coefficient represents results from a separate regression. We normalize variables such that a higher impatience measure value corresponds to greater impatience, and a higher health or behavior measure value corresponds to healthier behavior. Panel A shows the impatience index and its five components. Panel B shows the predicted index and its three index components. Daily steps are from the phase-in period pedometer data. Daily exercise is self-reported. The health index is as in Table 4. 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, vegetable, and rice; spoonfuls of sugar; fruit, junk food, and sweets intake; and whether one avoids unhealthy foods. Data are at the individual level and include the full sample. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.

Appendix	Table A.2	: Missing	Pedometer	Data	During	the	Intervention	Period
					0			

Dep. variable:	No Steps data	Reason no steps data		<b>Reason</b> 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	-0.0140 [0.0174]	-0.0287** [0.0142]	0.0155 [0.0124]	-0.00203 [0.00511]	0.00571 [0.00731]	$0.0166^{**}$ [0.00694]	-0.00471 [0.00594]
Monitoring mean # Individuals # Observations	$.19 \\ 2,607 \\ 218,988$	.15 2,559 205,732	.047 2,607 218,988	.0049 2,607 218,988	.0099 2,607 218,988	.012 2,607 218,988	.02 2,607 218,988

Notes: Each observation is an individual  $\times$  day. The sample includes Incentives and Monitoring. Missing data have two sources: pedometer non-wearing (i.e., steps = 0) (column 2) or failure to retrieve pedometer data (column 3). Columns 2 + 3 = column 1 except column 2 conditions on there not being missing data (for consistency with our main step analyses, results are similar without this restriction), while columns 1 and 3 do not. Columns 4-7 summarize the reasons pedometer data in column 3 were missing. Controls are the same as in Table 2. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.

			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	0.057	0.056
4-Day Threshold	0.055	0.055	0.056	0.056	0.055
5-Day Threshold	0.059	0.059	0.059	0.059	0.058

Appendix Table A.3: Threshold Treatments Increase Cost-Effectiveness Relative to Base Case, With Similar Increases Among Those Who Are More and Less Impatient

Notes: The table displays the cost-effectiveness of different treatment groups (in rows) and different samples (in columns). Costeffectiveness equals average compliance divided by the average payment per day, in units of days complied per INR. The sample includes Base Case and Threshold (Threshold pools the 4- and 5-day Threshold). We test for differences in cost-effectiveness using a mathematically equivalent test for differences in the fraction of days complied on which participants earned payment, shown in column 4 of Table 2 and Figure 2b.

Dependent variable:		Exceede	ed step target	(×100)	
_	(1)	(2)	(3)	(4)	(5)
Impatience $\times$ Threshold	2.99 [-3.49, 9.48]	3.94 [-8.20,16.08]	$2.76^*$ [-0.35, 5.73]	$6.8^{**}$ [1.21, 13.24]	$8.07^{**}$ [1.35, 14.80]
Threshold	-5.29** [-10.06, -0.52]	$-6.79^{*}$ [-13.86, 0.28]	-3.97** [-7.12, -1.06]	$-6.5^{***}$ [-11.24, -3.16]	$-8.50^{***}$ [-13.71, -3.29]
Impatience	$-3.11^{*}$ [-6.68, 0.46]	-6.70 [-15.59, 2.20]	-1.56 [-3.74, 0.67]	-5.15** [-9.39, -0.30]	-4.06 [-9.23, 1.12]
Impatience measure:	Impatience index	Above-median impatience index	Predicted impatience index	Above-median predicted index	Simple CTB
Sample:	Late	Late	Full	Full	Full
Base Case mean	51.7	51.7	50.6	50.6	50.6
# Individuals	496	496	977	977	977
# Observations	39,562	39,562	78,096	78,096	78,096

Appendix Table A.4: Threshold Heterogeneity Results are Similar Among Naive Individuals

Notes: This table is the same as Table 3 but limited to the subsample of participants who did not demand commitment (that is they did not prefer both the 4-day and 5-day threshold contract relative to the base case contract). Controls are the same as in Table 2. Bootstrapped 95% confidence are in brackets; see the notes to Table 3 for a detailed description of the bootstrap procedure. Data are at the individual  $\times$  day level. The sample includes Base Case and Threshold. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.

Dependent variable:		Earned	l payment wh	en exceeded	target	
	(1)	(2)	(3)	(4)	(5)	(6)
Impatience $\times$ Threshold	-0.00625 [-0.03, 0.01]	-0.0109 [-0.04,0.02]	$\begin{array}{c} 0.00386\\ [-0.01,\ 0.01]\end{array}$	$\begin{array}{c} 0.0115\\ [-0.01, \ 0.02]\end{array}$	0.0202* [-0.00, 0.04]	-6.96e-05 [-0.01, 0.01]
Threshold	-0.114*** [-0.13, -0.10]	-0.109*** [-0.12,-0.10]	-0.115*** [-0.12, -0.11]	-0.119*** [-0.13, -0.11]	-0.127*** [-0.15, -0.11]	-0.115*** [-0.13, -0.10]
Impatience	$\begin{array}{c} 0.00208\\ [-0.00,\ 0.01]\end{array}$	0.00544 [-0.00, 0.01]	-0.000834 [-0.00, 0.00]	-0.00275 [-0.00, 0.00]	-0.00233 [-0.01, 0.00]	$0.00483^{*}$ [-0.00, 0.01]
Impatience measure:	Impatience index	Above- median impatience index	Predicted impatience index	Above- median predicted index	Chose commitment	Simple CTB
Sample:	Late	Late	Full	Full	Full	Full
Base Case mean	1	1	1	1	1	1
# Individuals	1,007	1,007	$1,\!846$	$1,\!846$	$1,\!681$	1,844
# Observations	$42,\!830$	$42,\!830$	$79,\!248$	$79,\!248$	$71,\!525$	$79,\!150$

### Appendix Table A.5: Thresholds Are Similarly Cost-Effective Among Those with Higher Impatience

Notes: This table shows heterogeneity in the impact of Threshold on the fraction of days on which participants received payment, conditional on meeting the step target, by different measures of impatience. A higher level of this outcome indicates lower cost-effectiveness among treatment groups that received the same payment per day (all groups except Small Payment). The impatience measure changes across columns. Controls are the same as in Table 2. Bootstrapped 95% confidence are in brackets; see the notes to Table 3 for a detailed description of the bootstrap procedure. Data are at the individual  $\times$  day level. The sample includes Base Case and Threshold. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.

		Post-intervention	
Dependent variable:	Exceeded step target	Daily steps	Daily steps (if $> 0$ )
	(1)	(2)	(3)
Incentives	$0.071^{***}$ $[0.01]$	$537.2^{**}$ [220.90]	$648.3^{***}$ [195.82]
No incentives mean # Individuals # Observations	$0.156 \\ 1,122 \\ 91,756$	4,674 1,122 91,756	6,773 1,122 62,858

Appendix Table A.6: The Effects of Incentives Persist After the Intervention Ends

Note: This table shows the average treatment effect of Incentives relative to Control and Monitoring (pooled) during the "post-intervention period" (i.e., the 12 weeks after the intervention ended). Each observation is a person-day. Columns 1 and 2 include all days, and column 3 only includes days where the participant wore the pedometer (i.e., had step count > 0). Controls are the same as in Table 2. The number of individuals differs from the total number recruited for the post-intervention period because roughly 11% of participants withdrew immediately. The likelihood of immediate withdrawal is not significantly different between the incentive and comparison groups. Standard errors, in brackets, are clustered at the individual level. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.

### **B** Theoretical Predictions Appendix

### B.1 Building Intuition: Time-Bundling for Sophisticates and Naifs

The effectiveness of time-bundled contracts decreases in  $\delta$  for both sophisticates and naifs. However, the exact conditions for compliance on day 1 differ. To build intuition, return to our 2-day example from Section 2.3 and note that, on day 1 of the threshold contract, the individual's motivation to comply is to have the opportunity to be paid 2m' for complying on day 2. Letting  $w_{2,1}$  be her prediction on day 1 about her compliance on day 2, the value of this opportunity to her is

$$(2m' - \delta e_2)w_{2,1}\big|^{w_1=1},\tag{11}$$

which is equal to the discounted (by d = 1) payment 2m' net of the discounted effort costs  $\delta e_2$  if the individual thinks she will comply on day 2 given compliance on day 1 (i.e., if  $w_{2,1}|^{w_1=1} = 1$ ). (This is simply a re-expression of equation (5).)

For sophisticates, who correctly predict their future preferences,

$$w_{2,1}|^{w_1=1} = \mathbb{1}\{e_2 < 2m'\}.$$
 (12)

Thus, for a sophisticate to place a positive value on a day 2 work opportunity (i.e., for expression (11) to be positive), it must be that  $e_2 < 2m'$ : the payment for day 2 work must be sufficiently large to entail a soft *commitment* for the day 2 self to follow through. The sophisticate complies on day 1 to give her future self strong incentives to comply.

In contrast, naifs believe that their day-2 selves have the same preferences as their day-1 selves. For them,

$$w_{2,1}\Big|^{w_1=1} = \mathbb{1}\{\delta e_2 < 2m'\}.$$
(13)

Thus, naifs place a positive value on the day 2 opportunity as long as it has positive net present value (NPV) from the day 1 perspective (i.e., as long as discounted payments net of discounted effort costs,  $2m' - \delta e_2$  are positive). That is, naifs positively value any lucrative day 2 *option* that they *want* their day 2 selves to execute. Naifs comply on day 1 to give their day 2 selves the *option* to follow-through.<sup>51</sup>

With time-bundled thresholds, these differences in motivation between sophisticates and naifs should not normally affect behavior. The day 2 opportunities that are lucrative enough options to motivate naifs to comply on day 1 are also generally associated with high enough day 2 payments to provide a soft commitment for day 2 compliance. Likewise, any day 2 opportunity that provides a soft commitment that motivates a sophisticate to comply on day 1 will also provide an option that motivates a naif to comply on day 1 (i.e., equation (12) implies equation (13)).<sup>52</sup> By pairing the options that motivate naifs with the commitment that both motivates

<sup>&</sup>lt;sup>51</sup>In either case, Prediction 1 still holds because the equation (11) value is still weakly decreasing in  $\delta$ . The equation (11) value is  $(dM - \delta e_2)\mathbb{1}\{e_2 < M\}$  for sophisticates and  $(dM - \delta e_2)\mathbb{1}\{\delta e_2 < dM\}$  for naifs, both of which are decreasing in  $\delta$ . To see this in the naive case, note that  $(dM - \delta e_2)\mathbb{1}\{\delta e_2 < dM\} = \max\{dM - \delta e_2, 0\}$ .

<sup>&</sup>lt;sup>52</sup>Equations (5), (12), and (13) show that the only difference between sophisticates and naifs is that, if  $e_1 + \delta e_2 < 2m$  and  $e_2 > 2m$ , naifs would comply on day 1 and then fail to follow-through on day 2, while sophisticates would not comply on day 1 because they know they would not follow-through on day 2. However, this behavior should

sophisticates and helps naifs follow through, thresholds work for both types.<sup>53</sup>

### B.2 Time-Bundled Threshold Contracts and Impatience Over Effort

In this section, we present a series of propositions that provide the theoretical underpinning for Prediction 1. The propositions demonstrate that, holding all else equal, both compliance and effectiveness in threshold contracts tend to decrease in  $\delta^{(t)}$ . In contrast, in time-separable contracts, compliance and effectiveness are flat in  $\delta^{(t)}$  (Observation 1). Thus, the lower  $\delta^{(t)}$  is, the higher compliance and effectiveness are in a threshold relative to time-separable contract.

Specifically, Proposition 1 examines threshold contracts with K = T (i.e., where one must comply on all days to receive payment). It shows that, for all T, regardless of the effort cost distribution, compliance is weakly decreasing in  $\delta$ . To gain tractability to examine threshold effectiveness and threshold contracts with K < T, we then make assumptions about the effort cost distribution. Proposition 2 examines effectiveness when K = T = 2 and shows that, under relatively general conditions, effectiveness in the threshold contract is weakly decreasing in  $\delta$ .

Proposition 3 shows that, if costs are perfectly positively correlated over time, both compliance and effectiveness under the threshold are decreasing in  $\delta^{(t)}$  for any  $K \leq T$  and any T. Finally, Proposition 4 examines a simplified model where costs are binary and known from day 1, K = 2 and T = 3. We show that compliance and effectiveness are higher when  $\delta^{(t)}$  is lower.

The propositions together suggest that Prediction 1 holds in many empirically-relevant conditions, and especially when either (a) K is high relative to T,<sup>54</sup> or (b) costs are positively correlated across periods. Both (a) and (b) hold in our empirical setting: our experiment uses relatively high levels of K relative to T, and costs are positively correlated across days.

We begin with a more formal presentation of the agent's problem from Section 2.1, which is helpful for proving our propositions. We next present the propositions underlying Prediction 1.

### B.2.1 Agent Problem

Given the notation and assumptions in Section 2.1, we can express the agent's problem as follows. Let  $w_{t,j}$  be the agent's prediction on day j about her compliance on day t > j. On day t, the agent chooses compliance,  $w_t$ , to maximize expected discounted payments net of effort costs:

<sup>54</sup>Thresholds where K/T is very low may not always be better for impatient naifs than patient people because they include more days where current and future effort are substitutes, which can cause naifs to procrastinate.

be rare, as it requires that  $e_2 > e_1/(1-\delta)$ , which implies that  $e_2 >> e_1$  and/or that  $\delta$  is very low. The intuition for why, conditional on day 1 compliance, it is rare to not follow through on day 2 is that people sink costs as they approach the threshold. Thus, the marginal incentive to comply is strictly higher on day 2 than on day 1.

<sup>&</sup>lt;sup>53</sup>In contrast, in some types of time-bundled contracts (other than thresholds), the differences in the decisions of naifs and sophisticates may be greater. For example, we have investigated the full class of 2-day time-bundled contracts. This class also includes contracts where the day 2 wage is not 0 in the absence of day 1 compliance (e.g., a contract paying \$5 for day 2 effort if the agent did not comply on day 1 and \$10 if she did). In such contracts, *options* and *commitment* can be less tightly linked. For example, some time-bundled contracts that are not thresholds function like commitment contracts and are more effective for sophisticates, as day 1 compliance generates a soft commitment for day 2 compliance but not a positive NPV option. In others, day 1 compliance generates an option but not a commitment, and they are more effective for naifs. We have explored all of the non-threshold time-bundled contracts in the 2-day contract space and, importantly, the result that performance is weakly decreasing in δ holds for all of them, not just for time-bundled thresholds.

$$\max_{w_t \in \{0,1\}} \mathbb{E}\left[ d^{(T-t)} m_T - \sum_{j=t+1}^T \delta^{(j-t)} w_{j,t} e_j \middle| e_1, .., e_t, w_1, .., w_t \right] - w_t e_t,$$
(14)

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:

$$V_t(0) < V_t(1) - e_t \tag{15}$$

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.2.2** Propositions underlying Prediction 1

**Proposition 1** (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 is weakly decreasing in  $\delta^{(t)}$  for all  $t \leq T - 1$ .

*Proof.* We provide the proof here for T = 2. The proof for T > 2 is in Online Appendix I.1.

Recall that the condition for complying on day 1 is to comply if  $e_1 < V_1(1) - V_1(0)$  (equation (15)). Let  $w_{t,j}$  be the agent's prediction on day j about her compliance on day t > j. With the threshold contract, we have that:

$$V_1(1) - V_1(0) = \mathbb{E}\left[(dM - \delta e_2)w_{2,1}|e_1, w_1 = 1\right] - \mathbb{E}\left[-\delta e_2 w_{2,1}|e_1, w_1 = 0\right]$$
(16)

We examine this expression separately for sophisticates and naifs.

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:

$$V_{1}(1) - V_{1}(0) = \mathbb{E}\left[(dM - \delta e_{2})w_{2,1}|e_{1}, w_{1} = 1\right] - \mathbb{E}\left[-\delta e_{2}w_{2,1}|e_{1}, w_{1} = 0\right]$$
$$= \mathbb{E}\left[(dM - \delta e_{2})\mathbb{1}\left\{e_{2} < M\right\} + \delta e_{2}\mathbb{1}\left\{e_{2} < 0\right\}|e_{1}\right]$$
(17)

We show that this is weakly decreasing in  $\delta$  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  $\delta$  for all values of  $e_2$ . There are two cases:

- 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  $\delta$ .
- 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  $\delta$ .

Since equation (17) is weakly decreasing in  $\delta$ , day 1 compliance is decreasing in  $\delta$ . The same is true for day 2 compliance, since  $w_2 = 1$  if both  $w_1 = 1$  and  $e_2 < M$ , and  $w_1$  is weakly decreasing in  $\delta$ . Thus, compliance in the threshold contract is decreasing in  $\delta$  for sophisticates.

We now turn to naifs. For naifs, 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:

$$V_{1}(1) - V_{1}(0) = \mathbb{E}\left[(dM - \delta e_{2})w_{2,1}|e_{1}, w_{1} = 1\right] - \mathbb{E}\left[-\delta e_{2}w_{2,1}|e_{1}, w_{1} = 0\right]$$
  
$$= \mathbb{E}\left[(dM - \delta e_{2})\mathbb{1}\{\delta e_{2} < dM\} + \delta e_{2}\mathbb{1}\{\delta e_{2} < 0\}|e_{1}\right]$$
  
$$= \mathbb{E}\left[\max\{dM - \delta e_{2}, 0\} + \delta e_{2}\mathbb{1}\{e_{2} < 0\}|e_{1}\right]$$
(18)

Again, we show that this is decreasing in  $\delta$  by showing that the argument,  $\max\{dM - \delta e_2, 0\} + \delta e_2 \mathbb{1}\{e_2 < 0\}$ , is weakly decreasing in  $\delta$  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  $\delta$ .
- 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  $\delta$ .

Since equation (18) is weakly decreasing in  $\delta$ , day 1 compliance (and hence day 2 and total compliance) are also decreasing in  $\delta$  for naifs.

We now examine effectiveness when T = K. We examine the case where T = 2 and, to gain tractability, make a reasonable assumption on the cost function, assuming that  $e_2$  is weakly increasing in  $e_1$ , in a first order stochastic dominance sense.<sup>55</sup> This assumption flexibly accommodates the range from IID to perfect positive correlation, just ruling out negative correlation. Under this assumption, we show that effectiveness is weakly decreasing in  $\delta$  as long as there is not "too much" inframarginal behavior. When there is too much inframarginal behavior, not only will the effectiveness prediction not hold but incentives cease to be a cost-effective approach.

**Proposition 2** (T = 2, K = 2, Threshold Effectiveness and Impatience Over Effort). Let T = 2. Let  $e_2$  be weakly increasing in  $e_1$ , in a first order stochastic dominance sense. Fix all parameters other than  $\delta^{(t)}$ . Take any threshold contract with threshold level K = 2; denote the threshold payment M. As long as there is not "too much" inframarginal behavior,<sup>56</sup> the effectiveness of the threshold contract is weakly decreasing in  $\delta$ .

*Proof.* We first show that, if costs are positive, cost-effectiveness in the threshold is not increasing in  $\delta$ . Because Proposition 1 showed that compliance is decreasing in  $\delta$ , this establishes that effectiveness is decreasing in  $\delta$  when costs are positive. We then show sufficient conditions for threshold effectiveness to decrease in  $\delta$  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 (17) and (18), we know that the value of

 $<sup>{}^{55}</sup>F_{e_2|e_1}(x)$  is weakly decreasing in  $e_1$  for all x, with  $F_{e_t|e_{t'}}(x)$  the conditional CDF of  $e_t$  given  $e_{t'}$ .

<sup>&</sup>lt;sup>56</sup>See equation (22) for the exact condition. 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  $\delta$  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.

 $e^*$  will depend on the agent's sophistication and, importantly, decrease in  $\delta$ .

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

$$w_{1} = \mathbb{1}\{e_{1} < e^{*}\}$$

$$w_{2} = w_{1}\mathbb{1}\{e_{2} < M\} + (1 - w_{1})\mathbb{1}\{e_{2} < 0\}$$

$$= w_{1}\mathbb{1}\{0 < e_{2} < M\} + \mathbb{1}\{e_{2} < 0\}$$

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  $\delta$ . In that case,  $w_2 = w_1w_2$ . Therefore we have:

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

$$= \frac{1}{M} \left( \frac{1}{\mathbb{E}[w_2 | w_1 = 1]} + 1 \right)$$
(19)

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

**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  $\delta$ .

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  $\lambda C - P$ .)

$$2 (\lambda C - P) = \lambda \mathbb{E}[w_1 + w_2] - M \mathbb{E}[w_1 w_2] = \lambda (F(e^*) + \mathbb{E}[w_1 \mathbb{1}\{0 < e_2 < M\} + \mathbb{1}\{e_2 < 0\}]) - M \mathbb{E}[w_1 \mathbb{1}\{e_2 < M\}] = \lambda (F(e^*) + \mathbb{E}[\mathbb{1}\{e_1 < e^*\}\mathbb{1}\{0 < e_2 < M\} + \mathbb{1}\{e_2 < 0\}]) - M \mathbb{E}[\mathbb{1}\{e_1 < e^*\}\mathbb{1}\{e_2 < M\}] = \lambda (F(e^*) + Prob(e_1 < e^*, 0 < e_2 < M) + Prob(e_2 < 0)) - MProb(e_1 < e^*, e_2 < M).$$
(20)

We now take a derivative with respect to  $e^*$ . Let  $g(e^*) = Prob(e_1 \leq e^*, e_2 \in S)$ , where S is

some set. It is straightforward to show that  $g'(e^*) = f(e^*) \operatorname{Prob}(e_2 \in S | e_1 = e^*)$ .<sup>57</sup> 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  $\delta$ ) is:

$$\lambda(1 + Prob(0 < e_2 < M | e_1 = e^*)) \ge MProb(e_2 < M | e_1 = e^*)$$
(21)

or

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

or

$$Prob(e_2 < 0|e_1 = e^*) \le \frac{\lambda}{M} + \left(\frac{\lambda}{M} - 1\right) Prob(0 < e_2 < M|e_1 = e^*)).$$
 (22)

If  $\lambda > M$ , condition (22) will always hold. More broadly, the condition will be more likely to hold the greater  $\lambda$  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 (22) holds), the effectiveness of a threshold contract is decreasing in  $\delta$ .

We now turn to examine threshold contracts with K < T. To gain tractability, we begin with the case where costs are perfectly correlated across periods, showing that both compliance and effectiveness under the threshold are increasing in impatience for any threshold level  $K \leq T$ .

**Proposition 3** (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 threshold contract will be weakly decreasing in  $\delta^{(t)}$ .

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. Again, threshold compliance and effectiveness are higher among those who are more impatient.

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

Dividing by  $\epsilon$  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  $\epsilon$  go to 0 and using the definition of the derivative gives that  $g'(e^*) = f(e^*) Prob(e_2 \in S|e_1 = e^*)$ .

 $<sup>^{57}\</sup>mathrm{To}$  show this, note that

**Proposition 4.** 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  $e_H \ge e_L$ . 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  $\delta$ . 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  $\delta = 1$ .

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

**Proposition 5.** 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  $\delta$ . 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  $\delta$ .

Proof. See Online Appendix I.1.

### 

#### **B.3** The Effectiveness of Threshold and Linear Contracts

In this section, we compare the effectiveness of time-bundled threshold and time-separable linear contracts under a range of effort cost assumptions, paying particular attention to how the relative effectiveness of thresholds depends on  $\delta$ . 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 (Proposition 6) examines the relative performance of the contracts in the limit as  $\delta$  goes to 0 under very general assumptions. It shows that, for sufficiently low  $\delta$ , for any linear contract, there exists a threshold contract that achieves substantially higher costeffectiveness 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.

The next four propositions (Propositions 7a - 8b) examine the full range of  $\delta$ , not just the case where  $\delta$  is sufficiently low. While we make additional assumptions on the effort cost distributions for tractability, the 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. IID effort costs is a common assumption in the literature (e.g., Garon et al., 2015). In each case, we begin with a testable comparison between threshold and linear contracts that offer the same payment per day before moving to more abstract comparisons that teach us about whether the optimal threshold contract or the optimal linear contract is more effective (and how that relationship depends on  $\delta$ ).<sup>58</sup>

<sup>&</sup>lt;sup>58</sup>Predictions about optimal contracts are hard to test since most policymakers do not have sufficient information about the cost function and  $\delta$  to solve for the optimal contracts.

**Proposition 6.** Let d = 1 and T = 2. Fix all parameters other than  $\delta$ , and take a linear contract that induces compliance C > 0.

(a) If agents are naive and  $e_2$  is weakly increasing in  $e_1$ , in a first order stochastic dominance sense,<sup>59</sup> then for sufficiently small  $\delta$ , 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  $\delta$ , 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 I.2.

The potential improvements from threshold contracts demonstrated by Proposition 6 are quantitatively large. For example, when costs are IID and agents are naive with sufficiently low  $\delta$ , for a linear contract that generates C = .9, there exists a threshold contract that generates 95% as much compliance but for less than half the cost.

**Proposition 7a** (Perfect Correlation, M = 2m). Let T = 2. Fix all parameters other than  $\delta$ . Consider a linear contract with payment m and a threshold contract with payment 2m. Then, regardless of agent type, the threshold contract is more effective than the linear contract if  $\delta < 2d - 1$ . If  $\delta \geq 2d - 1$ , then the linear contract may be more effective.

*Proof.* See Online Appendix I.2.

**Proposition 7b** (Perfect Correlation, Optimal Contracts). Let T = 2. Fix all parameters other than  $\delta$ , 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. Hence, if  $\delta < 2d - 1$ , the most effective contract will always be a threshold contract.

*Proof.* See Online Appendix I.2.

**Proposition 8a** (IID Uniform, M = 2m). Let d = 1. Fix all parameters other than  $\delta$ . Let costs be independently drawn each day from a uniform[0,1] distribution. Take any threshold contract paying M < 2 and compare it with the linear contract paying  $m = \frac{M}{2}$ .

(a) If M < 1, the threshold contract is always more cost-effective, but whether it has higher compliance (and hence whether it is more effective) depends on  $\delta$ . Define  $\frac{2M^2}{1+M}$  as the cutoff value for naifs and  $2 - \frac{2}{M+M^2}$  as the cutoff value for sophisticates. If  $\delta$  is less than the cutoff value for a given type, then the threshold contract is more effective, as it generates greater compliance. (b) If 1 < M < 2,<sup>60</sup> then the threshold contract is more effective.

*Proof.* See Online Appendix I.2.

 $\square$ 

 $<sup>{}^{59}</sup>F_{e_2|e_1}(x)$  is weakly decreasing in  $e_1$  for all x, with  $F_{e_t|e_{t'}}(x)$  the conditional CDF of  $e_t$  given  $e_{t'}$ . This assumption flexibly accommodates the range from IID to perfect positive correlation, just ruling out negative correlation.

<sup>&</sup>lt;sup>60</sup>Note that the principal would never pay M > 2 since M = 2 achieves 100% compliance regardless of  $\delta$ .

**Proposition 8b** (IID Uniform, Optimal Contracts). Let d = 1. Fix all parameters other than  $\delta$ . Let costs be independently drawn each day from a uniform[0,1] distribution. Whether the most effective threshold contract is more effective than the most effective linear contract depends on  $\delta$ as well as  $\lambda$ , the principal's marginal return to compliance. For a wide and plausible range of values of  $\lambda$ ,<sup>61</sup> there exists a cutoff value of  $\delta$  such that the threshold contract is more effective when  $\delta$  is below the cutoff, and the linear contract is more effective when  $\delta$  is above the cutoff. For the remaining values of  $\lambda$ , either the threshold contract is always more effective, or the linear contract is always more effective, but in either case the effectiveness of the threshold relative to linear is decreasing in  $\delta$ .

*Proof.* See Online Appendix I.2.

### **B.4** Proofs of Predictions Regarding Frequency

In this subsection, we first prove Prediction 3 from Section 2.4. Next, we present and prove a second prediction that follows Kaur et al. (2015) in showing an additional way to use empirical data to make inferences about the discount factor over payments, which we use in Section 5.4.

**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.<sup>62</sup>

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 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.

<sup>&</sup>lt;sup>61</sup>See proof in Online Appendix I.2 for specific ranges for both naifs and sophisticates.

<sup>&</sup>lt;sup>62</sup>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.

### C Measuring The Effort and Payment Discount Factors

This section provides additional detail on measurements of impatience in our sample. We first describe how we validate the impatience index — our primary measure of effort discounting — using an incentivized effort task. We then present multiple estimates of the discount factors over effort and payment from our experimental context, showing substantial discounting of effort but not of payment. Finally, we show that there is limited correlation between the discount factors over effort and payment.

### C.1 Validating the Impatience Index

We begin by describing the incentivized effort task data used for the validation exercise, along with other data collected. Next, we describe two effort discount rate measures obtained from these data. Third, we use these measures to validate our impatience index.

#### C.1.1 Data Collection in the Validation Sample

We validate our impatience index using a separate sample of 71 people who are very similar to our experimental sample (hereafter: the "validation sample").<sup>63</sup> The validation sample was randomly selected from a later evaluation of a similar incentive program for exercise (Dizon-Ross and Zucker, 2023) with nearly identical recruitment criteria,<sup>64</sup> and observable characteristics are balanced across the validation sample and experimental sample: walking levels, demographic characteristics, BMI, etc., are statistically indistinguishable (Online Appendix Table F.18).

In the validation sample, we collected the same *impatience index* described in this study and incentivized two tasks to measure impatience over effort and recharges, respectively.

Effort Task Respondents were incentivized to perform an effort task, which we call the "Effort Choice by Date" task, following the methodology of Augenblick (2018) and Augenblick and Rabin (2019), which John and Orkin (2022) previously adapted to a field setting. The task was to call into a toll-free automated phone line, listen to a useless 30-second recording, and answer a simple question to confirm that they listened. On the survey date (day 0), individuals chose how many calls to complete at time t for a piece rate w, where t is 0 (i.e., the same day), 1, 7, or 8 days from the time of the decision, and the piece rate is INR 10, 6, 2, or 0.65 One choice was then randomly selected for implementation, and respondents received both the piece rate for the implemented choice as well as an additional 100 INR if they completed all the tasks they chose (in addition to one "mandatory task"). We refer to the measures we construct from these data as effort impatience measures.

Patterns in the data indicate that respondents understood the exercise. For example, the average number of tasks chosen increases with the piece rate, with respondents choosing an average of 5.6, 7.1, 7.6, and 8.0 tasks when the piece rates were 0, 2, 6, and 10 INR, respectively. Our field team also reported limited respondent misunderstanding.

<sup>&</sup>lt;sup>63</sup>The sample size is comparable to the number of people who completed choices in the two seminal papers measuring impatience with effort tasks: 99 in Augenblick (2018) and 100 in Augenblick and Rabin (2019).

<sup>&</sup>lt;sup>64</sup>Both studies targeted participants from Coimbatore, Tamil Nadu, using public screening camps as the primary recruitment tool, and both focused on individuals aged 30-65 who were literate, comfortable using mobile phones, capable of receiving mobile recharge payments, and had or were at high risk of lifestyle disease. However, the later study enrolled participants with high blood pressure in addition to high blood sugar.

<sup>&</sup>lt;sup>65</sup>We include a 0 INR piece rate following guidance from John and Orkin (2022) that it helped their model converge. However, our structural model does not converge with the 0 INR piece rate choice, so we exclude it when estimating the structural parameters.

**Recharge Choices** A secondary goal for the validation sample was to assess the relationship of the impatience index with recharge impatience. To do so, we measure impatience over recharges with a multiple price list (MPL) (Andreoni and Sprenger, 2012a; John and Orkin, 2022). Participants made 10 choices between receiving a recharge today and a later date (either 7 and 14 days from today). For simplicity, the recharge today was always 50 INR, and the later recharges were larger whole numbers: 60, 70, 90, 100, and 150 INR. One choice from the MPL was also randomly selected for implementation.

The MPL choices are not ideal for estimating a structural recharge discount factor: the later payment amounts are all meaningfully larger than the earlier payment (we cannot distinguish between one-week discount factors in the range from  $\frac{50}{60} = 0.83$  to 1), and, as with all MPLs, any mistrust in receiving the payment will push participants toward earlier payment and bias implied discount factors downwards (Halevy, 2008; Andreoni and Sprenger, 2012b). Instead, we construct a reduced-form *recharge impatience measure* as the proportion of choices where the individual chose the smaller recharge on the sooner date.

### C.1.2 Structural and Reduced-Form Effort Impatience Measures

The data from the effort task are consistent with positive discounting of future effort with some present bias. Consistent with positive discounting, the number of tasks chosen on days with t > 0 are all significantly greater than on t = 0. (Specifically, participants chose 7.4, 7.0 and 7.5 tasks on days 1, 7, and 8, respectively, and only 6.4 tasks on day 0.) Consistent with present bias, the biggest jump in task allocations appears between "today" and "tomorrow".

We thus parameterize a constant discount factor for all future days:  $\delta^{(t)} = \delta_{QH}$  if t > 0. This is equivalent to a  $\beta - \delta$  model in which  $\delta = 1$ . We use the effort task data to construct two measures, one structural and one reduced-form, for this parameter.

**Structural Measure and Evidence** Our structural estimation follows John and Orkin (2022).<sup>66</sup> (The estimating equation is in the notes to Table C.1.) We structurally estimate  $\delta_{QH}$  at the group level. As in John and Orkin (2022), individual-level structural estimates converge for less than half of our sample.

Column 1 of Table C.1 shows that, in the full validation sample, we estimate a  $\delta_{QH}$  of 0.565, which is significantly different from 1 and suggests a high degree of effort impatience. In column 2, we follow Augenblick and Rabin (2019) and remove "problematic" individuals with limited effort choice variation or effort choices that are not primarily monotonic in wage offers.<sup>67</sup> The discount factor estimate is similar and still significantly different from 1.

**Reduced-Form Measure and Evidence** Our reduced-form measure is based on the excess number of tasks chosen on future dates relative to day 0 at a given piece rate, following Augenblick (2018) and Augenblick and Rabin (2019). Specifically, for all task allocations made on future days (t > 0) at piece rate w, we construct a measure at the individual  $\times$  choice level equal to the tasks allocated on day t minus the tasks allocated on day 0 at the same piece rate w. People who are more impatient (lower  $\delta_{QH}$ ) will choose more tasks on future days than today, and thus have higher average values of this measure.

 $<sup>^{66}</sup>$ John and Orkin (2022) assumes quasilinear utility and a power effort cost function following Augenblick (2018), and includes a non-monetary per-task reward s in addition to the piece rate following DellaVigna and Pope (2018).

 $<sup>^{67}</sup>$ We remove 28 of 71 respondents in a field setting; Augenblick and Rabin (2019) remove 28 of 100 in a lab setting for the same reasons. Our removal rates are not significantly different for those with below- vs. above-median impatience index.

	Full valid	ation sample	Below impatie	v-median nce sample	Above impatien	-median ace sample
-	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{\delta_{QH}}$	$0.565 \\ [0.133]$	0.556 [0.153]	0.997 [0.008]	0.998 [0.007]	0.178 [0.157]	$0.367 \\ [0.208]$
P-value: $\delta_{QH} = 1$	0.001	0.004	0.737	0.803	< 0.001	0.002
P-value: $\delta_{QH} = \delta_{QH}^{\text{Below}}$	< 0.001	0.004			< 0.001	0.002
P-value: $\delta_{QH} = \delta_{QH}^{Above}$	< 0.001	0.061	< 0.001	0.002		
Sample	All	Changers + Monotone	All	Changers + Monotone	All	Changers + Monotone
# Individuals	71	43	32	24	39	19
# Observations	852	516	384	228	468	228

Appendix Table C.1: Structural Estimates of the Effort Discount Factor

Notes: This table displays structural estimates of the effort discount factor,  $\delta_{QH}$ , in the validation sample, estimated using data from the Effort Choice by Date task of Augenblick (2018) using an estimation approach similar to John and Orkin (2022). The optimal allocation of effort is given by:  $e^* = argmax(s + d^{(11)} \cdot \phi \cdot w) \cdot e - \delta^{(t)}(\frac{1}{\gamma}e^{\gamma})$ , where t is the time of effort provision,  $\gamma$  captures the convex cost of effort, s is a parameter that captures the non-monetary reward for each task, w is the monetary piece rate,  $d^{(11)}$  captures the monetary discounting of the payment in 11 days, and  $\phi$  is a slope parameter. We parametrize  $\delta^{(t)} = \delta_{QH}$  (equivalent to a quasihyperbolic model with  $\delta = 1$ ) and  $d^{(11)} = 1$  and estimate s,  $\phi, \delta_{QH}$ , and  $\gamma$ . We present results using the full validation sample and the subsamples with below- and above-median impatience index, with or without inclusion restrictions from choice patterns. Columns 1, 3, and 5 have no inclusion restriction; columns 2, 4, and 6 restrict to individuals who changed their effort choice at least once and had at most 1 choice non-monotonicity in payment levels.

Overall, participants chose to complete 13% fewer tasks in the present than the future, suggesting meaningful effort discounting. The result is similar if we again remove problematic individuals: the restricted sample allocates 15% fewer tasks in the present than the future. Our results mimic Augenblick (2018) and Augenblick and Rabin (2019) which find that participants choose to complete 16% and 10-12% fewer tasks in the present than the future, respectively.

### C.1.3 The Impatience Index Correlates with Effort Impatience Measures

In this section, we show that our impatience index correlates with the incentivized effort impatience measures in the validation sample. In contrast, it does not correlate with recharge impatience. Overall, this provides evidence that the impatience index proxies for impatience in the effort, but not payment, domain.

Correlation with Effort Impatience Measures Columns 3-6 of Table C.1 show that structural estimates of  $\delta_{QH}$  are substantially higher among those with lower impatience index. Specifically, among individuals with below-median impatience index, our estimate of  $\delta_{QH}$  is 0.997 and statistically indistinguishable from 1 (column 3). In contrast, we estimate that  $\delta_{QH}$  is 0.178 for those with above-median impatience (column 5). We can reject equality of this estimate with 1 and with the corresponding estimate of  $\delta_{QH}$  for those with below-median impatience. Columns 2, 4, and 6 show similar results after removing problematic respondents: our estimates of  $\delta_{QH}$ are again significantly different for those with above- and below-median impatience index. We summarize the reduced-form effort impatience measure separately for those with aboveand below-median impatience index in Figure C.1(a). The above-median impatience sample has substantially higher average values of the reduced-form effort impatience measure: they allocate an average of 1.3 more tasks to future dates than today across piece rates, while those with below-median impatience index allocate only 0.4 more tasks to future days.

Appendix Figure C.1: Higher Impatience Index Predicts Higher Effort Impatience but Not Higher Recharge Impatience



Notes: Data come from the validation sample and are at the individual level. Panel (a) displays the average difference between the number of tasks chosen on all future dates minus the number of tasks chosen on the survey day (for the same payment amount) separately for the below- and above-median impatience index samples. In Panel (b), we display the average proportion of recharge MPL choices where the individual chose to get a smaller recharge today rather than a larger recharge in the future separately for the below- and above-median impatience index samples.

To test the significance of this difference, we estimate the following regression:

$$EffortImpatience_{itw} = \beta_0 + \beta_1 ImpatienceIndex_i + \beta_2 y_{i0w} + \tau_w + \tau_t + \varepsilon_{itw}$$
(23)

where  $EffortImpatience_{itw}$  is the reduced-form effort impatience measure for individual *i* allocating tasks on day *t* at piece rate *w*,  $ImpatienceIndex_i$  is either the impatience index or an indicator for having an above-median impatience index, and  $y_{i0w}$  is the number of tasks chosen by individual *i* at piece rate *w* on day 0; controlling for this allows the effort impatience measure to vary with the overall number of chosen tasks and improves precision.<sup>68</sup>  $\tau_w$  and  $\tau_t$  are fixed effects for the piece rate and task day, respectively. The coefficient of interest is  $\beta_1$ .

Consistent with Figure C.1, Column 1 of Table C.2 shows that the difference in reduced-form effort impatience between those with above- and below-median impatience index is roughly 1.0 task, significant at the 10% level. Column 2 shows that the relationship is even stronger excluding

<sup>&</sup>lt;sup>68</sup>Define  $y_{itw}$  as the number of tasks chosen by individual *i* on day *t* for piece rate *w*. Since  $EffortImpatience_{itw} = y_{itw} - y_{i0w}$ , the coefficients from this regression are exactly equivalent to a regression with  $y_{itw}$  as the dependent variable that includes the same controls. The specification in equation (23) allows the mean value of the dependent variable to be comparable to Figure C.1.

problematic individuals: the gap is 1.7 tasks, significant at the 5% level. Columns 3 and 4 show qualitatively similar but less precise patterns with the impatience index as the regressor.

	Effort impatience			Recharge impatience				
	Future tasks - day 0 tasks			Chose earlier date				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Above-median impatience index	$1.000^{*}$ [0.513]	$1.708^{**}$ [0.798]			-0.0457 $[0.102]$	-0.0397 $[0.109]$		
Impatience index			0.763 [0.629]	$2.666^*$ [1.490]			-0.0425 $[0.0911]$	-0.0559 $[0.114]$
P-value: Impatience	0.055	0.038	0.229	0.081	0.655	0.716	0.642	0.625
Sample	All	Changers + Mono- tone	All	Changers + Mono- tone	All	No vio- lations	All	No vio- lations
Dep. var. mean (below-median impatience)	0.445	0.596	0.445	0.596	0.469	0.455	0.469	0.455
Correlation (dep var, Impatience index)	0.15	0.19	0.13	0.25	-0.05	-0.05	-0.05	-0.06
# Individuals # Observations	$71\\852$	$\begin{array}{c} 43\\516\end{array}$	71 852	$\begin{array}{c} 43\\516\end{array}$	71 710	$\begin{array}{c} 64 \\ 640 \end{array}$	71 710	64 640

Appendix Table C.2: Impatience Index Correlates With Effort (But Not Recharge) Impatience

Notes: This table shows the relationship between the effort and recharge impatience measures and the impatience index in the validation sample. Each observation is an individual  $\times$  effort or recharge choice. The dependent variable in columns 1–4 is the difference between the tasks allocated in the choice and the tasks allocated on day 0 (the survey date) for the same piece rate; controls include fixed effects for the piece rate and task day, as well as the number of tasks chosen for that same piece rate on day 0. The dependent variable in columns 5–8 is an indicator for choosing recharges today rather than in the future; controls include fixed effects for how many weeks in the future the individual will be paid for the later recharge option (either 1 or 2 weeks) and for the relevant payment amount. The "Changers + Monotone" sample restricts to individuals who changed their effort choice at least once and had fewer than two choice non-monotonicities in payment levels. The "No violations" sample represents people who do not switch multiple times on either price list. The regressor in columns 1, 2, 5 and 6 is an above-median impatience index dummy, while in columns 3, 4, 7 and 8 the regressor is the continuous index. Correlations shown at the bottom of each column are between the individual-level average of the dependent variable and the version of the impatience index used in that column. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.

Excluding problematic individuals, the magnitudes of the correlations between effort impatience and the impatience index are relatively high for the (noisy) domain of effort impatience -0.3 and 0.2 for the continuous and binary indices, respectively. In comparison, Augenblick et al. (2015) and Augenblick (2018) find correlations of 0.2 and 0-0.2 between effort impatience estimates and demand for commitment or qualitative discounting questions, respectively. Lack of Correlation with Recharge Impatience Measure Figure C.1(b) summarizes the recharge impatience measure separately for those with above- and below-median impatience index. Recharge impatience (i.e., choosing a smaller, sooner recharge over a larger, later recharge) is very similar across the subsamples; in fact, those with above-median impatience index have slightly lower recharge impatience. Columns 5 and 7 of Table C.2 confirm that there is no meaningful or significant relationship between recharge impatience and the impatience index using regression analysis. Columns 6 and 8 replicate the results without problematic respondents.

### C.2 Additional Estimates of the Discount Factors Over Effort and Payment

In this section, we present two estimates of the discount factor over payment (recharges), and one additional estimate of the discount factor over effort, all from our main experimental sample. We then summarize these estimates alongside the effort discount factor estimated in the validation sample (the Section C.1.2 estimate based on the Effort Choice by Date data). While both estimates of the discount factor over effort are meaningfully below 1, both payment discount factor estimates are close to 1 and significantly higher than either effort discount factor estimate. We begin by describing the additional estimation procedures.

"Simple CTB" Estimates of the Discount Factors Over Effort and Payment Following Augenblick et al. (2015), we estimate the discount factors for effort and money using the "Simple CTB" choices in each domain described in Section 4.2. Our primary specifications parametrize each discount factor as a single quasihyperbolic discount factor on future events (e.g.,  $\delta^{(t)} = \delta_{QH}$ ) but we estimate exponential parameterizations for robustness (e.g.,  $\delta^{(t)} = \delta_{Exp}^t$ ).<sup>69</sup>

Paycycle Estimates of the Discount Factor Over Payment Since impatience over payment will lead effort to increase as the payday approaches, one can use the pattern of effort over the pay cycle to estimate the payment discount factor. We follow Kaur et al. (2015), which calculates the discount factor using the elasticity of walking to payment and the pattern of effort as the payday approaches. We calculate the payment discount factor with the equation  $\frac{1}{d_{QH}} - 1 = \frac{1}{\varepsilon} \frac{w_T - w_t < T}{w_t < T}$ . where  $\varepsilon$  is the elasticity of walking to payment,  $w_t$  is compliance in period t, day T is payday, and days t < T all occur before payday. We calculate the percentage increase in compliance on payday,  $\frac{w_T - w_t < T}{w_t < T}$  from the estimated "payday spike" in the base case group (column 1 of Online Appendix Table F.12), and we estimate  $\varepsilon$  from the compliance response to the payment variation between the small payment and base case groups.

**Comparing the Discount Factors over Effort and Payment** Figure C.2 shows the payment discount factor estimates from both the Simple Recharge CTB and the paycycle effects, as well as the effort discount factors estimated from the Simple Effort CTB and the Effort Choice by Date. In all cases, the figure presents the estimates with the discount factors parametrized as a single discount factor ( $\delta_{QH}$ ) applied to all future periods.

<sup>&</sup>lt;sup>69</sup>The quasihyperbolic CTB discount factor over recharges is estimated with the equation  $\ln\left(\frac{c_t+\omega_1}{c_{t+k}+\omega_2}\right) =$ 

 $<sup>\</sup>frac{\ln(d_{QH})}{\alpha-1}\mathbb{1}_{t=0} + \frac{1}{\alpha-1}(1+r)$  where  $c_t$  is money in the earlier period,  $c_{t+k}$  is money in the later period,  $\omega_1$  and  $\omega_2$  captures background consumption, and r is the interest rate for each choice. The estimating equation for the discount factor over effort is similar:  $c_t$  and  $c_{t+k}$  are replaced by  $e_t$  and  $e_{t+k}$  (minutes of walking on days t and t+k),  $\omega_1$  and  $\omega_2$  are background walking effort (10 minutes), and 1+r captures the marginal rate of substitution between sooner and later effort. Following Augenblick et al. (2015), we choose  $\omega_1 = \omega_2 = \omega$ , as a function of the base recharge consumption or base walking effort (we set the  $\omega$ 's at 50% of the base level for recharges and walking, so  $\omega = 50$  INR and  $\omega = 10$  minutes of walking, respectively), but the results are robust to a range of values from 25% to 200% of the base level.

The estimates of the payment discount factor are both near 1, with the payday effect estimate greater than (but not statistically significantly different from) 1, and the CTB estimate close to 1 (0.962) but significantly different from it. In contrast, both estimates of the effort discount factor are substantially smaller, at 0.565 from the validation sample and 0.845 from the Simple CTB in our main sample. Both are significantly less than either estimate of the payment discount factor (*p*-values for tests of equality are in the notes for Figure C.2.)

Appendix Figure C.2: The Discount Factors Over Effort Are Significantly Lower Than the Discount Factors Over Money



Notes: This figure presents four structural estimates of the discount factors over effort (blue bars) and payment (orange bars). From left to right, the estimates come from the Simple Effort CTB data from the experimental sample, the Effort Choice by Date data from the validation sample, the pay cycle method in the experimental sample, and the Simple Recharge CTB data from the experimental sample. The discount factor is parameterized as a single quasihyperbolic discount factor on the future ( $\delta^{(t)} = \delta_{QH}$  or  $d^{(t)} = d_{QH}$ ). The *p*-values for tests of equality between the effort discount factor ( $\delta_{QH}$ ) from the Effort Choice by Date methodology and the two monetary discount factors ( $d_{QH}$ ) estimated via the Simple Recharge CTB and payday effects are 0.042 and 0.052, respectively. The *p*-values for tests of equality between the effort discount factor ( $\delta_{QH}$ ) from the Simple Effort CTB and the two monetary discount factors ( $d_{QH}$ ) estimated via Simple Recharge CTB and payday effects are both <0.001. The respective samples for bars 1, 2, 3, and 4 include 852 choices of 71 individuals, 6,380 choices of 3,190 individuals, 71,672 days of 890 individuals, and 16,146 choices of 2,307 individuals.

Results are similar if we estimate exponential discount factors. We estimate daily exponential effort discount factors of 0.976 and 0.953 using Simple Effort CTB and Effort Choice by Date, respectively. Both are significantly less than 1 and significantly less than either estimate of the exponential payment discount factor (1.009 and 0.992 for Pay Cycle and Simple Recharge CTB estimates, respectively).

### C.3 Measures of Effort and Recharge Impatience Are Uncorrelated

This section summarizes two types of evidence from our setting suggesting that discount factors over effort and recharge are relatively uncorrelated. First, survey measures of effort and recharge impatience are uncorrelated. Second, measures of effort impatience do not correlate with pay cycle effects.

	Direct measure		Proxies for recharge impatience			
	Simple CTB (Recharge)	Negative mobile balance	Negative yesterday's talk time	Prefers daily (=1)	Prefers monthly (=-1)	# Individuals
	(1)	(2)	(3)	(4)	(5)	(6)
Impatience index	0.004	0.032	-0.068	-0.038	0.034	1740
Predicted impatience Index	0.000	0.021	-0.014	-0.005	-0.003	3192
Chose commitment	-0.006	0.009	-0.001	0.005	0.010	2871
Simple CTB	0.006	-0.011	-0.037	0.001	0.041	3190

Appendix Table C.3: No Correlation Between Measures of Impatience over Effort and Recharges

Notes: This table displays the correlations in our experimental sample between our various measures of impatience in the effort domain (in the rows) and measures and proxies for impatience in the recharge domain (in columns). The "Simple CTB (Recharge)" measure is the average of the share of money allocated to today from the questions used in the Simple Recharge CTB. Proxies for recharge impatience in columns 2–5 were all measured at baseline. For columns 4 and 5: 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. Data are at the individual level. The sample in each row is the subset of participants we have each impatience measure for. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.

Table C.3 shows that there is no significant or meaningful correlation between any of the measures of impatience over effort and impatience over payment collected in the experimental sample. Similarly, we find that the correlation of the individual-level averages of the recharge impatience and effort impatience measures in the validation sample is only -0.05, which is statistically indistinguishable from 0 (*p*-value = 0.66).

As discussed in Section C.2, pay cycle effects also measure impatience over payments. Thus, we can test whether participants' impatience over payment relates to our measures of impatience over effort by testing whether effort impatience measures predict pay cycle effects.

Panel A of Online Appendix Figure F.4 shows that there are no meaningful payday spikes even among those with above-median impatience index. Moreover, the patterns across the pay cycle are very similar for those with below-median impatience, depicted in Panel B. Results are similar for the other measures of effort impatience (i.e., the predicted impatience index, demand for commitment, and simple CTB). Regression analysis confirms that there are no large or significant differences in pay cycle effects across any measure of effort impatience.

### **D** Distributional Impacts of Thresholds

This section assesses the effect of thresholds on the distributions of weekly and interventionaverage compliance. We first assess whether thresholds decrease intermediate effort just below the threshold. Panels (a) and (b) of Figure D.3 show histograms at the individual × week level of the number of days the individual met their step target in that week, for the 4-day or 5-day threshold group, respectively, relative to Base Case and Monitoring (confidence intervals are relative to Monitoring). Indeed, the threshold contracts do modestly decrease effort just below the threshold: the prevalence of walking 3 or 4 days is lower in 5-Day Threshold than either Base Case (*p*-value <0.001) or Monitoring (*p*-value = 0.008), and the prevalence of walking 2 or 3 days is lower in 4-Day Threshold than either reference group (*p*-values <0.001 for both Base Case and Monitoring).<sup>70</sup>



(a) 4-Day Threshold

(b) 5-Day Threshold

Appendix Figure D.3: Thresholds Modestly Decrease Compliance Right Below the Threshold

Notes: Figures show histograms of the number of days a participant exceeded the step target each week during the intervention period in the Base Case, 4-Day or 5-Day Threshold, and Monitoring. Data are at the respondent-week level. Confidence intervals represent a test of equality between Monitoring and each other group from regressions with the same controls as Table 2, except that, because the data are at the weekly level, we do not control for day-of-week fixed effects. See Online Appendix Table F.17 for the regressions.

However, the magnitude of these differences are relatively small (especially compared to the differences from Monitoring), leading to only slight differences between Base Case and Threshold in the overall distribution of weekly compliance. Specifically, Panels (a) and (b) of Figure D.4 show the cumulative distribution functions (CDFs) of weekly compliance in 4-Day and 5-Day Threshold, respectively, relative to Base Case and Monitoring. While the distributions of weekly compliance in Base Case and both threshold groups all differ markedly from the distribution in Monitoring, the differences between Base Case and the threshold groups are small. Panels (c) and (d) of Figure D.4 shows similar results for the distribution of individual-level (instead of individual × week-level) compliance. Quantile regressions reveal no significant differences between the

<sup>&</sup>lt;sup>70</sup>Notably, neither threshold increases the likelihood of walking exactly the threshold number of days. Our model suggests this may reflect that the contracts pay for above-threshold compliance (e.g., the 4-day threshold pays for the 5th day of compliance). Additional explanations outside of the model include habit formation or that it is easier to schedule walking every day in a given week than on a subset of days.

threshold groups and Base Case in the 25th, 50th, or 75th percentiles of the distributions of either individual  $\times$  week-level or individual-level compliance (see Online Appendix Table F.7). Kolmogorov-Smirnov (KS) tests for the equivalence of the individual-level distributions also fail to reject the null of equal distributions (*p*-values 0.238 and 0.852 for the 4- and 5-Day Threshold, respectively, relative to Base Case).





(d) 5-Day Threshold, Overall Compliance

#### Appendix Figure D.4: Threshold and Base Case Have Similar Compliance Distributions

Notes: This figure shows the cumulative distribution functions (CDFs) of the distributions of weekly compliance (i.e., the number of days the individual exceeded the step target in a week) in Panels (a) and (b), and intervention-average compliance (i.e., the percentage of days the individual exceeded the step target during the intervention period) in Panels (c) and (d). All CDFs are plotted separately by treatment group for the monitoring, base case, 4-day (Panels (a) and (c)), and 5-day (Panels (b) and (d)) threshold groups. For Panels (a) and (b), data are at the individual  $\times$  week level, limited to weeks where the individual has at least 4 days of data. For Panels (c) and (d), data are at the individual level, limited to individuals who had at least 21 days of data over the 12-week intervention period. Both weekly and intervention-average compliance are residualized using the same controls as in Table 2 except that we do not include day-of-week fixed effects because data are not at the day level.

### **E** Predicting Impatience with Policy Variables

This appendix provides proof of concept that a policymaker could use hard-to-manipulate observable characteristics to predict impatience and effectively target the threshold contract.

Our Section 5.3 results suggest that a policymaker could improve our program's effectiveness by targeting threshold contracts only to more impatient individuals. However, impatience is challenging to observe; even were policymakers to field surveys on impatience, participants might game their responses to avoid a specific contract — especially a financially dominated one.

To address this concern, we construct a "policy prediction" of impatience: a prediction of the impatience index using demographics (e.g., age, labor force participation) and medical information (e.g., HbA1c, fatigue) that health policymakers would likely have access to. We show that there is significant heterogeneity in the effect of the threshold by the policy prediction. Hence, the policy prediction could be used to personalize contract assignment.

To prevent overfitting, we use a split sample approach. First, in a randomly-selected training sample, we fit a LASSO model to predict the impatience index with the variables listed in the Table E.1 notes. We then use the model to predict impatience out of sample for all other participants (the "regression sample"). Finally, in the regression sample, we estimate the heterogeneity in Threshold performance by the policy prediction using equation (9). To sufficiently power this regression, we allocate 2/3 of participants to the regression sample.

The results, in Table E.1, are similar to Table 3: Threshold has a higher treatment effect among people with higher predicted impatience. This suggests that personalizing thresholds using a policy prediction could significantly improve the effectiveness of incentives at scale.

Dependent variable:	Exceeded step target				
	(1)	(2)			
Impatience $\times$ Threshold	0.03**	0.06**			
	[0.00, 0.06]	[0.00, 0.12]			
Threshold	-0.01	-0.03**			
	[-0.04, 0.02]	[-0.07, -0.00]			
Impatience	-0.02**	-0.05**			
	[-0.04, -0.00]	[-0.09, -0.01]			
τ		Above-median policy			
Impatience measure:	Policy prediction	prediction			
Base Case mean	.502	.502			
# Individuals	1,969	1,969			
# Observations	157,946	157,946			

Appendix Table E.1: Threshold Effect Varies with Policy Prediction of Impatience

Notes: This table replicates Table 3 with an impatience index predicted out-of-sample with the following variables (and their interactions with above-median age, gender, and individual and household income): age; gender; labor participation; personal and household monthly income; household size; HbA1c; RBS; systolic and diastolic BP; BMI; waist circumference; walking speed; diagnosed diabetic or hypertensive; overweight; owns home; number of rooms and running water in home; has a bank account; hired help; number of scooters, cars, computers, smartphones, and mobile phones; mobile balance; hours of work on a weekday; consumes alcohol and cigarettes/bidis; has foot ulcer, rapid deterioration in eyesight, and pain or numbness in legs or feet. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%.