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

Shilpa Aggarwal — Indian School of Business

Rebecca Dizon-Ross — University of Chicago

Ariel Zucker * — UC Santa Cruz

August 21, 2023

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, time-bundled contracts generate as much effort as linear contracts, yet at a reduced cost. Moreover, time-bundled contracts perform meaningfully better among individuals with greater impatience over effort, suggesting that impatience is a contributing mechanism. In contrast, increasing the frequency of payment – which should be effective if individuals are impatient over payment rather than effort – has no effect, suggesting limited impatience over payments. Overall, the incentive program is effective, increasing daily steps by roughly 20 percent (13 minutes of brisk walking) and improving health.

---

*Aggarwal: Indian School of Business, shilpa_aggarwal@isb.edu. Dizon-Ross: University of Chicago Booth School of Business, rdr@chicagobooth.edu. Zucker: University of California, 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). 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 (e.g., Gertler et al., 2019), exercising (e.g., Carrera et al., 2020), and studying (e.g., Fryer, 2011). One motivation for these incentives is to offset underinvestment due to impatience, a common trait (e.g., Mahajan et al., 2020; Augenblick and Rabin, 2019; O'Donoghue and Rabin, 1999a). Impatient individuals (i.e., those who heavily discount the future) place low value on the future benefits, which leads them to undertake fewer of these beneficial behaviors. Structuring the incentive contracts so that they work well for impatient individuals could therefore be highly valuable, yet our understanding of how to do so remains limited.

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. We use a randomized controlled trial (RCT) to compare time-bundled contracts to a more standard separable contract (in which the payment for current effort depends only on current effort). We show that time-bundled contracts meaningfully improve contract performance for impatient individuals and hence are an effective strategy for adapting incentives for impatience. Our RCT also evaluates an incentive program for exercise among diabetics and prediabetics, showing that it could be a powerful tool in the global fight against chronic disease.

We begin by showing theoretically that, relative to separable contracts, time-bundled contracts are more effective when individuals have a higher discount rate over effort. The relevant discount rate for this prediction is the primitive discount rate in an individual's utility function, as opposed to the discount rate over payment, which instead reflects the available borrowing and savings opportunities (Augenblick et al., 2015).

To illustrate the intuition for why time-bundled contracts work well when people highly discount their future effort costs, 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 works both days. For the contract to induce two days of work, the total payment must exceed the worker's present discounted cost of effort.[1] For example, if her daily cost of effort is \$10, and she discounts future effort by 50%, the payment only needs to be \$15: \$10 for the first day plus a discounted \$5 for the second. In contrast, if you pay her separately for each day of work, a larger minumum payment of \$20 is required to induce two days of work: \$10 per day of effort. Thus, time-bundled threshold contracts can generate the same amount of effort for a lower cost by exploiting the fact that, when individuals have high effort discount rates, it is "cheaper" to buy their future (discounted) effort than their current effort.

[1]This example assumes a zero short-run interest rate on payments for simplicity.

Time-bundled thresholds should be effective for all types of people with high discount rates over effort: time-consistent or time-inconsistent and, among time-inconsistent, sophisticated or "naïve" (or unaware) about their own present bias. We consider thresholds' effectiveness for naïfs to be an important aspect of their potential effectiveness for the impatient. Naïve time inconsistency is common (for example, Mahajan et al. (2020) estimate that 50% of a sample of Indian adults are naïfs), but naïfs are difficult to motivate (Bai et al., 2020). The effectiveness for naïfs differentiates time-bundled threshold contracts from commitment contracts, another approach used to motivate time-inconsistent people.[2] For commitment contracts to be effective, people must be sophisticated about the differences between their preferences and discount rates in the future relative to the present-day. In contrast, time-bundled contracts directly leverage *present-day* discount rates, which even naïfs understand. That is, even naïfs discount their future effort and will sell it at a discount today.

We explore the effects of time-bundled contracts using an RCT evaluating an incentive program for exercise among diabetics and prediabetics. The healthy behaviors (e.g., exercise) that help prevent and manage diabetes and other lifestyle diseases feature short-run costs but only long-run benefits, making impatient people more likely to underinvest in them. Indeed, evidence suggests that people with diabetes (and other chronic lifestyle diseases) are more impatient than the general population (Reach et al., 2011; Wainwright et al., 2022), making them particularly well-suited for exploring the impact of tailoring incentives for 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, our experiment varied whether payment was a linear function of the number of days the participant complies with the step target or was instead a time-bundled threshold function. The time-bundled threshold function 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.

The primary contribution of our evaluation is to demonstrate the effectiveness of using time-bundled threshold contracts to incentivize the impatient. We present two main findings. First, we show that, on average, across the full sample, the time-bundled threshold contracts perform better than the linear contract—they achieve the same sample-average level of compliance as the linear contract, but do so at a lower cost. For example, the five-day threshold contract pays out nearly 20% 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. We show that this improves the performance of the contract from the perspective of a policymaker who

[2]Commitment contracts provide people with the option to undertake dominated actions in order to compel their future selves into a specific action. For example, a commitment contract for day 2 work might, on day 1, collect money from workers, and then only return the money to the workers if they worked on day 2.

wants to maximize the benefits of compliance net of the incentive costs.[3]

The second finding is that high levels of impatience in our sample are an important mechanism underlying the effectiveness of the time-bundled threshold contracts, as time-bundled threshold contracts are meaningfully more effective for those who are more impatient over effort. Specifically, heterogeneity analysis using a measure of impatience taken from the psychology literature shows that, relative to linear contracts, 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 below-median impatience. This difference is large relative to 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. Although our analyses exploit non-random variation in impatience across the population, we provide evidence suggesting that confounding factors do not drive our results.[4] The threshold also improves cost-effectiveness (i.e., the payout per day of compliance) among both less and more impatient populations. The threshold thus clearly improves performance among those with greater impatience, while having an ambiguous effect for those with lower impatience.

These results imply that policymakers or firms may be able to improve the performance of incentives by customizing whether people receive linear or time-bundled threshold contracts based on their impatience. These improved incentives could be leveraged for a range of policy goals, such as savings (e.g., Beaman et al., 2014; Breza and Chandrasekhar, 2019), preventive health (e.g., Hussam et al., 2022; Banerjee et al., 2010), and school attendance (e.g., Barrera-Osorio et al., 2011). Policymakers could customize the incentives at the population level by using time-bundled thresholds for populations that are particularly impatient, such as people with chronic disease or younger people (Read and Read, 2004). Customization could also occur at the individual level. Individual-level customization can be challenging to implement since impatience is often not observable; however, we provide multiple pieces of evidence suggesting that such personalization would be feasible, for example by showing that a principal could use more easily observed characteristics as proxies for impatience.

To place our findings on time-bundling in context, we also assess a more standard strategy for adjusting incentives for impatience: increasing the frequency of payment. 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) describe more frequent payment as the

[3] We discuss other potential objectives (e.g., welfare maximization) later in the paper. The statement depends on the assumption that the benefits of compliance are linear in compliance, which evidence suggests may hold in many settings, including exercise (Warburton et al., 2006).

[4] For example, our main findings are robust to controlling for other observable characteristics interacted with the threshold. Machine learning tools designed for studying heterogeneity also show that, relative to other covariates, impatience has a particularly strong signal in predicting the impact of time-bundled thresholds.

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 often do not do (Augenblick et al., 2015).

We find that increasing the frequency of payment has no impact in our setting, indicating that participants have low discount rates over the contract payments (mobile phone credits).[5] The lack of impact of high-frequency payments in our setting makes it important to identify other methods to adjust incentives for impatience and highlights the significance of our finding that time-bundled contracts are one such method.

The 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 growing burden is to promote lifestyle changes such as better exercise and diet (World Health Organization, 2009), the existing evidence-based interventions promoting lifestyle change in this population are prohibitively expensive (Howells et al., 2016). Governments are thus interested in scalable interventions to promote lifestyle change among diabetics. Indeed, we conducted our RCT in partnership with the Government of Tamil Nadu, one of the most populous states in India, who funded the project to identify an intervention to scale up across their state to address their exploding diabetes epidemic.

We find that our relatively low-cost incentives program substantially increases exercise and improves the health of diabetics. Providing an incentive of just 20 INR (0.33 USD) per day increases compliance with the step target by 20 pp off of a base of 30%. Average daily steps increase by 1,300, equivalent to 13 additional minutes of brisk walking, roughly a 20 percent increase. Importantly, nearly 60% of the treatment effect on steps continues for several months after the intervention ends. The increase in exercise induced by incentives translates to improvements in blood sugar, cardiovascular health, and mental health. These impacts on exercise and health are promising for policy, especially since, unlike existing evidence-based exercise interventions for diabetics, our program is scalable and low cost.

1.1 Contributions to the Literature

This paper contributes to three strands of literature: time preferences, nonlinear incentives, and incentives for health behaviors.

[5] While it is possible that people would have been more impatient over payments delivered with a different modality, limited impatience over payments is not rare (Augenblick et al., 2015; Andreoni and Sprenger, 2012).

Time Preferences Our first contribution is to show theoretically and empirically that time-bundled threshold contracts are effective for a wide range of people who are impatient over effort. Researchers have primarily motivated impatient agents with commitment devices (e.g., Royer et al., 2015; Ashraf, Karlan, and Yin, 2006). Commitment is a useful tool, but it is not a panacea. Take-up of commitment devices is modest (Laibson, 2015), undermining their use as a broad policy solution. Moreover, commitment devices are only effective for sophisticated time-inconsistents; they are less effective—and can even be harmful—for naïfs (e.g., Bai et al., 2020; John, 2020). In contrast, time-bundled contracts do not require sophistication; if anything, we show that naïvete opens up another channel for time-bundled contracts to be effective. Our theoretical insight that time-bundling can motivate impatient people relates to work by Jain (2012), who shows that firms can theoretically increase productivity by offering multi-period quotas to salespeople who are present-biased over both payments and effort.

We add to the literature that examines alternative approaches to commitment contracts to motivate impatient agents. O'Donoghue and Rabin (1999b) and Carrera et al. (2020) both examine ways to help time-inconsistent procrastinators avoid delay in completing a single task.[6] Andreoni et al. (2018) customize contracts to agent time preferences with the goal of making agents exert the same effort on two different days. DellaVigna and Pope (2018) examine whether decreasing the lag between effort and payment increases effort. Our distinctiveness from these related studies lies in the novel approach (time-bundled contracts) used to increase effort.

A secondary contribution to the time preferences literature is to study the implications of domain-specific discounting for contract design. Although it is well known that there is a distinction between discount rates over payment and effort (Augenblick et al., 2015), the vast majority of papers examining dynamic contracts assume the same discount rate for both (e.g., Lazear, 1981; Chassang, 2013). We show that allowing these discount rates to differ has interesting implications: while more frequent payment is effective for those who discount payment highly, time-bundling is effective for those who discount effort highly.

Nonlinear Incentives Although there is a theoretical literature showing that many optimal dynamic contracts display nonlinearities over time (e.g., Lazear, 1981; Lambert, 1983), the empirical literature using exogenous variation to compare dynamically linear and nonlinear contracts is small. Moreover, most existing experiments focus on differential selection into nonlinear contracts (e.g., Larkin and Leider, 2012; Kaur et al., 2015). Our experiment contributes to the literature by providing a rare empirical comparison of the two types of contracts, showing

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

that dynamically nonlinear contracts can in fact increase cost-effectiveness and documenting how the dynamic structure of incentives interacts with time preferences. DellaVigna and Pope (2018) also randomize contract linearity but do not examine cost-effectiveness or the role of time preferences in determining the effectiveness of non-linear contracts.

Health Finally, we show that incentives for exercise are a scalable, effective intervention that can help decrease the burden of chronic disease in resource-poor settings. Prior evaluations of incentives for diabetics have targeted non-exercise outcomes with limited success (Sen et al., 2014; VanEpps et al., 2019; Desai et al., 2020); for example, Long (2012) provides diabetics in the US with incentives to lower their blood sugar and finds no impact. In contrast, building on previous work showing that incentives increase walking among healthy populations (e.g., Bachireddy et al., 2019; Finkelstein et al., 2008, 2016; Patel et al., 2016), we find that incentives increase not just walking but also health among those with chronic disease. Relative to other exercise incentive programs, ours stands out for its relatively large and persistent effect on behavior, measurable impacts on downstream health outcomes, and low cost. The success of our targeted exercise incentive program contrasts with the lack of impact of more broad-based, comprehensive wellness programs in the US (Jones et al., 2019).

The paper proceeds as follows. Section 2 presents our theoretical predictions. Sections 3 and 4 discuss the study setting and design. Section 5 explores the empirical relationship between time-bundled contracts and impatience, and Section 6 presents the impacts of incentives on exercise and health. 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 not time-bundled. 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 incentivizes effort only if a threshold level of effort is reached.

We present two key results applicable to various types of impatience, including time-inconsistent sophistication and time-inconsistent naivete. The first result is that the time-bundled threshold contract's effectiveness is normally 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, conversely, 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 discount rates over payment are high.

2.1 Set-Up

Each day, an individual chooses whether to complete a binary action. The principal then gives the individual a payment whose amount depends on the individual's past and present actions. Define w_t as an indicator for whether the individual "complies" (i.e., completes the action) on day t. Let m_t be the payment made to the individual on day t.

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

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

where e_t is the effort cost of complying on day t, $\delta^{(t)}$ is the discount factor over effort t days in the future, and $d^{(t)}$ is the discount factor over payments received t days in the future (for notational simplicity, we denote $\delta^{(1)}$ as δ 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 if either $\delta^{(t)}$ or $d^{(t)}$ are non-exponential functions of t or if $d^{(t)} \neq \delta^{(t)}$. Among time-inconsistent agents, we follow O'Donoghue and Rabin (1999a) in distinguishing sophisticates, who are aware of their discount factors (over both effort and money), from naïfs, who "believe [their] future selves' preferences will be identical to [their] current self's." Thus, letting $w_{t,j}$ be the agent's prediction on day j about her compliance on day $t > j$, sophisticates accurately predict how their future selves will behave ($w_{t,j} = w_t$) while naïfs may not ($w_{t,j} \geq w_t$).

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 the beginning of day t. Note that e_t can be negative, reflecting that 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 constraints.[7] 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 λC 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). With linear benefits, effectiveness becomes $\lambda C - P$.

We want to compare the effectiveness of different contracts even when we do not know λ. 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.[8]

2.2 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:

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

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

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

Holding all else constant, compliance is thus independent of $\delta^{(t)}$.[9]

[7] 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.

[8] This is true assuming effectiveness is positive. To see this, rewrite effectiveness as $C\left(\lambda - \frac{1}{(C/P)}\right)$.

[9] In particular, 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 $\delta^{(t)}$.

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

Observation 1. In a linear contract, holding all else constant, neither compliance, cost-effectiveness, nor effectiveness depend on $\delta^{(t)}$.

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

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 below which no incentive is received, and above which payment is a linear function of the number of days of compliance. Total payment in the threshold contract is thus:

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

An important question is how the effectiveness of threshold contracts depends on impatience over effort. Appendix B.2 presents a series of propositions investigating this comparative static. We summarize the takeaway in the following prediction:

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

The propositions underlying Prediction 1 show that, holding all else equal, compliance and effectiveness in time-bundled threshold contracts tend to decrease in $\delta^{(t)}$ under a broad range of assumptions. In contrast, in linear contracts, both compliance and effectiveness are flat in $\delta^{(t)}$ (Observation 1). Thus, the lower $\delta^{(t)}$ is, the higher compliance and effectiveness tend to be in a time-bundled threshold relative to linear contract.

Specifically, Proposition 1 examines threshold contracts with $K = T$ (i.e., where one must comply on all days in order to receive payment). It shows that, for all T, *regardless of the effort cost distribution,* compliance is weakly decreasing in δ. To gain tractability to examine threshold effectiveness and threshold contracts with $K < T$, we then make additional 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 δ.[10] 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 version of the model where costs can either be high or low and are known from day 1, $K = 2$ and $T = 3$. We show that both compliance and effectiveness are higher when $\delta^{(t)}$ is lower.

Overall, the propositions suggest that, when either (a) K is high relative to T,[11] or (b) costs are positively correlated across periods, Prediction 1 tends to hold. 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.[12]

While Prediction 1 speaks to the heterogeneity in the performance of threshold relative to linear contracts by $\delta^{(t)}$, it is also important from a policy perspective to understand whether threshold or linear contracts perform better for any given level of $\delta^{(t)}$. To explore this, Appendix B.3 presents a series of propositions comparing threshold and linear contracts, restricting to the case where $T = 2$ and making additional assumptions (e.g., about the cost functions) for tractability. The propositions compare optimized threshold and linear contracts, as well as comparing threshold and linear contracts that offer the same payment per day (as in our experimental setting), as the latter comparison is easier to implement in practice.[13]

Prediction 2 (Threshold versus Linear Effectiveness, $T = 2$)**.** *Holding all else equal, under many conditions:*
(a) When δ is sufficiently low, threshold contracts tend to be more effective than linear contracts that offer the same payment amount per day ($m = m'$), while when δ is sufficiently high, the reverse is true.
(b) When δ is sufficiently low, the most effective contract tends to be a threshold contract, while when δ is sufficiently high, the most effective contract tends to be linear.

While the assumptions underlying Prediction 2 are more restrictive than those underlying

[10]Proposition 2 also makes the reasonable assumption that e_2 is weakly increasing in e_1, which flexibly accommodates the range from IID costs to perfect positive correlation but rules out negative correlation. We show that, as long as there is not "too much" inframarginal behavior, effectiveness in the threshold contract is weakly decreasing in δ. When there is too much inframarginal behavior, not only will the effectiveness prediction not hold but incentives become a less cost-effective approach.

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

[12]Individually-demeaned steps in a group that did not receive incentives have a correlation of 0.4 across days. Raw (i.e., not demeaned) steps have a correlation of 0.7 across days.

[13]Empirically optimizing contracts involves knowledge of both the discount rate and the distribution of costs.

Prediction 1, the prediction suggests that the comparison between threshold and linear contracts is empirically relevant (as one will not always dominate the other) and that principals will often prefer threshold to linear contracts when individuals are sufficiently impatient.

Intuition for the Result that Threshold Effectiveness Decreases in δ We illustrate the intuition by considering 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's motivation to comply is to have the opportunity to be paid $2m'$ for complying on day 2. The value of this opportunity to her is

$$(2m' - \delta e_2) w_{2,1}\Big|^{w_1=1}, \tag{5}$$

which is equal to the discounted (by $d = 1$) payment $2m'$ net of the discounted effort costs δe_2 if the individual thinks she will comply on day 2 given compliance on day 1 (i.e., if $w_{2,1}\Big|^{w_1=1} = 1$). Importantly, because the future effort cost is discounted, the value is weakly decreasing in δ for both sophisticates and naïfs: impatient people value the opportunity more.

The fact that impatient people value the day 2 opportunity highly underlies the threshold's greater effectiveness for them. Considering compliance on each day in turn, individuals comply on day 1 if the value of the day 2 opportunity outweighs their day 1 effort cost:

$$w_1 = \begin{cases} 1 & \text{if } e_1 < (2m' - \delta e_2) w_{2,1}\Big|^{w_1=1} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Since $(2m' - \delta e_2) w_{2,1}\Big|^{w_1=1}$ is weakly decreasing in δ, impatient people comply more on day 1. On day 2, individuals comply if $w_1 = 1$ and the payment exceeds their effort costs:

$$w_2 = \begin{cases} 1 & \text{if } e_2 < 2m' \text{ and } w_1 = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

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.[14]

To highlight the intuition from the introduction, note that one can rewrite equation (6) as

$$w_1 = \begin{cases} 1 & \text{if } e_1 + \delta e_2 < 2m' \text{ and } w_{2,1}\Big|^{w_1=1} = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

For individuals to comply, the payment ($2m'$) must outweigh the present discounted cost of effort ($e_1 + \delta e_2$). The more the person discounts effort, the lower the present discounted cost is.

[14]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 $\delta^{(t)}$.

Sophisticates and Naïfs Although Predictions 1 and 2 hold for both sophisticates and naïfs, the exact compliance conditions differ (specifically, the terms projecting future behavior, $w_{j,t}$). In the two-day example, for sophisticates, who correctly predict their future preferences,

$$w_{2,1}\Big|^{w_1=1} = \mathbb{1}\{e_2 < 2m'\}. \tag{9}$$

Thus, for a sophisticate to place a positive value on a day 2 work opportunity (i.e., for expression (5) 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, naïfs believe that their day-2 selves have the same preferences as their day-1 selves. For them,

$$w_{2,1}\Big|^{w_1=1} = \mathbb{1}\{\delta e_2 < 2m'\}. \tag{10}$$

Thus, naïfs 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, naïfs positively value any lucrative day 2 *"option"* that they *want* their day 2 selves to execute. Naïfs comply on day 1 to give their day 2 selves the *option* to follow-through.[15]

With time-bundled thresholds, these differences in motivation between sophisticates and naïfs should not normally affect behavior. The day 2 opportunities that are lucrative enough options to motivate naïfs 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 (9) implies equation (10)).[16] By pairing the options that motivate naïfs with the commitment that both motivates sophisticates and helps naïfs follow through, thresholds work for both types.

In contrast, as discussed in Online Appendix F, in some types of time-bundled contracts (other than thresholds) naïfs and sophisticates often make different decisions,[17] as *options* and

[15]In either case, Prediction 1 still holds because the equation (5) value is still weakly decreasing in δ. The equation (5) 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 naïfs, both of which are decreasing in δ. To see this in the naïve case, note that $(dM - \delta e_2)\mathbb{1}\{\delta e_2 < dM\} = \max\{dM - \delta e_2, 0\}$.

[16]Equations (6), (9), and (10) 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, precisely because they know they would not follow-through on day 2. However, this behavior should be rare, as it requires that $e_2 > e_1/(1-\delta)$, which implies that e_2 is substantially higher than e_1 and/or that δ 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 move toward the threshold. Thus, the marginal incentive to comply is strictly higher on day 2 (where it is $2m'$) than on day 1 (where it is $2m' - \delta e_2$).

[17]Online Appendix F investigates the full class of 2-day time-bundled contracts. This class also includes

commitment are less tightly linked. For example, some time-bundled contracts that are not thresholds function like commitment contracts and are effective for sophisticates only, as day 1 compliance generates a soft commitment for day 2 compliance but not a positive NPV option.[18]

2.4 Payment Frequency and Impatience over Payment

We now briefly explore a potential strategy for improving the performance of incentives in the case that people are impatient over payment rather than effort: increasing payment frequency. Specifically, we return to the base case separable linear contract from equation (2) and analyze compliance under different payment frequencies by changing the length of the payment period T. 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.*

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 will have higher effectiveness than the linear when discount rates are high — 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 then test for heterogeneity in the effect of the threshold relative to the linear contract based on a baseline measure of impatience over effort (because Predictions 2 and 1 hold for both sophisticates and naifs, we do not attempt to separate these two types). 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. These hypothesis tests were all specified *ex ante*.[19]

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).

[18]Specifically, generating a commitment means that the payment for day 2 compliance is greater than e_2 if and only if $w_1 = 1$. Generating an option means that the payment for day 2 compliance is greater than δe_2 (rather than e_2) if and only if $w_1 = 1$. In contracts where day 1 compliance generates a commitment but not an option, sophisticates might comply on day 1 even when their effort cost exceeds the maximum potential financial benefit of day 1 compliance in order to induce their day 2 self to comply, while naïfs will not.

[19]Before launching our experiment, we prepared a pre-analysis plan that guided our design and power calculations. While we did not polish it to post publicly as part of our AEA registry, one can find it at https://www.chicagobooth.edu/-/media/faculty/rebecca-dizon-ross/research/PAP_NCD_2015.pdf.

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 higher both in southern states and in urban areas. We selected our sample through a series of public screening camps held in various locations including the government hospital, 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."[20] After screening, we contacted eligible individuals by phone and invited them to participate in a program encouraging walking.

Surveyors visited the participants at their homes or workplaces to conduct a baseline health survey, deliver lifestyle modification advice, and enroll them in a one-week phase-in period designed to collect baseline walking data and to familiarize participants with program procedures. Surveyors gave participants pedometers to use for the duration of the program; throughout the study, we gathered step data by syncing the pedometers with a central database. Because syncing requires an internet connection, which most participants did not have, pedometer step data were not available in real time. Instead, to have realtime data, we asked participants to report their daily step count to an automated calling system which called them every evening and prompted them to enter the number of steps recorded on their pedometer. During the pre-intervention visit, surveyors demonstrated how to wear a pedometer properly, report steps, and check text messages from our reporting system. Surveyors asked respondents 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. We exclude from the sample all participants who withdrew or were found ineligible prior to randomization, leaving a final experimental sample of 3,192 individuals. The sample represents 41% of the screened, eligible population (see Table A.1 for the share of people dropped in each stage of enrollment). We

[20] 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.

screened and enrolled the sample on a rolling basis from Oct. 2016 to Oct. 2017.

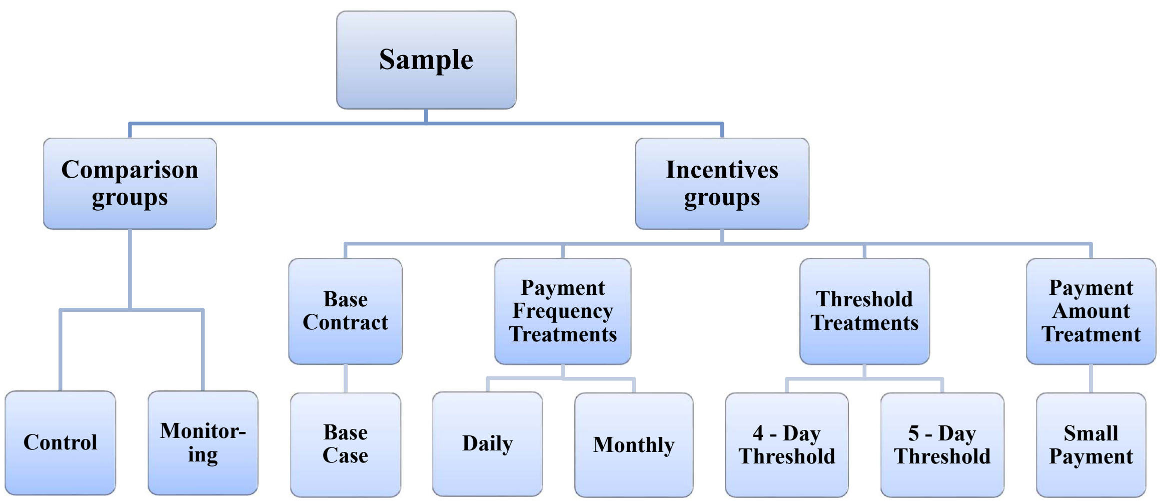
3.2 Experimental Design and Contract Launch

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 into one of six incentive contracts for walking, as shown in Figure 1 and described next.



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

Figure 1: Experimental Design

3.2.1 Incentive Groups

All incentive groups received payments for accurately reporting steps above the daily 10,000-step target through the automated step-reporting system. We delivered all incentive payments as mobile recharges (credits to the participant's mobile phone account).[21] After reporting steps, participants immediately received text-message confirmations of their step report, payment earned, and the payment date. We also sent participants weekly text messages summarizing their walking behavior and total payments earned.

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

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

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

Time-Bundled Threshold Contracts The *threshold* treatment groups differ from the base case incentive group only in linearity: while the base case is a linear contract, the threshold contracts use time-bundled threshold payment functions. The *4-day threshold group* 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, the *5-day threshold group* 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 groups to walk at least four or five days per week when initially explaining the contracts.[22] To maximize statistical power, we follow our *ex ante* analysis plan and pool the *4-day threshold* and *5-day threshold* groups for our main analyses. We sometimes also show the results for the two thresholds separately as exploratory analyses.[23]

[21] 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).

[22] For those in the threshold groups, the target days-per-week was the same as their assigned threshold level; for those in the other groups, it was randomly assigned in the same proportion as the threshold groups were divided between the 4- and 5-day groups.

[23] 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.

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

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

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

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.

Monitoring Monitoring group participants were treated identically to the incentive groups except that they did not receive incentives. They received pedometers and were encouraged to wear the pedometers and report their steps every day. They also received 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 The pure control group received neither pedometers nor incentives during the intervention period (they returned their pedometers at the end of the phase-in period). Because most incentive programs bundle the "monitoring" effect of a pedometer with the effect of incentives, the pure control group is a useful benchmark from a policy perspective.[24]

[24] To accommodate a request from our government partners, we also tested one additional intervention: weekly text message reminders to engage in healthy behaviors (the "SMS treatment"). Ten percent of the sample, cross-randomized across all other treatments, received this treatment, which we control for in our regressions.

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 H.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.[25]

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 group participants) for 12 weeks after the intervention period had ended. We hereafter refer to this period as the post-intervention period. Participants no longer reported steps daily or received incentive payments, but surveyors still returned every four weeks to sync their pedometers.

4 Data and Outcomes

This section first describes our baseline data sources — a health survey, a week of pedometer data, and a time-preference survey — and presents summary statistics. Next, it describes our two sources of outcomes data: pedometer data and a health survey.

4.1 Baseline Data: 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 the previous three months; random blood sugar (RBS), a measure of more immediate blood sugar control; body mass index (BMI) and waist circumference, two measures of obesity; blood pressure, a measure of hypertension; and a short

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

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 Baseline Data: Time Preferences

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." Since our sample was elderly and had low levels of education, our primary measure of impatience over effort is an index of responses to simple survey questions from the psychology literature on impatience and procrastination that our full sample could comprehend. This simple index gave us more reliable data than the screen-based convex time budget (CTB) measure of Andreoni and Sprenger (2012), which we also implemented, but which our sample had difficulty understanding.[26] The index questions, listed in Panel A of Table A.2, are a subset of the Tuckman (1991) and Lay (1986) scales chosen *ex ante* by our field team as translating well to our setting. Each question asks respondents to respond on a Likert scale of agreement with statements such as "I'm continually saying 'I'll do it tomorrow'." We construct the index (hereafter: the impatience index) by averaging the standardized question responses, as we pre-specified when including the questions in the survey.

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 values of the index walk less and have worse diets at baseline (Table A.2).

In Online Appendix J, we further validate our impatience index by showing that it predicts an incentivized measure of impatience over effort. Following our experiment, we elicited incentivized choices from a sample of similar participants regarding the number of effort tasks they wanted to complete on different days (e.g., the same day, a week later) 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, we find that those with higher values of the impatience index also make more effort-impatient incentivized choices, signing up for relatively more tasks in the future than the present.[27]

[26] Because respondents did not understand the CTB method, we have an order of magnitude more law-of-demand violations than lab-based studies with college students. Moreover, as described in Online Appendix K, our CTB estimates do not converge for 44% of the sample, they do not correlate in the expected direction with any behaviors, and respondents did not follow through with their chosen allocations. These issues make the CTB estimates unusable for analysis.

[27] Specifically, Online Appendix Figure J.2(a) shows that those with above-median impatience index have over twice as large a gap between tasks chosen for the future versus the present than those with below-median impatience index. We also structurally estimate the discount factor following Augenblick (2018), DellaVigna and Pope (2018), and John and Orkin (2021), and find that the discount factor varies significantly with the impatience index. Specifically, we estimate the average discount factor over effort 1, 7, and 8 days in the future. In the full sample, the estimated discount factor of 0.601 is both economically and statistically different from 1. However, only those with above-median impatience index appear to have a discount factor less than 1: it is

We began collecting our impatience index partway through the experiment,[28] so it is only available for the latter 55% of the sample. Luckily, that sample size is sufficiently powered to conduct heterogeneity analyses. That said, to check the robustness of our results in the full sample, we create a "predicted index" using a LASSO prediction fit with three similar survey questions on self-control that were collected from all participants. Panel B of Table A.2 lists the questions and shows that the predicted index correlates in the expected direction with behavior measures such as the health risk index.

Although impatience measures tend to be noisy (Kremer et al., 2019), and ours may be particularly so, measurement error would bias us against finding heterogeneity. Thus, the heterogeneity we would find with a more precise measure is likely even larger than what we find with our noisy measure.

Impatience over Payments Although we did not specify heterogeneity tests by impatience over payments *ex ante*, our data does contain some proxies for impatience over recharges, such as recharge balances and recharge usage.[29] We also collected additional data on impatience over recharges as part of our validation exercise after the experiment. These data suggest that our proxies do correlate with impatience over payment, and that discount rates over effort and over payment are relatively independent in our setting.[30]

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 and has slightly more males than females. The average monthly household income is approximately 16,000 INR (about 200 USD) per month, close to the median for an urban household in India (Ministry of Labour and Unemployment, 2016). Panel B shows that our sample is at high risk for diabetes and its complications: 65% of the sample has been diagnosed with diabetes by a doctor, 81% have HbA1c levels that indicate diabetes, and the RBS measures show poor blood sugar control. The sample also has high rates of comorbidities: 49% have hypertension and 61% are overweight. Panel C shows that, on average, participants

economically and statistically indistinguishable from 1 among those with below-median impatience index.

[28]We initially planned to only use the CTB measures. We added the impatience index after challenges surfaced in our data collection which made the CTB estimates unusable for analysis.

[29]Higher balances and/or usage indicate a person is less constrained and so the discount rate over recharges is likely to be lower, closer to the interest rate than the discount rate over consumption.

[30]Specifically, in our validation exercise we collected an incentivized measure of impatience in the payment (mobile recharge) domain using incentivized choices on a multiple price list (Andreoni and Sprenger, 2012). Online Appendix J suggests that those with higher balances and usage make less impatient incentivized choices over payment, suggesting that these are indeed appropriate proxies for [a lack of] impatience over recharges. However, none of the proxies for impatience over payment correlate consistently with our impatience index, which focuses on the effort domain (Table A.3). There is also no correlation between the incentivized measures of impatience over recharges and either the incentivized choices over effort or the impatience index.

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

	Full sample	Control	Monitoring	Incentives pooled	Daily	Base case	Monthly	Threshold	Small payment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Demographics									
Age (from BL)	49.56	49.78	50.28	49.44	49.57	49.60	48.80	49.41	49.11
	(8.51)	(8.19)	(8.95)	(8.55)	(8.60)	(8.33)	(8.94)	(8.71)	(7.84)
Female (=1)	0.42	0.46	0.43	0.41	0.44	0.41	0.38	0.41	0.48
	(0.49)	(0.50)	(0.50)	(0.49)	(0.50)	(0.49)	(0.49)	(0.49)	(0.50)
Labor force participation (=1)	0.74	0.73	0.72	0.75	0.75	0.74	0.81	0.75	0.70
	(0.44)	(0.45)	(0.45)	(0.43)	(0.43)	(0.44)	(0.39)	(0.43)	(0.46)
Per capita income (INR/month)	4465	4488	4620	4447	4068	4477	4599	4461	4341
	(3641)	(4483)	(3160)	(3447)	(2765)	(3496)	(3235)	(3570)	(2615)
Household size	3.91	3.94	3.82	3.91	3.92	3.89	3.74	3.96	3.58
	(1.62)	(1.54)	(1.51)	(1.64)	(1.45)	(1.70)	(1.59)	(1.65)	(1.29)
B. Health									
Diagnosed diabetic (=1)	0.67	0.67	0.68	0.66	0.62	0.68	0.62	0.67	0.59
	(0.47)	(0.47)	(0.47)	(0.47)	(0.49)	(0.47)	(0.49)	(0.47)	(0.50)
Blood sugar index	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(0.02)	(0.03)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)
Hba1c (mmol/mol)	8.69	8.69	8.74	8.69	8.59	8.73	8.68	8.70	8.35
	(2.33)	(2.36)	(2.40)	(2.32)	(2.37)	(2.28)	(2.45)	(2.33)	(2.14)
Random blood sugar (mmol/L)	192.52	191.32	196.07	192.51	195.58	193.26	193.30	192.23	177.38
	(89.44)	(88.73)	(86.67)	(89.87)	(91.54)	(88.25)	(98.14)	(90.42)	(77.00)
Systolic BP (mmHg)	133.37	133.20	134.08	133.35	135.12	133.29	134.05	132.88	135.62
	(19.16)	(20.28)	(17.72)	(19.01)	(21.35)	(19.10)	(19.19)	(18.38)	(21.42)
Diastolic BP (mmHg)	88.49	88.47	88.54	88.49	89.47	88.20	88.51	88.48	90.00
	(11.11)	(11.51)	(10.12)	(11.09)	(12.68)	(10.77)	(10.13)	(11.11)	(13.19)
HbA1c: Diabetic (=1)	0.82	0.82	0.81	0.82	0.77	0.84	0.79	0.81	0.77
	(0.38)	(0.38)	(0.39)	(0.38)	(0.42)	(0.36)	(0.41)	(0.39)	(0.42)
BP: Hypertensive (=1)	0.49	0.46	0.51	0.49	0.53	0.49	0.51	0.49	0.45
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Overweight (=1)	0.61	0.62	0.66	0.60	0.57	0.60	0.57	0.60	0.67
	(0.49)	(0.49)	(0.47)	(0.49)	(0.50)	(0.49)	(0.50)	(0.49)	(0.48)
BMI	26.38	26.43	26.46	26.36	26.46	26.45	26.38	26.24	26.99
	(4.30)	(4.24)	(3.63)	(4.36)	(5.33)	(4.51)	(4.82)	(4.01)	(4.10)
C. Walking - phase-in									
Exceeded step target (=1)	0.25	0.25	0.24	0.25	0.25	0.23	0.27	0.26	0.27
	(0.32)	(0.31)	(0.32)	(0.32)	(0.32)	(0.30)	(0.33)	(0.33)	(0.34)
Average daily steps	7014	7081	6906	7007	7068	6823	7446	7084	7018
	(3983)	(3953)	(3701)	(4015)	(4198)	(3966)	(3869)	(4037)	(4195)
D. Impatience over effort									
Impatience index (SD's)	0.09	0.00	0.05	0.12	0.04	0.14	0.18	0.09	0.26
	(0.99)	(1.00)	(0.89)	(0.99)	(0.95)	(1.05)	(0.91)	(0.97)	(0.91)
Predicted index (SD's)	-0.05	0.00	-0.15	-0.06	-0.09	-0.02	-0.02	-0.08	-0.12
	(1.00)	(1.00)	(0.94)	(1.01)	(1.02)	(1.00)	(1.09)	(1.00)	(0.97)
E. Mobile recharges									
Current mobile balance (INR)	29.34	30.80	29.48	28.98	28.61	29.69	28.55	28.45	30.05
	(49.59)	(48.79)	(48.68)	(49.88)	(38.54)	(52.08)	(63.65)	(47.96)	(36.59)
Yesterday's talk time (INR)	6.58	7.22	6.47	6.44	5.86	6.58	7.67	6.31	4.94
	(8.76)	(10.14)	(8.95)	(8.36)	(6.25)	(8.77)	(9.19)	(8.28)	(5.77)
Prefers daily payment (=1)	0.17	0.18	0.16	0.17	0.20	0.17	0.20	0.17	0.18
	(0.38)	(0.38)	(0.37)	(0.38)	(0.40)	(0.37)	(0.40)	(0.38)	(0.39)
Prefers monthly payment (=1)	0.24	0.25	0.28	0.24	0.27	0.24	0.23	0.24	0.26
	(0.43)	(0.43)	(0.45)	(0.43)	(0.45)	(0.43)	(0.42)	(0.43)	(0.44)
F-tests for joint orthogonality									
P-value (relative to control)	N/A	N/A	0.78	0.41	0.71	0.48	0.41	0.51	0.54
P-value (relative to monitoring)	N/A	0.78	N/A	0.90	0.91	0.85	0.61	0.97	0.66
P-value (relative to base case)	N/A	0.48	0.85	N/A	0.50	N/A	0.77	0.96	0.47
Sample size									
Number of individuals	3,192	585	203	2,404	166	902	164	1,106	66
Percent of sample	100.0	18.3	6.4	75.3	5.2	28.3	5.1	34.6	2.1
Number of ind. with ped. data	2,559	–	200	2,359	163	890	163	1,079	64

Notes: Standard deviations are in parentheses. BMI is body mass index, and BP is blood pressure. Overweight means BMI above 25. Hypertensive means systolic BP above 140 or diastolic BP above 90. The Threshold column pools both the 4-day and 5-day threshold groups. In the incentive and monitoring groups, the number of individuals with pedometer data ("Number of ind. with ped. data") differs from the total number of individuals because a few participants withdrew immediately. The likelihood of immediate withdrawal is not significantly different between incentive and monitoring (p-value > 0.7, Table A.4 column 5).

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 2–4 of Table 1 show means for the pure control, monitoring, and incentive groups, while columns 5–9 show means separately for each incentive subgroup. To explore balance, we jointly test the equality of all characteristics in each of our three "comparison" groups (control, monitoring, and the base case incentive groups—the reference group for all incentive subgroups) with each of the treatment groups. All tests fail to reject the null that all differences are zero. Online Appendix Table H.2 shows covariate balance in the subsample for whom we have post-intervention period data.

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.

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 70% of days. As a result, pedometer wearing rates are high, and the difference between treatment groups is small: 85% in monitoring versus 88% in incentives. However, the difference is statistically significant (Table A.4, column 2). To address the imbalance, we show robustness to Lee (2009) bounds accounting for missing step data due to not wearing pedometers.[31] 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.

[31] We do not have participant pedometer data (e.g., because the pedometer broke or the sync was unsuccessful) on 6% of days. Missing pedometer data is balanced across incentive and monitoring groups (column 3, Table A.4). While our main specifications drop days with missing pedometer data, Table A.5 shows robustness to alternate specifications and Lee bounds. While missing data is balanced overall, one specific source of missing data (mid-intervention withdrawals) is imbalanced (column 6 of Table A.4), but results are robust to Lee bounds accounting specifically for that source (column 5 of Table A.5).

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 daily by each pedometer on the times that the participant put it on and took it off.[32] Reassuringly, Panel B of Table A.6 shows that these times are balanced across groups.

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% (relative to 87% in the intervention period), post-intervention wearing rates are balanced across arms, and our results are robust to a Lee bounds exercise (Online Appendix Tables H.3 and H.4).

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 to participants' homes. In 99.6% of visits, participants were not sharing their pedometers. Second, we check whether participants' minute-wise step counts exceed expectations given their age. This happened very rarely and is balanced across incentive and monitoring groups (Table A.6).

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 of the treatment groups (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 we pre-specified HbA1c as our only blood sugar measure, we had some problems measuring it in the field.[33] As such, we also decided to measure RBS, which is also strongly associated with diabetes severity (Bowen et al., 2015).[34] RBS is much easier to reliably measure in the field. Our measures of RBS and HbA1c both have predictive power for the other.[35] As a result, our preferred measure incorporates both the HbA1c and RBS measurements, but we also present the measures separately as pre-specified.

Since exercise is also associated with improvements in hypertension and cardiovascular health, we measured blood pressure, BMI, and waist circumference as secondary health out-

[32]Specifically, for a subset of days, the pedometers record data on minute-wise (instead of day-wise) step counts, allowing us to back out the first and last minute the pedometer was worn.

[33]The only available measurement tool (the SD A1cCare analyzer from SD Biosensor) was temperature-sensitive and error prone, and its measurements did not line up with lab measurements (the gold standard).

[34]The main downside of RBS as a clinical measure is that it is more sensitive to recent activity such as eating; however, proper glycemic control involves minimizing RBS spikes and so, on average, across the sample, RBS can give us a good measure of the glycemic control of our sample (Dandona, 2017).

[35]Online Appendix Table H.5 shows that baseline RBS has strong predictive power for endline HbA1c in the control group even conditional on baseline HbA1c, and that the reverse is true as well.

comes. 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 the more impatient members of our sample, in line with our theoretical prediction that impatience is a mechanism for its effectiveness. Finally, we find that higher-frequency payments do not increase effectiveness, suggesting that the discount rate over payment is low.

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:

$$y_{it} = \alpha + \beta Incentives_i + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\lambda + \varepsilon_{it}, \tag{11}$$

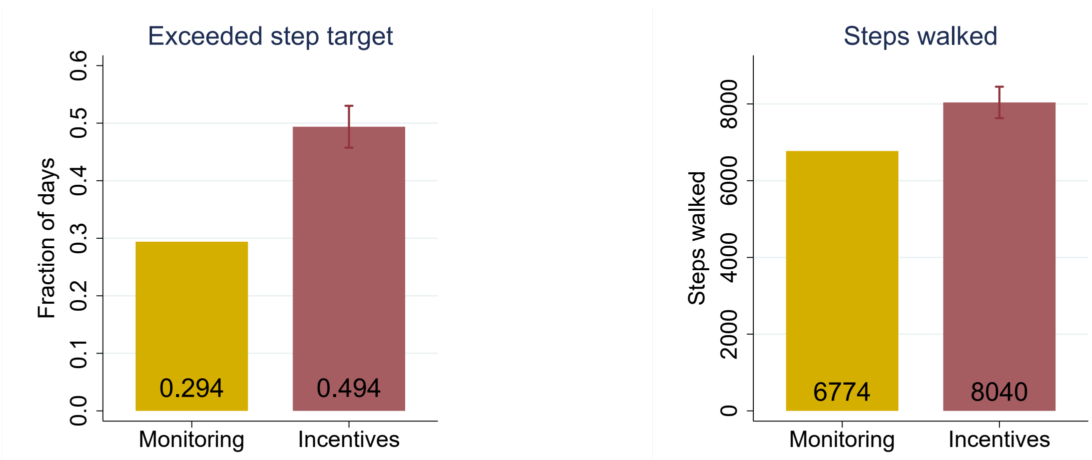
where y_{it} is either individual i's steps on day t during the intervention period or an indicator for individual i surpassing the 10,000-step target on day t; $Incentives_i$ is an indicator for being in the incentive group; and $\boldsymbol{X}_i$ and $\boldsymbol{X}_{it}$ are vectors of individual- and day-level controls, respectively, described in the notes to Table 2. We cluster the standard errors at the individual level. The coefficient of interest, β, is the average treatment effect of incentives relative to monitoring only. Table 2 shows the results. Figure 2 also displays the results graphically.

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 and Figure 2(a)). This effect does not simply reflect participants shifting steps from one day to another: column 2 of Table 2 and Figure 2(b) show 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

Table 2: Incentives Increase Average Walking

Dependent variable:	Exceeded step target	Daily steps	Daily steps (if > 0)
	(1)	(2)	(3)
Incentives	0.200***	1266.0***	1161.5***
	[0.0186]	[208.7]	[188.5]
Monitoring mean	0.294	6,774	7,986
# Individuals	2,559	2,559	2,557
# Observations	205,732	205,732	180,018

Notes: This table shows the treatment effect of incentives (relative to monitoring) on walking. The columns show coefficient estimates from regressions based on equation (11) using intervention-period pedometer data. In column 1, "Exceeded step target" is an indicator variable equal to 1 if the individual exceeded their step target. Individual-level controls are a second order polynomial of age and weight, gender, height, and the average of daily steps during the phase-in period (before randomization). Day-level controls are month-year and day-of-week fixed effects. Online Appendix Table H.6 shows robustness to excluding controls or using controls selected by double-lasso. The sample includes the incentive and monitoring groups. The omitted category in all columns is the monitoring group. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.



(a) Probability Exceeded Step Target

(b) Average Daily Steps

Figure 2: Incentives Increase Average Walking

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

walking each day. This treatment effect is at the high end of effect sizes 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).

This analysis excludes the control group, for whom we have no pedometer data. Because monitoring itself may have a positive impact, these estimates are likely conservative for the overall impact of incentives. That said, a comparison of monitoring group steps between the baseline and intervention periods (controlling for time) suggests that, while monitoring may increase the likelihood of exceeding the step target, it does not increase steps (Online App. L).

The treatment 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 Table A.5 reports Lee bounds which account for the non-random patterns of missing data. Both strategies find similar effects. The estimates are also robust to excluding the control variables from the regression, and to using controls selected by double-Lasso (Online Appendix Table H.6).

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

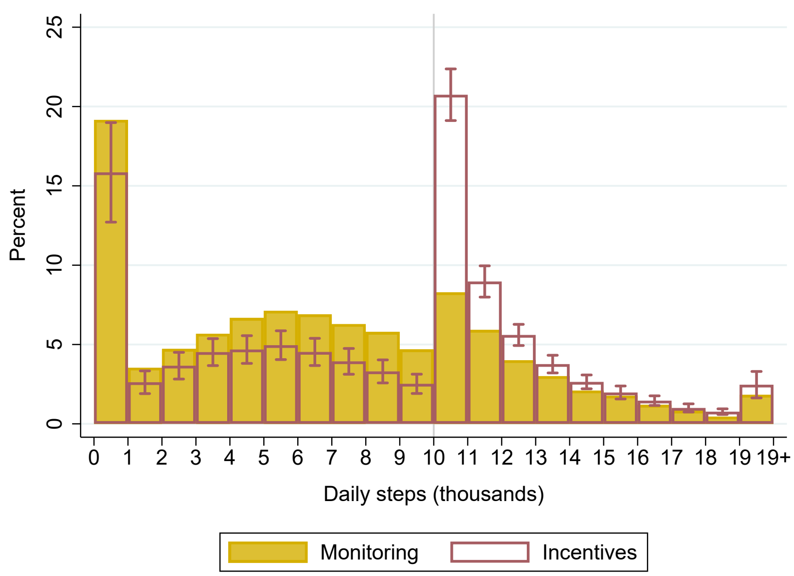


Figure 3: Incentives Shift the Distribution of Steps Walked per Day

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

5.2 Time-bundled Threshold Contracts Increase Average Effectiveness

We begin our analysis of time-bundled thresholds by comparing the sample-average performance of the threshold and linear contracts. Prediction 2 suggests that, when the effort discount rate is sufficiently high, as it may be in our population with chronic disease (e.g., Wainwright et al., 2022), time-bundled threshold contracts tend to be more effective overall than linear contracts.

In order to establish that the time-bundled threshold contracts are effective on average, we can show that they result in weakly more compliance and weakly higher cost-effectiveness than linear contracts in the full sample, with one inequality strict, as described in Section 2. We thus examine compliance 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:

$$y_{it} = \alpha + \sum_j \beta_j \times \left(incentives^j\right)_i + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\theta + \varepsilon_{it}, \tag{12}$$

where y_{it} are daily walking outcomes and $\left(incentives^j\right)_i$ is an indicator for whether individual i is enrolled in incentive treatment group $j \in$ (daily, base case, monthly, threshold, small payment). The β_j coefficients capture the average effect of each incentive treatment group relative to the monitoring group. Table 3 displays the results.

The effect of the threshold treatment on compliance is very similar to the effect of the base case (linear) treatment on compliance, with the estimates within 1.3 pp of each other and the difference not statistically significant (p-value=0.36). Figure 4(a) displays the result graphically. It also shows the 4-day threshold group and 5-day threshold groups separately—neither has meaningfully different compliance than the base case.

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 group only receive payment for exceeding the step target if they do so on at least four or five days in a given week; when they comply on fewer days, they are not rewarded. As shown in Figure 4(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. As a result, the cost-effectiveness of the threshold contracts are 11% and 17% higher than that of the base case contract (Table A.7).

Because the threshold contracts have the same compliance and are more cost-effective than the base case, they are more effective overall. For comparison, the small payment treatment is

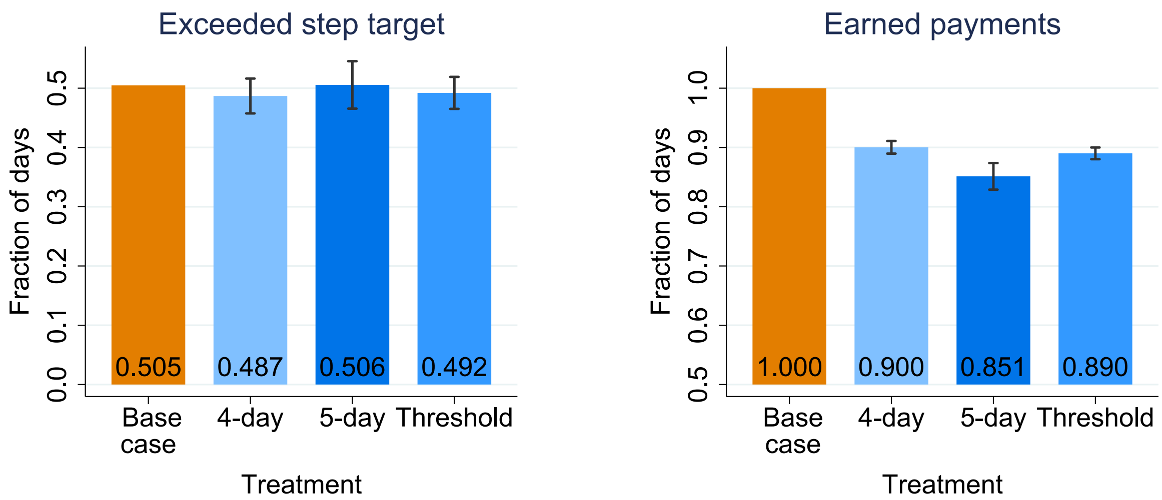
Table 3: All Incentive Contracts Increase Walking

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

Notes: We report incentive effects (relative to monitoring) separately by each incentive treatment group. The columns show coefficient estimates from regressions based on equation (12) using daily intervention-period pedometer data. Each column uses the same controls as in Table 2; the results are robust to excluding controls or using controls selected by double-Lasso (Online Appendix Table H.7). The sample includes the incentive and monitoring groups. The omitted category in all columns is the monitoring group. The Threshold group pools the 4- and 5-day Threshold groups. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

also more cost-effective than the base case (it pays half as much per day complied), but this comes at the cost of reduced compliance, as shown in Table 3. The fact that the threshold contracts achieve the same compliance as the base case for lower cost implies that a budget-neutral threshold (i.e., a threshold contract with the same average cost as the base case) would have higher compliance than the base case.

Variance 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 health benefits of compliance appear approximately linear in our setting (Warburton et al., 2006), there are many settings with nonlinear benefits. In those settings, effectiveness depends



(a) Probability Exceeded Target

(b) Earned Payment When Step Target Met

Figure 4: Thresholds Do Not Affect Average Walking But Increase Cost-Effectiveness

Notes: The figure compares the time-bundled threshold treatments with the base case (linear) incentive treatment, all during the intervention period. Panel A shows the average probability of exceeding the daily step target; Online Appendix Figure H.1 shows the same figure for average daily steps. Panel B shows the fraction of days on which the participants received payments, conditional on meeting the step target. The confidence intervals represent tests of equality between the base case incentive group and each other treatment group. Both panels use the same controls as in Table 2. The Threshold group pools the 4- and 5-day threshold groups. The results for the 4-day and 5-day thresholds are shown in Online Appendix Table H.8.

not just on average compliance but on the variance of compliance levels.

Theory suggests that thresholds could increase the variance of compliance by decreasing intermediate effort just below the threshold (Grant and Green, 2013). This would decrease the effectiveness of thresholds for principals who particularly value compliance improvements among those with low levels of compliance (i.e., principals with concave benefits to compliance). To assess this prediction empirically, panels A and B of Figure A.2 show the treatment effect of the base case contract and of the 4-day and 5-day threshold contracts, respectively, on the share of weeks in which an individual met their step target exactly 0 times, 1 time, etc. All treatment effects are relative to the monitoring group. The threshold contracts decrease effort just below the threshold: the 5-day threshold decreases the prevalence of walking 3 or 4 days relative to either the base case (p-value <0.001) or the monitoring group (p-value $= 0.008$), and the 4-day threshold decreases the prevalence of walking 2 or 3 days relative to either reference group (p-values <0.001 relative to both the base case and monitoring groups).[36]

[36]One may be surprised that neither threshold increases the likelihood of walking exactly the threshold number of days (e.g., 4 days for the 4-day threshold). Within the context of our model, this may partly reflect that the contracts pay on the margin for above-threshold compliance (e.g., the 4-day threshold paid for the 5th day of compliance), which reduces heaping at the threshold level. 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.

However, the magnitude of these differences are relatively small, leading to only relatively small differences between the base case and threshold contracts in the overall distribution of weekly compliance, as shown in Figure A.3. Specifically, the figure shows quantile regression coefficients of the effects of the base case and of the 4-day and 5-day threshold contracts, respectively, on the percentiles of the distribution of weekly compliance, relative to the monitoring group. While there are some differences, the threshold and base case contracts have relatively similar patterns of impacts across the distribution, with both impacts peaking around the 70th percentile of the distribution. Figure A.4 shows similar results for the distribution of individual-level (instead of individual×week-level) compliance. The differences between the base case and threshold quantile regression coefficients are significant at roughly the rate that would be expected due to chance (i.e., 3% of coefficients are significant at the 5% level and 12% at the 10% level, see Online Appendix Table H.9). Overall, the relatively minor distributional differences between threshold and linear contracts imply that thresholds would be the preferred option even by many principals with concave benefit functions, provided their benefit functions are not too concave.

5.3 Mechanisms: Impatience Over Effort 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 link, as we show that the threshold is more effective for more impatient individuals. Specifically, relative to the base case, the threshold 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 find that the estimated heterogeneity is robust to controlling for many covariates interacted with the threshold, suggesting that this condition holds. We also use machine learning to demonstrate the importance of impatience for predicting the treatment effect of thresholds on compliance, conditional on other covariates. Moreover, to tie our data to our theory more precisely, we present a model calibration that further bolsters the link between discount rates and threshold compliance given the parameters in our particular setting.

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

$$y_{it} = \alpha + \beta_1 Impatience_i \times Thresh_i + \beta_2 Thresh_i + \beta_3 Impatience_i + \boldsymbol{X}'_i \pi + \boldsymbol{X}'_{it} \theta + \varepsilon_{it}, \quad (13)$$

where y_{it} is an indicator for whether individual i exceeded the 10,000-step target on day t and $Thresh_i$ is an indicator for being in the threshold group. 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 β_1, which captures how the effect of the threshold (relative to the base case) varies with impatience. Our prediction is that $\beta_1 > 0$.

Table 4: Time-Bundled Thresholds Increase Compliance More for the Impatient

Dependent variable:	Exceeded step target (×100)			
Impatience measure:	Impatience index	Above median impatience index	Predicted impatience index	Above median predicted index
Sample:	Late	Late	Full	Full
	(1)	(2)	(3)	(4)
Impatience × Threshold	3.80^{**}	5.97^{*}	3.12^{***}	5.94^{**}
	[0.57, 7.03]	[-0.86,12.81]	[0.89, 5.14]	[0.20, 9.59]
Threshold	-1.30	-3.81	-1.18	-3.41^{**}
	[-4.36, 1.76]	[-8.89,1.28]	[-3.33, 0.75]	[-5.95, -0.58]
Impatience	-2.97^{**}	-4.68^{*}	-2.38^{***}	-5.3^{**}
	[-5.36, -0.57]	[-9.46,0.10]	[-3.84, -0.75]	[-8.03, -0.97]
# Individuals	1,075	1,075	1,969	1,969
# Observations	86,215	86,215	157,946	157,946
Base case mean	50.4	50.4	50.2	50.2

Notes: This table shows heterogeneity by impatience in the effect of threshold contracts relative to linear contracts. The impatience measure changes across columns; its units in columns 1 and 3 are standard deviations. The sample includes the base case and threshold incentive groups only. The "Late" sample includes only participants who were enrolled after we started measuring the impatience index; the Full sample includes everyone. The Threshold group pools the 4- and 5-day threshold groups. See Online Appendix Table H.11, Panel B for results with the Threshold group disaggregated (unpooled). (Panel A of that table shows results using daily steps as the outcome.) Bootstrap draws were clustered at the individual level, and bootstrapped 95% confidence intervals are in brackets. For the regressions that use the predicted impatience index, to construct the 95% confidence interval, we conduct three steps in each bootstrap sample: 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 (13). The Gaussian standard errors and p-values for the column 1 $Impatience \times Threshold$ coefficient are 1.9 and 0.046, respectively; for column 2, the corresponding values are 3.78 and 0.114. Controls are the same as in Table 2. Significance levels: * 10%, ** 5%, *** 1%.

Table 4 shows that, consistent with the theory, 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 the threshold group relative to the linear group by 4 pp (statistically significant at the 5% level). Column 2 uses a dummy for having an above-median value of the impatience measure. We include this estimate because it is easier to interpret, although it has lower statistical power since it does not leverage all the underlying variation in the data. Relative to the base case, the threshold generates 6 pp higher compliance for those with above-median impatience than those with below-median (p-value < 0.10). This represents a large increase relative to the sample-average effect of either contract (20 pp). Recall that we only have the impatience index for the sample enrolled later in the experiment; to 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-value < 0.01 and < 0.05 in columns 3 and 4, respectively.

Figure 5 presents a visualization of column 4; it shows that, relative to the linear contract, the threshold contract increases compliance among the more impatient while decreasing it among the less impatient. The difference between the effects is the significant 6 pp effect.

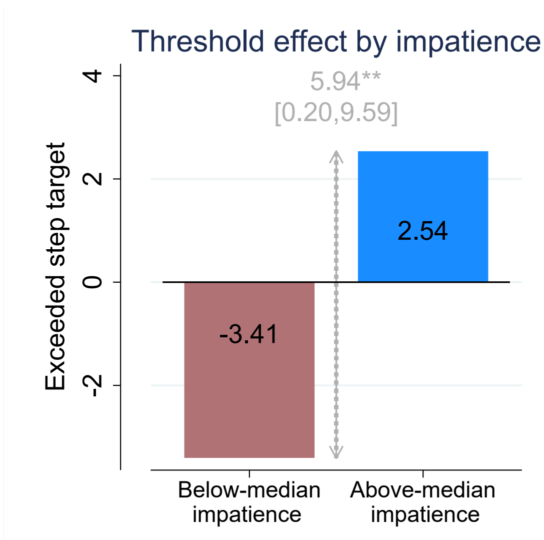
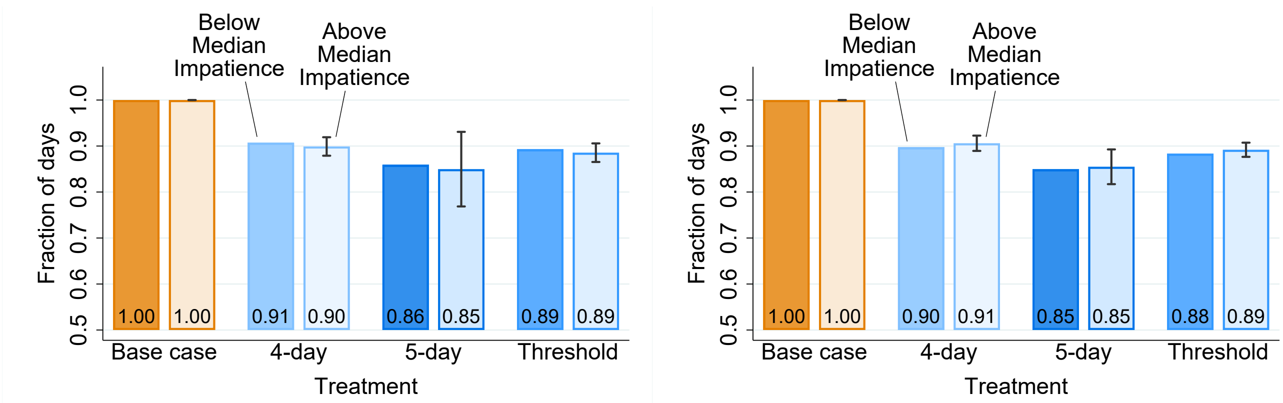


Figure 5: Impatience is Pivotal to Compliance Under the Time-Bundled Threshold

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

Cost-Effectiveness and Effectiveness Prediction 1 suggested that, in addition to increasing compliance more among the impatient, threshold contracts should also increase *effectiveness* more among the impatient. Since we have already established the compliance result, to demonstrate the effectiveness result, it is sufficient to show that, relative to the base case, thresholds do not increase cost-effectiveness more among the patient than the impatient. Figure 6 (as well as Online Appendix Table H.12) show that this is true. Paired with the compliance result, this implies that the threshold increases effectiveness more for those with higher impatience than lower impatience.



(a) Actual Impatience

(b) Predicted Impatience

Figure 6: Thresholds Are Similarly Cost-Effective For Those with Lower and Higher Impatience

Notes: The figures show the fraction of days on which the participants in the base case and threshold groups received payments, conditional on meeting the step target, in different subsamples. In each treatment group, the left bar is the below-median impatience sample and the right bar is the above-median impatience sample; Panel A splits the sample using the actual impatience index while panel B uses the predicted impatience index. The confidence intervals represent bootstrapped tests of equality between the above/below median groups within each treatment group, using the same bootstrap procedure described in the notes to Table 4. Both panels use the same controls as in Table 2. Threshold pools the 4- and 5-day threshold groups. Results in table form are in Online Appendix Table H.12

It is also important to understand whether the threshold has higher or lower effectiveness than the linear contract for each group. 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 the base case, the 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 (λ from Section 2).

Robustness of the Compliance Heterogeneity Results 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 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 7 examines the robustness of the Table 4 estimates to controlling for other baseline covariates and their interactions with the 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), which could both influence the performance of threshold contracts, among other controls.

Reassuringly, Figure 7 shows that the coefficients on the interaction of impatience and the threshold remain stable as we add these additional controls. Panel A shows stability of the coefficient from column (1) using the actual impatience index as the measure of impatience, and Panel B shows stability of the coefficient from column (3) using the predicted impatience index. The stability of both coefficients suggests that it is likely impatience itself (and not its correlates) driving the estimated relationships.

Moreover, even if omitted variables were affecting our Table 4 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.

Another potential confound that was difficult to measure at baseline (and hence which we do not control for) is the individual-level propensity for habit formation. However, we can measure the propensity for forming habits at endline by assessing how much of the treatment effect of incentives persists after payments stop. Table H.13 in the Online Appendix reassuringly suggests that the propensity to form habits is not correlated with impatience in our setting, as impatience does not predict the persistence of incentive effects after payments stop.

Compliance Machine Learning Results A machine learning methodology also shows that the impatience index has particular significance relative to alternative covariates in predicting the effect of the Threshold. Specifically, we model the individual-level treatment effects of the Threshold using a causal forest (Wager and Athey, 2018; Athey et al., 2019). Figure 8 shows that the impatience index is one of the most important predictors among the controls in Figure 7, where importance is a weighted sum of the number of splits of the causal forest at each depth. This suggests that impatience has a particularly strong signal in predicting the impact of the Threshold.

Finally, we use an alternative, machine learning based approach to heterogeneity analysis

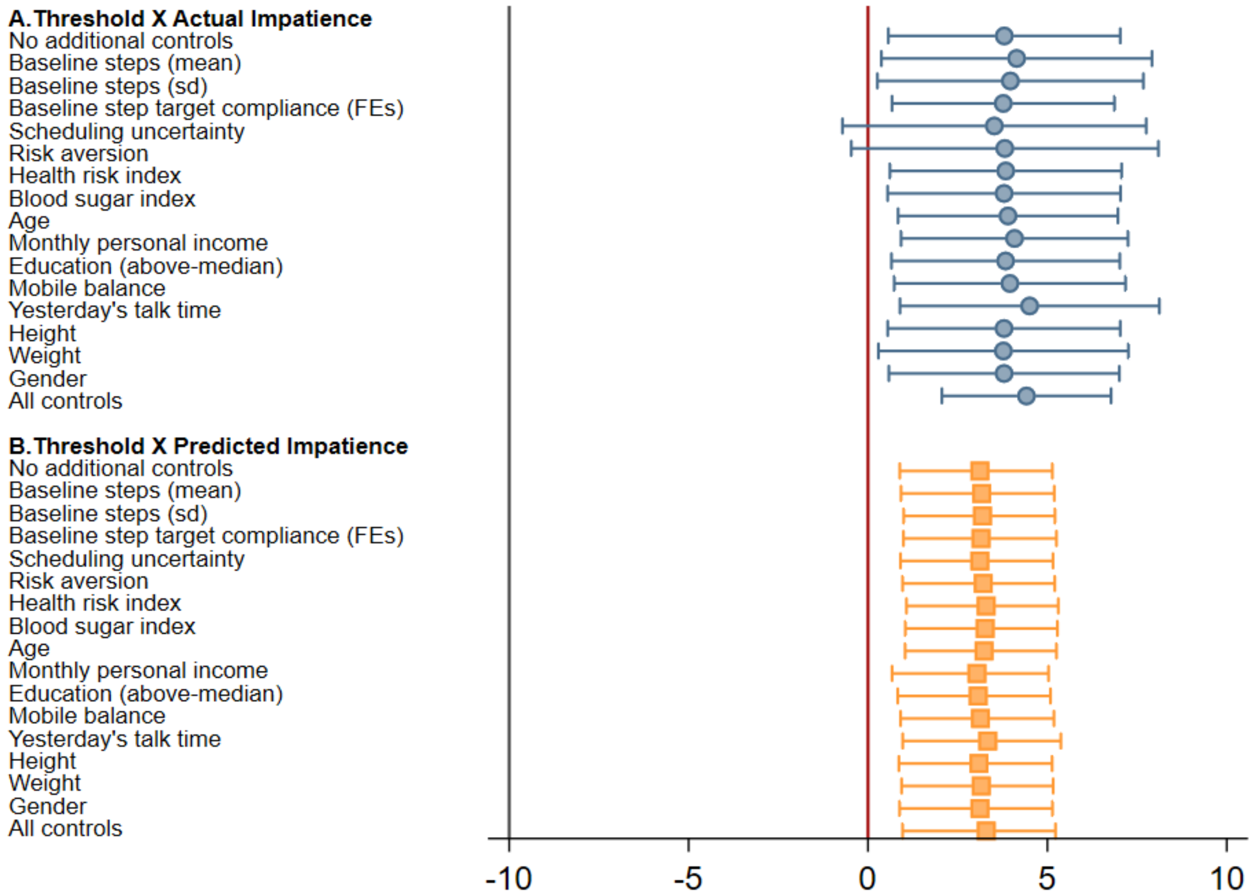


Figure 7: Threshold Heterogeneity by Impatience is Robust to a Variety of Controls

Notes: Panel A displays robustness of the $Threshold \times Impatience$ coefficient from column (1) of Table 4 to including various additional controls, interacted with $Threshold$, in the regression. As a reference, the first "No additional controls" row just displays the $Threshold \times Impatience$ coefficient, and 95% confidence interval, from column (1) of Table 4. The next 15 rows show estimates of the $Threshold \times Impatience$ coefficient from the same regression model, each estimated with two additional controls: a control for the main effect of the covariate listed in the row title, and a control for that same 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 15 rows (both main effects and interactions with $Threshold$). Panel B is analogous but based on column (3) of Table 4. Thus, Panel A shows robustness of the $Threshold \times Impatience$ coefficient when the actual impatience index is the measure of $Impatience$ whereas Panel B shows robustness when the predicted impatience index is the measure of $Impatience$. Baseline steps (mean) and baseline steps (sd) represent the mean and standard deviation, respectively, 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. Risk aversion is an incentivized measure from a multiple price list. Scheduling uncertainty represents the individual's stated frequency of facing unexpected events (such as business duties) that would prevent them from walking for 30 minutes in a given day. Income is winsorized at the 5th and 95th percentiles. The unit of observation is a respondent × day. All confidence intervals are constructed via bootstrap, with bootstrap draws done at the individual level, as in Table 4.

that is robust to multiple hypothesis testing and overfitting concerns and show that this approach also identifies impatience as a significant predictor of threshold effectiveness. We follow

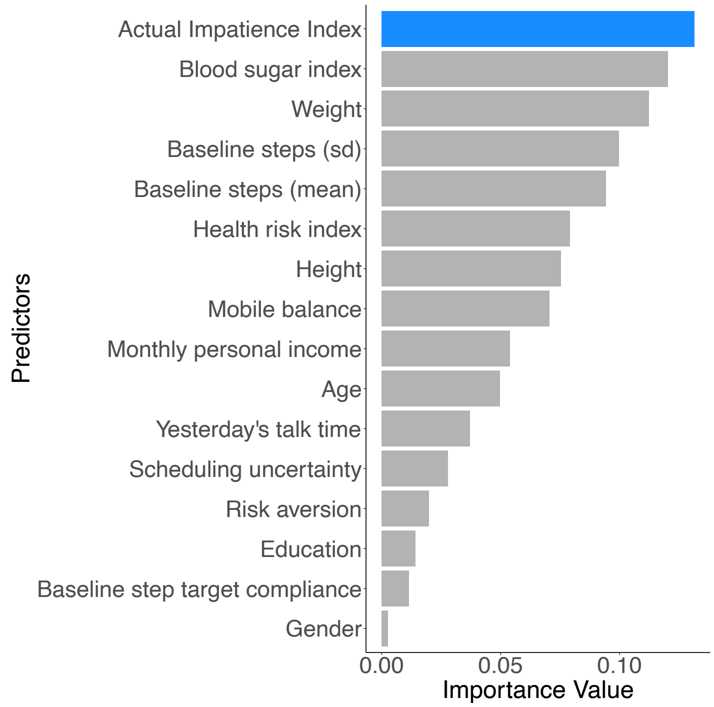
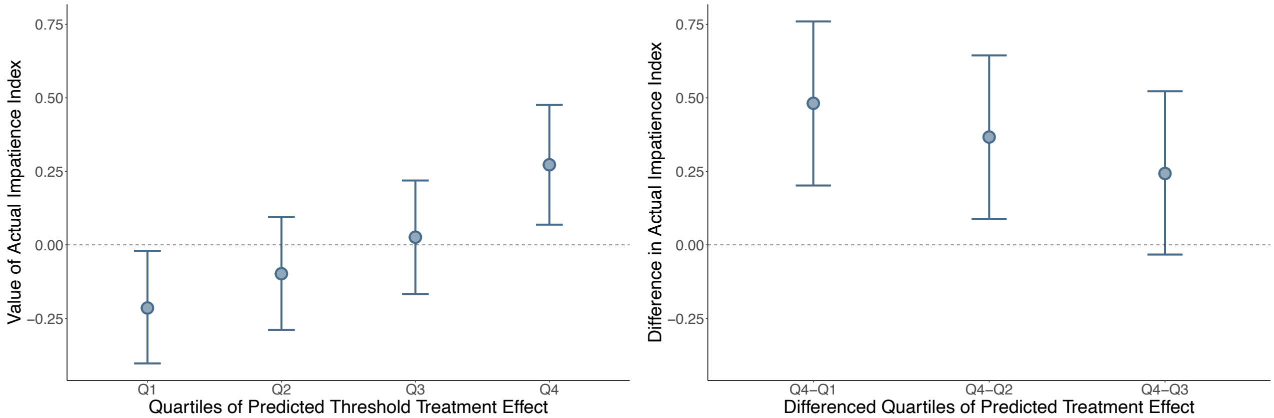


Figure 8: The Causal Forest Selects the Impatience Index as an Important Predictor of the Threshold Effect

Notes: This figure displays the variable importance for each predictor included in a causal forest prediction of the threshold treatment effect on average daily steps at the individual level. Variable importance is a weighted sum of the number of splits of the causal forest at each depth. We limit the sample to those from whom we collected the impatience index. Predictors include the actual impatience index and controls shown in Panel A of Figure 7, except that this analysis uses continuous versions of the baseline compliance and education variables, instead of indicator variables for each value and for being above median, respectively, as the importance analysis more naturally handles continuous variables. Missing values of predictor variables are imputed with the treatment-group mean, and then we include in the predictor list a vector indicators for whether each variable is missing (each of which the analysis assigned importance values of 0, and hence which we do not depict for brevity). We implement the Causal Forest using the GRF package in R (Tibshirani et al., 2023).

the classification analysis methodology of Chernozhukov et al. (2018) to split our sample into quartiles of predicted treatment effects of the threshold relative to the base case, and we compare the impatience levels of the most and least "affected" groups (i.e., the groups with the highest and lowest predicted treatment effects). The results are shown in Figure 9: the average level of the impatience index is increasing in the predicted average threshold effect (Panel A), and the most affected quartile has impatience levels that are a statistically significant 0.5 standard deviations above the least affected quartile (Panel B). Note that our use of the Chernozhukov et al. (2018) approach to inference is not strictly necessary: the method is primarily designed to avoid overfitting when researchers do not prespecify the dimension for heterogeneity analysis. In our case, we prespecified impatience as the critical dimension of heterogeneity. Nevertheless, we find it reassuring that even this more robust method identifies impatience as



(a) Average Impatience Increases with Predicted Threshold Treatment Effect

(b) Impatience Significantly Higher in Highest vs. Lowest Quartiles of Predicted Threshold Effect

Figure 9: Classification Analysis Shows Impatience Varies With Predicted Threshold Effect

Notes: The figure displays the heterogeneity in the actual impatience index across quartiles of predicted conditional average treatment effects of the threshold. Panel A shows the value of the actual impatience index in each quartile of predicted threshold treatment effects with 95% confidence intervals. Panel B shows the difference in the actual treatment effect between the most-affected quartile and the other three quartiles. Confidence interval bars represent tests for equality between the compared groups at the 95% confidence level. 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 7 and indicators for whether each variable is missing.

a significant predictor of the threshold effect. Figure A.5 shows qualitatively similar, although less statistically robust, results for the predicted impatience index.

Compliance Model Calibration To tie our data to our theory more precisely and truly hold all other factors constant when analyzing the effect of impatience, in Appendix D we calibrate a model to determine whether the gap in predicted compliance between the threshold and linear contracts varies meaningfully with the discount rate over effort. We find that it does: projected compliance in the most effective time-bundled contract increases by 3 pp relative to the linear contract for each 10 pp decrease in the discount factor.

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

These findings have important policy implications, suggesting that time-bundled thresholds are a useful tool to adjust incentives for impatience over effort. Policymakers could tailor time-bundled thresholds at the population level, using them when incentivizing groups known for greater impatience, such as those with chronic disease (Wainwright et al., 2022) or younger individuals (Read and Read, 2004). They could also personalize the assignment of time-bundled thresholds within a population, for example by assigning them based on observable measures of impatience. Such personalization is likely feasible: Andreoni et al. (2018) use discount rates estimated through a simple effort allocation experiment to successfully personalize incentive contracts with the goal of equalizing worker effort across days.

While such an assignment mechanism might give participants incentive to misreport their discount rates, Appendix C demonstrates that it is possible to predict impatience — and, critically, heterogeneity in the threshold effect — using characteristics that are easily observable and more difficult to manipulate (e.g., BMI and gender). Policymakers could hence assign contracts based on these predictors of impatience rather than impatience itself. Moreover, manipulation of impatience measures may not be a significant concern for two reasons. First, although for simplicity our experiment compared a linear contract to a time-bundled threshold contract that was financially dominated, the optimal time-bundled threshold may in many cases not be financially dominated by the optimal linear contract.[37] Using sets of contracts where neither financially dominates the other may decrease the incentive for manipulation. Second, evidence from a nearly identical setting to this experiment's shows that, in the domain of incentives for behavior change, participants tend not to manipulate their observables to avoid assignment to financially dominated contracts (Dizon-Ross and Zucker, 2022).[38]

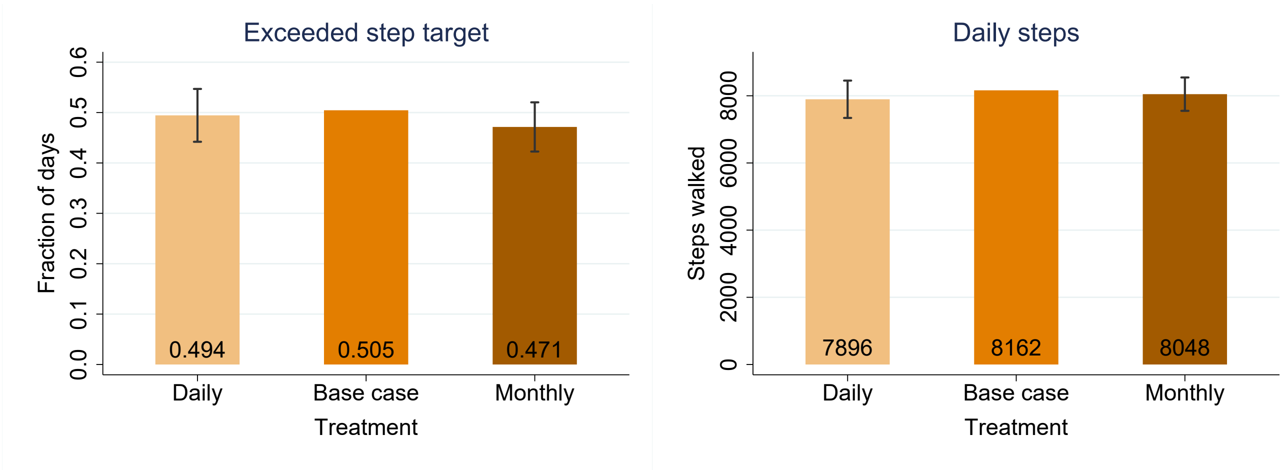
5.4 Payment Frequency Does Not Meaningfully Affect Effectiveness

To understand the role of payment frequency and the discount rate over financial payments in incentive design, we compare average compliance in the daily, weekly (base case), and monthly groups. Figure 10 and Table 3 both 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 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 11 shows that walking behavior is remarkably steady across the payment cycle. The estimates here are more precise,

[37] e.g., the optimal threshold might pay \$100 for complying every day in a week whereas the optimal linear could pay \$10 per day of compliance.

[38] The reason appears to be that participants recognize that such manipulation might not benefit their health.

allowing us to rule out even small effects of payment discount rates on compliance.[39]



(a) Probability Exceeded Step Target

(b) Average Daily Steps

Figure 10: Payment Frequency Does Not Significantly Impact Walking

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

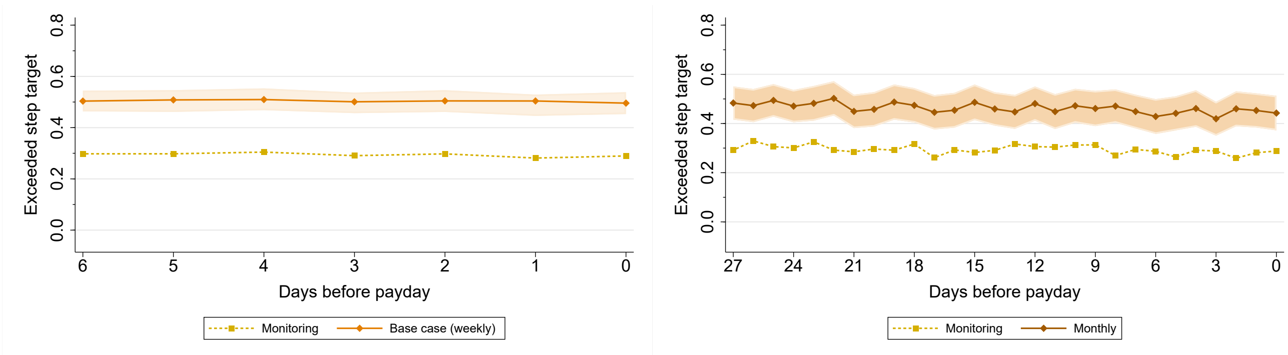
While it is possible that people would have been more impatient over payments delivered with a different modality, limited impatience over payments is not rare (Augenblick et al., 2015; Andreoni and Sprenger, 2012).[40] Thus, increasing payment frequency may not always be an effective way to adjust incentives for impatience.

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 concepts 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

[39]Specifically, Online Appendix Table H.14 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 not significantly different from zero, and the confidence intervals are tight, allowing us to rule out even small effects. For example, if we assume linearity of compliance in lag to payment, then the confidence interval around the slope in the weekly treatment rules out the possibility that, because of monetary discounting, daily payments would generate a mere 0.3 pp more compliance than weekly.

[40]Augenblick et al. (2015) and Andreoni and Sprenger (2012) find limited impatience over payments made in the US via cash and via check, respectively. That said, Kaur et al. (2015) find evidence of payday spikes for payments made to data entry workers in Mysore India, highlighting that payment discount rates may vary.



(a) Weekly Payment Cycle

(b) Monthly Payment Cycle

Figure 11: The Probability of Exceeding the Step Target Is Stable over the Payment Cycle

Notes: The figures show the probability of exceeding the daily 10,000-step target among individuals receiving the base case (i.e., weekly) incentive (Panel A) and a monthly incentive (Panel B) relative to the monitoring group, according to days remaining until payday. Effects control for payday day-of-week fixed effects, day-of-week fixed effects, day-of-week relative to survey day-of-week fixed effects, and the 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.

behavior outweighs the marginal social cost in the "base case" version of a program, then variations that increase compliance and effectiveness have high potential to increase social welfare. This is likely the case here since the estimated social benefits of walking are large relative to the private costs and incentive amounts (Reiner et al., 2013).

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 naïve individuals who pay upfront for commitment but fail to follow through (e.g., Bai et al., 2020).

Even though individuals do not pay upfront for threshold contracts, there is a potentially analogous issue. Naïfs 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, as described in Section 2.3, there are theoretical reasons to doubt that this would happen much in practice.[41] Two pieces of empirical evidence also suggest that our threshold contract did not reduce participants' welfare. First, at endline, we asked participants whether they were interested in continuing the program. The vast majority said that they were, with no significant difference between the threshold group and other groups and, within the threshold group, no significant difference between the more and less impatient (Online App. Table H.15).

[41] 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 Section 2.3 and especially footnote 16 for more detail.

Second, impatient people are no more likely (and in fact are less likely) than patient people to comply and *not* be paid for it under threshold contracts (Online App. Table H.16).

It is important to note that Predictions 1 and 2 only hold for the specific types of time-bundled contracts highlighted in Section 2 (e.g., threshold contracts as defined in equation 4). The results for contracts that are time-bundled but take a different form can differ.[42]

6 Empirical Results: Program Evaluation

The impacts of an incentive program on health and behavior are of policy interest, especially among a population like ours that has a high risk of complications from noncommunicable disease. This section delves into the impact of incentives on exercise patterns and health. We first examine how our exercise impacts changed over time, both during and after the intervention. We then show that the program improved cardiovascular and mental health. Finally, we interpret the overall behavioral and health impacts in light of the literature on related programs.

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

Since chronic disease management requires ongoing lifestyle changes, it is critical to find programs that can lead to sustained improvements in exercise. In light of this, we analyze the evolution of exercise impacts over time, beginning with their evolution during the intervention. Panels A and B of Figure A.7 estimate equation (11) separately by week of the intervention. After an initial spike at week 1, the effect of incentives on walking remains stable during the full intervention period. This suggests that policymakers could extend the program further with similar effects, an encouraging finding as insurers and governments are increasingly rolling out longer-term (and even permanent) incentive programs.

Do the effects of incentives also persist after the payments stop? Studies of similar exercise programs find mixed results (e.g., Royer et al., 2015; Charness and Gneezy, 2009). To examine persistence, we estimate equation (11) using the pedometer data from the 12 weeks after the intervention ended. While we have pedometer data from the control group during this period,

[42] For example, Carrera et al. (2022) analyze a contract that pays a fixed amount for compliance that exceeds a given threshold and show that it decreases welfare for present-biased participants relative to a linear contract. However, their contract differs in two important ways from the contracts we analyze. First, their contract uses a relatively low threshold level (K/T level in our Section 2 terminology). In their main treatment, participants only need to comply on 43% of periods, or 12 out of 28, to receive payment. As highlighted in our Section 2 discussion, Prediction 1 only holds when K/T is sufficiently high. Low K/T can lead to procrastination among naifs. Second, the Carrera et al. (2022) contract does not pay "on the margin" for any compliance strictly exceeding the threshold. That is, its payment function is $m_T = M\mathbb{1}\{\sum_{t=1}^{T} w_t \geq K\}$ (for some constant M) instead of our payment function of $m_T = m' \sum_{t=1}^{T} w_t \mathbb{1}\{\sum_{t=1}^{T} w_t \geq K\}$. Not paying on the margin increases the motivation for participants to wait until later periods to comply, since the participants do not want to accidentally "overshoot" and comply above the threshold level K. Waiting until later can decrease welfare if those later periods turn out to have higher cost realizations.

sample size is limited: we only collected post-intervention period data from a third of our sample. We pool the comparison groups for power, so the *Incentives* coefficient represents the effect of incentives relative to the control and monitoring groups pooled. [43]

Table 5 shows that the incentive group walks significantly more then the comparison groups even after incentives end. The treatment effect on steps is statistically significant and large: around 10% of the comparison group mean (columns 2 and 3). For comparison, the treatment effect of incentives relative to monitoring during the intervention period was 20% of the monitoring group mean. Hence, a meaningful portion of the treatment effect appears to have persisted.[44] Panels C and D of Figure A.7 suggest that increases in walking persisted 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.

Table 5: The Effects of Incentives Persist After the Intervention Ends

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

Note: This table shows the average treatment effect of incentives relative to the control and monitoring groups (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. Online Appendix Table H.17 shows that the results are robust to excluding controls and using controls selected by double Lasso. The number of individuals differs from the total number of individuals recruited for the post-intervention period because roughly 10% of participants withdrew immediately. The likelihood of immediate withdrawal is not significantly different between the incentive and comparison groups (Online Appendix Table H.3 column 5), and Online Appendix Table H.4 shows that the results are robust to a Lee bounds exercise. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

[43] The results are similar when we compare incentives with control alone (Online Appendix Figure H.2); with only 72 people, the post-intervention monitoring group is too small to analyze alone.

[44] Note that we are comparing the effect of incentives relative to control in the post-intervention period with the effect of incentives relative to *monitoring* in the intervention period. While this comparison overstates the degree of persistence if there is a positive effect of monitoring on steps, Online Appendix L suggests that monitoring does not affect steps.

6.2 Incentives Moderately Improve Health

We now examine whether the incentives program measurably improves health. Our experiment was powered to detect the difference between the incentive and pure control groups. While we lack statistical power to compare health outcomes (which are relatively noisy) in the monitoring group with the other groups, we show it for completeness. Table 6 reports results from regressions of the following form:

$$y_i = \alpha + \beta_1 Incentives_i + \beta_2 Monitoring_i + \boldsymbol{X}_i'\gamma + \varepsilon_i, \tag{14}$$

where y_i is an endline health outcome for individual i and $\boldsymbol{X}_i$ is a vector of controls (shown in the table notes). β_1 represents the effect of incentives relative to the control group.

Table 6 shows that the incentive program moderately improves blood sugar and cardiovascular health. Column 1 presents the treatment effect on our preferred blood sugar measure, the standardized index incorporating both the HbA1c and RBS measures of blood sugar control. 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 the overall health risk index by 0.05 standard deviations, significant at the 10% level.

Since health outcomes among those with more severe diabetes might be more responsive to exercise, our *ex ante* analysis plan included an analysis of the health impacts separately among those with higher blood sugar. We thus estimate the following regression:

$$\begin{aligned} y_i =& \alpha + \beta_1 Incentives_i + \beta_2 Incentives_i \times LowBloodSugar_i + \beta_3 Monitoring_i \\ &+ \beta_4 Monitoring_i \times LowBloodSugar_i + \beta_5 LowBloodSugar_i + \boldsymbol{X}_i'\gamma + \varepsilon_i, \end{aligned} \tag{15}$$

with $LowBloodSugar_i$ an indicator for having below-median baseline values of the blood sugar index (i.e., less severe diabetes). β_1 is the coefficient of interest, telling us the treatment effect of incentives among those with above-median blood sugar (i.e., with $LowBloodSugar = 0$). β_2 further allows us to test if the effect is significantly different among those with above- and below-median baseline blood sugar.

The results, shown in columns (5)-(8) of Table 6, indicate that the program improves health more among those with more severe diabetes, although the differences from those with less severe diabetes are not statistically significant. Among the sample with above-median blood sugar, incentives decrease the blood sugar index by 0.09 standard deviations and decrease RBS by 12 mg/DL, both significant at the 5% level.

Both the full sample and subsample treatment effects on blood sugar are moderately-sized but meaningful from a clinical perspective.[45] In addition, an exploratory analysis of the treat-

[45] For example, to interpret the RBS result, note that, for RBS measured in the morning, a value of less than

Table 6: Incentives Moderately Improve Blood Sugar and Cardiovascular Health

	Full sample effects				Heterogeneity by baseline blood sugar			
Dependent variable:	Blood sugar index (1)	HbA1c (2)	Random blood sugar (3)	Health risk index (4)	Blood sugar index (5)	HbA1c (6)	Random blood sugar (7)	Health risk index (8)
Incentives	-0.05*	-0.07	-6.1*	-0.05*	-0.09**	-0.1	-11.8**	-0.08**
	[0.03]	[0.07]	[3.5]	[0.02]	[0.05]	[0.1]	[5.9]	[0.04]
Incentives × below-median blood sugar					0.09	0.1	11.3	0.07
					[0.05]	[0.1]	[7.0]	[0.05]
Monitoring	-0.03	-0.1	1.8	0.01	-0.05	-0.3*	1.3	-0.05
	[0.05]	[0.1]	[6.6]	[0.04]	[0.08]	[0.2]	[10.5]	[0.07]
Monitoring × below-median blood sugar					0.05	0.3	-0.7	0.1
					[0.09]	[0.2]	[12.6]	[0.09]
p-value: I = M†	0.573	0.534	0.188	0.138	0.539	0.278	0.160	0.600
Control mean‡	0.0	8.4	193.8	0.0	0.6	10.1	248.3	0.5
# Individuals	3,067	3,066	3,067	3,068	3,067	3,066	3,067	3,068

Notes: † Incentives = Monitoring. ‡ In columns 1-4 we report means of the full control group and in columns 5-8 we report means of control individuals with above-median values of the baseline blood sugar index.
Observations are at the individual-level. Columns 1-4 display OLS estimates of equation (14). Columns 5-8 display display OLS estimates of equation (15); note that $LowBloodSugar_i$ in equation (15), which is the indicator for having below-median baseline values of the blood sugar index (i.e., less severe diabetes), is labeled as "below-median blood sugar" in the table. (Online App. Table H.18 shows that the estimates are nearly quantitatively identical if we analyze heterogeneity based on baseline HbA1c instead of the baseline blood sugar index, and Online App. Table H.19 shows that we reach similar conclusions, particularly for the high blood sugar sample, when, instead of using OLS to analyze the treatment effects, we use an instrumental variables analysis, using the dummies for each of the different incentive subtreatments as instruments for intervention-period steps.) HbA1c is the average plasma glucose concentration (%). Random blood sugar is the blood glucose level (mg/dL). The blood sugar index is constructed by taking the mean of endline HbA1c and random blood sugar 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, random blood sugar, mean arterial blood pressure, body mass index, and waist circumference standardized by their average and standard deviation in the control group. See Online Appendix Table H.20 for treatment effects on the components of the index 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 squared (or index components squared for indices), a dummy for the SMS treatment, and the following controls: age, weight, height, gender, and their second-order polynomials, as well as endline completion date, month-year and day-of-week fixed effects for endline completion dates, hour of endline completion, and dummy for late completion. Columns 5-8 additionally control for the indicator for below-median blood sugar. Online Appendix Table H.21 shows that the estimates are similar, just less precise, when we omit the control variables from the regressions or use controls selected by double-Lasso. Robust standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

ment effects on RBS suggests that the effects may amplify over time. Specifically, since we measured RBS (but not HbA1c) every 3 weeks during the intervention period, we can track how the RBS treatment effects evolve. Figure A.8 shows that the treatment effect of incentives increases at each subsequent measurement. This suggests that the effects might continue to grow if either the program were extended or (as we show) the exercise effects persist.

100 mg/dl would be normal, values of 100-125 mg/dl would indicate prediabetes, while values above 126 mg/dl indicate diabetes. Thus an improvement of 6 or 12 mg/dl would bring someone near the diabetes threshold either a quarter or half of the way towards normal (healthy) blood sugar.

Table A.8 examines whether the intervention had coincident impacts on mental health or fitness. Incentives improve the mental health index by 0.09 SD. In contrast, we find no effects on physical fitness, perhaps because we could only 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 H.22).

6.3 Incentives and Chronic Disease: Results Summary and Discussion

Overall, these results are promising from a policy perspective. The exercise results show that incentives substantially increase exercise throughout the entire intervention period. Some of the effect even persists after the intervention period ends. Exercise has important long-run health benefits for diabetics (Praet and van Loon, 2009; Qiu et al., 2014; Lee et al., 2012), and even in the short run we find that incentives translate to meaningful improvements in blood sugar, cardiovascular health, and mental health.

Our work thus provides a rare example of an effective lifestyle intervention that can be scaled in resource-poor settings with limited health infrastructure. Interventions previously shown to improve exercise among diabetics and prediabetics have required highly trained staff to engage in frequent and personally-tailored interactions with participants (Aziz et al., 2015; Qiu et al., 2014), and hence have had limited scalability. Developing scalable approaches to promote exercise among those with diabetes and other chronic diseases is a crucial policy priority.

Our intervention is scalable and relatively low-cost. The per-person program cost of the incentive program is 1,700 INR or 26 USD. That is equal to just 7% of the estimated annual direct cost of care for a diabetic in Tamil Nadu, or 21% of the direct cost of care during the 3-month intervention period (Tharkar et al., 2010). Interventions generating similar levels of exercise among diabetics in other contexts have produced cost savings of at least the same order of magnitude, even without effects that persist like we find (Nguyen et al., 2007, 2008). Thus, incentive programs could be an important tool to help decrease the burden of chronic disease.

7 Conclusion

This paper makes two important contributions. First, we provide new insights into how to adjust incentives for impatience. We show both theoretically and empirically that, relative to linear contracts, the performance of time-bundled contracts is 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 naïve, thus broadening the arsenal for motivating impatient or time-inconsistent individuals. The intuition behind the prediction is that people who discount their future effort more place a higher value on future work opportunities. Time-bundled contracts link better future work opportunities with effort today, thus providing particular motivation for the impatient to exert more effort today. The success of the time-bundled contract in improving performance in the face of impatience is particularly notable when contrasted with the failure in our sample of the conventional strategy for adjusting incentives for impatience: higher-frequency payment. To be effective, more frequent payment requires people to be impatient over *payment*, which even those with high primitive discount rates may not be. In contrast, the success of time-bundled contracts relies on high primitive discount rates.

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 allows us to make our second contribution: we show that an incentive program for walking improves health and leads to a large and persistent increase in walking among a population suffering from chronic disease. Existing evidence-based interventions promoting lifestyle change in similar populations are intensive and prohibitively expensive (Howells et al., 2016). Our study provides 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 she does not exert enough effort today, so effort today increases her future payoff to effort. While standard models show one reason such contracts enhance effort is simply the high stakes of job loss, our work suggests that these contracts have extra bite if the agent discounts her future effort.

Our 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 linear contracts in other populations with high discount rates of effort. A final topic to explore is how to personalize time-bundled contracts at scale at the individual level. One option is to use targeting based on observables, as Dizon-Ross and Zucker (2022) shows can work even for a dominated contract characteristic. Together, the answers to these questions will allow policymakers to effectively employ time-bundled contracts to motivate impatient people.

References

Andreoni, J., M. Callen, Y. Khan, K. Jaffar, and C. Sprenger (2018). Using preference estimates to customize incentives: An application to polio vaccination drives in Pakistan. *NBER Working Paper 22019*, 1–66.

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

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

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

Banerjee, A. V., E. Duflo, R. Glennerster, and D. Kothari (2010, may). Improving immunisation coverage in rural india: Clustered randomised controlled evaluation of immunisation campaigns with and without incentives. *BMJ (Clinical research ed.) 340*, 1–9.

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.

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

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

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*.

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.

Dandona, P. (2017). Minimizing glycemic fluctuations in patients with type 2 diabetes: approaches and importance. *Diabetes technology & therapeutics 19*(9), 498–506.

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 (2022). Can price discrimination incentivize behavior change? Evidence from a randomized field experiment. Technical report, Working paper, University of Chicago.

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

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.

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.

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.

Hussam, R., A. Rabbani, G. Reggiani, and N. Rigol (2022). Rational habit formation: experimental evidence from handwashing in india. *American Economic Journal: Applied Economics 14*(1), 1–41.

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 (2021). Can simple psychological interventions increase preventive health investment? *Journal of the European Economic Association*.

Jones, D., D. Molitor, and J. Reif (2019). What do workplace wellness programs do? evidence from the illinois workplace wellness study. *The Quarterly Journal of Economics 134*(4), 1747–1791.

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.

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

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

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

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

Lee, I. M., E. J. Shiroma, F. Lobelo, P. Puska, S. N. Blair, et al. (2012, 7). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The Lancet 380*, 219–229.

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

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.

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., R. T. Ackermann, E. M. Berke, A. Cheadle, B. Williams, et al. (2007). Impact of a managed-medicare physical activity benefit on health care utilization and costs in older adults with diabetes. *Diabetes Care 30*(1), 43–48.

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.

Patel, M. S., D. A. Asch, R. Rosin, D. S. Small, S. L. Bellamy, et al. (2016, mar). Framing financial incentives to increase physical activity among overweight and obese adults. *Annals of Internal Medicine 164*(6), 385.

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

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).

Reach, G., A. Michault, H. Bihan, C. Paulino, R. Cohen, and H. Le Clésiau (2011). Patients' impatience is an independent determinant of poor diabetes control. *Diabetes & Metabolism 37*(6), 497–504.

Read, D. and N. L. Read (2004). Time discounting over the lifespan. *Organizational behavior and human decision processes 94*(1), 22–32.

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

Royer, H., M. Stehr, and J. Sydnor (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics 7*(3), 51–84.

Sen, A. P., T. B. Sewell, E. B. Riley, B. Stearman, S. L. Bellamy, et al. (2014). Financial incentives for home-based health monitoring: a randomized controlled trial. *Journal of general internal medicine 29*, 770–777.

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). *gif: 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.

VanEpps, E. M., A. B. Troxel, E. Villamil, K. A. Saulsgiver, J. Zhu, et al. (2019). Effect of process- and outcome-based financial incentives on weight loss among prediabetic new york medicaid patients: a randomized clinical trial. *American Journal of Health Promotion 33*(3), 372–380.

Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242.

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.

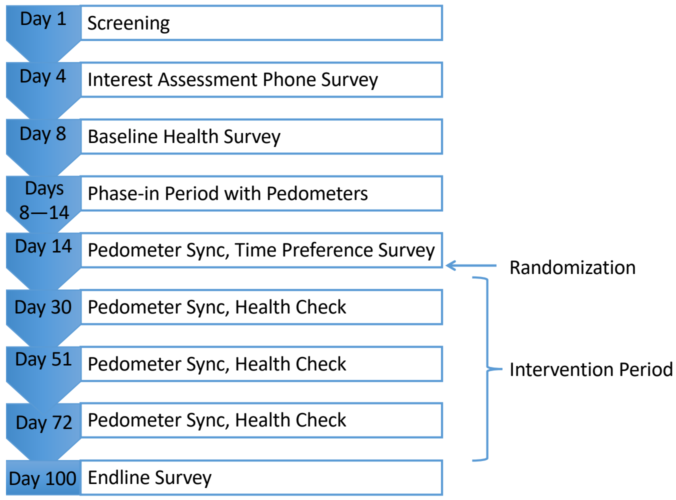
Warburton, D. E., C. W. Nicol, and S. S. Bredin (2006). Health benefits of physical activity: The evidence. *Canadian Medical Association Journal 174*(6), 801–809.

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.

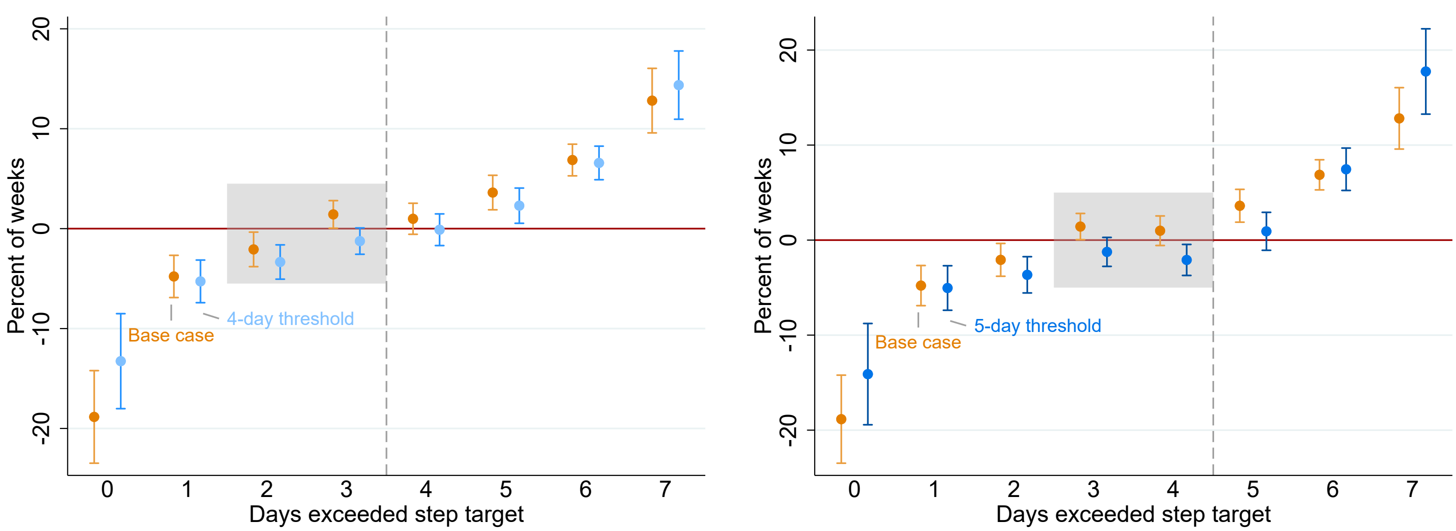
Appendices

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 - D. The Online Appendix contains Appendices E - M 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 into the timing of incentive delivery by delaying the start of the intervention period by one day for randomly selected participants. The intervention period was exactly 12 weeks for all participants.

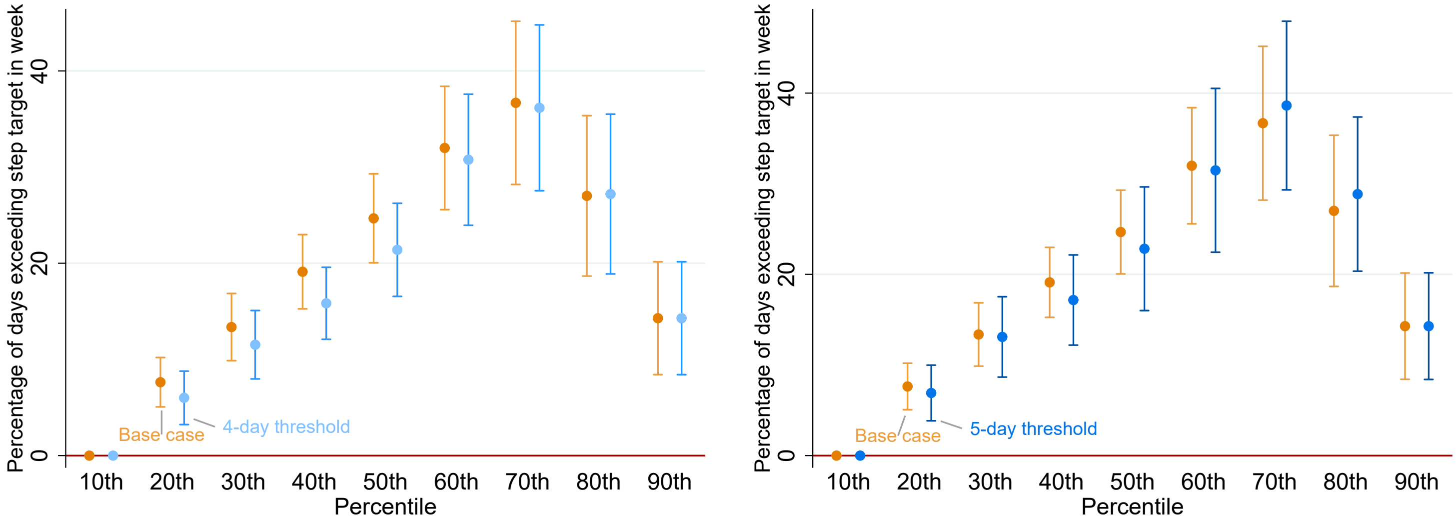


(a) 4-Day Threshold

(b) 5-Day Threshold

Appendix Figure A.2: Thresholds Modestly Decrease Compliance Right Below the Threshold

Notes: This figure shows the treatment effects of base case, 4-day threshold, and 5-day threshold on the number of days walked each week during the intervention period relative to monitoring. Data are at the respondent-week level. The vertical dashed lines in both panels indicates the threshold levels. Compliance levels to the right of the dashed lines are paid and those to the left are not. The two compliance levels just below the threshold are highlighted in both panels. Confidence intervals are relative to the monitoring group. Controls are the same as in Table 2. See Online Appendix Table H.10 for table version.

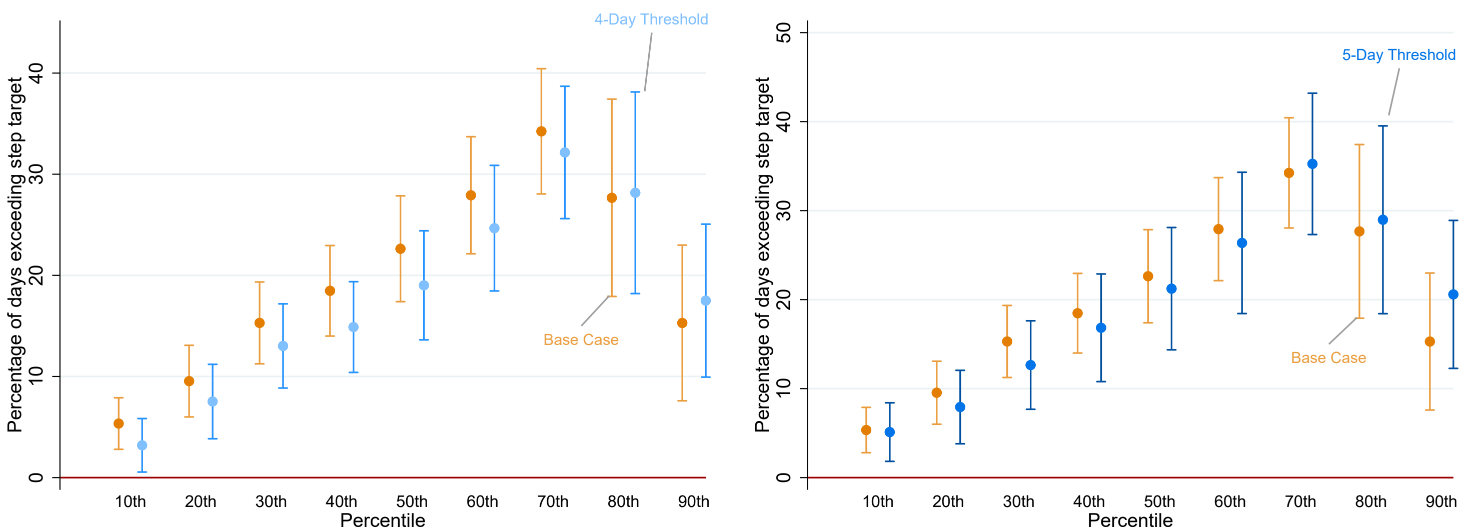


(a) 4-Day Threshold

(b) 5-Day Threshold

Appendix Figure A.3: Thresholds Only Modestly Affect the Distribution of Weekly Compliance

Notes: This figure shows quantile regression coefficients of the effects of the base case, 4-day, and 5-day thresholds (relative to monitoring) on the 10th through 90th percentiles of the distribution of weekly compliance (i.e., the percentage of days the participant exceeded the step target in the week) during the intervention period. Data are at the respondent-week level. Confidence intervals are relative to the monitoring group. Controls are the same as in Table 2. See Online Appendix Table H.9 for table version of both.

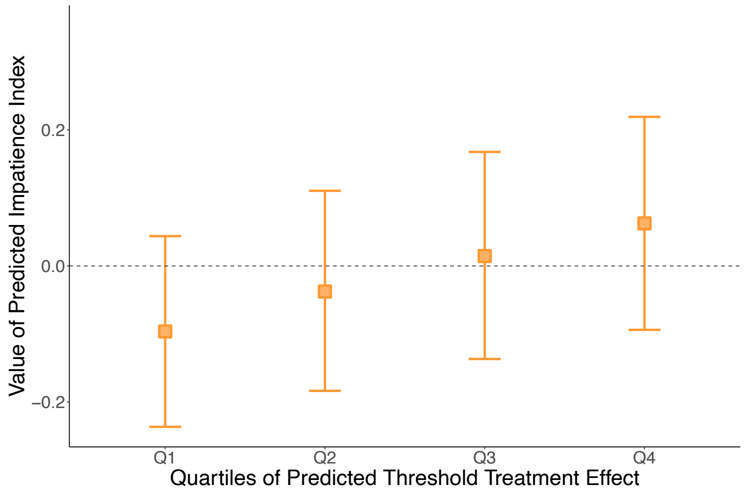


(a) 4-Day Threshold

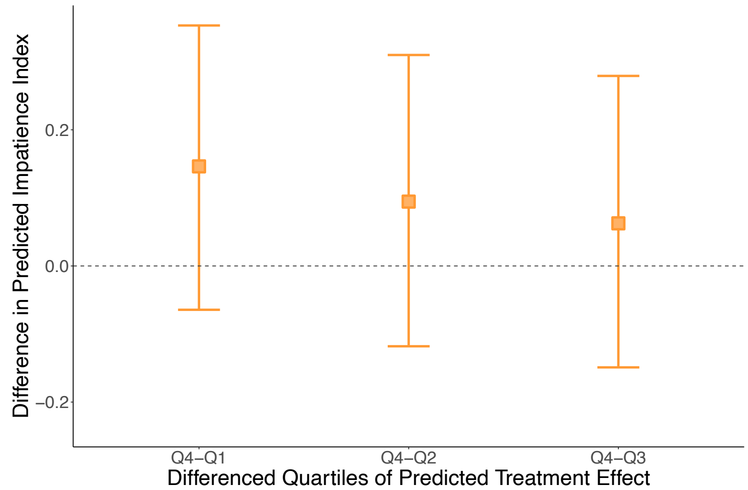
(b) 5-Day Threshold

Appendix Figure A.4: Threshold and Base Case Have Similar Impacts Across the Distribution of Individual-Level Compliance

Notes: This figure shows the quantile regression coefficients of the effects of the base case, 4-day, and 5-day thresholds (relative to monitoring) on the 10th through 90th percentiles of the distribution of individual-level compliance (i.e., the percentage of days the individual exceeded the step target during the intervention period). Data are at the individual level. Confidence intervals are relative to the monitoring group. Controls are the same as in Table 2. See Online Appendix Table H.9 for table version.



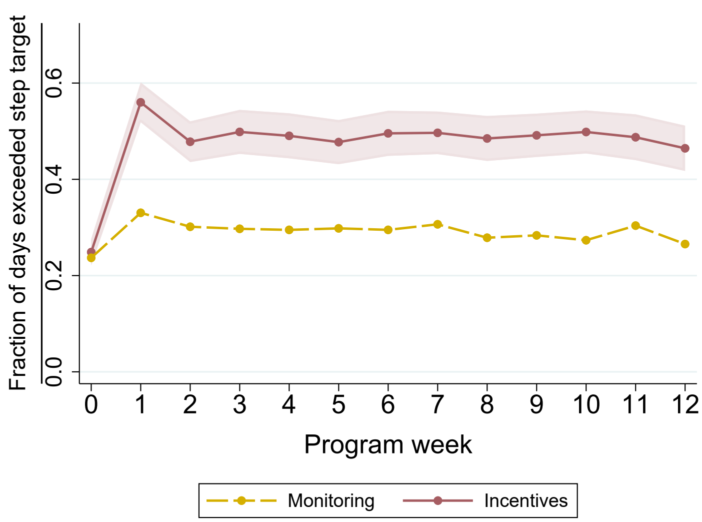
(a) Value of Predicted Impatience Across Quartiles of the Predicted Threshold Treatment Effect



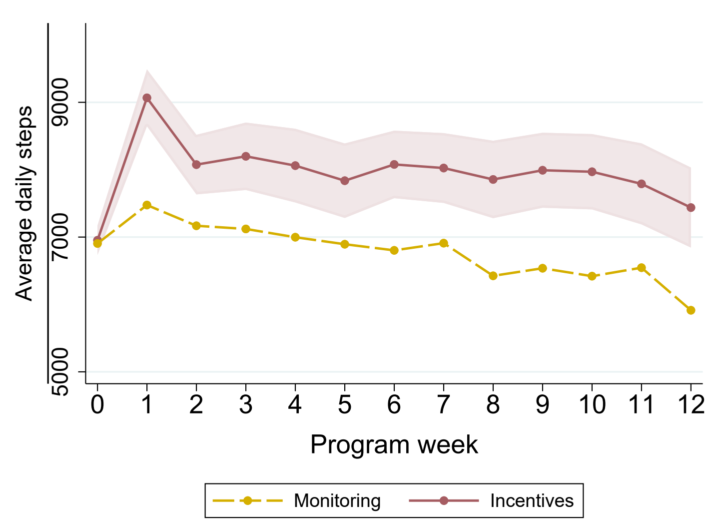
(b) Difference in Predicted Impatience Across Quartiles of the Predicted Threshold Treatment Effect

Appendix Figure A.5: Classification Analysis Results for Predicted Impatience Index

Notes: Figure replicates Figure 9 using the predicted impatience index instead of the actual impatience index.



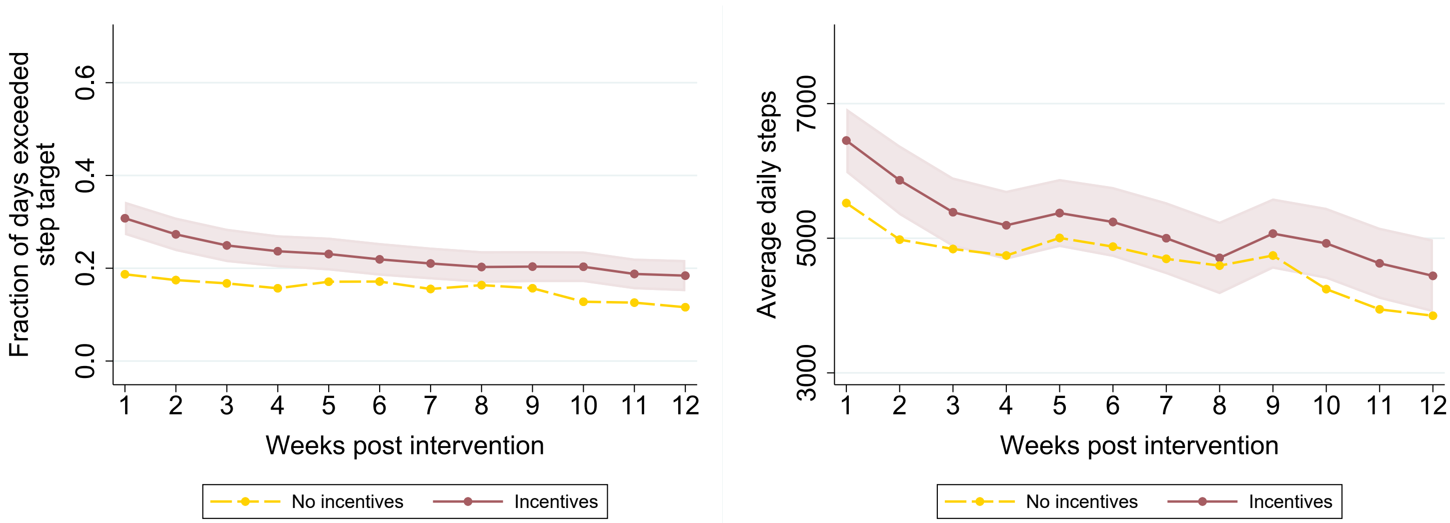
(a) Step-Target Compliance



(b) Daily Steps Walked

Appendix Figure A.6: 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 comparison groups from regressions with the same controls as in Table 2. Both graphs are unconditional on wearing the pedometer. See Online Appendix Figure H.3 for versions of the figures that condition on wearing the pedometer; they suggest that the reason that steps trend downwards in all groups over time in panel B is that pedometer wearing rates declined over time.

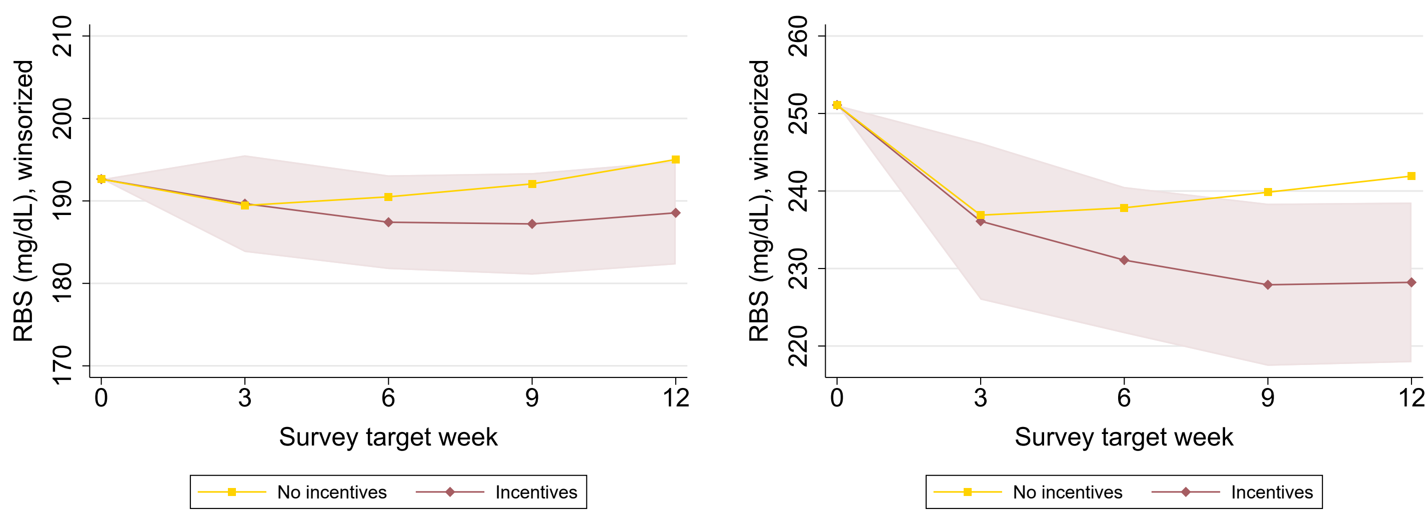


(a) Step-Target Compliance

(b) Daily Steps Walked

Appendix Figure A.7: Incentive Effects Persist After the 12-Week Program

Notes: Panel C shows the average probability of exceeding the step target and Panel D shows the average daily steps walked, both in the 12 weeks subsequent to the intervention. "No incentives" represents the pooled monitoring and control groups; the panels look very similar when we compare with the control group only (Online Appendix Figure H.2). The shaded areas represent a collection of confidence intervals from tests of equality within each weekly period between the incentive and comparison groups from regressions with the same controls as in Table 2. All graphs are unconditional on wearing the pedometer. See Online Appendix Figure H.3 for versions of the figures that condition on wearing the pedometer; they suggest that the reason that steps trend downwards in all groups over time in panel D is that pedometer wearing rates declined over time.



(a) Full sample

(b) Above-median blood sugar sample

Appendix Figure A.8: 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; see Table H.23 in the Online Appendix. 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 6. 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 .06 and .01 for the Panels A and B, respectively (see Table H.23 in the Online Appendix).

Appendix Table A.1: Enrollment Statistics

Total screened: 57,599

Total eligible: 7,781

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

Appendix Table A.2: Measures of Impatience Over Effort Correlate in the Expected Direction with Baseline Measures of Exercise, Health, and Behavior

Covariate type:	Exercise		Baseline Indices			
Dependent variable:	Daily steps	Daily exercise (min)	Negative health risk index	Negative vices index	Healthy diet index	# Individuals
A. Impatience Index Measures						
Impatience index	-0.080***	-0.070***	-0.016	-0.052	-0.181***	1,740
1. I'm always saying: I'll do it tomorrow	-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	-0.054	-0.052	-0.012	-0.043*	-0.149***	1,740
3. I postpone starting on things I dislike to do	-0.042*	0.004	0.004	-0.052	0.050	1,740
4. I'm on time for appointments	-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	-0.039	-0.069***	-0.009	-0.043*	-0.207***	1,740
B. Predicted index measures						
Predicted index	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?	0.015	-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?	-0.001	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?	-0.027	-0.062***	-0.018	0.031*	-0.038	3,192

Notes: This table displays the correlations between our impatience measures and a number of baseline health and behavior measures. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize health and behavior measures so that higher values correspond to healthier behavior; hence we expect all correlations to be negative. Panel A displays the impatience index along with the five questions from which it is generated. Panel B shows the predicted index along with the three questions from which it is generated. See Online App. Table H.24 for summary statistics on the components of each index. The health index includes an individual's measures of HbA1c, random blood sugar, blood pressure, body mass index, and waist measurement. The vices index includes an individual's daily cigarette, alcohol, and areca nut usage. The healthy diet index includes an individual's daily number of wheat meals, vegetable meals, rice meals, spoonfuls of sugar, and fruit, junk food, and sweets intake, as well as whether a respondent goes out of his or her way to avoid unhealthy foods. Significance levels: * 10%, ** 5%, *** 1%.

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

Covariate type:	Recharge variables				Credit constraint proxies		
Dependent variable:	Negative mobile balance	Negative yesterday's talk time	Prefers daily payment (=1)	Prefers monthly payment (=-1)	Negative wealth index	Negative monthly household income	# Individuals
Impatience index	0.032	-0.068	-0.038	0.034	0.047*	0.037*	1740
Predicted impatience index	0.021	-0.014	-0.005	-0.003	-0.036*	0.023	3192

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

Appendix Table A.4: Missing Pedometer Data during the Intervention Period

Dep. variable:	No Steps data	**Reason** no steps data		**Reason** no data from Fitbit			
		Did not wear Fitbit	No data from Fitbit	Lost data entire period	Immediate withdrawal	Mid-intervention withdrawal	Other reasons
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Incentives	-0.0140	-0.0287**	0.0155	-0.00203	0.00571	0.0166**	-0.00471
	[0.0174]	[0.0142]	[0.0124]	[0.00511]	[0.00731]	[0.00694]	[0.00594]
# Individuals	2,607	2,559	2,607	2,607	2,607	2,607	2,607
# Observations	218,988	205,732	218,988	218,988	218,988	218,988	218,988
Monitoring mean	0.19	0.15	0.05	0.00	0.01	0.01	0.02

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

Appendix Table A.5: Lee Bounds on the Impacts of Incentives on Exercise

Definition of missing:	No steps data	Did not wear Fitbit	No data from Fitbit	Lost data entire period	Withdrew immedi-ately	Mid-period withdrawal	Other reasons
A. Daily steps							
Regression estimate	1269	1269	1338	1338	1338	1338	1338
(conditional on nonmissing data)	[245]	[245]	[261]	[261]	[261]	[261]	[261]
Lee lower bound	1053	882	1230	1315	1297	1226	1303
	[357]	[214]	[267]	[290]	[195]	[320]	[304]
Lee upper bound	1426	1571	1572	1351	1430	1581	1358
	[342]	[307]	[305]	[291]	[230]	[321]	[276]
B. Met 10k step target							
Regression estimate	0.223	0.223	0.205	0.205	0.205	0.205	0.205
(conditional on nonmissing data)	[0.024]	[0.024]	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]
Lee lower bound	0.215	0.208	0.200	0.204	0.203	0.200	0.204
	[0.025]	[0.019]	[0.023]	[0.019]	[0.024]	[0.024]	[0.021]
Lee upper bound	0.232	0.242	0.216	0.206	0.209	0.217	0.206
	[0.025]	[0.028]	[0.025]	[0.020]	[0.025]	[0.023]	[0.021]
# Individuals	2,607	2,559	2,607	2,568	2,598	2,561	2,566
# Observations	218,988	205,732	218,988	206,488	209,008	211,551	206,320

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

Appendix Table A.6: Summaries from the Minute-level Pedometer Data

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

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

Appendix Table A.7: Threshold Treatments Increase Cost-effectiveness Relative to the Base Case, With Similar Increases among Those who are More and Less Impatient

		Sample defined by impatience indices			
	Full sample	Below median (actual)	Above median (actual)	Below median (predicted)	Above median (predicted)
Treatment group	(1)	(2)	(3)	(4)	(5)
Base case	0.050	0.050	0.050	0.050	0.050
Threshold	0.056	0.056	0.057	0.057	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

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

Appendix Table A.8: Incentives Also Improve Mental Health

Dependent variable:	Mental health index	Fitness time trial index
	(1)	(2)
Incentives	0.094**	0.014
	[0.045]	[0.044]
Monitoring	0.17**	0.056
	[0.074]	[0.074]
p-value: incentives = monitoring	0.233	0.527
Control mean	0.0	0.0
# Individuals	3,068	2,890

Notes: Observations are at the individual-level. Both specifications control for the baseline value of the index components, the index components squared and the same set of additional controls as in Table 6. The Mental health index averages the values of seven questions adapted from RAND's 36-Item Short Form Survey (SF-36). A large value of Fitness Time Trial Index indicates low fitness: it is an index created by the average two trials of endline seconds to walk four meters, and the seconds to complete five sit-stands standardized by their average and standard deviation in the control group. See Online Appendix Table H.25 for treatment effects on the individual components of the mental health and fitness indices. We follow World Health Organization guidelines to trim biologically implausible fitness time trial index components (i.e., z-scores < -4 or > 4). Robust standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

B Theoretical Predictions Appendix

B.1 Agent Problem

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

$$\max_{w_t \in \{0,1\}} \mathbb{E}\left[d^{(T-t)} m_T - \sum_{j=t+1}^{T} \delta^{(j-t)} w_{j,t} e_j \middle| e_1, .., e_t, w_1, .., w_t \right] - w_t e_t, \tag{16}$$

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

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

$$\begin{gathered} V_t(0) < V_t(1) - e_t \\ \text{or} \\ e_t < V_t(1) - V_t(0). \end{gathered} \tag{17}$$

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 Threshold Contracts and Impatience Over Effort

In this section, we present a series of propositions that provide the theoretical underpinning for Prediction 1 from Section 2.3. In particular, the propositions demonstrate that, holding all else equal, both compliance and effectiveness in threshold contracts tend to decrease in $\delta^{(t)}$. We begin by examining compliance in threshold contracts with $T = K$.

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 (17)). With the threshold contract, we have that:

$$V_1(1) - V_1(0) = \mathbb{E}\left[(dM - \delta e_2) w_{2,1} | e_1, w_1 = 1 \right] - \mathbb{E}\left[-\delta e_2 w_{2,1} | e_1, w_1 = 0 \right] \tag{18}$$

We examine this expression separately for sophisticates and naïfs.

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

$$\begin{aligned} V_1(1) - V_1(0) &= \mathbb{E}\left[(dM - \delta e_2) w_{2,1} | e_1, w_1 = 1 \right] - \mathbb{E}\left[-\delta e_2 w_{2,1} | e_1, w_1 = 0 \right] \\ &= \mathbb{E}\left[(dM - \delta e_2) \mathbb{1}\{e_2 < M\} + \delta e_2 \mathbb{1}\{e_2 < 0\} | e_1 \right] \end{aligned} \tag{19}$$

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

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

Thus, equation (19) is weakly decreasing in δ. That means that day 1 compliance is decreasing in δ. Hence, day 2 compliance is as well since $w_2 = 1$ if both $w_1 = 1$ and $e_2 < M$, and w_1 is weakly decreasing in δ. Thus, compliance in the threshold contract is decreasing in δ for sophisticates.

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

$$\begin{aligned} V_1(1) - V_1(0) &= \mathbb{E}\left[(dM - \delta e_2)w_{2,1} | e_1, w_1 = 1\right] - \mathbb{E}\left[-\delta e_2 w_{2,1} | e_1, w_1 = 0\right] \\ &= \mathbb{E}\left[(dM - \delta e_2)\mathbb{1}\{\delta e_2 < dM\} + \delta e_2 \mathbb{1}\{\delta e_2 < 0\} | e_1\right] \\ &= \mathbb{E}\left[\max\{dM - \delta e_2, 0\} + \delta e_2 \mathbb{1}\{e_2 < 0\} | e_1\right] \end{aligned} \tag{20}$$

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

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

Thus, equation (20) is weakly decreasing in δ. Hence day 1 compliance (and hence day 2 and total compliance) are also decreasing in δ for naïfs. □

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.[46] 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 δ 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,[47] the effectiveness of the threshold contract is weakly decreasing in δ.*

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

[47] See equation (24) 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

Proof. We first show that, if costs are positive, cost-effectiveness in the threshold is not increasing in δ. Because Proposition 1 showed that compliance is decreasing in δ, this establishes that effectiveness is decreasing in δ when costs are positive. We then show sufficient conditions for threshold effectiveness to decrease in δ when costs can be negative.

To simplify notation, let e^* be the agent's cutoff value for complying in period 1, such that agents comply in period 1 if $e_1 < e^*$. From equations (19) and (20), we know that the value of e^* will depend on the agent's sophistication and, importantly, decrease in δ.

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

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

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

$$\begin{aligned} C/P &= \frac{1}{M}\frac{\mathbb{E}[w_1 + w_2]}{\mathbb{E}[w_1 w_2]} = \frac{1}{M}\frac{\mathbb{E}[w_1 + w_1 w_2]}{\mathbb{E}[w_1 w_2]} = \frac{1}{M}\left(\frac{\mathbb{E}[w_1]}{\mathbb{E}[w_1 w_2]} + 1\right) = \frac{1}{M}\left(\frac{\mathbb{E}[w_1]}{\mathbb{E}[w_1]\mathbb{E}[w_2|w_1 = 1]} + 1\right) \\ &= \frac{1}{M}\left(\frac{1}{\mathbb{E}[w_2|w_1 = 1]} + 1\right) \end{aligned} \tag{21}$$

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

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

First, we rewrite the expression for effectiveness under the threshold given what we know

δ increases compliance by making people more likely to comply on day 1. The benefit is extra compliance and the cost is extra payment. The cost will be particularly large if there is a lot of inframarginal behavior on day 2, because now the principal has to pay out for all of the day 2's on which day 1 compliance was induced, which the principal used to get for free.

about C and P. (For notational simplicity, we examine $2(\lambda C - P)$ instead of $\lambda C - P$.)

$$
\begin{aligned}
2(\lambda C - P) &= \lambda \mathbb{E}[w_1 + w_2] - M\mathbb{E}[w_1 w_2] \\
&= \lambda \left(F(e^*) + \mathbb{E}[w_1 \mathbb{1}\{0 < e_2 < M\} + \mathbb{1}\{e_2 < 0\}]\right) - M\mathbb{E}[w_1 \mathbb{1}\{e_2 < M\}] \\
&= \lambda \left(F(e^*) + \mathbb{E}[\mathbb{1}\{e_1 < e^*\}\mathbb{1}\{0 < e_2 < M\} + \mathbb{1}\{e_2 < 0\}]\right) - M\mathbb{E}[\mathbb{1}\{e_1 < e^*\}\mathbb{1}\{e_2 < M\}] \\
&= \lambda \left(F(e^*) + Prob(e_1 < e^*, 0 < e_2 < M) + Prob(e_2 < 0)\right) - MProb(e_1 < e^*, e_2 < M).
\end{aligned} \tag{22}
$$

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 $g'(e^*) = f(e^*)\,Prob(e_2 \in S | e_1 = e^*)$.[48] Thus, we have

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

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

$$\lambda(1 + Prob(0 < e_2 < M|e_1 = e^*)) \geq MProb(e_2 < M|e_1 = e^*) \tag{23}$$

or

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

or

$$Prob(e_2 < 0|e_1 = e^*) \leq \frac{\lambda}{M} + \left(\frac{\lambda}{M} - 1\right) Prob(0 < e_2 < M|e_1 = e^*)). \tag{24}$$

If $\lambda > M$, condition (24) will always hold. More broadly, the condition will be more likely to hold the greater λ relative to M. The condition essentially guarantees that there not be "too much" inframarginal behavior, which generally decreases the efficacy of incentives. For example, when $\lambda > M/2$, which is a reasonable condition as it guarantees that the payment to the agent for two days of compliance is less than the benefits to the principal, a sufficient condition is

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

We have thus showed that, as long as there is not "too much" inframarginal behavior (i.e, as long as equation (24) holds), the effectiveness of a threshold contract is decreasing in δ. $\square$

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*

[48] To show this, note that

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

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

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

Proof. See Online Appendix I.1. □

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.

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

Proof. See Online Appendix I.1. □

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

Proposition 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 δ. Consider a threshold contract with $K = 2$, where the agent must thus comply on at least 2 days in order to receive payment. For sophisticates, compliance and effectiveness in the threshold contract are weakly decreasing in the discount factor δ.*

Proof. See Online Appendix I.1. □

B.3 The Effectiveness of Threshold and Linear Contracts

In this section, we compare the effectiveness of threshold and linear contracts under a range of effort cost assumptions, paying particular attention to how the relative effectiveness of thresholds depends on δ. For simplicity, throughout the section, we assume that $T = 2$ and that $K = 2$ and denote the threshold payment as M (i.e., $M = 2m'$).

Our first proposition (Proposition 6) examines the relative performance of the contracts in the limit as δ goes to 0 under very general assumptions. It shows that, for sufficiently low δ, for any linear contract, there exists a threshold contract that achieves substantially higher cost-effectiveness with relatively little—and potentially even no—loss in compliance. In contrast, for any linear contract, one can always construct another *linear* contract with substantially higher cost-effectiveness by decreasing the payment amount, but the loss in compliance may be arbitrarily large.

The next four propositions (Propositions 7a - 8b) examine the full range of δ, not just the case where δ 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 δ).[49]

Proposition 6. *Let $d = 1$ and $T = 2$. Fix all parameters other than δ, and take a linear contract that induces compliance $C > 0$.*
(a) If agents are naïve and e_2 is weakly increasing in e_1, in a first order stochastic dominance sense,[50] *then for sufficiently small δ, there exists a threshold contract with $K = 2$ that has at least two times higher cost-effectiveness (and $1 + \frac{1}{C}$ times higher cost-effectiveness if costs are IID) and that generates compliance $\frac{1+C}{2}$ of the linear contract.*
(b) If agents are sophisticated and costs are IID, then for sufficiently small δ, there exists a threshold contract with $K = 2$ that has at least $1 + C$ times higher cost-effectiveness and that generates compliance at least $\frac{1+C}{2}$ of the linear contract.

Proof. See Online Appendix I.2. □

The potential improvements from threshold contracts demonstrated by Proposition 6 are quantitatively large. For example, when costs are IID and agents are naïve with sufficiently low δ, 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 δ. 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) Let $T = 2$. Fix all parameters other than δ, and take any linear contract that induces compliance $C > 0$. Let there be perfect correlation in costs across days ($e_1 = e_2$). Then, regardless of agent type, there exists a threshold contract that induces compliance of at least C and that has approximately $2\frac{d}{1+\delta}$ times greater cost-effectiveness than*

[49] Predictions about optimal contracts are hard to test since most policymakers do not have sufficient information about the cost function and δ to solve for the optimal contracts.

[50] $F_{e_2|e_1}(x)$ 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.

the linear contract. Hence, if $\delta < 2d - 1$, the most effective contract will always be a threshold contract.

Proof. See Online Appendix I.2. $\square$

Proposition 8a (IID Uniform, $M = 2m$)**.** *Let $d = 1$. Fix all parameters other than δ. 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 δ. Define $\frac{2M^2}{1+M}$ as the "cutoff value" for naïfs and $2 - \frac{2}{M+M^2}$ as the "cutoff value" for sophisticates. If δ is less than the cutoff value for a given type, then the threshold contract is more effective, as it generates greater compliance.
(b) If $1 \leq M < 2$,[51] then the threshold contract is more effective.

Proof. See Online Appendix I.2. $\square$

Proposition 8b (IID Uniform, Optimal Contracts)**.** *Let $d = 1$. Fix all parameters other than δ. 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 δ as well as λ, the principal's marginal return to compliance. For a wide and plausible range of values of λ,[52] there exists a "cutoff" value of δ such that the threshold contract is more effective when δ is below the cutoff, and the linear contract is more effective when δ is above the cutoff. For the remaining values of λ, 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 δ.*

Proof. See Online Appendix I.2. $\square$

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.*[53]

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 $d^{(T-t)}$

[51] Note that the principal would never pay $M > 2$ since $M = 2$ achieves 100% compliance regardless of δ.

[52] See proof in Online Appendix I.2 for specific ranges for both naifs and sophisticates.

[53] 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.

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. □

C Predicting Impatience (and Associated Treatment Effect Heterogeneity) with Policy Variables

This appendix provides proof of concept that it is possible to use hard-to-manipulate observable characteristics that policymakers are likely to have access to in order to predict impatience and effectively target the threshold contract.

Since the results from Table 4 show that the threshold treatment effects vary with impatience over effort, policymakers can potentially improve the effectiveness of the program by targeting the threshold treatment to more impatient individuals. However, true impatience is hard for policymakers to observe; even when policymakers are able to ask impatience-related questions in a survey, participants may have an incentive to game their responses to achieve assignment to a specific contract — especially if one contract financially dominates the other. To address this potential concern, we construct a "policy prediction of impatience" by predicting our impatience index using demographics (e.g., age, labor force participation, bank account ownership) and medical information/records/history (e.g., Hba1c, fatigue) that health policymakers are likely to have access to (see the complete list of predictors in the notes of Table C.1). We then show that the policy prediction successfully predicts heterogeneity in the effect of the threshold, and hence could be used for personalizing assignment of the threshold contract at the individual level.

To prevent overfitting, we adopt a split sample approach to generate our policy prediction of impatience. First, we fit a LASSO prediction model within a randomly-selected training sample drawn from individuals with the actual impatience index. The model predicts our actual impatience index based on the predictors listed in the notes of Table C.1, as well as interactions of all of the predictors with indicators for having above-median values of the following: age, gender, individual income and household income. We then use that LASSO prediction model to generate the value of the policy prediction for the individuals who were not in the training sample (the "regression sample"). Finally, within the regression sample only, we estimate the heterogeneity in threshold treatment performance by the impatience index (i.e., estimate equation 13).[54]

Table C.1 shows the estimates from the heterogeneity regression. The results are similar to Table 2: the threshold has a meaningfully higher treatment effect among people with higher values of the policy predictiction of impatience. This suggests that using a similar policy prediction to personalize assignment of the threshold target could significantly improve the effectiveness of the policy at scale.

Appendix Table C.1: Threshold Treatment Effect Varies with Predicted Impatience Measure Constructed From Variables Available to Policymakers

Dependent variable:	Exceeded step target (×100)	
Impatience measure:	Policy predicted impatience index	Above median policy predicted impatience index
Sample:	Excluding training sample	
	(1)	(2)
Impatience × Threshold	2.81**	6.59**
	[0.26, 5.36]	[0.84, 12.34]
Threshold	-1.16	-4.56**
	[-4.04, 1.71]	[-8.60, -0.51]
Impatience	-0.97	-2.66
	[-2.82, 0.88]	[-6.90, 1.58]
# Individuals	1,746	1,746
# Observations	140,017	140,017
Base case mean	50.2	50.2

Notes: This table repeats the analysis in Table 4 but with a "policy" predicted impatience index constructed with the following variables: age, gender, labor participation, personal monthly income, household monthly income, household size, Hba1c, random bloog sugar, systolic BP, diastolic BP, BMI, waist circumference, walking speed, diagnosed diabetic, diagnosed hypertensive, overweight, owns home, home has running water, having a bank account, hired help at home, number of scooters owned, number of cars owned, number of computers owned, number of smartphones owned, number of mobile phones owned, number of rooms in house, mobile balance, hours of work on a weekday, whether consumes alcohol, whether smokes cigarettes or bidis, whether recently experienced fatigue, tingling in hands and feet, pain in legs or feet, back pain, headaches, fever, dizziness, or severe headaches, whether has foot ulcer, rapid deterioration in eyesight, pain or numbness in legs or feet. The prediction includes all of those variables and their interactions with having above-median values of age, gender, individual income and household income. The sample excludes individuals who were in the training sample used to generate the policy prediction of impatience. Controls are the same as in Table 2. Significance levels: * 10%, ** 5%, *** 1%.

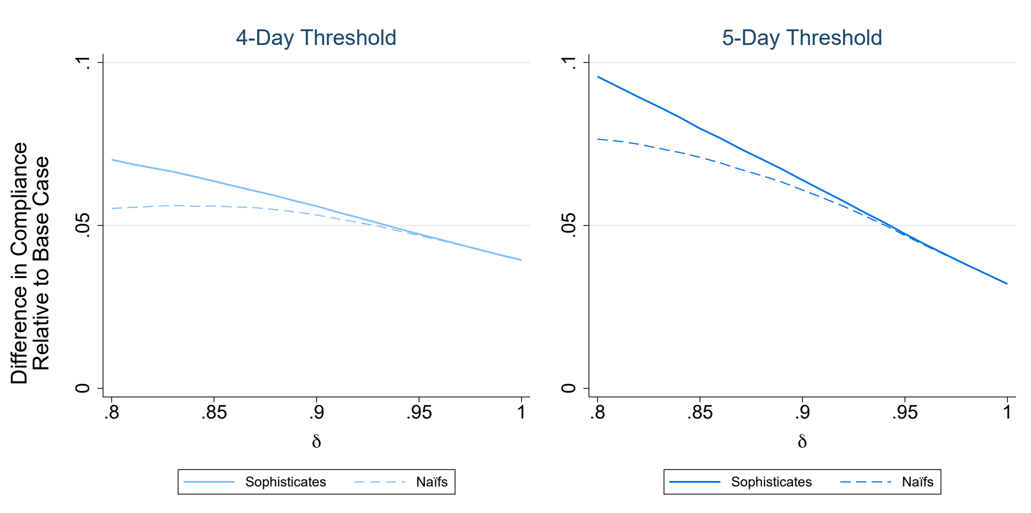
D Model Calibration for Threshold vs. Base Case

We calibrate a model using the empirical distribution of walking costs to show that, in this setting, the predicted performance of the threshold treatment increases meaningfully with

[54]To ensure power for our heterogeneity regression, we defined the size of the training sample to be such that the regression sample has twice as many observations as the training sample.

impatience over effort. We begin with the Section 2 framework. To tractably examine contracts with 7-day payment periods and with 4- and 5-day thresholds, we simplify the model by assuming that the effort discount rate is exponential with discount factor δ (i.e., that $\delta^{(t)} = \delta^t$), that $d^{(t)} = 1$, and that all future effort costs are known on day 1.

We first estimate the CDF of effort costs, as described in Online Appendix G. We then use the estimated CDF to calibrate the model and predict how relative compliance in the base case and threshold contracts would vary with δ. Figure D.1 displays the results, with δ on the x-axis and the gap between compliance in the threshold and base case linear contract on the y-axis (shown separately for the 4- and 5-day thresholds).



Appendix Figure D.1: Threshold Relatively More Effective for More Impatient

Notes: The figure shows the difference between compliance in each Threshold contract relative to the Base Case as predicted by our calibrated model. δ represents the exponential discount factor over effort.

The downward-sloping curves in the figure confirm the theoretical intuition from our model: for people who are more impatient over effort (smaller δ), there are larger compliance gains from thresholds. This is true for both naïfs and sophisticates with moderate levels of impatience.[55] In addition, the increase in performance of the threshold contract as impatience increases is quantitatively important, especially for the 5-day threshold contract, where the threshold has more bite, and where we see stronger results empirically as well (Online Appendix Table H.11, Panel B). For example, decreasing the effort discount rate from 1 to 0.9 increases relative compliance in the 5-day threshold contract by 3 pp among both sophisticates and naïfs.[56]

[55] As naïfs become more impatient ($\delta < 0.85$), the linear contract starts to gain relative to the 4-day threshold, as naïfs begin to procrastinate in early periods under the threshold contract. However, even very impatient naïfs still do better with the threshold than completely patient people ($\delta = 1$), which is our theoretical prediction when the threshold level is less than the number of days (Proposition 4 in App. B.2).

[56] The calibration overestimates the average effect of the threshold, which in practice we found to be zero. This is likely because our model does not incorporate uncertainty regarding future effort costs. However, our interest is heterogeneity by impatience, which we do not believe will change by incorporating uncertainty.