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

July 8, 2022

Abstract

Promoting lifestyle changes such as regular exercise is critical for the global fight against diabetes. One barrier to lifestyle change is impatience (i.e., heavy discounting of the future), which makes short-run financial incentives for lifestyle change a promising approach and also makes it important to ensure the incentives work well in the face of impatience. We evaluate whether providing incentives for exercise to diabetics can help address the problem of diabetes in India. We also test a novel prediction, namely that "time-bundled" contracts, where the payment for future effort is increasing in current effort, are more effective when agents are impatient. We find positive results on both fronts. First, incentives increase daily steps by roughly 20 percent (13 minutes of brisk walking) and improve blood sugar. Second, consistent with our prediction, time-bundled contracts work better for more impatient people.

*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, Yashna Nandan, Varun Satish, and Emily Zhang for outstanding research assistance. We are grateful to Ned Augenblick, Abhijit Banerjee, Marianne Bertrand, Esther Duflo, Pascaline Dupas, Alex Frankel, Rick Hornbeck, Seema Jayachandran, Anett John, Supreet Kaur, Jonathan Kolstad, Ted O'Donoghue, 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

Chronic lifestyle diseases such as type 2 diabetes and hypertension represent a severe threat to health and development in low and middle income countries (LMICs). Chronic diseases cause 70% of the deaths in LMICs and are associated with substantial morbidity, premature mortality, and lost productivity (World Health Organization, 2020). 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).

There is widespread agreement that the key to addressing the growing health and economic burden of chronic disease is to promote three lifestyle changes: more physical activity, healthier diet, and less tobacco and alcohol use (World Health Organization, 2009). Each of these changes can prevent disease onset, decrease the rate of costly complications, and avert premature mortality. However, a large portion of patients diagnosed with chronic lifestyle diseases do not adopt the recommended lifestyle changes despite the high personal stakes (Carpenter et al., 2019), and existing evidence-based interventions promoting lifestyle change are prohibitively expensive (Howells et al., 2016). Policymakers are thus particularly interested in scalable interventions to promote lifestyle change.

One potential barrier to lifestyle change is that people heavily discount the future — that is, they are impatient (e.g., Mahajan, Michel, and Tarozzi, 2020; Augenblick and Rabin, 2019). People with chronic lifestyle diseases may be especially impatient (Reach et al., 2011; Wainwright et al., 2022). Lifestyle changes, from exercising to forgoing dessert, involve short-run costs but only longer-run benefits. As a result, impatient people will tend not to make these changes even if the eventual benefits are large relative to the costs. A common approach to address impatience in other contexts is to provide short-run positive reinforcement, such as small financial incentives (e.g., Banerjee, Duflo, Glennerster, and Kothari, 2010; Royer, Stehr, and Sydnor, 2015). However, whether incentives can in fact promote lifestyle change among those with chronic lifestyle diseases is not well understood. Moreover, we have little knowledge of how to best structure incentives so that they work well when people are impatient — in any context, lifestyle change or otherwise.

This paper makes two distinct contributions. First, we test whether a scalable program that provides incentives for exercise to diabetics can help decrease the burden of chronic disease in LMICs. Second, we shed light on how to improve the performance of incentives when agents are impatient over effort. To make the latter contribution, we test a novel prediction that "time-bundled contracts," which make the payment for future effort increase in current effort, are more effective when agents discount effort costs more. Notably, this prediction is about the "primitive" discount rate from an individual's utility function, which applies to all utility costs and benefits, including effort costs. This primitive discount rate is distinct from the discount rate over payment, which instead reflects the available borrowing and saving opportunities. As shown in Augenblick, Niederle, and Sprenger (2015), even if individuals have a high primitive discount rate over utility costs like effort, their discount rate over payment can be low since the payment discount rate should equal the market interest rate for people with access to borrowing and saving.

To illustrate the intuition for why time-bundled contracts work well when people are impatient over effort costs, imagine you need a worker to perform two days of work. Consider first a time-bundled "threshold" contract that pays 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, the minimum payment to induce two days of work must be higher, at \$20: \$10 per day of effort. Time-bundled contracts thus exploit the fact that, when individuals have high effort discount rates, it is "cheaper" to buy their future (discounted) effort than their current effort.

Time-bundled contracts should be effective for all types of people with high discount rates over effort: time-consistent or time-inconsistent and, among time-inconsistent, both those who are sophisticated and even those who are "naïve" (or unaware) about their own present bias. We consider the effectiveness for naïfs to be an important feature. Naïve time inconsistency is common (for example, Mahajan et al. (2020) estimate that 50% of a sample of Indian adults are naïfs). Naïfs are also difficult to motivate (Bai, Handel, Miguel, and Rao, 2020). The effectiveness for naïfs differentiates time-bundled 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 will sell their future effort at a discount today.

To make our two contributions, we partnered with the Government of Tamil Nadu, one of the most populous states in India, to conduct a randomized controlled trial (RCT) evaluating an incentive program for exercise among a sample of diabetics and prediabetics in urban areas. With a diabetes prevalence rate of 24% among adults in urban Tamil Nadu (Ranasinghe et al., 2021), the government funded this study to try to identify a lifestyle intervention that they could scale up across the state to address their exploding diabetes epidemic.

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

[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.

The program monitors participants' walking for 3 months using pedometers and provides them with small financial incentives in the form of mobile phone credits if they achieve a daily step target of 10,000 steps. To evaluate the program, we randomly assign participants to a treatment group that receives the program or a control group that does not. To test the prediction that time-bundled contracts work better for impatient people, within the treatment group, we randomly vary whether payment is a linear function of the number of days the participant complies with the step target or whether payment is instead a time-bundled threshold function. The time-bundled threshold contract only rewards compliance with the step target if the step target is met a minimum number of days that week. We compare the different contracts to assess their average efficacy and to test our prediction that time-bundling will have heterogeneous impacts by impatience over effort.

Our evaluation demonstrates that incentives for exercise could be a cost-effective intervention to help decrease the burden of chronic disease in India and beyond. The incentives program substantially increases exercise. Providing an incentive of just 20 INR (0.33 USD) per day of compliance with the step target increases compliance by 20 percentage points (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. Moreover, roughly 40% of the treatment effect on walking continues for several months after the intervention ends. This is more persistence than many other exercise interventions (e.g., Royer et al., 2015; Acland and Levy, 2015) and suggests that people formed healthy habits that could last (Hussam, Rabbani, Reggiani, and Rigol, 2022).

The treatment effects on exercise are of standalone interest: our sample has high rates of diabetes and hypertension, and, in the long run, regular exercise has been shown to prevent complications from both (Lee et al., 2012). Moreover, we find that the improvements in exercise induced by incentives translate to modest improvements in blood sugar, the primary clinical marker for diabetes, even in the short run. The program also improves an index of cardiovascular health that includes blood pressure and body mass index (BMI) in addition to blood sugar, and increases an index of mental health. Taken together, these impacts on exercise and health are promising for policy, especially since – unlike previous evidence-based interventions promoting lifestyle change among diabetics – our program is scalable and low cost.

Since chronic disease is associated with impatience, we next examine strategies to improve the performance of incentives for impatient people. Our headline finding is that time-bundled contracts meaningfully increase relative effectiveness for those who are impatient over effort, consistent with our theoretical prediction. Specifically, heterogeneity analysis using a baseline measure of impatience shows that, relative to linear contracts, time-bundled threshold contracts increase compliance with the step target by 6 pp more for people with above-median impatience than for those with below-median impatience, a large difference relative to the sample-average effect of either contract (20 pp). The 6 pp difference represents the gap between a 3 pp positive effect among those with above-median impatience and a 3 pp negative effect among those with below-median impatience. The level of impatience is thus pivotal to whether linear or time-bundled contracts generate more effort.

We also find that, on average, across the full sample, the time-bundled threshold contract performs better than the linear contract—it achieves the same sample-average level of compliance as the linear contract, but does so at a lower cost.[3] We show that this improves the performance of the contract from the perspective of a policymaker who wants to maximize the benefits of compliance net of the incentive costs.[4] While it is theoretically ambiguous whether time-bundled contracts will outperform linear contracts in general, we show that this result is more likely to hold when discount rates over effort are high, as they may be in our sample given the high levels of chronic disease.

These results imply that policymakers can improve the performance of incentives by customizing whether people receive linear or time-bundled contracts based on their impatience. This type of personalization is likely feasible. Andreoni, Callen, Khan, Jaffar, and Sprenger (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. One concern with personalizing time-bundled contracts is that some are dominated by linear contracts from the participant perspective, paying out weakly less payment for any level of effort. This creates concerns that participants will game the system to avoid assignment to time-bundled contracts. However, Dizon-Ross and Zucker (2022) demonstrate that, in the context of walking incentive contracts, people are unlikely to manipulate observable characteristics in order to avoid assignment to dominated contracts.[5] Our results also indicate a role for customizing the use of time-bundled contracts at the population level by using time-bundled contracts among particularly impatient populations such as people with chronic disease.

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 main way to adjust incentives for present bias. However, they also acknowledge that increasing

[3]This finding contributes to a small literature comparing linear to nonlinear contracts. While most existing experiments focus on differential selection into nonlinear contracts (e.g., Larkin and Leider, 2012; Kaur, Kremer, and Mullainathan, 2015), our experiment is one of the few to randomize whether incentives are linear or nonlinear. DellaVigna and Pope (2018) also randomize contract linearity but do not examine cost-effectiveness.

[4]This statement depends on the assumption that the benefits of compliance are linear in compliance. We also discuss other potential objectives later in the paper.

[5]Specifically, Dizon-Ross and Zucker (2022) find that few people manipulate observed baseline walking levels downward in order to avoid assignment to dominated exercise incentive contracts, even when manipulation is nearly effortless, as they recognize that such manipulation might dampen their own long-run exercise.

payment frequency should only be effective if people are impatient over *payments*, which even those who are impatient over effort may not be.

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). 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; DellaVigna and Pope, 2018).[6] Thus, increasing payment frequency may not always be an effective way to adjust incentives for impatience. This makes it important to identify other ways to adjust incentives for impatience and highlights the significance of our finding that time-bundled contracts are one such way.

Contributions to the Literature Our first contribution is to show that incentives for exercise are a scalable, effective lifestyle 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; for example, Long (2012) provides diabetics in the US with incentives to lower their blood sugar and finds no impact.[7] In contrast, building on previous work showing that incentives increase walking among healthy populations in the developed world (e.g., Bachireddy et al., 2019; Finkelstein et al., 2008, 2016; Patel et al., 2016), we find that exercise incentives deliver positive behavioral and health impacts for diabetics in the developing world. Moreover, relative to other exercise incentive programs, our program stands out for its relatively large and persistent effect on behavior, for its measurable impacts on downstream health outcomes, and for its low cost. The success of our targeted exercise incentive program contracts with the lack of impact of more broad-based, comprehensive wellness interventions in the US (Jones, Molitor, and Reif, 2019).

Our second contribution is to the time preferences literature: we show theoretically and empirically that time-bundled threshold contracts are effective for a wide range of people with impatient preferences 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 (Bai et al., 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

[6]Augenblick et al. (2015) and Andreoni and Sprenger (2012) find limited impatience over payment via cash and via check, respectively, and DellaVigna and Pope (2018) find that decreasing the lag until a payment is made into an mTurk account has limited quantitative impact on effort.

[7]Sen et al. (2014) show that incentives for glucose monitoring improve monitoring but not blood sugar. VanEpps et al. (2019) and Desai et al. (2020) evaluate a program in multiple US states that provides prediabetics with incentives for attendance in a health program and for weight loss, and do not find clear impacts on weight.

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, Royer, Stehr, and Sydnor (2020) both examine ways to help time-inconsistent procrastinators avoid delay in completing a single task.[8] Andreoni, Callen, Khan, Jaffar, and Sprenger (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.

The paper proceeds as follows. Section 2 presents our theoretical predictions. Sections 3 and 4 discuss the study setting and design. Section 5 presents the impacts of incentives on exercise and health. Section 6 explores time-bundled contracts and impatience. 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 agents have high discount rates over effort. We first specify the agent'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 effect of making the contract time-bundled (i.e., making the payment for future effort increase in current effort). We show that, for multiple types of impatience, including time-inconsistent sophistication and time-inconsistent naivete, the effectiveness of the time-bundled contract is often increasing in the discount rate over effort. While one can come up with parameter assumptions under which the result does not hold, we show that it holds in many typical and empirically relevant cases.

[8] 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.

Finally, we briefly analyze a potential strategy for adjusting incentives for impatience over payment rather than effort: increasing the frequency of payments. We show that this strategy is effective if agents have high discount rates over payment.

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.[9]

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. 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 ($w_{j,t} \geq w_t$).

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

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

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.[10] 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.[11]

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 $\delta^{(t)}$.[12]

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

[10] This is a distinct objective from maximizing welfare, but is often used in practice. For example, in health, policymakers and insurance companies often want to maximize the total health benefits of a program relative to the program costs. We discuss the appropriateness of this objective in Section 6.4.

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

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

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

We will see that this result contrasts with 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 the 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 when threshold contracts are more effective than linear contracts. Appendix B.2 compares the overall effectiveness of threshold and linear contracts formally, yielding two main takeaways. First, threshold contracts can substantially increase effectiveness when the discount rate over effort is high. Second, the discount rate over effort can be pivotal to whether threshold or linear contracts are more effective.

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

Prediction 1 (Threshold Effectiveness and Impatience Over Effort)**.** *Holding all else equal, 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.*

Prediction 1 is based on a series of propositions, presented in Appendix B.3. The propositions 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 $\delta^{(t)}$ (Section 2.2). 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 4 examines threshold contracts with $K = T$ (i.e., where one must comply on all days in order to receive payment). We show that, for all T, *compliance* is weakly decreasing in δ. To examine effectiveness, Proposition 5 considers the case where $T = 2$ and makes the reasonable assumption that e_2 is weakly increasing in e_1, an assumption which flexibly accommodates the range from IID costs to perfect positive correlation and just rules out negative correlation. We show that, in that case, under relatively general conditions,[13] *effectiveness* in the threshold contract is also weakly decreasing in δ.

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

Thus, overall, the propositions suggest that, when either (a) K is high relative to T,[14] or (b) costs are positively correlated across periods, our prediction 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.[15]

Intuition We illustrate the intuition for Prediction 1 by considering a simplified case, with $d = 1$, $T = 2$, $K = 2$, and with 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 she places on that day 2 opportunity is

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

which is equal to the discounted (with $d = 1$) payment $2m'$, net of the discounted effort costs δe_2, in the states of the world where the individual thinks she will comply on day 2 if she complies on day 1 (i.e., where $w_{2,1}\big|^{w_1=1} = 1$). 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.

[13] See Proposition 5 for the exact condition; it entails there not being "too much" inframarginal behavior. When there is too much inframarginal behavior, not only will the effectiveness prediction not hold but incentives also become less likely to be a cost-effective approach.

[14] 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.

[15] 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.

The fact that impatient people value the day 2 opportunity more is what underlies the threshold's greater effectiveness for them. 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.[16]

Sophisticates and Naïfs Although Prediction 1 holds 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,

$$w_{2,1}\Big|^{w_1=1} = \mathbb{1}\{e_2 < 2m'\}. \tag{8}$$

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,

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

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 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.[17]

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

[17] 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 $(dM - \delta e_2)\mathbb{1}\{\delta e_2 < dM\} = \max\{dM - \delta e_2, 0\}$.

With time-bundled thresholds, these differences 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 (8) implies equation (9)).[18] 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 E, there are other types of time-bundled contracts (other than thresholds) in which naïfs and sophisticates can make very different decisions.[19] In those contracts, *options* and *commitment* are less tightly linked. For example, there are time-bundled contracts where compliance on day 1 generates a soft *commitment* for day 2 compliance but does not generate a positive NPV *option*. These contracts function exactly like commitment contracts and are effective for sophisticates only.[20] Importantly, thresholds never fall in this category.

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. See Appendix B.4 for proofs.

Prediction 2 (Frequency)**.** *If agents are impatient over the receipt of financial payments (i.e., if $d^{(t)} < 1$ for $t > 0$ and is weakly decreasing in t), then the compliance and effectiveness of the base case linear contract are weakly increasing in the payment frequency. If agents are*

[18] To be more specific, comparison of equations (8) and (9) shows that the only potential difference between sophisticates and naifs is that, if $\delta e_2 < 2m' < e_2$, naifs could comply on day 1 while sophisticates would not. However, this behavior should be rare in practice. To see the reason, compare equations (6) and (7): because people sink costs as they move toward the threshold, the marginal incentive to comply is strictly higher on day 2 (where it is $2m'$) than on day 1 (where it is $2m' - \delta e_2$). Thus, if a naïf complies on day 1, they generally follow-through on day 2. Failure to follow-through for the naif 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.

[19] Online Appendix E investigates the full class of 2-day time-bundled contracts. This class also includes contracts where the day 2 wage is not 0 in the absence of day 1 compliance (e.g., a contract paying $5 for day 2 effort if the agent did not comply on day 1 and $10 if she did).

[20] 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 $w_1 = 1$. Thus, if a contract pays between δe_2 and e_2 for day 2 work if $w_1 = 0$ but more than e_2 for day 2 work if $w_1 = 1$, day 1 compliance generates a commitment without generating an option. In such contracts, 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.

patient over the receipt of financial payments ($d^{(t)} = 1$), then payment frequency does not affect compliance or effectiveness.[21]

In addition to containing a policy-relevant prediction for how to improve contract effectiveness in the face of impatience over payment, Prediction 2 also provides a way to use the empirical variation in compliance across payment frequencies to make inferences about the discount factor over payments. Our final prediction follows Kaur et al. (2015) to show an additional way to use empirical data to make inferences about the discount factor over payments, which will be useful for interpreting our results.

Prediction 3 (Payday Effects)**.** *If the discount factor over payments $d^{(t)}$ is decreasing in t, then the probability of complying in the base case linear contract increases as the payday approaches. If the discount factor over payments $d^{(t)}$ is constant in t, then the probability of complying is constant as the payday approaches.*

2.5 Empirical Tests

Predictions 1 and 2 informed the design of our experiment. To assess Prediction 1, among participants who receive incentives in our experiment, we randomly vary whether the contract is linear or has a threshold. 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 (we address potential confounds to the impatience measure in Section 6.2). To shed light on the potential role of payment frequency, we randomize the payment frequency, allowing us to understand both the effect of changing payment frequency and (per Prediction 2) whether agents are meaningfully impatient over payments. Both of these hypothesis tests were specified *ex ante*.[22]

3 Experimental Design

3.1 Sample Selection and Pre-Intervention Period

We conducted our experiment in an urban area of South India. India is facing a diabetes epidemic, and prevalence is higher both in southern states and in urban areas. We selected our sample through a series of public screening camps in the city of Coimbatore, Tamil Nadu. To recruit diverse socioeconomic groups, we held the camps in various locations including the government hospital, markets, religious institutions, and parks. During the camps, trained surveyors took health measurements, discussed each individual's risk for diabetes and hypertension, and conducted an eligibility survey. To be eligible for the study, individuals needed to

[21] 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.

[22] 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 in the AEA registry, one can find it at https://www.chicagobooth.edu/-/media/faculty/rebecca-dizon-ross/research/PAP_NCD_2015.pdf.

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."[23] After screening, we contacted eligible individuals by phone and invited them to participate in a program encouraging walking.

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

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 told participants what treatment group they had been randomly assigned to 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 the enrollment process). We 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.

[23]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.

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).[24] 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.

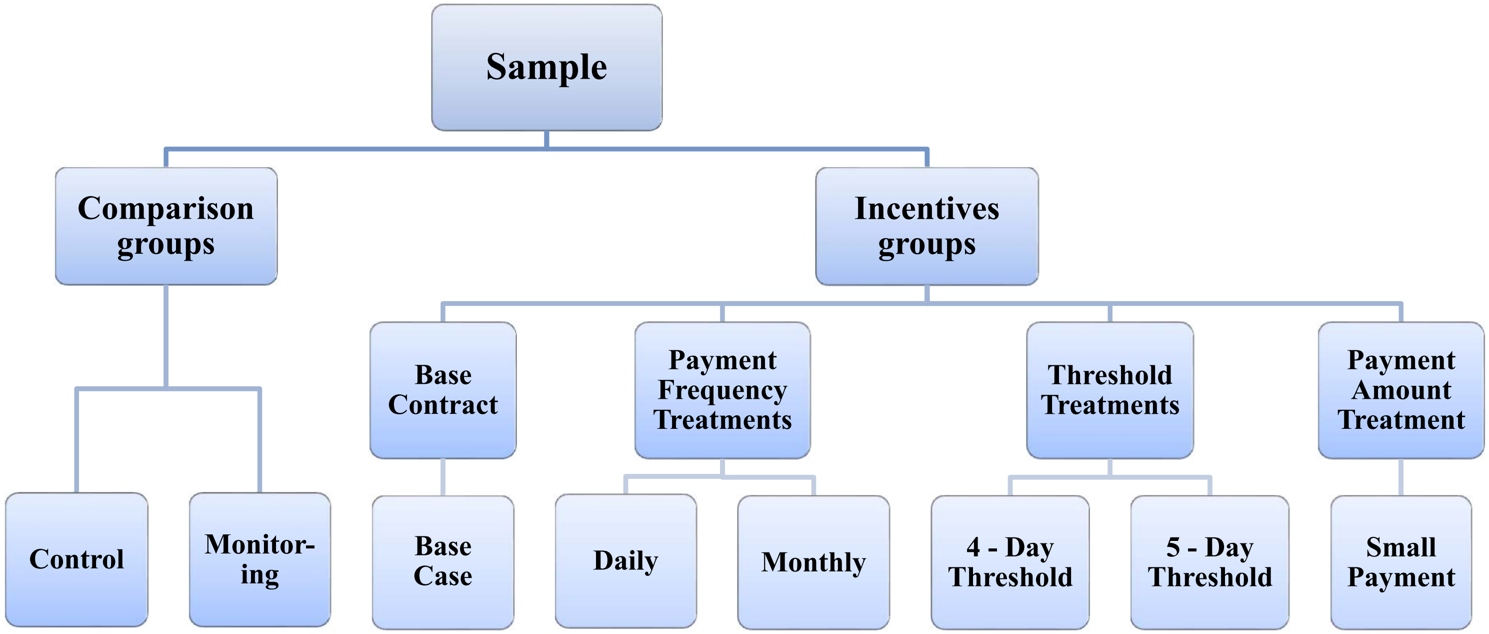
We call this the *base case* contract because all other contracts differ from it in exactly one dimension: linearity, payment frequency, or payment amount. We can compare any other group to the base case group to assess the effect of changing a single contract dimension.

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. 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. We follow our *ex ante* analysis plan and pool the *4-day threshold* and *5-day threshold* treatment groups for our main analyses in order to maximize statistical power.[25]

[24] 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).

[25] We sometimes also show the results for the two thresholds separately as exploratory analyses. 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.



	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

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

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

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 2 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 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.[26]

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

3.2.3 Contract Understanding

To ensure participants understood their contracts, a few days after each participant was told about their contract, a surveyor would call them to ask several questions testing their understanding of their contracts. 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 G.1).

3.3 The Intervention Period and After

To measure steps, we gave monitoring and incentive group participants Fitbit Zip pedometers for the duration of the intervention. Since most participants did not have regular internet access to sync their pedometer data, these data were not available in real time. Instead, we asked participants to report their daily step count to an automated calling system, which called participants every evening and prompted them to enter their daily steps from the pedometer. Incentive payments were based on these reports. To verify the reports, we visited participants every two to three weeks to manually sync their pedometers, cross-check the pedometer data against the reported data, and discuss any discrepancies. Anyone found to be chronically overreporting was suspended from the program. All empirical analysis is based on the synced pedometer data, not the reported data.[27]

We visited all participants three times during the 12-week intervention period. The primary purpose was to sync pedometers, but we also conducted short surveys to collect biometric and mobile phone usage data (we conducted these visits even with pure control group participants who did not have a pedometer in order to hold survey visits constant across participants). At the end of the 12-week intervention period, we conducted an endline survey. Figure A.1 shows the intervention timeline.

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 and presents summary statistics. Next, it describes our two sources of outcomes data: pedometer data and a health survey.

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

4.1 Baseline Data: Health, Walking, and Time Preference

We have three sources of baseline data: a baseline health survey, a week of pedometer data, and a time-preference survey. The baseline health survey, conducted at the first household visit, contains information on respondent demographics, health, fitness, and lifestyle. Health measures include HbA1c, a measure of blood sugar control over the previous three months; random blood sugar (RBS), a measure of more immediate blood sugar control; body mass index (BMI) and waist circumference, two measures of obesity; blood pressure, a measure of hypertension; and a short mental health assessment. During the phase-in period (between the baseline health survey and randomization), we collected one week of baseline pedometer data.

Finally, following the phase-in period, we conducted a baseline time-preference survey to measure impatience over effort. Since we will present average treatment effects before we present heterogeneity based on measures of effort impatience, we defer our description of the effort impatience measures until Section 6.2.1, after we present the average treatment effects.

4.2 Summary Statistics

The first column of Table 1 displays the baseline characteristics of our sample. The sample is, on average, 49.4 years old and has slightly more males than females. The average monthly household income is approximately 16,000 INR (about 200 USD) per month, close to the median for an urban household in India (Ministry of Labour and Unemployment, 2016). Panel B shows that our sample is at high risk for diabetes and its complications: 65% of the sample has been diagnosed with diabetes by a doctor, 81% have HbA1c levels that indicate diabetes, and the RBS measures show poor blood sugar control. The sample also has high rates of comorbidities: 49% have hypertension and 61% are overweight. Panel C shows that, on average, participants walked 7,000 steps per day in the phase-in period, comparable to average daily steps in many developed countries (Bassett et al., 2010). 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 G.2 shows that covariates are also balanced within the subsample for whom we have data during the post-intervention period.

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

	Full sample	Control	Monitoring	Incentives pooled	Daily	Base case	Monthly	Threshold	Small payment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Demographics									
Age (from BL)	49.54	49.78	50.28	49.44	49.57	49.60	48.80	49.41	49.11
	(8.52)	(8.19)	(8.95)	(8.55)	(8.60)	(8.33)	(8.94)	(8.71)	(7.84)
Female (=1)	0.42	0.46	0.43	0.41	0.44	0.41	0.38	0.41	0.48
	(0.49)	(0.50)	(0.50)	(0.49)	(0.50)	(0.49)	(0.49)	(0.49)	(0.50)
Labor force participation (=1)	0.75	0.73	0.72	0.75	0.75	0.74	0.81	0.75	0.70
	(0.44)	(0.45)	(0.45)	(0.43)	(0.43)	(0.44)	(0.39)	(0.43)	(0.46)
Per capita income (INR/month)	4463	4488	4620	4447	4068	4477	4599	4461	4341
	(3638)	(4483)	(3160)	(3447)	(2765)	(3496)	(3235)	(3570)	(2615)
Household size	3.91	3.94	3.82	3.91	3.92	3.89	3.74	3.96	3.58
	(1.62)	(1.54)	(1.51)	(1.64)	(1.45)	(1.70)	(1.59)	(1.65)	(1.29)
B. Health									
Diagnosed diabetic (=1)	0.67	0.67	0.68	0.66	0.62	0.68	0.62	0.67	0.59
	(0.47)	(0.47)	(0.47)	(0.47)	(0.49)	(0.47)	(0.49)	(0.47)	(0.50)
Blood sugar index	0.01	0.01	0.05	0.01	0.00	0.02	0.01	0.01	-0.15
	(0.92)	(0.93)	(0.91)	(0.92)	(0.95)	(0.90)	(0.99)	(0.93)	(0.83)
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)	192.42	191.32	196.07	192.51	195.58	193.26	193.30	192.23	177.38
	(89.39)	(88.73)	(86.67)	(89.87)	(91.54)	(88.25)	(98.14)	(90.42)	(77.00)
Systolic BP (mmHg)	133.34	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)	88.48	88.47	88.54	88.49	89.47	88.20	88.51	88.48	90.00
	(11.12)	(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	26.37	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	7009	7081	6906	7007	7068	6823	7446	7084	7018
	(3982)	(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)	29.26	30.80	29.48	28.98	28.61	29.69	28.55	28.45	30.05
	(49.42)	(48.79)	(48.68)	(49.88)	(38.54)	(52.08)	(63.65)	(47.96)	(36.59)
Yesterday's talk time (INR)	6.61	7.22	6.47	6.44	5.86	6.58	7.67	6.31	4.94
	(8.79)	(10.14)	(8.95)	(8.36)	(6.25)	(8.77)	(9.19)	(8.28)	(5.77)
Prefers daily payment (=1)	0.17	0.18	0.16	0.17	0.20	0.17	0.20	0.17	0.18
	(0.38)	(0.38)	(0.37)	(0.38)	(0.40)	(0.37)	(0.40)	(0.38)	(0.39)
Prefers monthly payment (=1)	0.24	0.25	0.28	0.24	0.27	0.24	0.23	0.24	0.26
	(0.43)	(0.43)	(0.45)	(0.43)	(0.45)	(0.43)	(0.42)	(0.43)	(0.44)
F-tests for Joint Orthogonality									
P-value (relative to control)	N/A	N/A	0.81	0.41	0.77	0.58	0.41	0.50	0.54
P-value (relative to monitoring)	N/A	0.81	N/A	0.95	0.86	0.85	0.59	0.98	0.71
P-value (relative to base case)	N/A	0.58	0.85	N/A	0.55	N/A	0.81	0.96	0.51
Sample size									
Number of individuals	3,192	585	203	2,404	166	902	164	1,106	66
Percent of sample	100.0	18.3	6.4	75.3	5.2	28.3	5.1	34.6	2.1
Number of ind. with ped. data	2,582	–	200	2,359	163	890	163	1,079	64

Notes: Standard deviations are in parentheses. BMI is body mass index, and BP is blood pressure. Overweight means BMI above 25. Hypertensive means systolic BP above 140 or diastolic BP above 90. The 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.2 column 5).

4.3 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.

4.3.1 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 significant at the 10% level (Table A.2, column 2). To address the imbalance, we report Lee (2009) bounds accounting for missing step data due to not wearing pedometers.[28] Our primary specifications do not condition on wearing the pedometer (instead setting steps and compliance to 0 on days when the pedometer was not worn), but we show that our results are robust to conditioning on wearing.

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 time that the participant put it on and the time that they took it off.[29] Reassuringly, Panel B of Table A.4 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 (Tables A.5 and A.6).

[28] We do not have participant pedometer data (e.g., because the pedometer broke or the sync was unsuccessful) on 6% of days. Missing pedometer data is balanced across incentive and monitoring groups (column 2, Table A.2). While our main specifications drop days with missing pedometer data, Table A.3 shows robustness to alternate specifications and Lee bounds. While missing data is balanced overall, one specific source of missing data (mid-intervention withdrawals) is imbalanced (column 5 of Table A.2), but results are robust to Lee bounds accounting specifically for that source (column 5 of Table A.3).

[29] 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.

Another concern is that participants might give their pedometers to someone else. Our data suggest that this concern is limited. First, we performed 836 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.4).

4.4 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.[30] As such, while piloting, we also decided to measure RBS, which is also strongly associated with diabetes severity (Bowen et al., 2015).[31] RBS is much easier to reliably measure in the field. Our measures of RBS and HbA1c both have predictive power for the other. (Table A.7 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.) 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 outcomes. We use these three measures to construct a standardized "health risk index" that also includes the two blood sugar measures.

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

[30] 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).

[31] 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).

5 Incentives and Chronic Disease

In this section, we explore whether incentives for exercise deliver results in the global fight against chronic disease. First, we test whether incentives increase exercise, both during and after the intervention. Exercise is a critical intermediate health outcome for this population since, in the long term, it reduces complications from diabetes and hypertension and averts premature mortality (World Health Organization, 2009). Second, we directly test whether incentives for exercise improve blood sugar and cardiovascular health.

5.1 Incentives and 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:

$$y_{it} = \alpha + \beta Incentives_i + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\lambda + \varepsilon_{it}, \tag{10}$$

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.

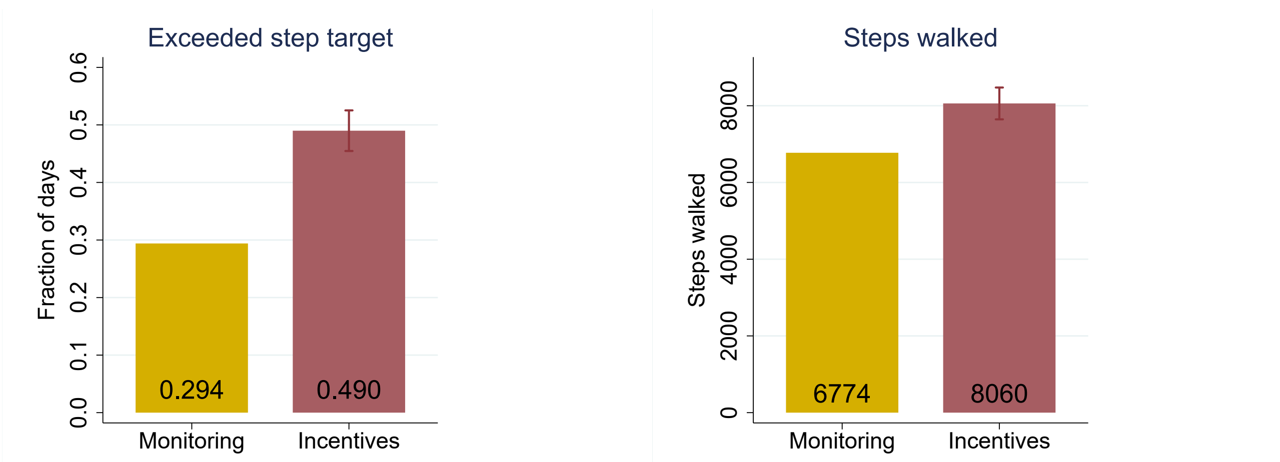
Incentives have large impacts on walking, increasing the share of days that participants reach their 10,000-step target by 20 pp (column 1 of Table 2 and Figure 2(a)). This effect does not simply reflect participants shifting steps from one day to another: column 2 of Table 2 and Figure 2(b) show that incentives increase walking by 1,266 steps per day, roughly a 20 percent increase that is equivalent to approximately 13 minutes of extra brisk walking each day, on average. 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 during the intervention period. Because monitoring itself may have an independent positive impact, these estimates are likely conservative for the overall impact of incentives. That said, a comparison of steps within the monitoring group between the baseline and intervention periods (controlling for time effects) suggests that, while monitoring may modestly increase the likelihood of exceeding the step target, it does not appear to increase steps (Online Appendix K).

Table 2: Incentives Increase Average Walking

Dependent variable:	Exceeded step target	Daily steps	Daily steps (if > 0)
	(1)	(2)	(3)
Incentives	0.196***	1286.2***	1144.3***
	[0.0180]	[211.4]	[190.3]
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 (10) 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. All specifications control for the average of the dependent variable during the phase-in period (before randomization) and its second order polynomial, a dummy for the SMS treatment, and a set of controls selected separately for each specification using the post double selection lasso method of Belloni et al. (2017). The set of controls lasso selected over included the following individual-level controls: age, weight, height, gender, and their second-order polynomials, as well as the following day-level-controls: month-year and day-of-week fixed effects. In columns 1 and 2, lasso selected age and a subset of the year-month fixed effects and day-of-week fixed effects. In column 3, lasso selected female and a subset of the day-of-week fixed effects. 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%.



(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 selected by lasso in Table 2. Panel A shows the average probability of exceeding the daily step target; Panel B shows average daily steps walked.

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.3 reports Lee bounds which account for the non-random patterns of missing data. Both strategies find similar effects. The estimates are also robust to excluding the control variables from the regression (Table A.8).

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.

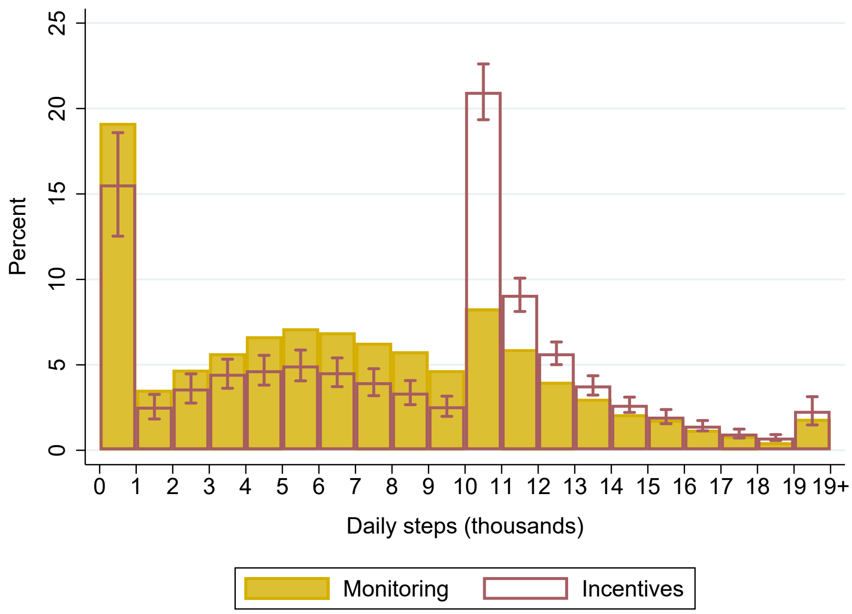
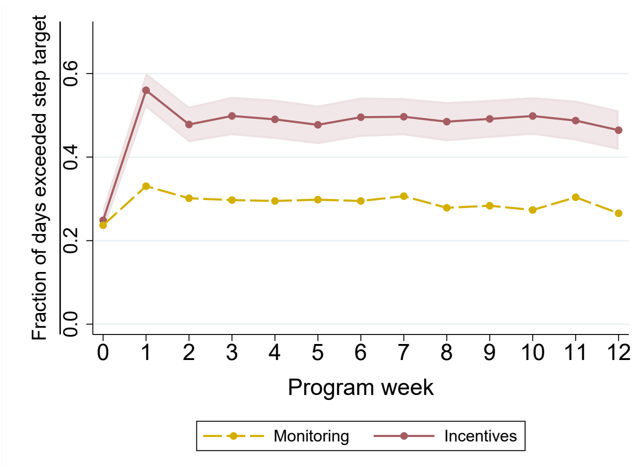


Figure 3: Incentives Shift the Distribution of Steps Walked per Day

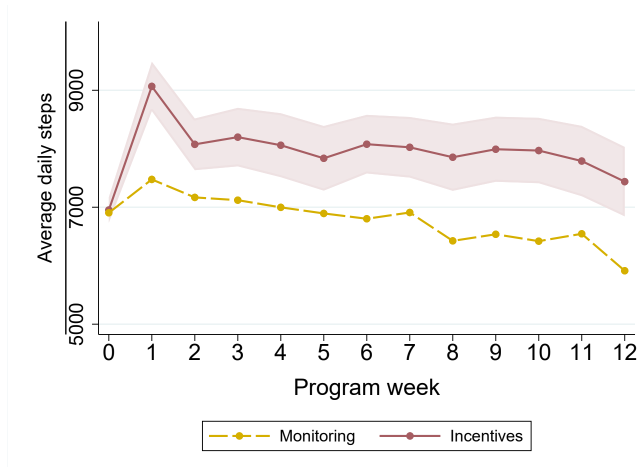
Notes: The figure displays the impact of the pooled incentive groups relative to the monitoring group during the intervention period. The confidence intervals represent tests of equality between the incentive and monitoring groups with the same controls selected by lasso in Table 2.

5.1.1 Evolution over Time and Persistence of Exercise Effects

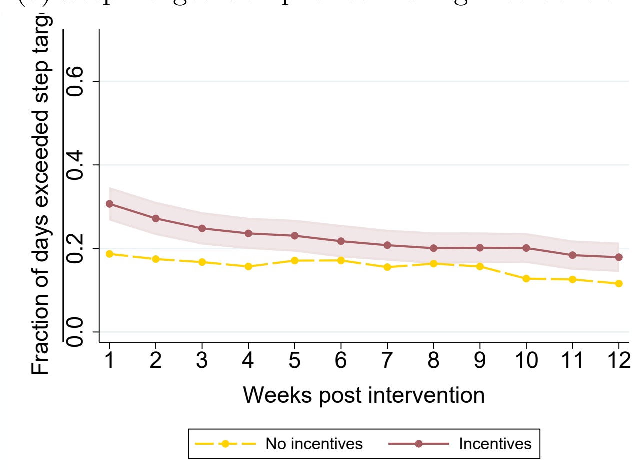
We now analyze how the exercise impacts evolve over time, both during and after the intervention. We begin with their evolution during the intervention. Panels A and B of Figure 4 estimate equation (10) separately by week of the intervention. After an initial spike at week 1, the effect of incentives on walking remains stable during the full intervention period. This suggests that policymakers could extend the program further with similar effects, an interesting finding since insurers and governments are increasingly rolling out longer-term (and even permanent) incentive programs.



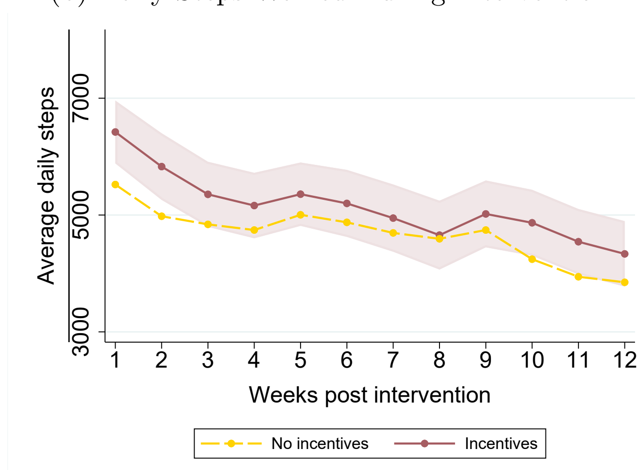
(a) Step-Target Compliance During Intervention



(b) Daily Steps Walked During Intervention



(c) Step-Target Compliance Post Intervention



(d) Daily Steps Walked Post Intervention

Figure 4: Incentive Effects are Steady through the 12-Week Program and Persist Afterward

Notes: Panels A and C show the average probability of exceeding the step target and Panels B and D show the average daily steps walked. Panels A and B depict the intervention period and Panels C and D depict the 12 weeks subsequent to the intervention. Week 0 in Panels A and B is the phase-in period (before randomization). "No incentives" in Panels C and D represents the pooled monitoring and control groups; the panels look very similar when we compare with the control group only (Online Appendix Figure G.1). 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 selected by lasso in Tables 2. All graphs are unconditional on wearing the pedometer. See Figure A.2 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 panels B and D is that pedometer wearing rates declined over time.

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 (10) using the pedometer data from the 12 weeks after the intervention ended. Since we have data from the control group as well during this period, the *Incentives* coefficient now represents the effect of incentives relative to the control and monitoring groups pooled. We pool the comparison groups for power given that sample size is limited because we only have data for the post-intervention period from a third of our sample.[32]

Table 3 shows that the incentive group walks significantly more then the pooled comparison groups even after incentives end. The treatment effect on steps is statistically significant and large: around 8% of the comparison group mean in both 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.[33] Panels C and D of Figure 4 suggest that some of the effect persisted until the end of the measurement period. Our short-run incentive program may thus induce habit formation, enabling long-term impacts.

Table 3: 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	0.063***	462.1**	611.5***
	[0.01]	[221.58]	[196.90]
No incentives mean	0.216	5,687	7,347
# 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 columns 1 and 2 only include days where the participant wore the pedometer (i.e., had step count > 0). Controls are selected separately for each column using the post-double selection lasso method of Belloni et al. (2017), where the set of controls that lasso selected from is the same as in Table 2. In each column, Lasso only selected a subset of the year-month and day-of-week fixed effects in all specifications (with the specific subset chosen varying by column). Table A.9 shows that the results are robust to excluding controls. 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 (Table A.5 column 5), and Table A.6 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%.

[32]The results are similar when we compare incentives with control alone (Online Appendix Figure G.1); with only 72 people, the post-intervention monitoring group is too small to analyze alone.

[33]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 K suggests that monitoring does not affect steps.

5.2 Health Effects

We now examine whether the incentives program measurably improves health in the short run. Our experiment was powered to detect the difference between the incentive and pure control groups. We lack statistical power to compare health outcomes (which are relatively noisy) in the monitoring group with the other groups, although we show it for completeness. Table 4 reports results from regressions of the following form:

$$y_i = \alpha + \beta_1 Incentives_i + \beta_2 Monitoring_i + \boldsymbol{X}'_i\gamma + \varepsilon_i, \tag{11}$$

where y_i is a health outcome at endline for individual i, and $\boldsymbol{X}_i$ is a vector of controls, shown in the table notes. The coefficient of interest is β_1, the effect of incentives relative to the control group.

Table 4 shows that the incentive program moderately improves blood sugar and overall cardiovascular health. Column 1 presents the treatment effect on our preferred blood sugar measure, the standardized index incorporating both the longer-term HbA1c and shorter-term RBS measures of blood sugar control. Incentives improve the index by 0.05 standard deviations. Columns 2 and 3 display HbA1c and RBS separately. Column 4 shows that incentives improve the overall health risk index by 0.05 SDs, 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. To do so, we estimate the following regression:

$$\begin{aligned} y_i =& \alpha + \beta_1 Incentives_i + \beta_2 Incentives_i \times LowBloodSugar_i + \beta_3 Monitoring_i \\ &+ \beta_4 Monitoring_i \times LowBloodSugar_i + \beta_5 LowBloodSugar_i + \boldsymbol{X}'_i\gamma + \varepsilon_i, \end{aligned} \tag{12}$$

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 values of the blood sugar index at baseline (i.e., with $LowBloodSugar = 1$). β_2 then allows us to test if the effect among those with above-median baseline blood sugar is significantly different from those without.

The results, shown in columns (5)-(8) of Table 4, indicate that the treatment effects are larger among those with more severe diabetes, although we can only reject equality for the blood sugar index (at the 10% level). Among the sample with above-median blood sugar, incentives decrease the blood sugar index by 0.10 SDs and decrease RBS by 12 mg/DL.

Both the full sample and subsample treatment effects on blood sugar are moderately-sized but meaningful from a clinical perspective.[34] In addition, an exploratory analysis of the treat-

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

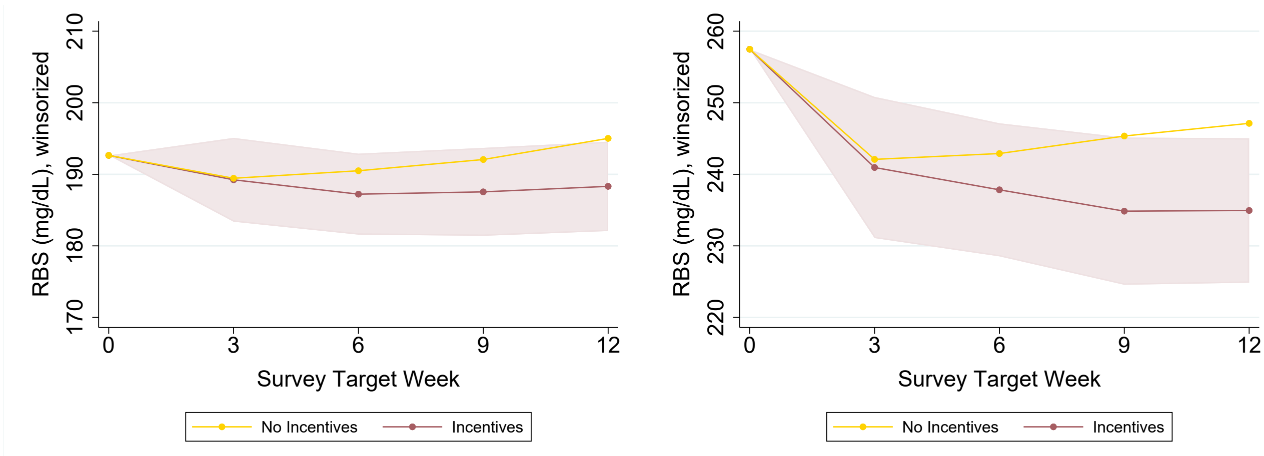
Table 4: Incentives Moderately Improve Blood Sugar and Cardiovascular Health

	Full sample effects				Heterogeneity by baseline blood sugar			
Dependent variable:	Blood sugar index	HbA1c	Random blood sugar	Health risk index	Blood sugar index	HbA1c	Random blood sugar	Health risk index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Incentives	-0.05**	-0.08	-6.0*	-0.05*	-0.10**	-0.1	-11.6**	-0.08**
	[0.03]	[0.07]	[3.5]	[0.03]	[0.05]	[0.1]	[5.9]	[0.04]
Incentives × below-median blood sugar					0.09*	0.1	11.0	0.08
					[0.05]	[0.1]	[7.0]	[0.05]
Monitoring	-0.03	-0.1	1.3	0.01	-0.07	-0.3	1.2	-0.05
	[0.05]	[0.1]	[6.6]	[0.04]	[0.08]	[0.2]	[10.5]	[0.07]
Monitoring × below-median blood sugar					0.08	0.3	-1.5	0.1
					[0.09]	[0.2]	[12.6]	[0.09]
p-value: I = M†	0.517	0.609	0.220	0.146	0.627	0.309	0.171	0.531
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 (11). Columns 5-8 display display OLS estimates of equation (12); note that $LowBloodSugar_i$ in equation (12), 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 G.3 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 G.4 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 G.5 for treatment effects on the other components of that index not shown here. Each physical health outcome that either appears as an outcome variable or is an index component is trimmed using World Health Organization guidelines to trim biologically implausible health outcome measurements (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 a set of controls selected separately for each specification using the post double selection lasso method of Belloni et al. (2017). The set of controls lasso selected over included the following controls: age, weight, height, gender, and their second-order polynomials, as well as month-year and day-of-week fixed effects for endline completion dates. The only control selected was for HbA1c, where lasso selected a subset of the month-year fixed effects for the endline completion date; no additional controls were selected for any other dependent variables. Columns 5-8 additionally control for the indicator for below-median blood sugar. Table A.10 shows that the estimates are similar, just less precise, when we omit the control variables from the regressions. 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 5 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.

Figure 5: Blood Sugar Treatment Effects Grow Over Time



(a) Full sample

(b) Above-median blood sugar sample

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 if we compare incentives with control alone, precision is just slightly lower; see Table G.6 in the Online Appendix. For each survey, we regress random blood sugar on the incentives dummy and control for the same controls selected by lasso for the random blood sugar specification in Table 4. The shaded areas represent a collection of 95% confidence intervals from those regressions. The p-values for the significance of the increase over time are .06 and .02 for the Panels A and B, respectively (see Table G.6 in the Online Appendix).

Table 5 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 G.7).

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 5: Incentives Also Improve Mental Health

Dependent variable:	Mental health index	Fitness time trial index
	(1)	(2)
Incentives	0.092**	0.015
	[0.045]	[0.045]
Monitoring	0.16**	0.071
	[0.072]	[0.074]
p-value: Incentives = Monitoring	0.255	0.386
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 a set of controls selected separately for each specification using the post double selection lasso method of Belloni et al. (2017). The set of controls lasso selected over included the following controls: age, weight, height, gender, and their second-order polynomials, as well as month-year and day-of-week fixed effects for endline completion dates. 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 G.8 for treatment effects on the individual components of the mental health and fitness indices. Each component of the fitness time trial index is trimmed using World Health Organization guidelines to trim biologically implausible health outcome measurements (i.e., z-scores < -4 or > 4). Robust standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

5.3 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 a scalable and effective lifestyle intervention that can be delivered 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 generate 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, our results suggest that incentive programs could be an important tool to help decrease the burden of chronic disease in India. Given these promising results, we now examine how to further improve the program in the face of impatience.

6 Incentives, Impatience, and Time-bundled Contracts

This section investigates the implications of impatience for the design of incentives, primarily exploring time-bundled thresholds, our contract variation designed to improve effectiveness in the face of impatience over effort. First, we compare the effectiveness of the time-bundled threshold and linear contracts in the full sample. Second, we test our main theoretical prediction: that time-bundled thresholds increase compliance and effectiveness more among those who are impatient over effort than among those who are not (Prediction 1). Third, we evaluate a more standard strategy for improving compliance and effectiveness in the face of impatience (over payment): increasing the frequency of payment. Finally, we discuss the potential policy implications of our findings and the welfare implications of improving contract effectiveness.

6.1 Average Effectiveness of Time-bundled Threshold Contracts

We first compare the sample-average performance of the threshold and linear contracts. Our theoretical analysis (e.g., Appendix Proposition 3) suggests that, under plausible conditions—such as the effort discount rates being sufficiently high—time-bundled thresholds can be more effective overall than linear contracts, making this an interesting comparison.

In order to establish that 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 (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. To test for differences across the incentive treatment groups, we estimate regressions of the following form:

$$y_{it} = \alpha + \sum_j \beta_j \times \left(incentives^j\right)_i + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\theta + \varepsilon_{it}, \tag{13}$$

where y_{it} are daily walking outcomes and $(incentives^j)_i$ is an indicator for whether individual i is enrolled in incentive treatment group $j \in$ (daily, base case, monthly, threshold, small payment). The β_j coefficients capture the average effect of each incentive treatment group relative to the monitoring group. Table 6 displays the results.

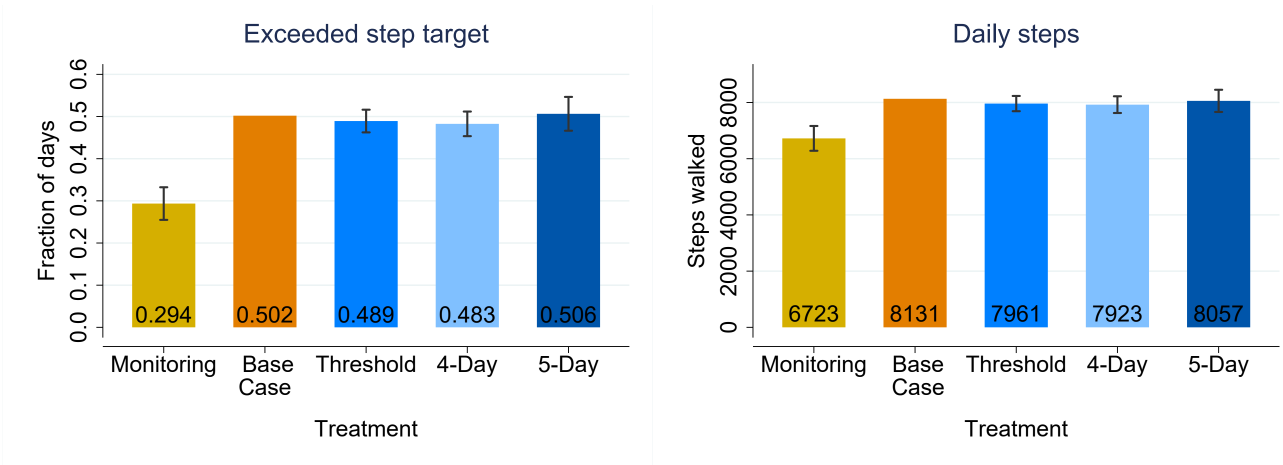
The effect of the threshold treatment on compliance is very similar to the effect of the base

Table 6: All Incentive Contracts Increase Walking

Dependent variable:	Exceeded step target	Daily steps	Daily steps (if > 0)
	(1)	(2)	(3)
Base case	0.207***	1411.4***	1197.1***
	[0.0196]	[225.0]	[201.8]
Daily	0.199***	1126.7***	1245.1***
	[0.0302]	[332.2]	[279.2]
Monthly	0.179***	1302.6***	1152.5***
	[0.0281]	[311.0]	[272.3]
Threshold	0.194***	1238.1***	1125.5***
	[0.0194]	[223.2]	[200.3]
Small payment	0.128***	740.8*	510.5
	[0.0382]	[381.0]	[331.3]
P-value for base case vs			
Daily	0.79	0.31	0.83
Monthly	0.27	0.67	0.84
Threshold	0.35	0.21	0.55
Small payment	0.03	0.05	0.02
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 (13) using daily intervention-period pedometer data. Each column uses the same controls selected by lasso in Table 2; the results are robust to excluding controls (Online Appendix Table G.9). 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%.

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 6 displays the result graphically. It also shows the 4-day threshold group and 5-day threshold groups separately—neither has meaningfully different compliance than the base case.



(a) Probability Exceeded Step Target

(b) Average Daily Steps

Figure 6: Adding a Time-Bundled Threshold Does Not Significantly Affect Average Walking

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

Cost-effectiveness and Overall Effectiveness However, 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. 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.11).

Because the threshold contracts have the same compliance and are more cost-effective than the base case, they are more effective overall. For comparison, the small payment treatment is also more cost-effective than the base case (it pays half as much per day complied), but this comes at the cost of reduced compliance (Table 6).

Variance and Effectiveness in Other Settings Equal compliance and higher cost-effectiveness only necessarily imply higher effectiveness when the benefits of compliance are linear (Section

2). While the health benefits of compliance appear to be linear in our setting (e.g., Warburton et al., 2006), there are many other settings of interest with nonlinear benefits. In those settings, effectiveness will depend not just on average compliance but on the variance of compliance across payment periods.

Theory suggests that thresholds can increase the variance of compliance by decreasing the likelihood of intermediate effort (i.e., effort just below the threshold) (Grant and Green, 2013). To assess this prediction empirically, Figure 7 displays histograms of the number of days the step target was met per week in the threshold and base case groups. The threshold contracts significantly increase variance, causing more individuals to achieve their step target zero or seven days in the week.[35] This implies that, relative to linear contracts, thresholds are likely to be less effective in settings where the benefits of compliance are concave, and more effective in settings where the benefits of compliance are convex.

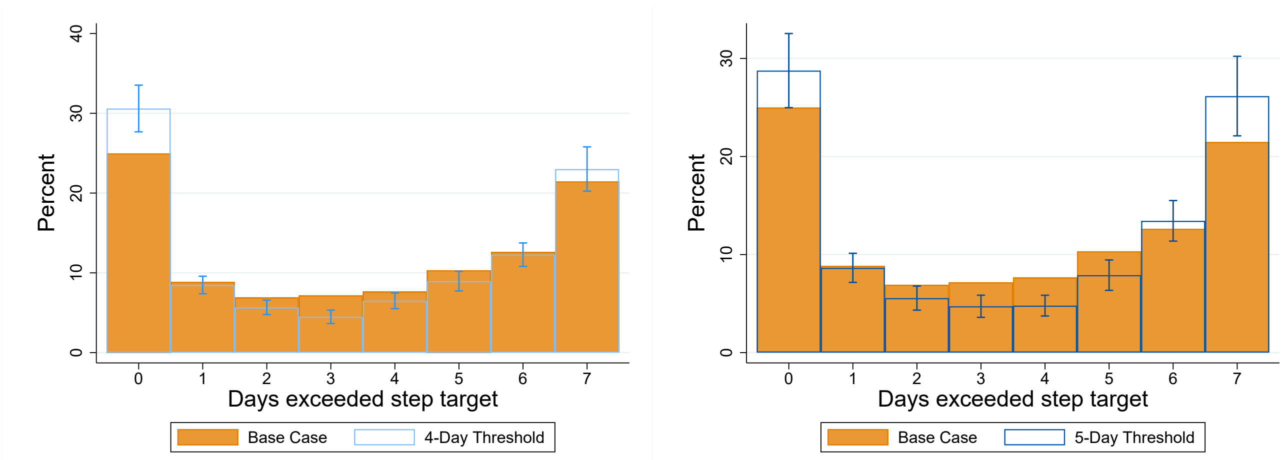


Figure 7: Time-Bundled Threshold Contracts Increase Variance Across Weeks

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

6.2 Heterogeneity in Threshold Effects by Impatience over Effort

We now assess whether time-bundled threshold contracts perform better relative to linear contracts when individuals are more impatient over effort (Prediction 1). To do so, we perform two analyses. First, we examine the heterogeneity by impatience in the effect of the threshold relative to the base case. Since Prediction 1 regards heterogeneity in the threshold effect *holding*

[35]Online Appendix Table G.10 substantiates this conclusion, presenting a formal test of the effect of the threshold on the variance of week-level compliance. While the increase in dispersion and in zeroes in the threshold treatment is consistent with theory, the increase in density at seven days in particular (instead of at the specific threshold level of four or five) is perhaps surprising. Potential explanations include habit formation or that it is easier to schedule walking every day in a given week than on a subset of days.

all else constant, this heterogeneity analysis will only be a direct test of the theory if impatience is not correlated with other variables that influence the effectiveness of the threshold. To shed light on whether this condition holds here, we control for many covariates interacted with the threshold and show that the estimated relationship is robust. Moreover, even if there were omitted variables affecting the estimate, our heterogeneity estimate may be the one that is more 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.

Second, to tie our data to our theory more precisely, in Appendix C, we calibrate a model to determine whether the gap in predicted compliance between the threshold and linear contracts varies 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.

We next describe how we measure impatience over effort before presenting our results on heterogeneity in the effect of thresholds by impatience over effort.

6.2.1 Measurement of 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 somewhat elderly and had difficulty with the more complicated screen-based measures often used in the literature (e.g., Andreoni and Sprenger 2012), we included simple measures that the full sample could comprehend.

Our primary measure of impatience over effort is an index of easy-to-comprehend, survey-based measures of impatience and procrastination taken from the psychology literature. Specifically, the questions are a subset of the Tuckman (1991) and Lay (1986) scales, with the subset chosen *ex ante* by our field team as being most appropriate for our setting. The questions, listed in Panel A of Table A.12, ask respondents to respond on a Likert scale of agreement with statements such as "I'm continually saying 'I'll do it tomorrow'."

These scales have been validated as being predictive of real behaviors such as poor academic performance (Kim and Seo, 2015). Indeed, this index (hereafter: 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.12). We construct the index by standardizing all question responses and taking the average, as we specified when we included the questions in the survey.

In Online Appendix I, we perform additional validation of our impatience index by showing that it predicts an incentivized measure of impatience over effort and that it does not predict an incentivized measure of impatience over payments. To do so, we gathered data from a separate sample of similar participants and elicited their incentivized choices 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 did this after the original experiment was complete; we were unaware of the Augenblick (2018) methodology when we conducted our experiment in 2016.) Reassuringly, we show that those with higher values of our impatience index also make more effort-impatient choices in the Augenblick (2018) exercise, signing up for relatively more tasks in the future than the present.[36] In contrast, we show that those with higher values of our impatience index do not make more impatient choices in the payment (mobile recharge) domain, as shown using incentivized choices on a multiple price list (Andreoni and Sprenger, 2012). Interestingly, there is also no correlation between the incentivized measures of impatience over effort and over recharges. The discount rates over the two domains may thus be relatively independent here. Indeed, in the data collected for our main study, we also find that our impatience index does not correlate with any proxies for impatience over recharges, such as recharge balances and recharge usage (Table A.13).

We began collecting our impatience index partway through the data collection,[37] so it is only available for the latter half of the sample. Luckily, that sample size is sufficient to achieve statistically significant results. That said, to check the robustness of our results in the full sample, we create a "predicted index" using a LASSO prediction based on three survey questions on self-control in specific domains (e.g., exercise, diet) that were similar in nature to the impatience index questions and were collected from all participants. Panel B of Table A.12 lists the questions used for prediction and shows that the predicted index correlates in the expected direction with behavior measures such as the health risk index.

[36]Specifically, Appendix Figure I.2(a) shows that those with above-median values of our impatience index have over twice as large a gap between the tasks chosen for the future relative to the present than those with below-median values of our impatience index. We also perform a structural estimation, following Augenblick (2018), DellaVigna and Pope (2018), and John and Orkin (2021), which also suggests that our impatience index meaningfully predicts the discount factor over effort. We estimate the average discount factor over effort costs incurred 1, 7, and 8 days in the future. Across the full sample, we estimate a discount factor of 0.9, which is both economically and statistically different from 1. However, only those with above-median values of the impatience index appear to have a discount factor less than 1. The estimated discount factor among those with below-median values of the impatience index is economically and statistically indistinguishable from 1.

[37]We initially planned to use the convex time budget (CTB) methodology of Andreoni and Sprenger (2012) to measure impatience over effort costs, as well as over mobile recharges, as described in Online Appendix J. However, these measures are difficult to implement in the field. Challenges surfaced early in our data collection which made the measures unusable for analysis, at which point we added our impatience index measure.

The key challenge for us in implementing CTB was that it was hard to get respondents to understand the paradigm, and likely as a result, we have an order of magnitude more law-of-demand violations than lab-based studies with college students. Other evidence suggesting a lack of understanding include our estimates not converging for 44% of the sample and respondents failing to follow through on their chosen allocations, as described in Online Appendix J. Further, the impatience measures estimated using this methodology do not correlate in the expected direction with any behaviors.

6.2.2 Heterogeneity by Baseline Impatience

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

$$y_{it} = \alpha + \beta_1 Impatience_i \times Thresh_i + \beta_2 Thresh_i + \beta_3 Impatience_i + \boldsymbol{X}_i' \pi + \boldsymbol{X}_{it}' \theta + \varepsilon_{it}, \quad (14)$$

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 7 shows that, consistent with our prediction, 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 below the median, a large increase relative to the sample-average effect of either contract (20 pp). The difference is significant at the 10% level. 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 slightly more precise) results.

Figure 8 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 6 pp effect. We previously showed that, theoretically, the discount rate over effort could be pivotal to whether the linear or threshold contract has higher compliance (e.g., Proposition 3 in Appendix B.2). The fact that that is the case here has important implications for policy: efforts by policymakers to personalize threshold assignment based on agent impatience could substantially increase compliance.

Impatience over effort is correlated with other factors, such as baseline exercise levels, that

Table 7: Time-Bundled Thresholds Increase Compliance More for the Impatient

Dependent variable:	Exceeded step target ($\times$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 $\times$ Threshold	3.67**	5.8*	3.1***	6.39*
	[0.15, 7.19]	[-0.62,12.23]	[0.87, 5.18]	[-0.40, 10.41]
Threshold	-1.51	-3.94	-1.1	-3.47**
	[-4.85, 1.82]	[-8.97,1.09]	[-3.33, 0.88]	[-5.93, -0.15]
Impatience	-3.62***	-6.14**	-2.05**	-4.74*
	[-6.06, -1.19]	[-11.09,-1.19]	[-3.54, -0.46]	[-7.77, 0.58]
# 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 G.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 (14). The Gaussian standard errors and p-values for the column 1 $Impatience \times Threshold$ coefficient are 1.9 and 0.053, respectively; for column 2, the corresponding values are 3.81 and 0.128. Controls are the same as selected by lasso in Table 2. Significance levels: * 10%, ** 5%, *** 1%.

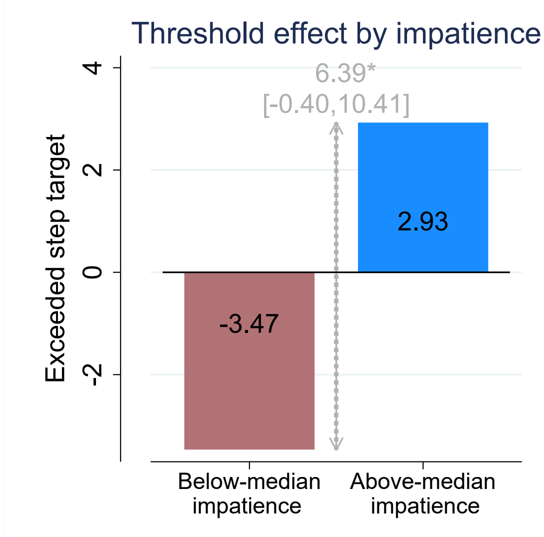


Figure 8: Impatience is Pivotal to Compliance Under the Time-Bundled Threshold

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 7 column 4.

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 9 examines the robustness of the Table 7 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 9 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.

Another potential confound that was difficult to measure at baseline (and hence which we

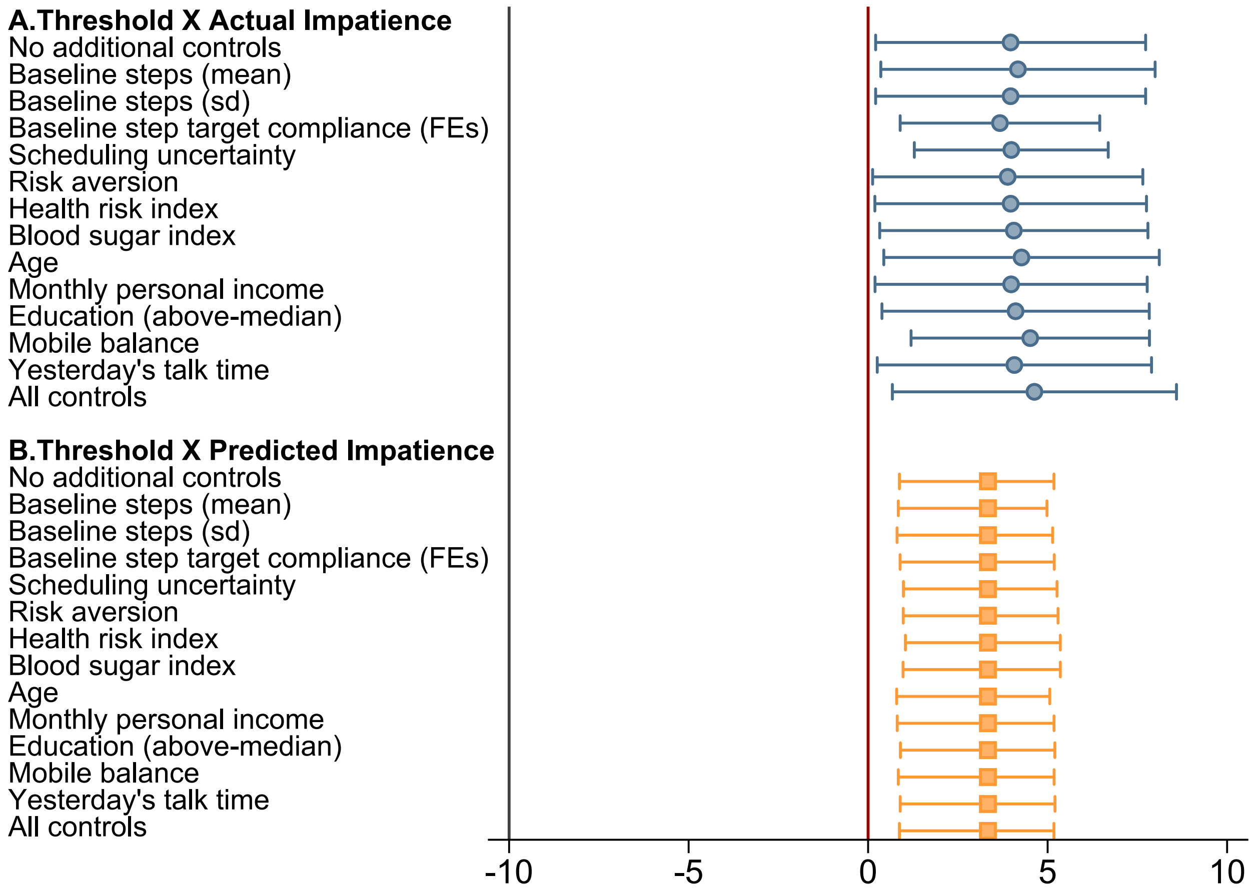


Figure 9: Threshold Heterogeneity by Impatience is Robust to a Variety of Controls

Notes: Panel A displays robustness of the $Threshold \times Impatience$ coefficient from column (1) of Table 7 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 7. The next 11 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 11 rows (both main effects and interactions with $Threshold$). Panel B is analogous but based on column (3) of Table 7. 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, resepctively, 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 constructed via bootstrap, with bootstrap draws done at the individual level, as in Table 7.

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 G.12 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.

Prediction 1 suggested that, in addition to increasing compliance more among the impatient, threshold contracts should also increase *effectiveness* more among the impatient. Since we have already demonstrated the compliance result, demonstrating the effectiveness result requires us to show that, relative to the base case, thresholds do not have lower cost-effectiveness among the impatient than the patient. Table A.11 shows that this is true.

6.2.3 Time-bundled Thresholds Result Summary

Consistent with our theoretical predictions, time-bundled thresholds generate meaningfully greater compliance and effectiveness among the impatient than the patient. In the full sample, they increase effectiveness by increasing cost-effectiveness without decreasing compliance. Taken together, these findings suggest that impatience in our sample could contribute to the good performance of the time-bundled threshold contract here.

These findings have important policy implications, suggesting that time-bundled thresholds are a useful tool to adjust incentives for impatience over effort. Policymakers could use time-bundled thresholds when they are incentivizing more impatient populations. They could also personalize the assignment of time-bundled thresholds within a population, for example by assigning them based on observable predictors of impatience.[38]

6.3 Payment Frequency

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

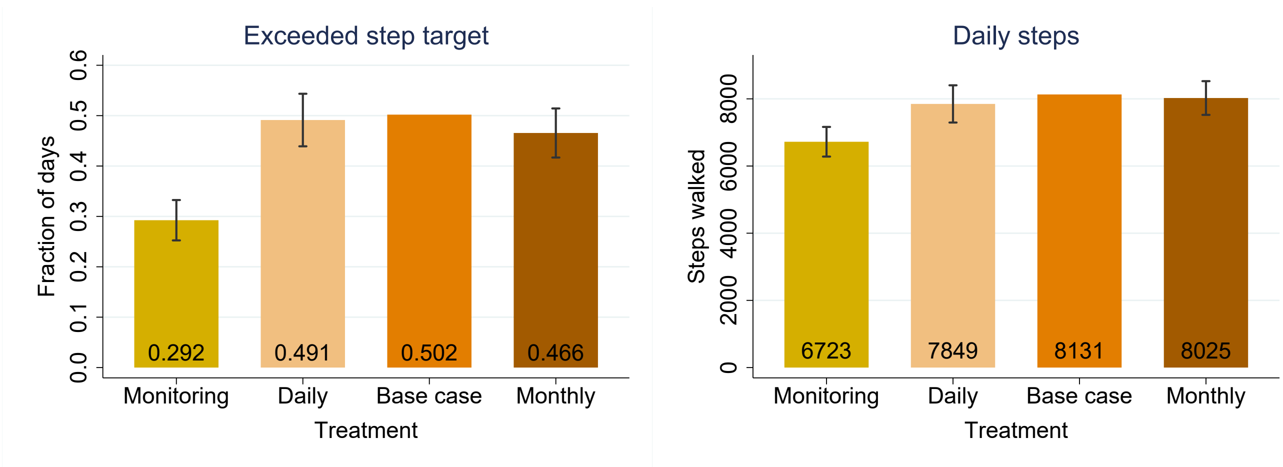
1. Between-treatment: We compare average compliance in the daily, weekly (base case), and monthly groups. We assess how payment frequency affects compliance and use Prediction 2 to shed light on the discount rate over payment.
2. Within-treatment: Within the base case and monthly groups, we examine how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

The approaches are complementary. The between-treatment approach answers the policy

[38]While such an assignment mechanism might give participants incentive to manipulate their observables, particularly if the time-bundled threshold contract is dominated by the linear contract as it is here, Dizon-Ross and Zucker (2022) shows that, in the domain of incentives for behavior change, participants do not manipulate observables to avoid assignment to dominated contracts.

question of whether payment frequency matters, while the within-treatment approach has more statistical power. The within-treatment approach also rules out potential confounds for making inferences about discount rates over payment using between-treatment effects.[39]

Between Treatment Figure 10 and Table 6 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.



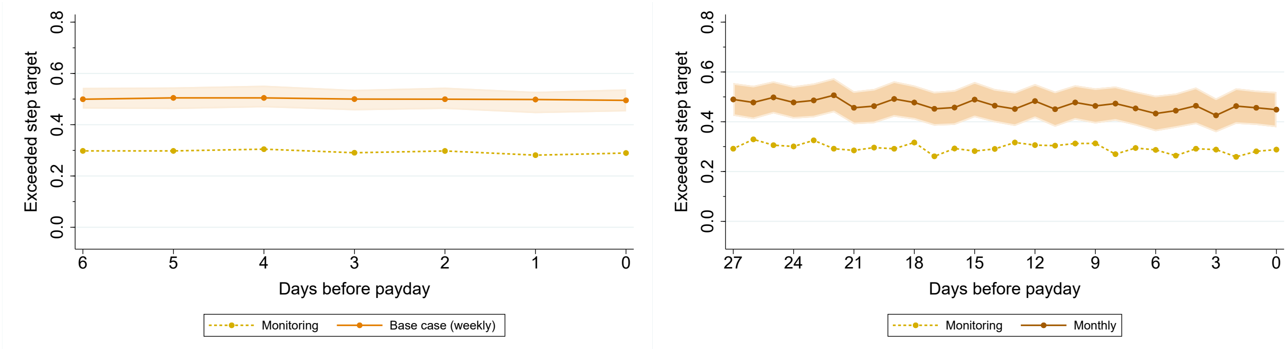
(a) Probability Exceeded Step Target (b) Average Daily Steps

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 selected by lasso in Table 2.

We thus do not find evidence that increasing payment frequency in the range from daily to monthly affects compliance—a perhaps surprising finding given the conventional wisdom. The lack of between-treatment frequency effects implies that the discount rate over our financial payments is small and has a relatively flat shape over the range from one day to one month. One important caveat to these results is that the between-treatment effects are somewhat imprecise, and we have limited power to reject large discount rates. We address this issue with the within-treatment analysis.

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



(a) 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.

Within treatment The within-treatment analysis confirms the suggestive evidence of flat and low discount rates from the between-treatment analysis. Figure 11 shows how compliance within the base case weekly (Panel A) and monthly (Panel B) treatments changes as the payment date approaches. If agents are impatient over payments, compliance should increase as the payday approaches (Prediction 3), yet we find that walking behavior is remarkably steady across the payment cycle. We also estimate 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, shown in Online Appendix Table G.13, are not significantly different from zero and suggest that, if anything, compliance *decreases* as the payment date approaches.

Our estimates are also more precise here than in the between-treatment analysis, 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 0.5 pp more compliance than weekly.

Although these results are consistent with recent evidence from other settings in showing limited discounting over payments (e.g., Augenblick et al. 2015; DellaVigna and Pope 2018), the absence of payday spikes conflicts with Kaur et al. (2015). The reasons why discount rates over payment vary across contexts are an open question for future work.

6.4 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 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 and/or social welfare but did not cause a Pareto improvement, instead decreasing the welfare of some individuals relative to a no-incentives benchmark. This concern is particularly vivid in light of evidence that pre-commitment contracts can decrease welfare among partially naïve individuals who pay upfront for commitment but fail to follow through (e.g., Bai et al., 2020).

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.[40] 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 G.14). 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 G.15).

7 Conclusion

This paper makes two important contributions. First, 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

[40] 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 18 for more detail.

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.

Second, 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 agents 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 agents. The intuition behind the prediction is that agents who discount their future effort more place a higher value on future work opportunities. Time-bundled contracts link better future work opportunities with effort today, thus providing particular motivation for the impatient to exert more effort today. 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 people with high primitive discount rates may not be. In contrast, the success of time-bundled contracts relies on high primitive discount rates.

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 interesting topic to study is how to personalize time-bundled contracts at scale at the individual level. One option is to use targeting based on observables as done in Dizon-Ross and Zucker (2022) for a different dominated contract characteristic. Together, the answers to these questions will allow policymakers to effectively employ time-bundled contracts to motivate impatient people.

References

Acland, D. and M. Levy (2015). Naiveté, projection bias, and habit formation in gym attendance. *Management Science 61*(1), 146–160.

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.

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

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

Banerjee, A. V., E. Duflo, R. Glennerster, and D. Kothari (2010). Improving immunisation coverage in rural india: Clustered randomised controlled evaluation of immunisation campaigns with and without incentives. *BMJ (Online) 340*(7759), 1291.

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.

Belloni, A., V. Chernozhukov, I. Fernandez-Val, and C. Hansen (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica 85*(1), 233–298.

Bommer, C., E. Heesemann, V. Sagalova, J. Manne-Goehler, R. Atun, T. Bärnighausen, and S. Vollmer (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.

Carpenter, R., T. DiChiacchio, and K. Barker (2019). Interventions for self-management of type 2 diabetes: An integrative review. *International Journal of Nursing Sciences 6*(1), 70–91.

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.

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

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

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

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, J. E. Garrett, T. Gilmer, H. Vue-Her, S. Rinn, K. Engel, 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, E. E. K. Nang, and K. R. Evenson (2016). Effectiveness of activity trackers with and without incentives to increase physical activity (trippa): A randomised controlled trial. *The Lancet Diabetes and Endocrinology 4*(12), 983–995.

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.

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

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

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

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.

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, P. T. Katzmarzyk, J. R. Alkandari, L. B. Andersen, A. E. Bauman, R. C. Brownson, F. C. Bull, C. L. Craig, U. Ekelund, S. Goenka, R. Guthold, P. C. Hallal, W. L. Haskell, G. W. Heath, S. Inoue, S. Kahlmeier, H. W. Kohl, E. V. Lambert, G. Leetongin, R. J. Loos, B. Marcus, B. W. Martin, N. Owen, D. C. Parra, M. Pratt, D. Ogilvie, R. S. Reis, J. F. Sallis, O. L. Sarmiento, and J. C. Wells (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, E. Lin, M. L. Maciejewski, and J. P. LoGerfo (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, J. Heuer, S. Sproat, C. Hyson, N. Haff, S. M. Lee, L. Wesby, K. Hoffer, D. Shuttleworth, D. H. Taylor, V. Hilbert, J. Zhu,

L. Yang, X. Wang, and K. G. Volpp (2016, mar). Framing financial incentives to increase physical activity among overweight and obese adults. *Annals of Internal Medicine 164*(6), 385.

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).

Ranasinghe, P., R. Jayawardena, N. Gamage, N. Sivanandam, and A. Misra (2021, 6). Prevalence and trends of the diabetes epidemic in urban and rural india: A pooled systematic review and meta-analysis of 1.7 million adults. *Annals of Epidemiology 58*, 128–148.

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.

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, M. F. Hu, Y. Tao, J. Zhu, J. D. Park, G. Loewenstein, D. A. Asch, and K. G. Volpp (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.

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, J.-Y. Chin, J. Matson, J. Anarella, P. Roohan, F. Gesten, 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.

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.

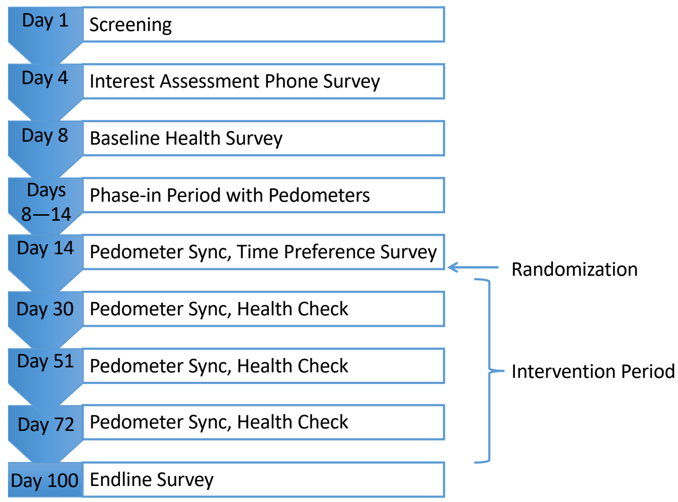
World Health Organization (2009). Global health risks, mortality and burden of disease attributable to selected major risks.

World Health Organization (2020). Global health estimates 2020: deaths by cause, age, sex, by country and by region, 2000–2015. *Geneva: WHO*, 2020.

Appendices

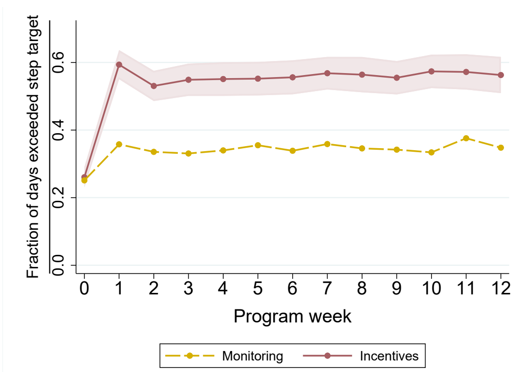
This section contains all appendix tables and appendix figures labeled with the prefix "A" (e.g., Table A.1, Figure A.1). It also contains Appendices B and C. The Online Appendix contains Appendices D - L and is available at: faculty.chicagobooth.edu/-/media/faculty/rebecca-dizon-ross/research/incentivedesignapp.pdf

Appendix Figure A.1: Experimental Timeline for Sample Participant

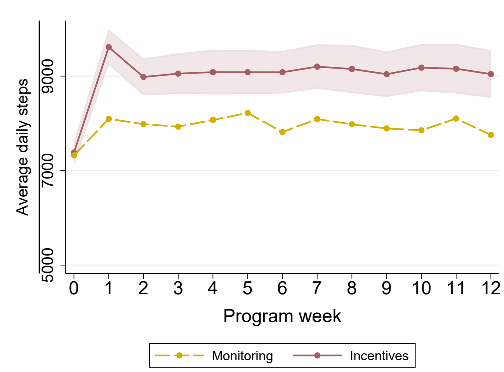


Notes: This figure shows an experimental timeline for a participant. Visits were scheduled according to the participants' availability. We introduced variation into the timing of incentive delivery by delaying the start of the intervention period by one day for randomly selected participants. The intervention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

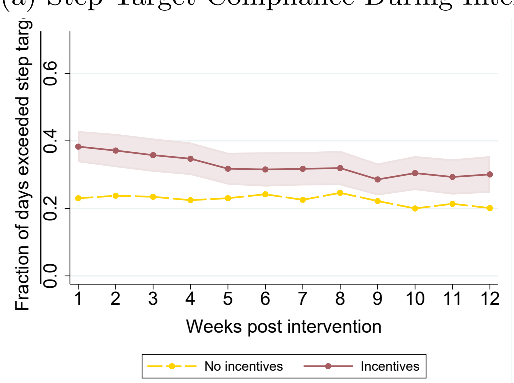
Appendix Figure A.2: Incentive Effects During and After Intervention, Conditional on Wearing the Pedometer



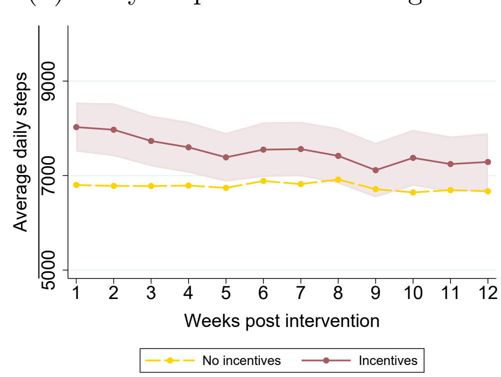
(a) Step-Target Compliance During Intervention



(b) Daily Steps Walked During Intervention



(c) Step-Target Compliance Post Intervention



(d) Daily Steps Walked Post Intervention

Notes: Figure is the same as Figure 4, except that the outcomes are all conditional on wearing the pedometer.

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: 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	-0.0128	-0.0279*	0.0158	-0.00145	0.00588	0.0164**	-0.00505
	[0.0175]	[0.0143]	[0.0124]	[0.00509]	[0.00738]	[0.00704]	[0.00592]
# 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 selected by lasso in Column 1 of Table 2. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.3: 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
	[255]	[286]	[273]	[291]	[279]	[275]	[253]
Lee upper bound	1426	1571	1572	1351	1430	1581	1358
	[333]	[307]	[280]	[300]	[304]	[279]	[227]
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
	[0.024]	[0.031]	[0.023]	[0.022]	[0.022]	[0.022]	[0.019]
Lee upper bound	0.232	0.242	0.216	0.206	0.209	0.217	0.206
	[0.025]	[0.030]	[0.023]	[0.022]	[0.022]	[0.023]	[0.019]
# 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.4: 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.5: Missing pedometer data during the post-intervention period

Dep. Variable:	No Steps data	**Reason** no steps data		**Reason** no data from Fitbit			
		Did not wear Fitbit	No data from Fitbit	Lost data entire period	Immediate withdrawal	Mid-intervention withdrawal	Other reasons
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Incentives	0.00334	0.00908	-0.00478	0.00193	0.00596	-0.00856*	-0.00411
	[0.0231]	[0.0208]	[0.0204]	[0.00389]	[0.0192]	[0.00499]	[0.00605]
# Individuals	1,254	1,122	1,254	1,254	1,254	1,254	1,254
# Observations	105,336	91,756	105,336	105,336	105,336	105,336	105,336
No incentives mean	0.40	0.31	0.13	0.00	0.10	0.01	0.02

Notes: This table reports the reasons that we do not have step data by treatment status in the post-intervention period. Each observation is an individual × day. Controls are the same as in Column 1 of Table 3. Column 1 reports Fitbit data missing for any reason, which can include mid-intervention withdrawal, Fitbit sync issues, and not wearing the Fitbit, amongst others. Columns 2 + 3 = Column 1, except that column 2 conditions on there being no missing data for consistency with our main step analyses whereas columns 1 and 3 do not. Columns 4-7 summarize reasons for why steps data might have been missing, and the variables in columns 4-7 sum up to the variable in column 3. Some people have no data during the entire measurement period, as summarized in columns 4 and 5. The omitted category is the pooled control and monitoring groups. "Did not wear Fitbit" takes value 1 when Fitbit data is non-missing and Fitbit steps = 0. As shown in Table G.2, there were 1,254 individuals who we enrolled in the post-intervention period and 1,122 whom we have pedometer data from. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.6: Lee Bounds Estimates of Exercise Effects in the Post-Intervention Period

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	765	765	471	471	471	471	471
(conditional on nonmissing data)	[238]	[238]	[246]	[246]	[246]	[246]	[246]
Lee lower bound	731	689	366	459	448	304	459
	[293]	[325]	[291]	[235]	[299]	[273]	[239]
Lee upper bound	840	934	503	515	554	522	515
	[320]	[282]	[269]	[254]	[280]	[266]	[246]
B. Met 10k step target							
Regression estimate	0.102	0.102	0.068	0.068	0.068	0.068	0.068
(conditional on nonmissing data)	[0.020]	[0.020]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]
Lee lower bound	0.101	0.099	0.063	0.067	0.067	0.060	0.067
	[0.025]	[0.021]	[0.017]	[0.017]	[0.022]	[0.019]	[0.017]
Lee upper bound	0.105	0.110	0.069	0.070	0.072	0.070	0.070
	[0.024]	[0.024]	[0.018]	[0.017]	[0.026]	[0.017]	[0.017]
# Individuals	1,254	1,122	1,254	1,128	1,248	1,122	1,128
# Observations	105,336	91,756	105,336	92,260	102,340	92,430	92,260

Notes: Table reports regression estimates and Lee bounds accounting for different types of missing pedometer data in the post-intervention period. The regression estimates condition on data not being missing, using different definitions of missing data in each column, and then the Lee bounds are estimated again allowing the definition of missing data to vary by column. Panel A reports results using average daily steps as the dependent variable, and Panel B reports results using proportion of days met 10k step target as the dependent variable. The omitted category is the monitoring and control groups pooled. The number of observations is reported for the Lee bounds regressions. The regression estimates are not directly comparable to those reported in Table 3 because each column conditions on the "type of missing" indicator in the first row being equal to 0 and does not include controls. The reason the column 1 and 2 estimates are so much larger than the column 3-7 estimates is that the column 1 and 2 estimates treat data as missing if the individual does not wear the pedometer. Thus, these estimates are more analogous to the treatment effect estimates on steps conditional on wearing the pedometer, e.g., column 3 of Table A.9. In contrast, the column 3-7 estimates treat data as missing only if it was missing by the definition of missing at the top of the column, and so instead treat days where the individual does not wear the pedometer as 0's. These estimates are thus more analogous to the treatment effect estimates unconditional on wearing the pedometer, e.g., column 2 of Table A.9.

Appendix Table A.7: HbA1c and RBS are predictive of each other

Dependent variable:	Endline HbA1c	Endline RBS
	(1)	(2)
Baseline HbA1c (SDs)	0.60***	0.33***
	[0.045]	[0.057]
Baseline RBS (SDs)	0.25***	0.37***
	[0.044]	[0.055]
# Individuals	560	561

Notes: This table reports estimates from regressing standardized HbA1c at endline (in column (1)) and standardized RBS (in column (2)) at endline on standardized HbA1c and standardized RBS at baseline. Standard errors in parentheses. The sample is the control group only. Significance levels: * 10%, ** 5%, *** 1%

Appendix Table A.8: Impacts of Incentives on Walking, Without Baseline Controls

Dependent variable:	Exceeded step target	Daily steps	Daily steps (if > 0)
	(1)	(2)	(3)
Incentives	0.196***	1286.2***	1144.3***
	[0.0180]	[211.4]	[190.3]
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 replicates the Table 2 estimates without including the baseline controls. The Threshold group pools the 4- and 5-day threshold groups. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.9: Impacts of Incentives on Post-Intervention Walking, Without Baseline Controls

	Post-intervention		
Dependent variable:	Exceeded step target	Daily Steps	Daily Steps (if > 0)
	(1)	(2)	(3)
Incentives	0.068***	470.8*	765.6***
	[0.02]	[246.30]	[239.11]
No incentives mean	0.22	5687.35	7347.39
# Individuals	1,122	1,122	1,122
# Observations	91,756	91,756	62,858

Notes: Notes: This table replicates the Table 3 estimates without including the baseline controls. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.10: Impacts of incentives on health, without baseline controls

	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	-0.044	-0.068	-5.53	-0.055	-0.092*	-0.15	-11.4*	-0.13**
	[0.043]	[0.11]	[4.37]	[0.047]	[0.053]	[0.14]	[6.12]	[0.060]
Incentives × below- median blood sugar					0.088	0.14	11.1	0.15*
					[0.062]	[0.16]	[7.23]	[0.084]
p-value: Incentives = Monitoring	0.435	0.952	0.153	0.102	0.770	0.463	0.294	0.803
# Individuals	3,067	3,066	3,067	3,068	3,067	3,066	3,067	3,068

Notes: This table reports the results of the specifications displayed in Table 4 without controls. Standard errors are in brackets and clustered at the individual-level. We also control for the monitoring dummy and it's interaction with the below-median blood sugar dummy but suppress those coefficients for brevity. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.11: Threshold treatments increase cost-effectiveness relative to the base case, with similar increases among those who are more and less impatient

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

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

Appendix Table A.12: Measures of impatience over effort correlate in the expected direction with baseline measures of exercise, health, and behavior

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

Notes: This table displays the correlations between our impatience measures and a number of baseline health and behavior measures. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize health and behavior measures so that higher values correspond to healthier behavior; hence we expect all correlations to be negative. Panel A displays the impatience index along with the five questions from which it is generated. Panel B shows the predicted index along with the three questions from which it is generated. See Online App. Table G.16 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.13: No correlation between measures of impatience over effort and recharges

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

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

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:

$$\max_{w_t \in \{0,1\}} \mathbb{E}\left[d^{(T-t)} m_T - \sum_{j=t+1}^{T} \delta^{(j-t)} w_{j,t} e_j \middle| e_1, .., e_t, w_1, .., w_t \right] - w_t e_t, \tag{15}$$

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:

$$\begin{gathered} V_t(0) < V_t(1) - e_t \\ \text{or} \\ e_t < V_t(1) - V_t(0). \end{gathered} \tag{16}$$

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

B.2 The Effectiveness of Threshold and Linear Contracts

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

Proposition 1. *Let $d = 1$ and $T = 2$. Fix all parameters other than δ, and take a linear contract that induces compliance $C > 0$.*
(a) If agents are naïve and e_2 is weakly increasing in e_1, in a first order stochastic dominance sense,[41] *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

[41] This assumption flexibly accommodates the range from IID to perfect positive correlation, just ruling out negative correlation.

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 H.1. □

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

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

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

Proof. See Online Appendix H.1. □

Proposition 3. *[IID Uniform] Let $d = 1$ and $T = 2$. Fix all parameters other than δ. Let costs be independently drawn each day from a uniform[0,1] distribution.*
(a) Take any threshold contract paying $0 < M < 1$. Whether there exists a more effective linear contract depends on δ. Define $\frac{(M+1)^2}{2}$ as the "cutoff value" for naïfs and $2 - \frac{4}{(1+M)^2}$ as the "cutoff value" for sophisticates. If δ is less than the cutoff value for a given type, there does not exist a linear contract that is more effective for that type; any linear contract with at least as high cost-effectiveness will generate strictly lower compliance.[42] In contrast, if δ is greater than the cutoff value, there always exists a linear contract that is more effective than the threshold for that type; in particular, a contract with the same cost-effectiveness and strictly higher compliance.
(b) Take any threshold contract paying $1 \leq M < 2$.[43] Regardless of δ, there does not exist a linear contract that is more effective.

Proof. See Online Appendix H.1. □

B.3 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 $T = K$.

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

Proposition 4 ($T = K$, 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 H.2.

Recall that the condition for complying on day 1 is to comply if $e_1 < V_1(1) - V_1(0)$ (equation (16)). With the threshold contract, we have that:

$$V_1(1) - V_1(0) = \mathbb{E}\left[(dM - \delta e_2)w_{2,1}|e_1, w_1 = 1\right] - \mathbb{E}\left[-\delta e_2 w_{2,1}|e_1, w_1 = 0\right] \tag{17}$$

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:

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

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 (18) 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:

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

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 (19) 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 sence. 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 generally weakly decreasing in δ.

Proposition 5 ($T = 2, K = 2$, 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.*[44] *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,*[45] *the effectiveness of the threshold contract is weakly decreasing in δ.*

Proof. We first show that, if costs are positive, cost-effectiveness in the threshold is not increasing in δ. Because Proposition 4 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 (18) and (19), 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:

$$\begin{aligned}
C/P &= \frac{1}{M}\frac{\mathbb{E}[w_1 + w_2]}{\mathbb{E}[w_1 w_2]} = \frac{1}{M}\frac{\mathbb{E}[w_1 + w_1 w_2]}{\mathbb{E}[w_1 w_2]} = \frac{1}{M}\left(\frac{\mathbb{E}[w_1]}{\mathbb{E}[w_1 w_2]} + 1\right) = \frac{1}{M}\left(\frac{\mathbb{E}[w_1]}{\mathbb{E}[w_1]\mathbb{E}[w_2|w_1 = 1]} + 1\right) \\
&= \frac{1}{M}\left(\frac{1}{\mathbb{E}[w_2|w_1 = 1]} + 1\right) \qquad (20)
\end{aligned}$$

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

[44] This assumption flexibly accommodates the range from IID to perfect positive correlation, just ruling out negative correlation.

[45] See equation (23) 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 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.

our more general assumption that e_2 is weakly increasing in e_1, note that higher δ means that $w_1 = 1$ will be associated with lower values of e_1 (since e^* is decreasing in δ). This implies lower values of e_2 conditional on $w_1 = 1$, since we assume that e_2 is weakly increasing in e_1. Lower values of e_2 then mean that $p_2^* = E[w_2|w_1 = 1]$ will be weakly higher. Hence, p_2^* is weakly increasing in δ and the first term is weakly decreasing in δ. Thus, we have shown that, with positive costs, C/P is weakly decreasing in δ.

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

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

$$\begin{aligned}
2\,(\lambda C - P) &= \lambda\mathbb{E}[w_1 + w_2] - M\mathbb{E}[w_1 w_2] \\
&= \lambda\left(F(e^*) + \mathbb{E}[w_1 \mathbb{1}\{0 < e_2 < M\} + \mathbb{1}\{e_2 < 0\}]\right) - M\mathbb{E}[w_1 \mathbb{1}\{e_2 < M\}] \\
&= \lambda\left(F(e^*) + \mathbb{E}[\mathbb{1}\{e_1 < e^*\}\mathbb{1}\{0 < e_2 < M\} + \mathbb{1}\{e_2 < 0\}]\right) - M\mathbb{E}[\mathbb{1}\{e_1 < e^*\}\mathbb{1}\{e_2 < M\}] \\
&= \lambda\left(F(e^*) + Prob(e_1 < e^*, 0 < e_2 < M) + Prob(e_2 < 0)\right) - M Prob(e_1 < e^*, e_2 < M).
\end{aligned} \tag{21}$$

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 $g'(e^*) = f(e^*)\, Prob(e_2 \in S|e_1 = e^*)$.[46] 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:

$$\lambda(1 + Prob(0 < e_2 < M|e_1 = e^*)) \geq M Prob(e_2 < M|e_1 = e^*) \tag{22}$$

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

$$Prob(e_2 < 0|e_1 = e^*) \leq \frac{\lambda}{M} + \left(\frac{\lambda}{M} - 1\right) Prob(0 < e_2 < M|e_1 = e^*)). \tag{23}$$

If $\lambda > M$, condition (23) 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

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

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 (23) holds), the effectiveness of a threshold contract is decreasing in δ. □

We now turn to examine threshold contracts with $T < K$. To gain tractability, we begin with the case where costs are perfectly correlated across periods.

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

Proof. See Online Appendix H.2. □

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

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

Proof. See Online Appendix H.2. □

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

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

Proof. See Online Appendix H.2. □

B.4 Proofs of Predictions 2 and 3

Prediction 2 (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.*[47]

Proof. Equation (3) implies that, in a linear contract, $C = \frac{1}{T}\sum_{t=1}^{T} F(d^{(T-t)}m)$. Compliance is thus increasing in the discount factor over payment $d^{(T-t)}$. If agents are "impatient," then $d^{(T-t)}$ is weakly decreasing in the delay to payment $T-t$. Increasing payment frequency then decreases the average delay to payment, which weakly increases compliance. If agents are patient, then the discount factor is 1 irrespective of the delay to payment and increasing payment frequency has no effect on compliance. Effectiveness follows the same pattern as compliance since cost-effectiveness is invariant to payment frequency (it is always $\frac{1}{m}$). $\square$

Prediction 3 (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.*

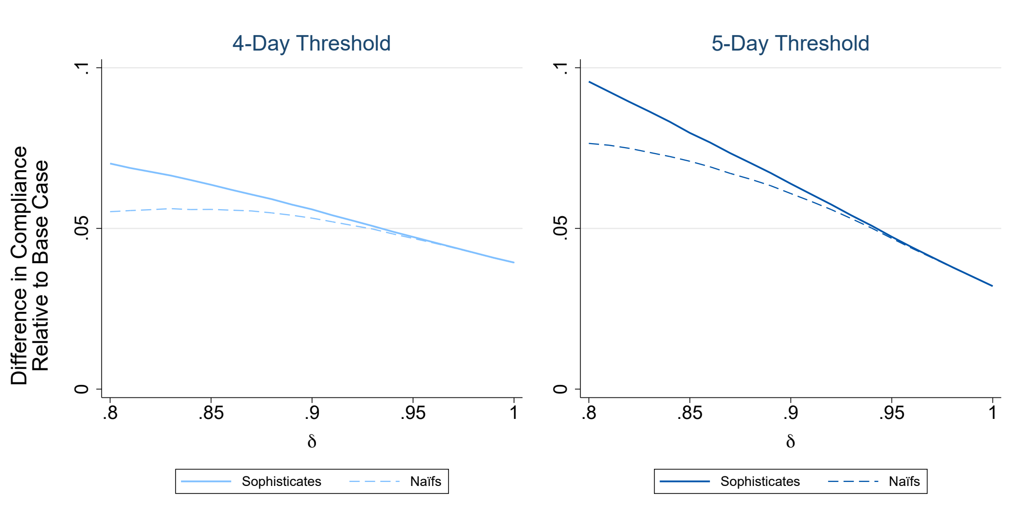
Proof. Recall that, on day t, agents comply if $e_t < d^{(T-t)}m$. As the payment date approaches, the time to payment $T-t$ decreases. If $d^{(T-t)}$ is decreasing, this increases $d^{(T-t)}$ and hence increases the likelihood that $e_t < d^{(T-t)}m$. If $d^{(T-t)}$ is flat, then the likelihood that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

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

C 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 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 F. 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 C.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 C.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.[48] 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 G.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.[49]

[48] 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 7 in App. B.3).

[49] 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.