Can price discrimination incentivize behavior change? Evidence from a randomized field experiment

Rebecca Dizon-Ross
University of Chicago

Ariel Zucker*
UC Santa Cruz

July 28, 2022

Abstract

Although personalizing policies can increase their effectiveness, it is difficult to personalize when beneficiaries' types are unobservable. Mechanism design offers two strategies to personalize with imperfect information — tagging on observables (i.e., 3rd-degree price discrimination), and offering a menu of policy choices (i.e., 2nd-degree price discrimination). Using a randomized controlled trial of incentives for exercise among 5,600 adults with diabetes and hypertension in urban India, we show that both 2nd-degree and 3rd-degree price discrimination substantially increase program impact, leading to a 75% increase in exercise relative to the effect of a one-size-fits-all benchmark. Moreover, we show that commitment motives, which are commonly present in policy domains such as preventive health and savings, play an important role in the success of both personalization mechanisms.

*Dizon-Ross: University of Chicago Booth School of Business, rdr@chicagobooth.edu. Zucker: University of California, Santa Cruz, arzucker@ucsc.edu. This study was funded by the Chicago Booth School of Business, the Tata Center for Development, and the Chicago India Trust. The study protocols received approval from the IRBs of Chicago, UC Berkeley, and the Institute for Financial Management and Research (IFMR). The experiment was registered on the AEA RCT Registry. We thank Rupasree Srikumar and Srinish Muthukrishnan for leading the fieldwork, and Emily Zhang and Katherine Daehler for outstanding research assistance. We are grateful to Abhijit Banerjee, Sydnee Caldwell, Josh Dean, Esther Duflo, Pascaline Dupas, Meredith Fowlie, Ben Golub, Seema Jayachandran, David Levine, Jeremy Magruder, Aprajit Mahajan, Ted Miguel, and Daniel Waldinger for helpful conversations and feedback and to numerous seminar and conference participants for insightful discussions. All errors are our own.

1 Introduction

Personalizing policy is a promising approach to increase its effectiveness. Individuals respond heterogeneously to policies, so assigning each individual to the right policy variant should outperform a one-size-fits-all approach. However, personalizing is challenging if beneficiary types are unobservable to the policymaker. The challenge is particularly severe if beneficiary and policymaker preferences are at odds, giving beneficiaries the incentive to misreport their types. For example, when policy variants are vertically differentiated with one more generous than the rest, all beneficiaries may want the most generous policy variant, while the policymaker may only prefer that variant for a subset of beneficiaries.

Consider the case of incentive programs for exercise, an increasingly common policy tool used by governments and insurance companies to promote health (Baicker et al., 2010). Many such policies pay people for meeting exercise targets, such as walking a certain number of steps in a day. For a given payment amount, the most effective step target may vary by individual. Lower step targets might induce more walking from those who walk less at baseline, while higher targets might induce more walking from those who walk more at baseline. Thus, the policymaker might wish to personalize the target to maximize exercise. However, lower targets are more generous than higher ones, as they pay weakly more for any level of walking. Thus, standard economic theory predicts that all beneficiaries would prefer the lowest target for themselves, putting their preferences at odds with the policymaker's.

Mechanism design offers a solution to this principal-agent problem: use the tools of price discrimination, traditionally used to personalize prices, to personalize policy variants instead. The policymaker could offer a menu of policy options – for example, a menu of incentive contracts with different step targets – and allow individuals their *choice* (second-degree price discrimination). To make the menu incentive-compatible, the policymaker could try to make the most generous policy unattractive to the "wrong" types, akin to a restaurant offering a low-priced early-bird special that is unattractive to rich customers. For example, in our step target case, the policymaker could design the contract menu such that contracts with lower step targets also offer lower payments for each day of compliance with the step target, so that lower step targets no longer dominate higher step targets from the beneficiary perspective. Alternatively, policymakers could assign policy variants based on observable characteristics or *tags* (third-degree price discrimination), akin to a restaurant offering discounts to seniors.

However, policymakers often view these strategies as risky or unpromising when policies are vertically differentiated; they worry that beneficiaries' strong preferences for the most generous policy will provide too much of a constraint. If the policymaker assigns the policy based on a *tag*, beneficiaries may try to game the system and manipulate their observable characteristics to access the most generous program (Fudenberg et al., 2005). If the policymaker tries to design an incentive-compatible *choice* menu, the policymaker might have to distort the most generous policy too much to dissuade certain beneficiaries from choosing it.

A central insight of this paper is that these constraints on personalized policy might be alleviated in domains where self-control problems are prevalent. In these domains, such as savings and health, behavioral biases such as present bias can cause individuals to underinvest relative to their private optimum. As a result, sophisticated beneficiaries who are aware of their own biases may demand commitment to increase their own investment. They might prefer a *less* generous policy for themselves if that policy induces more behavior change. In essence, it is possible that these beneficiaries are playing on the same team as the policymaker.

This paper provides proof of concept of the promise of personalized policy and evidence that commitment motives can play an important role in the success of personalization. We do so by performing a randomized controlled trial with 5,600 adults in urban India in which we evaluate two personalized treatments – one that uses *Choice* and one that uses a *Tag* – and compare them with a uniform, non-personalized approach.

We personalize a policy that offers incentives for physical activity to Indian adults with diabetes, prediabetes, or hypertension. Physical inactivity is a leading risk factor for diabetes, hypertension, and cardiovascular disease (Myers, 2008; Warburton et al., 2006), three diseases that are leading causes of mortality worldwide. A central justification for using incentives to promote preventive health is that it has large positive externalities (Dupas and Miguel, 2017; Finkelstein et al., 2009; Kenkel, 2000; Schwappach et al., 2007). However, since preventive health investments like exercise involve near-term costs but only long-run benefits, present bias may dampen investment, and individuals may demand commitment to increase their own investment. These characteristics make this an interesting setting to study the interactions between commitment and personalization.

Our program provides incentives to participants for achieving daily step targets, and we personalize the step target. Our Choice mechanism offers people a choice between a menu of step targets. To keep higher walkers from choosing the lowest step target, contracts with lower step targets also offer lower payments for each day of compliance with the step target, akin to degrading the "quality" of those options from the beneficiary perspective. The Tag mechanism tags participants based on their baseline activity levels (i.e., daily step counts) by asking people to wear a pedometer for six days before they are assigned a step target. We then assign higher step targets to people with higher step counts. Although this tag is easy for participants to manipulate, no other observable characteristics had sufficient explanatory power over daily steps.[1] In the absence of present-bias and demand for commitment, we

[1]Ideally tags would have good explanatory power over step target treatment effects, not baseline steps, but since we did not have data on step target treatment effects, we assumed that those with higher baseline

would thus expect agents to manipulate their steps downward to get an easier step target (Laffont and Tirole, 1988; Weitzman, 1980).

To try to decrease the tension between principal and agent objectives, we personalize step targets in a pre-period (i.e., before the intervention has begun), when commitment motives might push against agents' desires for an easier step target. During the pre-period, sophisticated time-inconsistent agents' objectives may be more aligned with the principal's, in that they may value step targets that will induce more walking effort from their future selves. This could affect their choices under Choice and their level of walking under Tag.

We then empirically evaluate the performance of the two personalization strategies. We randomly assign participants to either the Tag mechanism, the Choice mechanism, a randomly-assigned (not personalized) step target, or to a monitoring group that received a pedometer but no step target or incentives. To identify the effect of personalization, we compare walking under the Tag and Choice mechanisms to walking under a uniform step target that serves as a "one-size-fits-all" step target.[2]

We ask two primary questions: First, does personalization using price discrimination increase performance relative to the one-size-fits-all approach? Second, if so, does sophisticated agent present-bias and demand for commitment contribute to the success of personalization?

Our first result is that both the personalized Tag and Choice mechanisms substantially increase walking relative to the one-size-fits-all step target benchmark. Specifically, while incentivizing the one-size-fits-all step target increases activity by approximately 6 minutes[3] of brisk walking per day relative to monitoring with a pedometer alone, the personalized treatments increase walking by nearly 5 additional minutes per day, an improvement over the one-size-fits all treatment effect of approximately 75%. The Choice and Tag treatments achieve this increase in steps without an increase in incentive payments, thus increasing cost-effectiveness and steps achieved relative to the payout.

We show that the gains from personalization come from sorting individuals into the "right" step target for them, with gains seen across the full distribution of walking. When we randomly assign step targets in a non-personalized fashion, each target works better for some people but worse for others. In particular, lower targets work better for those at the bottom of the walking distribution but worse at the top, while higher targets do the opposite. Personalization achieves the advantages of the high and low targets at the top and bottom

steps would have more positive treatment effects for higher step targets relative to lower step targets. Our experimental data confirm this hypothesis.

[2] The uniform contract comparison is the contract we thought *ex ante* would be most effective, as specified in our pre-analysis plan (Dizon-Ross and Zucker, 2020). Among those receiving fixed step targets, we allocated twice as many participants to our "one-size-fits-all" target as to either of the other two step targets we considered to ensure power for this comparison.

[3] We estimate minutes of brisk walking a day using a conversion rate of 100 steps per minute.

of the distribution respectively, without having their downsides at the opposite ends of the distribution. We also see some differences between our Choice and Tag implementations. While both Choice and Tag achieve large increases at the bottom of the distribution, Choice does better in the middle, ensuring that most people do not receive a wildly inappropriate target, while Tag does better at the top, pushing more people into more aggressive targets.

Our second result is that commitment motives play a key role in the success of each of our personalization mechanisms. We describe the evidence on the role of commitment for the Tag mechanism first and the Choice mechanism second. In the Tag mechanism, all step targets pay the same amount (20 Indian Rupees (INR), roughly 0.33 USD) for step-target compliance, and so contracts that have higher step targets are weakly dominated from the agent's perspective. Thus, standard economic theory predicts that agents will manipulate their baseline steps downwards to try to get a lower step target. Instead, we find that participants assigned to the Tag mechanism modify their baseline steps *upward* on average in order to receive higher step targets, a form of paying now in effort to commit to future walking. We then show suggestive evidence that allowing participants to manipulate their steps in the Tag mechanism improved walking.

To investigate the role that commitment plays in the performance of the Choice mechanism, we elicit incentive-compatible choices over a second "commitment menu." This menu includes the same step targets as our primary menu, however all step targets pay the same 20 INR. As a result, unlike in our primary menu, higher step targets in the commitment menu are weakly dominated from the participants' perspectives. Despite this important difference, we find that most people choose the same step target on the commitment menu that they did on the primary Choice menu. For example, nearly as many people choose higher (dominated) step-target contracts from the commitment menu as choose higher step-target contracts in our primary Choice menu, when there is a financial incentive to choose higher targets. This suggests that much of the sorting into higher targets that we see in our primary Choice menu reflects a demand for commitment rather than the financial incentive to choose higher step targets, and thus that the demand for commitment plays a crucial role in the performance of the Choice treatment.

The main contributions of this paper are to the literature of empirical mechanism design. First, we provide a novel experimental evaluation of the use of second and third degree price discrimination to personalize policy. A small literature evaluates the use of price discrimination by firms to personalize prices. Dubé et al. (2019) experimentally test third-degree price discrimination for an online subscription, Levitt et al. (2016) experimentally test second-degree price discrimination in online gaming, and Leslie (2004) and Mortimer (2007) use simulations to compare second- and third-degree price discrimination strategies for the theater and DVD markets. In contrast, we evaluate the use of second- and third-degree price discrimination to personalize the assignment of public policies.

Our focus on personalized policies also ties us to a literature that has examined the personalization of *horizontally*-differentiated program variants, where there is often no tension between principal and agent objectives and hence no need for mechanism design and the tools of price discrimination.[4] As a result, this literature has not evaluated the price discrimination strategies that we do, in particular second-degree price discrimination.

Also related to our work is a literature that examines the targeting of programs on the extensive margin, or who gets a given program.[5] In contrast, we focus on targeting on the intensive margin, or who gets *what* program. This focus changes the type of strategies the policymaker should use, making price discrimination the appropriate toolkit. Just like a price-discriminating monopolist, our policymaker often wants to serve the entire market. In contrast, in the extensive margin work, the policymaker explicitly does not want to serve the entire market, and their goal is to screen out certain beneficiaries.[6]

Second, building on the insight from behavioral economics that time inconsistency can align principal and agent preferences (e.g., Kaur et al., 2015), we show that commitment motives can improve the personalization of policy. This has particularly interesting consequences for third-degree price discrimination. Economists have long postulated that manipulation of observables limits the ability of principals to target incentive contracts on observable behaviors (e.g., Fudenberg et al., 2005; Laffont and Tirole, 1988; Weitzman, 1980), and empirical evidence supports this idea.[7] We find that commitment motives are a previously unexplored mechanism that counter the negative effects of manipulation.

We make two additional contributions to the literature on commitment contracts. First,

[4]For example, Muralidharan et al. 2018 evaluates the effectiveness of personalizing educational instruction based on a child's current academic level, where both policymakers and households have the same objective: maximize outcomes.

[5]In particular, Alatas et al. (2012) compare targeting government transfers to the poor using a proxy-means test with community-based targeting (where communities identify the poorest members themselves) and unpack the mechanisms for each method's success. Jack (2013) experimentally evaluates the effectiveness auctions as a method of targeting payments for environmental services, and simulates the relative performance of targeting these contracts on observables. Other work evaluates the changes in targeting efficiency resulting from hassle costs (a form of choice) using experimental (Alatas et al., 2016; Finkelstein and Notowidigdo, 2019) and quasi-experimental (Deshpande and Li, 2019) methods.

[6]An exception to the extensive-margin focus is Andreoni et al. (2018), who test whether targeting contracts based on an observable, revealed preference measures of time preferences, can improve the principal's objective. However, the primary aim of Andreoni et al. (2018) is to test whether structurally estimated time preference parameters predict behavior. They therefore shut down key channels through which real-world targeting mechanisms might operate, such as manipulating observables.

[7]For example, Wolak (2006) finds that residential customers adjust electricity consumption upward when it is used as a baseline for calculating critical peak pricing bonus payments, resulting in easy conservation targets, and Cardella and Depew (2018) find that when future work expectations depend on current output, workers ratchet output downward.

while there are many examples of individuals selecting dominated contracts that may reduce their future rewards (see Ashraf et al. (2006); Royer et al. (2015), and Schilbach (2019) for examples from health settings), there is limited evidence showing that individuals are willing to make sacrifices in the present for a future commitment contract.[8] In contrast, we find that individuals assigned to the Tag group are "willing to pay" for commitment contracts in terms of costly effort now. Second, even if individuals demand commitment contracts, they do not always pay off. Previous work has found mixed results, with some opt-in commitment contracts appearing to be effective (e.g. Ashraf et al. (2006); Royer et al. (2015); Schilbach (2019)), and others failing to impact behavior or seeing large numbers of individuals fail to follow through on commitments (e.g. Bai et al. (2020); DellaVigna and Malmendier (2006)). We find that people in our setting are sufficiently sophisticated that commitment contracts may increase behavioral change.

The remainder of the paper proceeds as follows. Section 2 provides background information on the status of non-communicable diseases in India. Section 3 presents the theoretical framework for adapting second- and third-degree price discrimination to incentive contracts for achieving step targets. Section 4 describes the experimental design, and Section 5 describes the data. The results are shown in Sections 6 and 7, and Section 8 concludes.

2 Non-communicable Diseases in India

We conducted our experiment among adults in an urban area of South India who are living with, or at high risk of contracting, hypertension or diabetes. Encouraging physical activity is a critical health intervention in this population.

India is facing an explosion of lifestyle disease as sedentary lifestyles become more common (Gupta and Ram, 2019). Both hypertension and diabetes have reached epidemic proportions: it is estimated that 25% of adults in India had hypertension and 8.8% had diabetes in 2019, with similar numbers of adults at high risk of developing these diseases (Gupta and Ram, 2019; International Diabetes Federation, 2019).[9] This has large economic and social implications. Both diseases can lead to severe complications, such as stroke, heart attack, kidney failure, and (in the case of diabetes) blindness or amputations. In 2010, diabetes alone imposed an estimated cost of $38 billion—2 percent of GDP—on the healthcare system. Diabetes leads to approximately 1 million deaths per year in India (Tharkar et al., 2010), and hypertension leads to the death of approximately 1.6 million more individuals per year

[8] An exception is Milkman et al. (2014), who show that individuals are willing to pay in the present to only allow themselves access to television content at the gym as a commitment to exercise.

[9] Hypertension, or high blood pressure is typically defined as systolic blood pressure (BP) above 140 mmHg or a diastolic BP above 90 mmHg, or on treatment for hypertension. Diabetes occurs when a person's ability to produce or respond to insulin is impaired, and is defined by elevated blood glucose levels such that hemoglobin A1c is 6.5% or higher or fasting blood glucose is 126 mg/dl or higher.

(Gupta and Xavier, 2018). The prevalence of lifestyle disease is higher both in southern than northern states and in urban than rural areas (Anjana et al., 2011; Gupta et al., 2019).

Declines in physical activity are closely linked to the growing burden of lifestyle disease. A low level of physical activity is not only a risk factor for developing diabetes and hypertension in India (Bhansali et al., 2015; Little et al., 2016; Tripathy et al., 2017), but also leads to more rapid development of complications such as cardiovascular disease, stroke, and blindness for those who are already living with a lifestyle disease (Tandon et al., 2018).

There is widespread agreement that increasing physical activity is a critical pillar in the prevention and management of diabetes and hypertension (World Health Organization, 2013). The expected health benefits of additional exercise are large for most adults living in India, where activity levels are low.[10] Recent meta-analyses of the dose-response for all-cause mortality (Samitz et al., 2011) and onset of non-communicable disease (Kyu et al., 2016) suggest that the benefits of additional physical activity are large and approximately linear for activity levels up to 3,000-4,000 MET minutes per week—a level of activity nearly unheard of in urban India (Anjana et al., 2014). Developing programs that increase physical activity is a key priority for Indian governments.

Previous research has shown that incentives are a promising approach for increasing physical activity and decreasing the burden diabetes and hypertension. Specifically, Aggarwal et al. (2020) evaluate a program which offers incentives for achieving a daily step targets and is structured virtually identical to the non-personalized variant of the program we examine in this paper. They show that the program increased physical activity and decreased blood sugar and cardiovascular health risk among a sample of diabetics and prediabetics in India . This result is particularly promising given that the program is relatively low-cost and scalable, and that previous evidence-based approaches for increasing physical activity among populations with chronic disease are prohibitively expensive (Howells et al., 2016). This paper thus examines whether personalization can further improve that program for greater impact in the global fight against chronic disease.

3 Theoretical Framework

In this section, we present a simple model to illustrate participants' decisions over walking and incentive contracts in our setting.

In our experiment, participants receive incentives for compliance with a daily step target during a contract period. Prior to contract launch, some participants are informed that contracts will be assigned based on a "Tag" of walking in the pre-contract period, with higher step targets being assigned to those with higher walking, and all contracts paying

[10] Fewer than 10% of adults engage in any recreational physical activity, and fewer than 50% of adults reach the minimal level of overall physical activity recommended by the WHO (Anjana et al., 2014).

the same incentive for step target compliance. Other participants are offered a "Choice" between a menu of incentive contracts, where contracts with higher step targets pay weakly higher incentives for compliance.

Decisions over pre-contract period walking and contract menus are made in a time period *before* decisions over contract-period walking. In addition, the benefits of contract-period walking, including health payoffs and incentive payouts, are received in a time period *after* contract-period walking. Our model shows how the Tag and Choice mechanisms separate participant types, and how separation is mediated by time preference. In addition, the model indicates two empirical tests for commitment motives in our setting.

3.1 Model Setup

Utility and Discounting Assume participant i of type θ has the per-period utility function

$$u_t = h_t + y_t - (s_t)^\theta$$

where h_t are health benefits in period t, y_t is income in period t, s_t are steps in period t, and $(s_t)^\theta$ is the individual-specific cost of walking effort for θ types. We will focus below on the case where $\theta > 1$ to ensure that total utility satisfies the single crossing property: the marginal costs of steps $c'(\theta; s_t)$ is always *higher* for higher-θ types.[11] Note that a higher θ corresponds to higher walking costs and, as we will show, lower baseline walking levels.

We assume participants are quasi-hyperbolic (i.e., beta-delta) discounters. That is, participants discount future utility t periods in advance using the discount factor $\beta_i^{\mathbb{1}\{t\neq 0\}} \times \delta^t$, where $\beta_i = 1$ for time-consistent participants and $\beta_i < 1$ for time-inconsistent participants. We assume that all participants are sophisticated: they know the value of β_i. Furthermore, for expositional simplicity, we assume that $\delta = 1$.

Timing and Rewards The model has 3 periods.

Period 0 is the pre-contract period: participants undertake steps s_0, which the principal observes. Choice participants select contracts, and the principal assigns contracts to Tag participants based on s_0.

Period 1 is the contract period: participants undertake steps s_1, which the principal observes.

[11] In other words, the single crossing property (also known as the Spence/Mirrlees condition) says that the willingness to accept an increase in steps s_t is always smaller for higher types. The theory applies to more general $c(\theta; \cdot)$ such that total utility satisfies single crossing. While single crossing is not a necessary condition for Choice and Tag mechanisms to advantageously separate types, it is a sufficient condition. We invoke it here for analytical tractability, as is standard in the mechanism design literature.

Period 2 is the reward period: participants receive incentive income y_2 and health benefits h_2.

Incentive income for period-1 steps in each contract is defined by the incentive level W and the step target S: contracts pay W if $s_1 \geq S$ and 0 otherwise. Health benefits are also a function of period-1 steps, but also of period-0 steps. We assume for simplicity that health benefits increase linearly in both period 1 steps and period 0 steps: $h_2'(s_0) = h_2'(s_1) = b > 0$.[12]

The timing of the model replicates the fundamental stages of our personalization mechanisms: individuals first make decisions that influence their contract assignment and health, then they make decisions that influence their future income and health, and finally they reap the consequences of their decisions.

Walking in the Contract Period In period 1, each participant chooses their steps given their step target S and incentive level W. Specifically, they choose s_1 to maximizes the discounted income and health benefits in period 2 less walking costs in period 1:

$$s_1 = \arg\max_s \{\beta \times (bs + W^{\mathbb{1}\{s \geq S\}}) - s^{\theta}\} \tag{1}$$

There are two options for walking: either the step target binds and the person walks exactly the step target S, or the step target does not bind and instead steps are an interior solution characterized by the following first-order condition

$$s^* = \left(\frac{\beta b}{\theta}\right)^{\frac{1}{\theta - 1}} \tag{2}$$

Note that the incentive level does not enter the first-order condition (Equation 2). Individuals walk s^*, where marginal walking costs equal marginal walking benefits, in either of two cases: the incentive for reaching the target is so small that it does not encourage them to increase behavior (i.e. the contract is “out-of-reach”), or the step target is so small that they achieve it even without the incentive (i.e. the contract is “inframarginal”). In either case, the incentive contract does not influence walking.

Contracts do increase walking (i.e. the contract is “effective”) if two conditions are met: first, the step target S must be larger than s^*, the optimal walking level without a contract, and second, the additional discounted value of increasing walking to S from s^* must be larger than the additional costs. In this case, optimal walking is a corner solution at exactly S.

[12] Linear health benefits guarantee an interior solution and keep the model simple. However, the conclusions are equivalent as long as discounted step benefits are less convex than step costs. Otherwise, individuals would walk infinite steps each period.

Period-1 walking therefore takes the form:

$$s_1 = \begin{cases} S, & \text{if } S > s^* \text{ and } \beta\left(b(S - s^*) + W\right) > S^\theta - (s^*)^\theta \\ s^*, & \text{otherwise} \end{cases} \tag{3}$$

For time-inconsistent participants, period-1 walking is lower than their period-0 selves desire. This is a standard insight regarding present-biased agents: relative to their present self-interest, future selves under-perform actions which feature present costs and future benefits (e.g., Ainslie (1992)). Concretely, in period 0, participants desire their future walking under a contract with incentive W and step target S would take the following form:

$$s_1 = \begin{cases} S, & \text{if } S > s^+ \text{ and } b(S - s^+) + W > S^\theta - (s^+)^\theta \\ s^+, & \text{otherwise} \end{cases} \tag{4}$$

where $s^+ = \left(\frac{b}{\theta}\right)^{\frac{1}{\theta-1}} > s^*$.

Contract Choice and Walking in the Pre-Contract Period The participant's behavior in period 0 depends on how she values available contracts, which in turn depends on whether each contract is inframarginal, effective, or out-of-reach. The value of an inframarginal contract is simply the discounted value of the incentive, βW, which she will earn without modifying her behavior. The value of an out-of-reach contract is 0: it neither changes her behavior nor increases her rewards. The value of an effective contract is the discounted net benefit from achieving the step target S. More formally, in period 0, a participant of type θ with discount factor $\beta^{\mathbb{1}\{t \neq 0\}}$ values contract j with step target S and incentive W as:

$$V_0(\theta, \beta; j = \{S, W\}) = \begin{cases} \beta W, & \text{if } S \leq s^* \\ \beta\left(bS - bs^* + W - S^\theta + s^{*\theta}\right), & \text{if } s^* < S < \overline{s} \\ 0, & \text{if } S \geq \overline{s} \end{cases} \tag{5}$$

where s^* and $\overline{s}$ are both functions of θ and β. Specifically, $\overline{s}$ is the highest step target that an incentive W can induce type θ to walk.[13]

If a participant is assigned contracts via a "Choice" menu, she will choose the contract with the highest valuation. In addition, she will choose her period-0 steps to maximize discounted health benefits in period 2 less walking costs in period 0. Thus, for Choice participants, period 0 steps will satisfy the first order condition shown in Equation 2, and

[13] $\overline{s}$ solves: $s^\theta - \beta b s = \left(\frac{\beta b}{\theta}\right)^{\frac{\theta}{\theta-1}} + \beta W - \beta b\left(\frac{\beta b}{\theta}\right)^{\frac{1}{\theta-1}}$.

$s_0 = s^*$.

If she faces a "Tag" mechanism that assigns contracts based on s_0, she cannot choose a contract. However, she can modify s_0 in order be assigned her preferred contract. Modifying s_0 comes at a cost of $s_0^\theta - s^{*\theta} + \beta b(s^* - s_0)$. If there is an s_0 that will increase the value of her contract assignment by more than this cost, she will modify her steps in period 0. Otherwise, she will walk s^*.

The Principal We take the perspective of a principal trying to maximize *effectiveness*, defined as the benefit to the principal from steps less the payment to agents. This objective is analogous to the standard contract theory approach of maximizing output net of wage costs subject to incentive compatibility constraints.[14] As such, finding ways to increase steps for a given payout improves effectiveness. Increasing effectiveness in our experimental setting also has high potential to increase social welfare. Because of externalities to good health, the marginal social welfare of additional steps is likely large.

3.2 Personalizing Step Targets with a Tag

The Tag mechanism in our experiment holds incentives for step-target compliance fixed across contracts but personalizes the step target. The model highlights three channels through which the Tag mechanism can impact contract effectiveness relative to a single contract: market segmentation, tag endogeneity, and commitment. The results of our experiment will shed light on whether effectiveness of our Tag mechanism is influenced by each of these channels.

In the model, we consider a simple Tag mechanism with two contracts, both paying W. One contract has a higher step target, S_{high}, and the other has a lower step target, S_{low}. The Tag mechanism assigns S_{low} to anyone with period-0 steps below a cutoff $\hat{s_0}$: i.e., if $s_0 \leq \hat{s_0}$. If instead $s_0 > \hat{s_0}$, the mechanism assigns them S_{high}.

Market Segmentation We first demonstrate how the Tag mechanism would work in the simplest third-degree price discrimination setting, where the principal observes a signal of walking-cost type that the participant *cannot* manipulate. To do so, we focus on a stylized environment with only two agent types: a high-walking-cost type H and a low-walking-cost type L, so $\theta_H > \theta_L$. Our first proposition demonstrates that a principal could use this signal to improve contract effectiveness by segmenting the market.

Proposition 1. *Holding the incentive level constant, different step targets are differently effective for different types. In addition, for step targets that are effective for at least one*

[14] 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.

type, higher step targets are weakly more effective for lower-θ types (i.e. higher step targets are better for lower-walking-cost types).

We provide a proof in Online Appendix C. Here, we illustrate a simple case to show how a principal can use a non-manipulable Tag to segment the market.

We first rearrange Equation 3 to show how contract-period walking for a given incentive level W varies with the step target S.

$$s_1(\theta) = \begin{cases} s^*, & \text{if } S \leq s^* \\ S, & \text{if } s^* < S < \overline{s} \\ s^*, & \text{if } S \geq \overline{s}. \end{cases} \tag{6}$$

where s^* and $\overline{s}$ are both strictly decreasing in θ (holding all else constant).

Figure 1 plots an example of contract-period walking for the two types, $s_1(\theta)$, as a function of the step target S, holding the wage level constant at W.[15] The figure shows that the range of effective step targets for each type is between $s^*(\theta)$ and $\overline{s}(\theta)$. In this range, contract-period walking equals the step target. Everywhere else, contract-period walking is simply $s^*(\theta)$.

If the principal can tag using a non-manipulable signal of walking costs, she can assign step targets that will improve walking relative to offering either of the contracts to all participants. Specifically, a Tag algorithm will effectively segment the market if it assigns high-walking-cost types to a lower step target that is effective for them (i.e., $S_{low} \in (s^*(H), \overline{s}(H))$, where the solid blue line is increasing in Figure 1), and assigns low-walking-cost types to a higher step target that is effective for them (i.e., $S_{high} \in (s^*(L), \overline{s}(L))$, where the dashed red line is increasing in Figure 1). This will improve walking relative to offering either of the contracts alone.

Tag Endogeneity In practice, many tags can be manipulated by agents. This is the case in our experiment. A concern with endogenous tags is that agents of one type can imitate another type in order to receive a contract that is more valuable to the agent, but less effective in changing behavior. We next use our model to show one example of how this can make it difficult to segment the market.

Returning to Figure 1, imagine a Tag mechanism with $S_{low} \in (s^*(H), \overline{s}(H))$ and $S_{high} \in (s^*(L), \overline{s}(L))$, so the high step target is effective for low-walking-cost types and the low step target is effective for high-walking-cost types. Further, assume the step target assignment cutoff falls between the unmanipulated baseline walking levels of the two types:

[15] In the example, there is no overlap in effective step targets. An overlap is possible, but does not change the basic intuition.

$\hat{s_0} \in (\overline{s}(H), \overline{s}(L))$.

While this Tag mechanism would effectively segment the market if s_0 were fixed, it may fail to do so with endogenous s_0. Note that in this particular example, the high-cost type has no incentive to imitate the low-cost type. Equation 5 shows that the high-cost type strictly prefers S_{low}, which is effective, to S_{high}, which is out-of-reach and has value 0. However, for the low-cost type, the S_{low} is inframarginal and S_{high} is effective, and thus she may prefer S_{low}.[16] The low-cost type will reduce s_0 from $s^*(L)$ to the cut-off $\hat{s_0}$ if the net benefit of receiving the inframarginal contract with step target S_{low} exceeds the net cost of manipulating period-0 steps, i.e., if:

$$\beta(S_2^\theta - s^{*\theta} + bs^* - bS_2) > \hat{s_0}^\theta - s^{*\theta} + \beta(bs^* - b\hat{s_0}).$$

When this condition holds, endogeneity leads the Tag mechanism to fail to segment the two types.

Commitment Motive A final feature of the Tag mechanism is its timing: it uses a period-0 signal to assign a contract that rewards future effort with an even more distant payoff. The timing has important implications in the presence of time inconsistency. We next show that in period 0, time-inconsistent participants may place larger values on dominated contracts with higher step targets, while time-consistent participants will never do so. Time inconsistency can thus motivate high-cost types to imitate low-cost types by increasing s_0, suggesting a clean empirical test for the presence of time inconsistency in our setting. Moreover, we show that the upward manipulation of s_0 by time-inconsistents can improve the performance of the Tag mechanism relative to no manipulation.

Our next proposition motivates our test for commitment motives in the Tag group.

Proposition 2. *Time-consistent participants will never respond to a Tag by increasing s_0 above s^*. In contrast, time-inconsistent Tag participants may increase s_0 in order to be assigned a higher step target.*

We prove Proposition 2 in Online Appendix C. Here, we show the intuition with an example.

We again focus on a stylized environment with only two agent types. However, now we assume one is a time-consistent type TC with $\beta_{TC} = 1$ and the other is a time-inconsistent type TI with $\beta_{TI} < 1$.

Figure 2 shows how the time-consistent and time-inconsistent types value contracts that pay W for compliance with different step targets from the period-0 perspective. The contract

[16] As we show below, she always prefers S_{low} if she is time-consistent, but may prefer S_{high} if she is time-inconsistent.

valuation for time-consistent agents is weakly decreasing in the step target. Therefore, there is no value to be captured by increasing period-0 steps in order to get a higher step target, no matter what the step targets or baseline step cutoff.

For time-inconsistent agents, contract valuations increase for step targets just above s^*.[17] If a time-inconsistent agent values the contract with the higher step target more than the contract with the lower step target, and the cost of increasing baseline steps s_0 to the cutoff $\hat{s_0}$ is sufficiently low, then she will manipulate her period-0 steps upward. We can thus reject that participants are all time-consistent if any Tag participants modify their baseline steps upward.

The valuation curves in Figure 2 illustrate that commitment is a key channel through which endogenous manipulation of period-0 steps can improve Tag performance. Specifically, if the step cutoff is above s^*_{TI}, a non-manipulable Tag would assign both types S_{low}. On the other hand, if the tag is manipulable, the TI type might increase s_0 in order to get S_{high} if S_{high} is effective for her, leading her to walk more in the contract period. More generally, if enough participants adjust their baseline walking upward in order to receive higher, more effective step targets, then a Tag mechanism based on a manipulable, pre-period tag can outperform a non-manipulable tag.

3.3 Personalizing Step Targets with Choice

The primary Choice mechanism we consider allows individuals to choose among a menu of contracts, where contracts with higher step targets also pay more for step-target compliance. We next use the model to show two channels which influence the effectiveness of Choice relative to a single contract: earnings motives and commitment motives. The results of our experiment will also shed light on whether effectiveness of our Choice mechanism is influenced by each of these channels.

Market Segmentation through Earnings Motives Here, we show how a menu can segment the market into groups and assign each group a relatively more effective contract even without commitment motives. To do so, we return to the stylized environment with a high-walking-cost type H and a low-walking-cost type L, so $\theta_H > \theta_L$. Imagine a menu of two contracts, one with a higher step target and a higher payment level, and one with a lower step target and a lower payment level. The menu will sort types into different contracts and do better than either contract alone if it satisfies two conditions. First, the high-step-target contract must be preferred by the low-cost type L and the low-step-target contract must be preferred by the high-cost type H. Second, the high-step-target contract must be relatively

[17] While the exact shape of the contract valuation curve depends on β_i, θ_i, and b, Online Appendix C shows that the valuation curve is always weakly decreasing in S for time-consistents but always has a strictly increasing portion just above s^* for time-inconsistents.

more effective for type L and the low-step-target contract must be relatively more effective for type H.

Figure 3 shows an example menu that will segment the market simply because agents wish to increase their incentive earnings and reduce their walking costs. One contract pays a high wage, W_{high}, for compliance with the high step target S_{high}, and another contract pays a low wage, W_{low}, for compliance with the low step target, S_{low}. The figure shows how a high- and low-walking cost participant with time-consistent preferences will value these contracts. Note that time-consistent participants only value contracts because they increase future earnings; unlike time-inconsistent participants, they receive no value from increasing future walking.[18] The high-cost type prefers the low-wage, low-target contract since it is effective, while the other contract is out-of-reach. The low-cost type prefers the high-wage, high-target contract even though it induces her to walk more in the contract period: the higher wage is worth the added walking costs. Thus, both types prefer the contract that is most effective for them.

Commitment Motives Just like the Tag mechanism, the Choice mechanism is implemented in a pre-period. Thus, if commitment motives are present, they can also improve the effectiveness of Choice.

Returning to an environment with a time consistent and a time inconsistent type, we next show how we can use a modified Choice menu to test whether commitment motives influence contract selection. Specifically, imagine a menu where contracts with higher step targets *do not* pay more for step-target compliance.[19] Proposition 3 implies that only time inconsistent participants with a commitment motive will select higher step targets.

Proposition 3. *Suppose participants face a Choice menu in period 0 in which two contracts pay the same incentive amount W, but the contracts have different step targets such that $S_1 > S_2$, so the contract with the higher step target is dominated.*

Time consistent participants will never select the dominated contract, but time inconsistent participants may do so.

We prove Proposition 3 in Online Appendix C. However, the intuition can be seen by inspecting the valuation curves in Figure 2. Time consistent participants will never prefer a contract with a higher step target: their contract valuations are always decreasing in the step target. On the other hand, time inconsistent participants may prefer contracts with higher step targets as long as the higher targets are in the region where valuations are increasing

[18]The figure shows an example with time-consistent agents, but the menu would still effectively segment the market if the agents were time-inconsistent but had the same baseline walking levels, s_L^* and s_H^*.

[19]In our experiment, we will collect incentive-compatible preferences over just such a Choice menu, the Choice 20/20/20 menu.

in the step target (i.e. where the red dashed line is increasing in Figure 2). Therefore, if any participants prefer contracts with a higher step target, it is evidence that those participants are time inconsistent.

Like with Tag mechanisms, the ability of a Choice mechanism to segment the market interacts with the degree of time inconsistency. All else equal, it is easier to segment sophisticated time inconsistent participants into relatively more effective contracts with Choice. Time inconsistent participants value effective contracts not only because they increase earnings, but also because they commit their future selves to walk closer to s^+.

Figure 2 also helps shed light on how commitment motives can improve Choice performance. The figure shows that if the contract with the high wage has too high of a step target, a time-consistent agent of low-walking-cost type L will eventually prefer the low-wage contract (i.e. $(V_0(L, W_{high}, S_{high})$ will fall below $(V_0(L, W_{low})$ as S_{high} increases). There are many effective high step targets that are not preferred by L, even with the higher incentive. However, if the type-L agent were time-inconsistent, her valuations would be increasing in the step targets just above s^*, increasing the range of high step targets which she would value the high step-target contract. Thus, time inconsistency allows for a broader range of menus to effectively segment types.

4 Experimental Design

We conducted our experiment between May 2019 and December 2021.[20] We gave pedometers to adults living with or at high risk for developing non-communicable diseases in Coimbatore. Participants were asked to wear the pedometers for a six-day pre-contract period, after which we incentivized participants to achieve daily step targets during a four-week contract period. The pedometers record daily steps taken, which is our primary measure of physical activity.

The key elements of our experimental design are as follows. First, to assess the overall impact of personalization, we randomly assign some participants to have their step targets personalized (either by Tag or by Choice); other participants receive a fixed step target. Second, to study manipulation and demand for commitment under a Tag mechanism, we measure pre-contract period steps both for those in the Tag group, whose step targets are increasing in pre-contract period steps, and for those whose step target assignment is independent of pre-contract period steps. Third, to study demand for commitment under a Choice mechanism, we elicit selections from a menu with weakly dominated commitment

[20]We retroactively dropped all respondents whose participation was impacted by Covid-19 pandemic lockdowns from our sample. During these lockdowns, we were unable to conduct in-person surveys, or distribute, sync, or collect Fitbits, leading to high rates of data loss and attrition. Moreover, health and legal concerns may have made walking particularly difficult during the lockdowns.

contracts from a subset of participants. Finally, to study the impacts of another potential barrier to effective personalization, imperfect information, we also introduced experimental variation in the information available to participants making choices over contracts. For brevity, we describe the additional treatment arms that we used to study imperfect information as well as the results in Online Appendix F.

Phases of the Experiment We implemented the experiment in three phases. In each phase, we tweaked the design slightly in order to answer additional research questions. We pre-registered the additional design elements of Phase 2 and Phase 3 in Dizon-Ross and Zucker (2020). All analyses control for the phase of the experiment in which participants were enrolled.[21]

The remainder of this section discusses the details of our experimental design. Section 4.1 outlines the experimental timeline and Section 4.2 describes the treatment groups. See Appendix B for further information on how the personalization interventions (Tag and Choice) were designed.

4.1 Experimental Timeline

Recruitment and Sample Selection Figure 4 shows the experimental timeline for a given participant. We recruited our sample through a series of public non-communicable disease screening camps in the city of Coimbatore, Tamil Nadu. To enroll diverse socioeconomic groups, we held the camps throughout the city in locations ranging from markets and business centers to religious institutions and parks. During the camps, trained surveyors took basic anthropometric measurements, discussed each individual's risk for diabetes and hypertension, and conducted a brief eligibility survey. To be eligible for the study, individuals needed to have either a diabetes or hypertension diagnosis or elevated blood pressure or blood sugar, have low risk of injury from regular walking, be capable with a mobile phone,

[21]The two additions in Phase 2 of the experiment were layered onto the original experiment. First, we randomized some enrollees into a new treatment group, the Baseline Choice group, who made choices at Baseline. The point of this treatment was to study imperfect information, as described in Online Appendix F. The remaining enrollees were enrolled as in Phase 1 without changing treatment balance or randomization lists. Second, we introduced cross-randomized variation in the amount of time participants had with the pedometer before eliciting contract preferences, again to study imperfect information. We began Phase 3 of the experiment only after reaching our pre-registered target sample sizes (Phase 3 could be considered a separate experiment; however, we pool results for efficiency). In Phase 3, we introduced a small group who received the contract they selected from the 20/20/20 Menu which included three contracts with step targets of 10K, 12K, and 14K, respectively, all of which paid 20 INR for each day of step target compliance. The purpose of this treatment group was to measure demand for commitment in an incentive-compatible way. We also changed the treatment balance among the remaining treatment groups: we increased the relative size of the 10K and 14K target groups and eliminated one treatment group (the Choice + Info group) that was used to study imperfect information.

and be able to receive payments in the form of mobile recharges.[22]

After screening, we contacted eligible individuals by phone and invited them to participate in a program to encourage walking, and scheduled enrollment visits with those who expressed interest. Surveyors visited interested individuals at their homes or workplaces to verify screening criteria, conduct a baseline survey, and deliver basic lifestyle modification advice.[23] Only those who completed the baseline were enrolled in the study. Enrollment was conducted on a rolling basis as eligible and interested individuals were identified.[24]

We exclude participants who withdrew or were found ineligible prior to completing the Baseline survey from all analyses, leaving a final analysis sample of 5,606 individuals.[25] The sample represents 30% of the screened, eligible population. However, we do not have pedometer data for some of these participants (because they withdrew over the course of the program), and so not all these individuals are included in analyses of walking (see Online Appendix Table G.1 for the share of people dropped in each stage of the enrollment process).

Pre-contract Period and Treatment Assignment Following the Baseline survey, surveyors launched the pre-contract period. This period was designed to measure baseline activity levels and familiarize participants with study procedures.

To measure steps, we gave all participants pedometers for the duration of the study. The pedometer step data were collected by syncing the pedometers with a central database using an internet connection. However, because most participants did not have internet access, pedometer step 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 the daily steps recorded on their pedometer.

To launch the pre-contract period, surveyors explained that we wanted to measure their steps for six days using pedometers. Surveyors then demonstrated how to properly wear a pedometer, report steps, and check text messages from our reporting system. While there

[22] The initial full list of eligibility criteria was: must be diabetic or have elevated random blood sugar (> 140 mg/dL); 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. Due to a rule change at the Indian Council of Medical Research mid-study, we were only able to collect random blood sugar from the first 1,571 respondents. We therefore adjusted the eligibility criteria to include non-diabetic individuals with a hypertension diagnosis, elevated blood pressure (systolic blood pressure > 120 or diastolic blood pressure > 80 mm Hg), or slightly lower elevated blood sugar (> 135 mg/dL) who met all other criteria.

[23] The advice was based on Government of Tamil Nadu guidelines for diabetics and hypertensives, and included increasing physical activity and healthy diet.

[24] Enrollment continued until fieldwork was ended abruptly by the advent of the COVID-19 pandemic. Enrollment resumed in the spring of 2021, and concluded in October 2021

[25] Some information regarding treatment group (i.e. whether participants were in the Tag, BL Choice, or some other group) was revealed to participants and surveyors following the completion of the Baseline survey, so we include all those who received this information in analysis where data are available.

were no financial rewards for achieving step targets in this period, respondents received 50 INR for wearing the pedometer and reporting steps for at least five of the six pre-contract period days. Surveyors emphasized to all participants, regardless of treatment group, that they should walk as normal during this period.

We scheduled a second visit with participants with a target date of either one or seven days following the end of the pre-contract period. At this visit, surveyors synced the data from the pedometers[26] and conducted a "Choice" survey to elicit choices over menus of contracts.

A key difference in timing between participants is integral to the experimental design. In particular, most respondents were informed of their treatment group (i.e. how their contract would be assigned) only after the Choice survey, after all baseline activity and choice data were collected. However, we informed respondents in the Tag group of their treatment group at the end of the Baseline, *before* launching the pre-contract period.[27]

The Contract Period Surveyors launched the contract period after completion of the Choice survey. We randomized participants into either an incentive group, in which the participant received a financial reward for achieving one of the three step targets, or the Monitoring group, in which steps were monitored but not incentivized (the treatment groups are explained in more detail in Section 4.2 below.) The contract launch included a clear explanation of treatment group assignment, which determined how the participant's specific contract was assigned. Surveyors then walked participants through their contract, describing the step target and payment level, and all monitoring and verification processes. Surveyors explained step targets in the context of health recommendations, saying, "Doctors recommend that you walk *at least* 10,000 steps a day, and more is always better![28] We recommend that you try to walk at least [step target] steps a day and build up."

All incentive groups received payments if they reported achieving their daily step target through the automated step-reporting system. We delivered incentive payments as mobile

[26]Surveyors first used the Fitbit web application to automatically sync the actual walking data from the pre-contract period to an online step database. They compared actual steps to reported steps, and reviewed the step-reporting processes as needed, before administering the Choice survey.

[27]Potential enrollees were randomized into treatment groups using list randomization (stratified by median age and gender) as soon as their enrollment visits were scheduled; however, surveyors and participants were blinded to treatment group until after the Baseline survey (for the Tag group and another group discussed below, the Baseline Choice group) or after the Choice survey (all other groups). We informed the Tag group of their treatment assignment prior to the pre-contract period because adjustments to walking by this group prior to contract launch is a key part of our analysis. We also informed the Baseline Choice group of their treatment, as these individuals selected their contract prior to the pre-contract period.

[28]Research organizations like the Center for Disease Control (CDC) and American Diabetes Association (ADA) recommend daily exercise sessions with no more than two consecutive days of rest. The recommendation of 10,000 steps approximates the number of steps that our average participant would take if he added the exercise routine recommended by the CDC and ADA to his existing behavior. It is also a widely quoted target among doctors and health advocates and a common benchmark in health studies.

recharges (credits to the participant's mobile phone account). Incentive payments were delivered at a weekly frequency along with weekly text messages summarizing walking behavior and total earnings. Participants also received text-message confirmations of their step report, payment earned, and the payment date immediately after reporting steps. Monitoring participants received similar text message confirmations and summaries.

In order to encourage accurate reporting, we paid a 100 INR bonus if participants accurately reported steps on 80% of the contract period days, and an additional 100 INR if they accurately reported steps on all 28 days. These bonuses also encouraged pedometer wearing, as we did not consider extremely low step counts to be accurate. We also conducted a number of random and targeted audits of walking during the contract period, and suspended participants who repeatedly misreported achieving their step target.[29]

At the end of the four-week contract period, we conducted an Endline survey and once again synced the pedometers. We then paid cash bonuses to participants who had correctly reported steps during the contract period.

4.2 Treatment Groups and Experiment Design

Our design features three main treatment groups (the Tag, Choice, and Fixed groups) that identify the effects of personalization, and a Monitoring group which received a pedometer but no incentives and plays the role of a control group. We also include several mechanism treatment groups that are variations on the Choice treatment; these groups allow us to gather incentive-compatible contract preferences and to shed light on the role that commitment and information play in the performance of Tag and Choice.

We begin by describing the main treatment groups and Monitoring group. Next, we describe how we elicit incentive-compatible contract preferences. Finally, we describe the remaining mechanism treatments. Figure 5 depicts our experimental design and summarizes the key features of all treatment groups.

4.2.1 Main Treatment Groups and Monitoring Group

Our three main treatment groups, the Tag, Choice, and Fixed groups, are designed to test whether personalized step targets increase daily walking relative to a fixed step target. In each treatment group, participants received a pedometer and incentives to reach a daily target of 10,000, 12,000, or 14,000 steps. However, while the Tag and Choice groups received personalized step targets, the Fixed group received a randomly-assigned uniform target.

Main Treatment 1: Tag The *Tag* group was assigned a contract paying 20 INR (about 0.33 USD) per day of compliance with a 10,000, 12,000, or 14,000 step target. The step

[29] We targeted audits at participants whose step reporting appeared suspicious and temporarily suspended those who were found to be over-reporting steps. We then re-audited those with temporary suspensions and permanently terminated their contracts if they were found to be over-reporting a second time.

target was assigned based on activity levels in the pre-contract period.[30] The step target assignment algorithm was based on our estimate of the most effective target for each activity level (see Appendix B for more detail): participants who walked fewer than 5,500 steps per day were assigned a 10,000 step target; those who walked 5,500-7,500 steps per day were assigned a 12,000 step target, and those walking more than 7,500 steps per day were assigned a 14,000 step target.

Average Daily Steps Pre-contract Period	Assigned Step Target
<5,500	10,000 steps
5,500-7,500	12,000 steps
>7,500	14,000 steps

Because we were interested in understanding the incentives to adjust baseline walking behavior introduced by the Tag mechanism, we informed Tag participants of their treatment assignment and carefully explained the Tag step-target assignment algorithm with a visual aid (see Online Appendix Figure G.1) *before* launching the pre-contract period.

Main Treatment 2: Choice *Choice* participants were assigned a contract according to their choice from a menu of three contracts, where the contracts with higher step targets also paid higher incentives for step-target compliance. The menu, which we call the "16/18/20 menu," included a contract paying 16 INR for compliance with a 10,000 step target; 18 INR for compliance with a 12,000 step target; or 20 INR for compliance with a 14,000 step target. We chose these specific payment values based on piloting various menus to see which would induce separation, as described further in Online Appendix B.

Main Treatment 3: Fixed Participants in each of the three *Fixed* groups were randomly assigned one of three possible uniform contracts. The Fixed groups received 20 INR for compliance with one of three step targets: a 14,000 step target (*14K Target* group), a 12,000 step target (*12K Target* group), or a 10,000 step target (*10K Target* group).

Our primary comparison group for determining the effectiveness of the Tag and Choice group is the 12K Target group. Among the three Fixed step targets, we guessed that the 12,000 step target would perform best on average. This guess, which we specified in our pre-analysis plan, was based on our model of heterogeneous step target treatment effects (described in Appendix B). We overweighted the fraction of participants assigned to the

[30] Specifically, we assigned step targets based on average daily steps taken on days that participants recorded at least 200 steps. If a person were to wear a pedometer consistently, it would be extremely unlikely that they would record fewer than 200 steps, and so we considered such days as missing data.

12K Target group relative to the other two Fixed step target groups to ensure power for comparisons to Tag and Choice.[31] We therefore also call the 12K Target group the "one-size-fits-all" contract group.

In order to decompose the mechanisms that influence the effectiveness of the Tag and Choice groups, we construct additional synthetic comparison groups by re-weighting individuals in the Fixed target groups.

The Monitoring group, which received pedometers but no incentives, plays the role of a control group in our study. This group allows us to evaluate the effect of incentivizing step targets relative to no incentive.

Monitoring Group Monitoring participants were treated identically to the incentive groups except that they did not receive incentives. They received pedometers and were encouraged to wear the pedometers and report their steps every day. They also received the same daily step report confirmations texts and weekly text message summaries that the incentive groups received. Finally, during the upfront explanation of the contract, surveyors also verbally encouraged the monitoring group to try to walk at least 10,000, 12,000, or 14,000 steps, with the verbal target given in the same proportion as participants were assigned to the fixed 10K, 12K, and 14K Target groups.

4.2.2 Eliciting Incentive Compatible Menu Choices

During the Choice Survey, surveyors elicited each participant's preferred contract among two contract menus, which are summarized in the table below.[32] Both menus listed three contracts: one each with a 10,000 step target, 12,000 step target, and 14,000 step target. The payments for step target compliance differed across the menus.

Contract Menu	Payment Levels (INR)			Experiment Phases in which Menu Incentive Compatible
	10,000 Step Target	12,000 Step Target	14,000 Step Target	
16/18/20	16	18	20	Phase 1, 2, and 3
20/20/20	20	20	20	Phase 3 only

[31] In Experiment Phases 1 and 2, we assigned 45% of the Fixed sample to the 12K Target group. In Experiment Phase 3, we split the Fixed sample equally between the three step target groups, as the focus of the experiment shifted from evaluating the impact of personalization to measuring demand for commitment.

[32] We also elicited incentive-compatible selections from a third menu, the 10/15/20 Menu, in all phases of the experiment. The 10/15/20 Menu compared a 10,000 step target paying 10 INR, a 12,000 step target paying 15 INR, and a 14,000 step target paying 20 INR. We assigned less than 1% of the sample to their selection from this menu. For simplicity of presentation, we do not discuss the menu in detail.

The contract selections made from each menu were incentive compatible for many sample participants. The selections were made while surveyors and participants in all treatment groups except the Tag group were still blinded to their treatment assignment. To elicit preferences from the 16/18/20 menu, during all phases of the experiment, surveyors informed all participants (except for the Tag group) that there was a chance they might be assigned to the Choice group, in which case they would receive their contract according to their selection from the 16/18/20 Menu. Thus, choices over the 16/18/20 Menu were incentive compatible in all phases of the experiment. In addition, during the third phase of the experiment, we added a *Choice 20/20/20* treatment group to our design (see Figure 5). In this group, participants were assigned the contract they chose from the 20/20/20 menu. During this phase of the experiment, all participants were informed that they might be assigned to this *Choice 20/20/20* group, in which case their contracts would be assigned based on their selection from the 20/20/20 Menu. Participants were instructed to take their menu selections seriously when they had a positive probability of implementation.[33] We therefore have incentive compatible selections from the 16/18/20 menu for all experiment phases, and from the 20/20/20 menu for Phase 3.

We elicited preferences over the contract menus with two main objectives. First, we aimed to measure participants' demand for weakly dominated commitment contracts to increase future steps. We measure this using incentive compatible choices over the 20/20/20 Menu. Second, by collecting contract selections prior to treatment assignment for a subset of participants, we are able to use these selections as baseline controls when making comparisons among these participants.

[33]During Phase 1 and 2, the 16/18/20 Menu was presented first, followed by the 10/15/20 Menu, and then the 20/20/20 Menu. Study participants making incentive-compatible choices were instructed to take the first two Menus seriously since each choice had a positive probability of being implemented; however we emphasized that the probability of being assigned the 16/18/20 Menu choice was relatively large by saying "There is a good chance that your answer to this question will determine what program you are actually enrolled in" (in practice the likelihood was 30%) while the likelihood of being assigned the 10/15/20 Menu choice was relatively small by saying "There's still a chance that you will be enrolled in the program you choose, it's just a smaller chance than for the previous question" (in practice, 6%). We informed Phase 1 and 2 participants that the 20/20/20 Menu choice would not be implemented. During Phase 3, we randomized whether we asked the 20/20/20 Menu or the 16/18/20 Menu first, and asked the 10/15/20 Menu last. We again emphasized to these participants that the first two choices had relatively large probabilities of being implemented and additionally noted that 1 in 4 individuals making each choice would receive their selection (the exact likelihood was 27% for each choice), while the likelihood of being assigned the 10/15/20 Menu choice was relatively small (in practice, 7%).

5 Data and Summary Statistics

We employ four sources of data in our analysis: the Baseline survey, the Choice survey, baseline activity data, and contract-period activity data. Section 5.1 describes these datasets. Section 5.2 summarizes the baseline characteristics of our sample overall and in each of our main treatment groups.

5.1 Data

Baseline The Baseline survey, conducted at the first household visit, contains information on respondent health, socio-economics, and demographics. Self-reported health measures include diabetes and hypertension diagnosis; physical activity levels; and a short mental health assessment.[34] Anthropometric measures include height and weight; body mass index (BMI) and waist circumference, two measures of obesity; and blood pressure, a measure of hypertension.[35] Demographic information includes age, gender, household size, and marital status. Socio-economic information includes level of education; employment and occupation; household income; and measures of wealth including home features and asset ownership.[36]

Choice The Choice survey, conducted during the second household visit, elicits respondents' preferred contract from among contract menus. As described in Section 4.2.2, each menu compares three 4-week contracts that incentivize a daily step target of 14,000, 12,000, and 10,000 steps, respectively, but the incentives for compliance with each step target differ across the menu. Respondents were presented with a visual aid for each menu to clarify the choice being presented (Online Appendix Figure G.2 shows an example). Data from the Choice survey also contains self-reports of which contract the participant thought would be most effective, the main objective driving their choices, and participant's confidence in whether they made the best choice.

Physical Activity Both baseline and contract-period activity data consist of daily step counts recorded on the respondent's pedometer. Baseline activity data are collected during the six-day pre-contract period, after the Tag group was informed of their treatment but before the other groups know their treatments. Because the Tag group was informed that their step target would depend on walking during the pre-contract period, Tag participants

[34] We measure mental health using seven questions from RAND's 36-Item Short Form Survey, a standard instrument that has been validated for use in India (Rajeswari et al., 2005; Sinha et al., 2013). We selected questions related to emotional health. Changes to the survey are our own.

[35] We initially measured random blood sugar, an indicator of diabetes, but stopped following new regulatory restrictions on taking blood samples instituted part way through the experiment.

[36] Home features include: an indicator for access to running water inside the home, number of rooms, and ownership status. Asset information includes the number of smartphones, regular mobile phones, computers, bicycles, motor bikes/scooters, and cars owned by the household; and whether the respondent has a bank account.

may have adjusted their baseline activity in response to their treatment assignment. Therefore, baseline activity is not an appropriate control for analyses involving the Tag group. Instead, we use baseline activity as an outcome in comparisons between the Tag group and other groups (for whom baseline activity did not impact contract assignment). To control for walking levels at baseline in analyses using the Tag group, we construct a prediction of average daily steps during the pre-contract period. We implement a cross-validated LASSO regression among all groups except the Tag group, regressing average daily pre-contract period steps on a variety of characteristics measured during the Baseline survey.[37] We then use the predictive coefficients from the LASSO regression to create a predicted value of baseline activity in all groups, including the Tag group.

The time-series of daily steps taken by each participant during the contract period is our primary outcome for testing the efficacy of our personalization interventions. One concern with the pedometer data is that participants might have "cheated" in order to increase their pedometer step counts without actually walking. We address this concern in two ways. First, the program design dulled the incentive for falsifying pedometer data. Incentive payments were rewarded based on self-reports through the phone system, rather than through real-time monitoring of the pedometers. The self-reported step counts were then audited during the contract period for a random subset of participants, as well as those with suspicious reporting patterns (e.g. always surpassing the step target, round number entries), and those who consistently over-reported were suspended from the program. Thus, while there was still an incentive to falsify pedometer data among the incentivized groups (since there was a risk of termination if participants over-reported relative to what was shown on the pedometer), the incentive was substantially less than if we had made payments based on the pedometer step counts themselves. Instead, a simpler mode of cheating was to simply over-report. Second, we monitored for what we saw as the most worrisome type of potential cheating: sharing the pedometer with another, potentially more active, individual. We monitored pedometer sharing through the same audit process as over-reports. Specifically, we visited participants unannounced at their homes and workplaces, and checked if the pedometer was with them or someone else, and then synced the pedometer data to check for over-reporting. During the 1705 audits we conducted, we witnessed only one example of pedometer sharing.[38]

A second concern with the pedometer data is that the incentives for walking also increased the incentives for participants to wear their pedometers. This could lead us to overestimate the effectiveness of treatments that lead to more frequent pedometer wearing. Our bonus payment at Endline for pedometer wearing and correctly reporting steps during the contract

[37] We implement this procedure using the cvlasso command from Ahrens et al. (2020). The Lasso procedure selects predictors of baseline activity from among the covariates in Panels A and B of Appendix Table A.3.

[38] Appendix Table A.1 contains statistics on pedometer sharing and misreporting.

period was designed to counter this issue: participants' bonus payments were conditioned on the number of days they achieved at least 200 steps *and* reported accurately. Appendix Table A.2 shows that while incentivized participants do wear their pedometers slightly more often than Monitoring participants, the difference is small (only 1.5% of days, or less than one day of pedometer-wearing per participant across the entire contract period) and statistically indistinguishable from zero. Appendix Table A.2 also shows that pedometer wearing is not significantly more common in the two primary personalization groups, Tag and Choice, than it is in the "one-size-fits-all" 12K Target group.

5.2 Summary Statistics and Balance Checks

Characteristics of the full sample of respondents who completed the Baseline (prior to randomization) are in column (1) of Table 1. Panel A summarizes demographic characteristics. Participants' average age was 49. 37% of the sample were female, and 58% had completed some secondary education. The average monthly income per capita was just over 5400 INR, making the step target incentive of 20 INR equivalent to approximately 11% of average daily per capita income.

Measures of participants' health, which are shown in Panel B, show that the sample has high rates of chronic disease and disease risk. 31% of the sample had been diagnosed with diabetes and 30% with hypertension. Average blood pressure and BMI levels are both extremely high. The average blood pressure measurement of 137 /92 mm Hg exceeds the hypertension cutoff of 130/80 mm Hg or greater. The average BMI of 26 kg/m2 is in the obese range for Indians.[39] During the pre-contract period (when there were no step target incentives), participants walked an average of 6288 steps per day, which is very similar to average steps taken by Fitbit pedometer users across India (Dube, 2020).[40] Because these baseline activity data were collected after the Tag and Baseline Choice groups were told their treatment, the baseline activity statistics (Panel C of Table 1) exclude the Tag and Baseline Choice groups.

We show that baseline characteristics are balanced across treatment groups in Columns (3) through (9) of Table 1.[41] Omnibus tests of balance across all covariates, at the bottom of the table, fail to reject that any treatment group with a personalized step target has the same baseline characteristics as the 12K Target group (Bruhn and Mckenzie, 2009).

[39]In India, normal BMI is considered 18.0-22.9 kg/m2, overweight is 23.0-24.9 kg/m2, and obesity is ¿25 kg/m2. Indians have lower BMI cut-offs for overweight and obesity than other racial groups, in part because they tend to be susceptible to non-communicable disease at lower BMI levels.

[40]Indian Fitbit users averaged 6,533 per day in 2019.

[41]For expositional clarity, Table 1 pools the Fixed step target groups and pools variations to the main Choice treatment. We report summary statistics separately for each subtreatment and experiment phase in Online Appendix Table G.2.

6 Results

This section empirically examines the impacts of personalization on incentive effectiveness. We first show that while incentives increase average walking even without personalization, the 10K, 12K, and 14K targets impact walking patterns in very different ways, suggesting room for personalization. Second, we test the impacts of our main personalization treatments on performance. Finally, we show evidence on the mechanisms driving the efficacy of the Tag and Choice treatments, focusing on the role of time inconsistency.

6.1 Incentive Impacts without Personalization

We first test whether providing financial incentives for each step target increases walking in the contract period. Specifically, we compare average steps taken in each of the fixed target groups with the monitoring group. This comparison isolates the impact of the financial incentives for each target, holding constant motivational effects of having access to a pedometer and other aspects of the incentive program. Our analysis takes the following form:

$$y_{itk} = \alpha + \beta_1 \times 10\text{K Target}_i + \beta_2 \times 12\text{K Target}_i + \beta_3 \times 14\text{K Target}_i + \boldsymbol{X}_i' \gamma + \boldsymbol{X}_{it}' \lambda + \mu_k + \varepsilon_{it}, \quad (7)$$

where i represents a participant, t represents a date, k represents the experiment phase, and μ_k are indicators for the experiment phase in with the participant was enrolled. The outcome, y_{itk}, is individual i's steps on day t during the contract period; and $\boldsymbol{X}_{it}$ is a vector of day-level controls including calendar month-year, contract-week, and day-of-week dummies. $\boldsymbol{X}_i$ represent a vector of individual-level controls selected from more than four hundred potential individual-level covariates using the post-double-selection LASSO method of Belloni et al. (2014).[42] The omitted group is the monitoring group. Standard errors are clustered at the participant level.

The coefficients of interest, β_1, β_2, and β_3, represent the average treatment effects of each step target across the full sample relative to the monitoring group. Table 2 shows the coefficients, and Panel (a) of Figure 6 shows the results graphically, with the 95% confidence interval depicted on the incentives bar representing a test for equality between each of the fixed step target group and the monitoring group (as is the case for all the graphs in this section). All three step targets have positive impacts on daily walking, ranging from 610-849 steps. This is equivalent to approximately 6–9 additional minutes of brisk walking, on

[42]We implement the covariate selection method using the pdslasso command in Stata developed by Ahrens et al. (2018). Column (1) of Appendix Table A.3 shows the set of baseline characteristics from which the covariates are selected. Appendix Table A.4 shows that the results are qualitatively similar without any individual-level controls, as well as when we include pre-contract period step controls in the pool of covariates from which we select controls with double-LASSO.

average, each day — roughly a 9-12% increase relative to the monitoring group.[43]

While each of the three step targets have similar and statistically indistinguishable impacts on average daily steps, other outcomes show key differences that suggest the targets may be differently effective for (and differently valued by) different participants. First, participants in each group earn very different amounts. Panel (b) of Figure 6 shows that participants randomly assigned the 10K Target earn nearly twice those assigned the 14K Target. Second, participants' daily step patterns clearly respond to the assigned step target. Panel (a) of Figure 7 shows histograms of daily contract-period walking in each step target: daily steps bunch steeply just above whichever step target is assigned. Third, the distribution of average walking across the contract period *between* participants varies greatly across the three contracts. Panel (b) of Figure 7 plots kernel densities of average daily walking in the contract period across participants in the different step target groups as well as the Monitoring group. The three targets appear to motivate different segments of the walker distribution to walk more on average. Compared to the Monitoring group, the 10K Target has special success at moving the lowest walkers to higher levels of walking, and the 12K and 14K target have more success shifting the highest walkers—but do not move the lowest walkers at all.

We next directly test whether higher step targets work better for those who walk more in the pre-contract period, as Proposition 1 suggests is likely to be the case. Specifically, we check whether unmanipulated baseline activity interacts positively with the size of the step target assignment in increasing steps. Among participants in the Fixed target groups only, we run a regression of the following form:

$$y_{itk} = \alpha + \beta_1 \times \text{Step Target}_i \times y_i^{BL} + \beta_2 \times y_i^{BL} + \textbf{Fixed Target}_i'\delta + \mu_k + \varepsilon_{it}, \qquad (8)$$

where y_i^{BL} is the average of steps taken in the pre-contract period (for days on which total steps exceeded 200), which is the variable we used to assign step targets in the tag algorithm, and Step Target$_i$ is a continuous measure of the step target assigned to participant i (in thousands). $\textbf{Fixed Target}_i'$ is a vector of indicators for whether individual i was assigned to each of the fixed step targets, where the 12K Target group is omitted. The remainder of the variables are defined as in Equation 7. The coefficient of interest, β_1, can be interpreted as the additional increase in daily contract-period steps from increasing the step target by one

[43] One concern is that incentives impacts might be driven by an excitement to try something new early in the contract period, but may not be sustained over time. To examine this possibility, we map out the evolution of incentive effects over time using regressions with the same overall specification of Equation 7 separately for each week of the contract period. The results, in Appendix Figure A.1, show that while incentives do have a particularly strong impact in the first week, they remain effective throughout the 4-week program. This alleviates concerns that our results might not be relevant for incentive programs with longer duration.

thousand steps for individuals whose average pre-contract period walking is 1 step higher. The results, shown in Column (3) of Appendix Table A.6, show that β_1 is positive and significant, confirming that higher step targets are indeed better for higher baseline walkers. The other columns of Appendix Table A.6 show that the result is robust to alternative specifications.[44] Altogether, the results show encouraging evidence that incentives for each of the three selected step targets are not only effective on average, but also differently effective for different people.

6.2 Impacts of Personalization

We next test whether personalizing step targets improves walking in the contract period. We estimate a regression of the following form:

$$y_{itk} = \alpha + \beta_1 \times \text{Tag}_i + \beta_2 \times \text{Choice}_i + \boldsymbol{T}_i'\delta + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\lambda + \mu_k + \varepsilon_{it}, \tag{9}$$

where Tag_i and Choice_i are indicators for being in the Tag and Choice groups, respectively. The omitted treatment is the 12K Target group. $\boldsymbol{T}_i$ is a vector of indicator variables for the Monitoring, 10K Target, and 14K Target groups. The remaining notation is the same as Equation 7. The coefficients of interest are β_1 and β_2, which can be interpreted as the intent-to-treat impact of the Tag and Choice treatments, respectively, on daily walking relative to the 12K Target group. We include the 10K Target and 14K target groups in the regression so that we can easily use the regression output to construct other comparison groups for the Tag and Choice groups (other than the 12K Target comparison group), as described in Section 7.1.

The results, shown in Table 3 and graphically in Panel (a) of Figure 9, indicate that both personalization treatments substantially improve average walking relative to the 12K Target group.[45] The Tag treatment increases walking by 529 steps beyond the 12K Target group, and the Choice treatment increases walking by 446 steps. These estimates represent increases of roughly 90% and 75% of the estimated 12K Target impact, respectively. Table A.5 shows that our findings are robust to alternative controls. Column (1) excludes all but experimental phase controls, with similar results. While the Table 3 results (reshown in column (2) of Table A.5) exclude pre-contract period walking controls as they are likely endogenous to Tag assignment, column (3) of Table A.5 re-estimates the model excluding the Tag group and allowing the double-LASSO to select pre-contract period walking controls.

[44] Alternative specifications include adding double-Lasso selected covariates (Columns (2) and (4)), and interacting pre-contract period walking with the 10,000 and 14,000 step target (Columns (1) and (2)).

[45] As discussed in our registered report, we designed our experiment to detect the effects of personalization among participants in Phase 1 and Phase 2 of our experiment. Table G.3 shows that the results are robust to excluding Phase 3 (which was primarily designed to shed light on time-inconsistency and Choice.)

The estimated impact of the Choice treatment is similar in this specification as well.

We designed our Tag and Choice treatments to try to maximize average daily steps. Although this decision in part reflected practical concerns (we did not have the information available to try to optimize a more nuanced metric), it is common practice to use total behavior as the primary metric of success in evaluation of programs for behavior change (e.g., Banerjee et al., 2010).[46] That said, in practice, principals may be concerned with other objectives. For example, some may care about maximizing average steps relative to the cost. Others may believe that the benefit of steps is non-linear, and hence care about maximizing a non-linear function of steps instead of the average. We discuss these two potential other objectives in turn.

For principals who want to maximize walking relative to the cost, we also find promising results: the Tag and Choice treatments improve walking outcomes without substantially increasing total program costs. Specifically, the two personalization methods also have minimal impacts on earnings: Panel (b) of Figure 9 shows that while earnings are slightly higher in both the Tag and Choice groups, the difference is not statistically different from zero. The minimal change in earnings from personalization contrasts with the impacts of shifting the fixed step target: Among the three step targets we tested, raising the step target had minimal impacts on average steps, but resulted in much lower payments (recall Panel (b) of Figure 6). Personalization treatments, on the other hand, have large impacts on average steps, but result in nearly the same payments. Thus, relative to moving between fixed step targets, adding personalization to the principal's toolbox dramatically improves the technology in the "production function" that translates incentives to steps. Figure 11 displays this intuition graphically with a scatter plot of average walking as compared to average payments in the different treatment groups. Average walking is higher in the Tag and Choice treatments than the other treatments, while average incentive payments are similar to those in the 12K Target group.

Regarding our focus on average daily steps: it is first worth noting that this focus, which implicitly values all steps equally (i.e. assumes there is a linear welfare return to steps), appears to be reasonable in our setting. While there is broad agreement that the health returns to physical activity tend to flatten after a point, the evidence on the shape of returns to fairly low levels of activity levels is mixed: while some studies find concave returns, others find linear returns (Foulds et al., 2014; Loprinzi, 2015; Warburton et al., 2006), especially at low levels of activity. Activity levels are low enough among nearly all adults in urban India—

[46] This objective is in fact consistent with welfare maximization as long as the marginal benefits to additional steps are sufficiently high. In our case, maximizing steps will also maximize welfare if the costs of our margin of adjustment (moving participants between the contracts we consider in our experiment) are smaller than the marginal social benefits of additional steps. In such a case, the net welfare gains from additional steps is always positive, and so maximizing average daily steps is equivalent to maximizing welfare.

including our sample—to warrant the approximation of linear returns. However, it is likely that future evidence will deepen our understanding of the shape of the physical activity dose-response curve. In order to shed light on how this might change our understanding of the welfare effects of the two personalization mechanisms, we next examine how each mechanism changes the distribution of physical activity across participants.

While Tag and Choice have similar impacts on average walking, participants in the two groups have quite different step targets. Panels (a) and (b) of Figure 8 show the fraction of individuals assigned to each step target within the Tag and Choice groups, respectively. Step targets tend to be higher among Tag participants: while nearly 50% of the Tag group was assigned the 14K step target, more than 50% of the Choice group chose the 10K step target. However, both mechanisms led to clear segmentation based on baseline walking levels, with higher step targets in general going to those who walked more in the pre-contract period. Panel (d) of Figure 8 shows overlapping histograms of baseline activity among Choice participants who chose each of the three step contracts: the 10K step target (paying 16 INR for compliance), the 12K step target (paying 18 INR for compliance), and the 14K step target (paying 20 INR for compliance). While the histograms of baseline activity show substantial overlap compared to the Tag group in Panel (c), where sorting was a mechanical function of baseline activity, participants who chose higher step targets clearly tend to have higher activity levels.[47]

The two personalization treatments also result in different distributions of walking. Panel (c) of Figure 9 plots kernel density plots of average contract-period steps across participants in each of our three main treatment groups, and Table 4 reports quantile regressions with the same controls as column (1) of Table 3. The results show that both personalization methods shift the very bottom of the average walking distribution up relative to the 12K Target. However, the impact of the Choice treatment on median walking is larger than the Tag treatment, while the Tag treatment has larger impacts than Choice on most larger quantiles of the distribution. One potential explanation is that, with Choice, people are less likely to end up with a step target that is wildly inappropriate for them, and so nobody is lost at the bottom of the distribution. In contrast, the Tag treatment can force more people into more aggressive targets that push up the top of the distribution. These results suggest that if returns to physical activity are in fact concave in our sample, Choice may be the better personalization option, while if the returns are convex, Tag may improve welfare more.

[47] To make a more apples-to-apples comparison of sorting by baseline activity in Tag and Choice, Appendix Figure A.2 compares the *predicted* baseline walking in each step target among Tag and Choice participants. The predictions of baseline walking are based on non-manipulable baseline survey covariates, and the predictive model is fit using participants who were not in the Tag or BL Choice groups and therefore had no incentive to manipulate baseline activity. A similar pattern emerges: Tag participants are much more sharply segmented based on even non-manipulable prediction of baseline walking.

The quantile treatment effects in Table 4 also provide a concrete illustration of how personalization improves upon a one-size-fits-all approach. First, echoing Figure 7 panel (b), the coefficients on the 10K and 14K target in Table 4 show that any given *Fixed* target has advantages and disadvantages. Lowering the Fixed target from 12K (the omitted group) to 10K substantially increases the lower quantiles of the distribution, but has a large negative impact on the 75th percentile. Correspondingly, raising the Fixed target from 12K to 14K positively impacts the top of the distribution but has negative (albeit insignificant) impacts at the lower quantiles. Thus, all Fixed targets have downsides. In contrast, the Choice treatment is able to achieve all of the positive impact of the low 10K step target at the bottom of the distribution without having nearly the same large negative impacts at the top. Correspondingly, the Tag treatment achieves much of the positive impact of the 14K target at the top of the distribution without having its negative impacts at the bottom. Personalization thus has the potential to achieve much of the upside of using more extreme Fixed targets without the downsides.

7 Channels for the Impact of Choice and Tag

This section examines the channels through which each personalization mechanism improves contract-period walking. We first show that the personalization is not driven by random reshuffling from the fixed 12K Target group into the other, more effective Fixed step target groups. We next show that the Tag mechanism operated through two additional channels: the tag predicted heterogeneity in step-target effectiveness, and—due to a commitment motive—the endogeneity of the tag allowed for further advantageous self-selection based on private information. Finally, we examine the mechanisms behind Choice. We show that, on average, participants self-selected into step targets that were most effective. We find evidence that time inconsistency improved the Choice effectiveness. We then turn to examine the role of information in Choice, and do not find evidence that limited private information about future walking costs hindered the effectiveness of Choice.

7.1 Reshuffling Step Targets

Section 6.2 shows that personalization improves average walking relative to the 12K Target, which was the one-size-fits-all target that we (the researchers) chose as a comparison group. One potential, and somewhat uninteresting, explanation for this observation is that because the 10K and 14K Targets are more effective than the 12K Target, we could achieve the same result by randomly reshuffling participants from the 12K into these targets.

Figure 3 shows that this is not the case. While the 12K Target does in fact perform worse than the 10K and 14K Target, the differences are not large enough to explain the effects of personalization. The Tag and Choice do better than any of the fixed targets on average, and statistically significantly better than the Fixed step target groups reshuffled so that each step target is represented in the same proportion as the respective personalized step target. For example, the Tag allocated 32% of participants to the 10K Target, 23% to the 12K Target, and 45% to the 14K Target. While reshuffling the Fixed Step Target groups in these proportions increases steps by 110 per day relative to the 12K Step Target only, the Tag leads to an *additional* 420 daily steps beyond reshuffling (p-value that the difference is 0 is .022). Similarly, the Choice does 386 average daily steps better than reshuffling the Fixed groups to represent the Choice step targets (p-value .041).

Having shown that the impacts of personalization are not explained by our particular guess of the most effective one-size-fits-all target, we now turn to the rest of the story.

7.2 Mechanisms for Tag

7.2.1 Sorting through Exogenous Tag

The premise of the Tag mechanism is that it can segment the market: it can sort participants by observed baseline activity, and assign each baseline activity group a step target that is more effective than the other step targets (as explained in Proposition 1). In order to explore how the sorting works, we decompose observed baseline activity into an exogenous component, or the baseline activity that would have occurred without an incentive to manipulate, and an endogenous component, or the manipulation to baseline activity in order to self select into step targets under the Tag mechanism.

What would the impact of the Tag algorithm have been if there were no manipulation? We answer this question next. To do so, we imagine an "Exogenous Tag" group composed of participants in the fixed 10K, 12K, and 14K Target groups who were randomly assigned the step target that the Tag algorithm would have assigned them to. Because the participants in these groups were randomly assigned one of the three step targets, they had no incentive to manipulate baseline activity. We then compare walking in the Exogenous Tag group to walking in the Fixed 12K Target group.

Because we assigned more Fixed target participants to the 12K Target than the 10K or 14K Target, the 12K target is over-represented in our Exogenous Tag group. This means that people who walked a medium amount during the pre-contract period (i.e., between 5,500 and 7,500 average daily steps) are overrepresented in the Exogenous Tag group. Because the Exogenous Tag group is selected, we cannot directly compare walking between the Exogenous Tag group to the 12K Target group. We therefore must reweight the members Exogenous Tag group to undo the selection effect. In practice, we achieve reweighting by estimating heterogeneous treatment effects of the Fixed step targets for three groups who are unequally selected into the Exogenous Tag group: people whose Tag-assigned target (or Tag Target) is the 10K, 12K, and 14K step target, respectively. The impact of the Exogenous Tag without selection is simply a re-weighting of the treatment effects of being assigned the 10K Target for the low walkers, the 12K Target for the medium walkers and the 14K Target for the high walkers, where the weights are the fraction of low, medium, and high walkers in the population. We estimate the heterogeneous treatment effects using the following regression:

$$y_{itk} = \alpha + \textbf{Fixed Target}_i \times \textbf{Tag Target}_i'\beta + \textbf{Tag Target}_i'\delta + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\lambda + \mu_k + \varepsilon_{it}, \quad (10)$$

where **Tag Target**$_i$ is a vector indicators for the step target our Tag algorithm would have assigned participant i based on her baseline activity (medium walkers who would have been assigned the 12K Target are ommited). **Fixed Target**$_i'$ is a vector of indicators for whether individual i was assigned to each of the fixed step targets, where the 12K Target group is omitted. The sample includes only Fixed step target group participants.

The results are shown at the bottom of Table 5: the reweighted Exogenous Tag group walks 300 more steps per day than the 12K Target group.[48] This accounts for 84% of the 529 additional daily steps walked by the Tag participants in the contract period (as shown in Table 3). While the estimated impact of the Exogenous Tag is not very precise, the magnitude suggests that without manipulation, our Tag algorithm would have achieved most, but not all, of its full impact.

The results above indicate that the exogenous component of the Tag accounts for a large portion of its success. However, we are not sufficiently powered to reject that the Exogenous Tag group is only as effective as the 12K Target. We next perform a more high-powered test of whether the Exogenous Tag assigned step targets better-than-randomly. Specifically, we test whether Fixed step target group participants who were assigned to the step target that our Tag algorithm suggests, i.e., the Exogenous Tag group, do better than Fixed step target participants who were assigned to *other* step targets (pooled together for power). Our test takes the following form:

$$y_{itk} = \alpha + \beta_1 \times \text{Exogenous Tag}_i + \textbf{Tag Target}_i'\delta + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\lambda + \mu_k + \varepsilon_{it}, \quad (11)$$

where Exogenous Tag$_i$ is an indicator variable equal to 1 if participant i is in the Exogenous Tag group. **Tag Target**$_i$ are indicators for the step target our Tag algorithm would have assigned participant i based on her baseline activity (this control guarantees that our comparison is within Tag Target group, and corrects for the increased selection of those with a 12K Tag Target in the Exogenous Tag group). The omitted group was not assigned the

[48]The coefficient, 299.5, is equal to the coefficient on Low walkers × 10K Target times the fraction of the Fixed target sample who are Low walkers, plus the coefficient on High walkers × 14K Target times the fraction of the sample Fixed target sample who are High walkers.

step target selected by the Tag algorithm. The remainder of the variables are defined as in Equation 7. The test of the Tag algorithm effectiveness absent manipulation is whether β_1 is significantly greater than zero. The results from this test, shown in Appendix Table A.7, indicate that the Tag algorithm indeed assigned targets better than randomly: participants in the Exogenous Tag group walk a statistically significant average of 361 more steps than participants who were randomly assigned alternative step targets.[49]

7.2.2 Manipulation and Time Inconsistency

In practice, baseline walking was endogenous to Tag assignment. Participants in the Tag group faced an "earnings" motive to manipulate baseline steps downward in order to more easily earn incentives during the contract period. In addition, they might have faced a "commitment" motive to manipulate baseline steps upward in order to commit to additional healthy walking during the contract period. In this section, we examine how manipulation of both types impacted the effectiveness of the Tag mechanism. We first show evidence from the baseline walking behavior of the Tag that commitment motives exist in our setting using Proposition 2. We then examine whether the commitments made by Tag participants were sufficiently effective that they outweighed the impacts of the earnings motive, and thereby improved Tag performance.

We first show that, on average, Tag participants walk more during the pre-contract period, presumably in order to be assigned a higher step target. Panel (a) of Figure 12 compares the distribution of baseline activity for individuals in the Tag group and other groups. A Kolmogorov-Smirnov test rejects that pre-contract period walking among the Tag group is drawn from the same distribution as the rest of the sample (P-value =.047). Moreover, walking in the Tag group appears visibly distorted at the 5,500 steps per day cutoff, above which Tag individuals were assigned the 12K Target. While we lack statistical power to detect the discontinuity in the main experimental sample (P-value .176), if we also include pre-contract period walking for individuals who completed the pre-contract period but whose contract period was later interrupted by Covid, we reject the null of no manipulation at the 5,500 daily pre-contract period step cut-off (P-value =.034) using the test of Cattaneo et al. (2020).[50] While there is no obvious manipulation of baseline walking at the higher 7,500 cutoff, above which individuals were assigned to the 14K target, there are significantly more individuals in the Tag group who walk at very high levels during the pre-contract period. Panel (b) of Figure 12 compares the fraction of participants assigned each step target assignment in the Tag group compared to the step targets that the Tag algorithm would have

[49]In Online Appendix E, we explore potential improvements to the Tag algorithm using causal forest estimates of the average treatment effect of each Fixed step target conditional on baseline activity levels.

[50]This testing procedure is based on local polynomial density estimators as implemented by the "rddensity" command in Stata.

assigned to non-Tag participants.[51] The results indicate that endogenous response in the Tag group led to a 2.9 percentage point (pp) increase in the fraction of participants assigned to the 14K step target, and a -2.6 pp and -.2 pp decrease in the fraction of participants receiving the 10K and 12K step targets, respectively.[52]

One concern is that Tag participants did not understand the step target assignment algorithm, but adjusted baseline walking for some other reason than to get a preferred contract. Two pieces of evidence suggest that this is not the case. First, following contract launch, we quiz participants on how their contracts were assigned. More than 90% of Tag participants correctly report that their contracts were assigned based on their walking during the pre-contract period, indicating that the step target assignment mechanism is broadly understood (see Online Appendix Table G.4). Second, the fact that we see a discontinuous jump in baseline walking at the cut-off for being assigned the 12K Target suggests that participants understood the step target assignment algorithm quite precisely.[53]

While this evidence suggests that many Tag participants make commitments to walk more in the future by increasing pre-contract period walking, we cannot guarantee that the commitments are effective. We also cannot rule out that the earnings motive caused other participants to shift walking downward in order to be assigned an easier, but potentially less effective, contract. Thus, the total impact of the endogenous response to the Tag remains an empirical question.

We next test whether the endogeneity of baseline walking improved the performance of the Tag. To do so, we compare walking in the Tag group to walking in the Exogenous Tag group.[54] Specifically, we run a regression of the following form:

$$y_{itk} = \alpha + \beta_1 \times \text{Tag}_i + \boldsymbol{X}_i'\gamma + \boldsymbol{X}_{it}'\lambda + \mu_k + \varepsilon_{it}, \tag{12}$$

where the variable definitions are the same as in Equation 9, but the omitted group is the Exogenous Tag group and the double-LASSO-selected control variables, X_i, exclude baseline activity levels (which are endogenous in the Tag group). In order to correct for the unequal probabilities of being assigned to each fixed step target group, observations are weighted by

[51] We exclude the Baseline Choice group from these analyses, as their pre-contract period walking was also endogenous to treatment.

[52] Tag endogeneity leads to a statistically significant increase in the average step target by 138 steps (P-value = .027)

[53] In addition, prior to contract launch, we ask participants if they walked more, less, or the same as usual during the baseline period and why. Ten percent of Tag participants who report walking more than usual say they did so to get a higher step target, even though we instructed people to walk as normal (Appendix Table A.8).

[54] Recall this group is composed of participants in the fixed 10K, 12K, and 14K Target groups who were randomly assigned the step target that the Tag algorithm would have assigned them to.

the inverse probability of treatment. The coefficient of interest is β_1, which shows the average impact of endogenous baseline walking adjustments on walking in the contract period.

Importantly, β_1 does not isolate the impacts of commitment motives on contract period walking in the Tag group. Instead, it represents the combined impact of commitment and earnings motives. While earnings motives will lead to lower contract period walking in most cases,[55] commitment motives have an ambiguous effect: they increase walking if people are sufficiently sophisticated, but can backfire and decrease walking if people are naive. Therefore, a positive β_1 is suggestive that commitment motives are effective, but not conclusive (we will present more conclusive evidence on the effectiveness of commitment motives in Section 7.3 below).

The results are reported in Table 6. Walking in the Tag group is somewhat higher than in the Exogenous Tag group, though the difference is not statistically significant. Therefore, the evidence indicates that Tag participants who modify their baseline walking upward are sufficiently sophisticated that their commitments do not backfire, and suggests that commitments improve contract-period walking.

7.3 Sorting in Choice

While the Tag mechanism combines sorting by the principal with an element of self selection to segment the market, sorting in Choice operates purely through self selection. Therefore, the effectiveness of Choice relative to the Fixed step targets is clear evidence that Choice participants self selected into more effective step targets.[56] This evidence is even more convincing given that the contracts assigned to Choice participants paid weakly less than the contracts in the Fixed step target groups. If Choice participants had chosen step targets randomly from the 16/18/20 Menu, we would expect the Choice treatment to do worse than the reshuffling of the Fixed target group because the Choice contracts paid less than the Fixed contracts for compliance with the 10K and 12K step targets (i.e., 16 and 18 INR instead of 20 and 20 INR, respectively). Instead, Figure 10 showed that allowing Choice leads to higher contract-period walking than a reshuffle of the Fixed step targets. showed that allowing Choice leads to higher contract-period walking than a reshuffle of the Fixed step targets.

[55]Specifically, earnings motives will reduce contract period walking when they motivate participants to self-select into a lower step target that is less effective. However, it is possible that for some participants, the Tag assigned step target without manipulation is out-of-reach, in which case the earnings motive can actually improve contract period walking.

[56]Online Appendix D examines an alternative behavioral mechanism that could improve Choice effectiveness without self selection into more effective step targets: autonomy effects from being allowed to choose. We do not find evidence that autonomy improves the effectiveness of Choice.

7.3.1 Time Inconsistency and Choice

Having found evidence that Choice participants self-sort into more effective step targets, we next investigate the role of time inconsistency in self-sorting. We test how much participants' commitments to future walking affect the choices they make. To do so, we we focus on the 20/20/20 Menu, which presented a choice over the same three contracts assigned by the Tag mechanism. All three contracts paid the same 20 INR incentive for compliance with each step target.

Like in the Tag, the contract with the 10K target weakly dominates the other contracts on the 20/20/20 Menu: for any walking level during the contract period, it pays weakly more than the other contracts. However, also like in the Tag, the dominated contracts may serve as commitment devices for those who are time-inconsistent: by choosing a higher step target, participants encourage their future selves to undertake a costly action (i.e., walk more) for a distant benefit (i.e., better health). If participants select either of the dominated contracts, it is evidence that they place a net-positive value on increasing their future walking above the status quo, leading to partial alignment between principal and agent preferences at the contract choice stage.

Figure 13 shows that many participants prefer dominated commitment contracts. The figure shows the distribution of incentive-compatible choices in the 20/20/20 Menu.[57] One third of participants making the choice selected a dominated contract: 15.4% selected the 12K target paying 20 INR and another 14.2% selected the 14K target paying 20 INR over the 10K target paying 20 INR. It appears that time-inconsistency plays an important role in how participants select contracts in our setting.

One concern is that the demand for commitment contracts that we observe in the 20/20/20 Menu choices are driven by a lack of understanding. Therefore, for a subset of participants, we asked questions to confirm that they understood that the contracts with the 12K and 14K step targets were dominated. Specifically, we first ask participants how much they would be paid if they selected the 10K target and then walked 10,100 steps. We then ask participants how much they would be paid if they selected the 12K step target and then walked 10,100 steps. Online Appendix Table G.4 shows that 90% of participants answered both questions correctly. This suggests that participants broadly understood that the 12K and 14K step targets were weakly dominated.

A second concern is that we might observe demand for commitment contracts because participants in our sample do not value the potential incentive payments very much. If participants don't care whether of not they receive incentives, then the selection of a domi-

[57] One potential concern with this strategy is that there may be a strong "priming" effect for those participants who make the choice over the 20/20/20 Menu just after the 16/18/20 Menu choice. To alleviate this concern, we randomized the order of the menu choice. We do not find that choice order affects choices.

nated contract is not very meaningful.[58] Three pieces of evidence suggest that participants do value the incentives. First, participants' contract selections indicate that payment level is a component of preferences. We collect incentive compatible preferences over three contract menus. Appendix Figure A.3 shows that participants are more likely to select low step targets from menus where the low step targets pay more. As the payment amount for low step targets decreases (i.e., moving from right to left), more participants select the high step targets. Second, participants' contract period walking responds to incentives. Participants in all incentives groups walk many more steps than the Monitoring group. In addition, Appendix Table A.9 shows that among incentivized participants with the same step target (and the same preferences), participants walk more when their payment for step target compliance is higher. Finally, we find that wealth is a key predictor of heterogeneous treatment effects, suggesting that the marginal value of the incentive plays an important role in contract effectiveness. Not only is baseline wealth one of the most important variables in causal forest estimates of heterogeneous treatment effects of the three step targets (Online Appendix Table G.5), but also Appendix Figure A.4 shows that the 14K step targets is relatively more effective for individuals with lower wealth at baseline, suggesting that they are willing to undertake more steps in order to receive the payment.[59]

The evidence thus indicates that commitment demand influences contract choice, and motivates some participants to choose higher step targets.

8 Conclusion

This paper investigates the possibility of adapting second- and third-degree price discrimination strategies to improve policy effectiveness. Our personalization mechanisms were designed imperfectly: we did not explicitly model heterogeneous walking costs, or gather experimental data on heterogeneous treatment effects of different step targets. Yet, we find that the approach of second- and third-degree price discrimination can be harnessed to increase policy impact: even our imperfect personalization of incentive contracts for walking improves upon a one-size-fits-all contract by about 90%. This finding suggests that personalized policy may have wide promise for policymakers, even when the policymaker has limited information.

This is in part because both personalization strategies interact with an important force that characterizes many settings with internalities like present bias: beneficiaries' demand for commitment. If people are time-inconsistent and sophisticated, they will demand incentive contracts that commit their future selves to increase health behaviors. Their preferences are

[58] A related concern is that participants do not trust that we will pay them. To build trust, we delivered a test recharge to all participants during the pre-contract period.

[59] The methodology of the causal forest estimation is described in Online Appendix E.

thus partially aligned with the objectives of the principal. This alignment motivates agents to use their private information to improve incentive effectiveness.

We show that the demand for commitment contributes to the success of both of our personalization strategies. Not only do 30% of participants select dominated commitment contracts from a menu, but participants are also willing to undertake costly additional steps in the pre-contract period in order to receive these dominated contracts. Moreover, they have sufficient sophistication and private information about future walking costs that self-selection into commitment contracts increases walking during the contract period.

Our results have broad implications for the design of personalized policy. First, commitment motives can alleviate concerns that individuals will self-select into contracts that result in more payments but less behavior change. Second, individuals often know what works for them even when the principal does not. Third, the timing of the choices and behavioral adjustments may play a critical role in their success. Commitment motives only exist for future behaviors. Therefore, personalization strategies that have an element of self-selection may be more effective if they are implemented sufficiently far in advance.

References

Aggarwal, S., Dizon-Ross, R., and Zucker, A. D. (2020). Incentivizing behavioral change: The role of time preferences. *NBER Working Paper*, page No. 27079.

Ahrens, A., Hansen, C., and Schaffer, M. E. (2018). pdslasso and ivlasso: Progams for Post-Selection and Post-Regularization OLS or IV Estimation and Inference. *Statistical Software Components, Boston College Department of Economics*.

Ahrens, A., Hansen, C., and Schaffer, M. E. (2020). lassopack: model selection and prediction with regularized regression in Stata. *The Stata Journal*, 20(1).

Ainslie, G. (1992). *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person*. Cambridge University Press, Cambridge, UK.

Alatas, V. et al. (2016). Self-Targeting : Evidence from a Field Experiment in Indonesia Abhijit Banerjee Rema Hanna Ririn Purnamasari Matthew Wai-Poi. *Journal of Political Economy*, 124(2).

Alatas, V. et al. (2012). Targeting the poor: Evidence from a field experiment in Indonesia - 3ie Impact Evaluation Report 12. *3ie Impact Evaluation Report*, 102(12):1–40.

Andreoni, J. et al. (2018). Using Preference Estimates to Customize Incentives: An Application to Polio Vaccination Drives in Pakistan. *NBER Working Paper*, 22019:1–66.

Anjana, R. M. et al. (2014). Physical activity and inactivity patterns in India - results from the ICMR-INDIAB study (Phase-1) [ICMR-INDIAB-5]. *International Journal of Behavioral Nutrition and Physical Activity*, 11(1):1–11.

Anjana, R. M. et al. (2011). Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: Phase I results of the Indian Council of Medical Research–INdia DIABetes (ICMR–INDIAB) study. *Diabetologia*, 54(12):3022–3027.

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

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2):1179–1203.

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

Baicker, K., Cutler, D., and Song, Z. (2010). Workplace wellness programs can generate savings. *Health Affairs*, 29(2):1–8.

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

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650.

Bhansali, A. et al. (2015). Prevalence of and risk factors for hypertension in urban and rural India: The ICMR-INDIAB study. *Journal of Human Hypertension*, 29(3):204–209.

Bruhn, M. and Mckenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, 1(4):200–232.

Cardella, E. and Depew, B. (2018). Output restriction and the ratchet effect: Evidence from a real-effort work task. *Games and Economic Behavior*, 107:182–202.

Cattaneo, M. D., Jansson, M., and Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115:1449–1455.

DellaVigna, S. and Malmendier, U. (2006). Paying not to go to the gym. *American Economic Review*, 96(3):694–719.

Deshpande, M. and Li, Y. (2019). Who Is Screened Out? Application Costs and the Targeting of Disability Programs. *SSRN Electronic Journal*, 11(4):213–248.

Dizon-Ross, R. and Zucker, A. D. (2020). Targeting Incentive Contracts in Heterogeneous Populations. *AEA RCT Registry*.

Dowd, T. E. (2002). Psychological reactance in health education and promotion. *Health Education Journal*, 61(2):113–124.

Dube, A. (2020). Indians Are The Least Active And Second Most Sleep Deprived Country In The World, Claims Fitbit Study. *Mashable India*, pages 2–4.

Dubé, J.-P., Booth, C., and Misra, S. (2019). Personalized Pricing and Customer Welfare. *Working Paper*.

Dupas, P. and Miguel, E. (2017). Impacts and Determinants of Health Levels in Low-Income Countries. *Handbook of Field Experiments*, 2:3–93.

Finkelstein, A. and Notowidigdo, M. J. (2019). Take-Up and Targeting: Experimental Evidence from SNAP. *The Quarterly Journal of Economics*, 134(3):1505–1556.

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., and Dietz, W. (2009). Annual medical spending attributable to obesity: Payer-and service-specific estimates. *Health Affairs*, 28(5):822–831.

Foulds, H. J. et al. (2014). Exercise volume and intensity: A dose-response relationship with health benefits. *European Journal of Applied Physiology*, 114(8):1563–1571.

Fudenberg, D. et al. (2005). Behavior-Based Price Discrimination and Customer Recognition.

Gupta, R., Gaur, K., and S. Ram, C. V. (2019). Emerging trends in hypertension epidemiology in India. *Journal of Human Hypertension*, 33(8):575–587.

Gupta, R. and Ram, C. V. S. (2019). Hypertension epidemiology in India: emerging aspects. *Current opinion in cardiology*, 34(4):331–341.

Gupta, R. and Xavier, D. (2018). Hypertension: The most important non communicable disease risk factor in India. *Indian Heart Journal*, 70(4):565–572.

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

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

Jack, B. K. (2013). Private information and the allocation of land use subsidies in Malawi. *American Economic Journal: Applied Economics*, 5(3):113–135.

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

Kenkel, D. S. (2000). Prevention. In *Handbook of Health Economics*, volume 1, chapter 31, pages 1675–1720. Elsevier.

Kyu, H. H. et al. (2016). Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: Systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. *BMJ (Online)*, 354(3857):1–10.

Laffont, J.-J. and Tirole, J. (1988). The Dynamics of Incentive Contracts. *Econometrica*, 56(5):1153–1175.

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

Leslie, P. (2004). Price Discrimination in Broadway Theater. *The RAND Journal of Economics*, 35(3):520.

Levitt, S. D., List, J. A., Neckermann, S., and Nelson, D. (2016). Quantity discounts on a virtual good: The results of a massive pricing experiment at king digital entertainment. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7323–7328.

Little, M. et al. (2016). Factors associated with glucose tolerance, pre-diabetes, and type 2 diabetes in a rural community of south India: A cross-sectional study. *Diabetology and Metabolic Syndrome*, 8(1):1–11.

Loprinzi, P. D. (2015). Frequency of moderate-to-vigorous physical activity (MVPA) is a greater predictor of systemic inflammation than total weekly volume of MVPA: Implications for physical activity promotion. *Physiology and Behavior*, 141:46–50.

Mavis, B. E. and Stöffelmayr, B. E. (1994). Multidimensional Evaluation of Monetary Incentive Strategies for Weight Control. *The Psychological Record*, 44(2):239–252.

Milkman, K. L., Minson, J. A., and Volpp, K. G. (2014). of Temptation Bundling. *Management Science*, 60(2):283–299.

Mortimer, J. H. (2007). Price discrimination, copyright law, and technological innovation: Evidence from the introduction of DVDS. *Quarterly Journal of Economics*, 122(3):1307–1350.

Myers, J. (2008). The health benefits and economics of physical activity. *Current Sports Medicine Reports*, 7(6):314–316.

Rajeswari, R., Muniyandi, M., Balasubramanian, R., and Narayanan, P. (2005). Perceptions of tuberculosis patients about their physical, mental and social well-being: a field report from south India. *Social Science & Medicine*, 60(8):1845–1853.

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

Samitz, G., Egger, M., and Zwahlen, M. (2011). Domains of physical activity and all-cause mortality: Systematic review and dose-response meta-analysis of cohort studies. *International Journal of Epidemiology*, 40(5):1382–1400.

Schilbach, F. (2019). Alcohol and self-control: A field experiment in India. *American Economic Review*, 109(4):1290–1322.

Schwappach, D. L., Boluarte, T. A., and Suhrcke, M. (2007). The economics of primary prevention of cardiovascular disease - A systematic review of economic evaluations. *Cost Effectiveness and Resource Allocation*, 5:1–12.

Sinha, R., van den Heuvel, W. J. A., and Arokiasamy, P. (2013). Validity and Reliability of MOS Short Form Health Survey (SF-36) for Use in India. *Indian journal of community medicine : official publication of Indian Association of Preventive & Social Medicine*, 38(1):22–6.

Tandon, N. et al. (2018). The increasing burden of diabetes and variations among the states of India: the Global Burden of Disease Study 1990–2016. *The Lancet Global Health*, 6(12):e1352–e1362.

Tharkar, S., Devarajan, A., Kumpatla, S., and Viswanathan, V. (2010). The socioeconomics of diabetes from a developing country: A population based cost of illness study. *Diabetes Research and Clinical Practice*, 89(3):334–340.

Tripathy, J. P. et al. (2017). Prevalence and risk factors of diabetes in a large community-based study in North India: results from a STEPS survey in Punjab, India. *Diabetology and Metabolic Syndrome*, 9(1):1–8.

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

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

Weitzman, M. L. (1980). The "Ratchet Principle" and Performance Incentives. *The Bell Journal of Economics*, 11(1):302–308.

Whitehead, D. and Russell, G. (2004). How effective are health education programmes - Resistance, reactance, rationality and risk? Recommendations for effective practice. *International Journal of Nursing Studies*, 41(2):163–172.

Wolak, F. A. (2006). Residential Customer Response to Real-Time Pricing: The Anaheim Critical-Peak Pricing Experiment. *Center for the Study of Energy Markets*, pages 1–50.

World Health Organization (2013). Global action plan for the prevention and control of noncommunicable disease. Technical report, WHO Press, Geneva, Switzerland.

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

	Full Sample		Monitoring	10K Target	12K Target	14K Target	Tag	Choice	Choice Variations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Mean	SD	Mean	Mean	Mean	Mean	Mean	Mean	Mean
A. Demographics									
Age	49.20	8.84	49.01	49.04	49.28	48.89	49.20	49.65	48.98
Female	0.37	0.48	0.39	0.37	0.37	0.37	0.36	0.36	0.35
Married	0.92	0.27	0.93	0.92	0.91	0.92	0.93	0.91	0.92
Household size	3.72	1.50	3.68	3.83	3.78	3.75	3.67	3.65	3.56
Monthly income/capita (INR)	5517	6798	5207	5530	5902	5141	5434	5509	5370
Wealth index	0.04	0.48	0.03	0.06	0.06	0.04	0.05	0.02	0.01
Any secondary education	0.58	0.49	0.60	0.58	0.58	0.57	0.57	0.56	0.65
B. Health statistics									
Diagnosed diabetic	0.32	0.47	0.29	0.34	0.33	0.30	0.32	0.33	0.27
Dianosed hypertensive	0.30	0.46	0.34	0.32	0.29	0.27	0.29	0.32	0.37
Diastolic BP	91.83	11.91	92.72	92.17	91.27	91.03	91.91	92.34	92.88
Systolic BP	137.16	19.87	138.17	137.68	136.38	136.08	136.97	138.24	138.45
BMI	26.48	4.65	26.10	26.30	26.59	26.46	26.55	26.52	26.47
Waist circumference (cm)	94.79	10.34	93.60	94.58	94.82	94.70	94.74	95.46	94.55
Mental health index	-0.04	0.66	0.01	-0.08	-0.05	-0.01	-0.05	-0.04	0.00
C. Baseline Activity									
Avg. baseline steps	6270	3927	6053	6127		6487	6576	6263	6061
Test for joint orthogonality of covariates (vs. 12K Target)									
F-stat			0.52	0.92		1.06	0.60	1.41	0.96
P-value			0.93	0.54		0.39	0.86	0.13	0.50
Sample size									
Number of individuals	5,606		191	854	1320	869	1025	912	435
Percent of sample	100.0		3.4	15.2	23.5	15.5	18.3	16.3	7.8

Notes: This table shows summary statistics for characteristics measured at Baseline for all participants who completed the Baseline survey, before they or surveyors were informed of their treatment group. Monthly income per capita is self reported. The wealth index is the simple average of the following standardized variables: number of scooters owned, number of cars owned, number of computers owned, number of smartphones owned, number of not-smart phones owned, number of rooms in house, a home-ownership dummy, whether the home has a private water connection, and whether the participant has a bank account. BMI is body mass index, and BP is blood pressure. The mental health index is a simple average of answers to seven mental health questions from RAND's 36-Item Short Form Survey standardized relative to the Monitoring group. Choice variations includes the Choice 10/15/20 and Choice 20/20/20 groups.
The F-statistic tests the joint orthogonality of all characteristics to treatment assignment relative to the 12k Fixed target group, holding constant the experiment phase. Each F-statistic is obtained by running a column-specific regression.

Table 2: Impacts of Incentives for Fixed Step Targets on Exercise

Dependent Variable:	Daily Steps (Contract Period)
	(1)
10K Target	637.8*
	[364.3]
12K Target	616.2*
	[360.5]
14K Target	859.3**
	[379.0]
Monitoring Mean	7125.39
# Individuals	2,762
Monitoring	164
10K Target	714
12K Target	1,141
14K Target	743
Controls	
Experiment Phase	Yes
Day-level Controls	Yes
Demographics (Lasso)	Yes
Predicted Activity (Lasso)	Yes

Notes: The dependent variable is daily steps measured using the intervention-period pedometer data. The columns show coefficient estimates from the regression shown in Equation 7. Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. Day-level controls include month-year, contract-week, and day-of-week fixed effects. Demographic controls are selected by double-LASSO from characteristics measured at baseline (Panels A and B of Appendix Table A.3), including all two-way interactions. Predicted activity controls are selected by double-LASSO from measures of predicted average daily pre-contract period walking (Panel C of Appendix Table A.3), and interacted with all demographic controls. Specifically, double-LASSO-selected controls are: predicted average daily pre-contract period steps, its interaction with age in years, and its sixth decile interacted with the number of days the participant reported exercising the week prior to Baseline. // The sample includes the fixed 10K, 12K, and 14K Target groups and the monitoring group. The omitted category in all columns is the Monitoring group. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Table 3: Personalizing Incentives Increases Exercise

Dependent Variable:	Daily Steps (Contract Period)
	(1)
Gameable Target Group	529.4***
	[199.7]
Choice	445.8**
	[208.7]
10K Target	21.29
	[193.3]
14K Target	227.4
	[219.0]
Monitoring	-593.3
	[361.3]
12K Target Mean	7,868
# Individuals	4,349
10K Target	714
12K Target	1,141
14K Target	743
Tag	843
Choice	744
Monitoring	164
Controls	
Experiment Phase	Yes
Day-level Controls	Yes
Demographics (Lasso)	Yes
Predicted Activity (Lasso)	Yes

Notes: The dependent variable is daily steps measured using the contract-period pedometer data. Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. Day-level controls include month-year, contract-week, and day-of-week fixed effects. Demographic controls are selected by double-LASSO from characteristics measured at baseline (Panels A and B of Appendix Table A.3, and Panel C in Column 1), including all two-way interactions. Activity controls are selected by double-LASSO from measures of average daily pre-contract period walking (Panel D of Appendix Table A.3), and interactions with all demographic controls.
The sample includes the monitoring, Tag, Choice, and Fixed groups. The omitted category in all columns is the Fixed 12K Target group. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Table 4: Personalization Achieves Upsides of 10K and 14K Targets Without Downsides

Quantile Treatment Effects

Dependent Var:	Daily Steps			Mean Daily Steps		
Percentile:	0.25	0.5	0.75	0.25	0.5	0.75
	(1)	(2)	(3)	(4)	(5)	(6)
Tag	311.1	404.1	597.5**	533.5*	318.7	593.4**
	[252.5]	[305.8]	[241.4]	[272.5]	[310.6]	[290.9]
Choice	594.3**	633.2**	-184.3	789.1***	493.7	141.4
	[276.4]	[313.9]	[220.4]	[291.8]	[303.6]	[291.0]
10K Target	512.7*	472.6	-939.9***	348.1	208.6	-595.4**
	[275.2]	[293.3]	[174.8]	[262.4]	[292.5]	[243.4]
14K Target	-178.7	-346.8	878.3***	-140.1	-201.8	456.5
	[277.5]	[308.8]	[319.8]	[262.7]	[322.3]	[404.3]
12K Target Quantile	3289	7732	12380	4515	7675	11537
# Individuals	4,152	4,152	4,152	4,152	4,152	4,152
# 10K Target	709	709	709	709	709	709
# 12K Target	1129	1129	1129	1129	1129	1129
# 14K Target	738	738	738	738	738	738
Exp. Phase Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Day-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
BL Walking Controls	No	No	No	No	No	No

Notes: This table shows quantile treatment effects on exercise during the Intervention Period. The dependent variables are daily steps daily steps across all participants in Columns 1-3, and average daily steps for each participant (Columns 3-6), measured using the intervention-period pedometer data. Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. Day-level controls include month-year, contract-week, and day-of-week fixed effects. Demographic controls are selected by double-LASSO. The sample includes individuals who were assigned their chosen contract only. All coefficients are interpretable relative to the 12K Target group, which is the omitted category. The mean in the 12K Target group is shown below the coefficients. Standard errors, in brackets, are clustered at the individual level in Columns 1-3. Significance levels: * 10%, ** 5%, *** 1%.

Table 5: Impact of Exogenous Tag Assignment

Dependent Variable:	Daily Steps (Contract Period)
	(1)
Low walkers × 10K Target	113.9
	[434.4]
High walkers × 14K Target	781.0
	[493.6]
High walkers × 10K Target	-61.13
	[443.1]
Low walkers × 14K Target	132.5
	[482.6]
Low walkers	-1832.2***
	[290.0]
High walkers	2768.7***
	[295.9]
Medium Walkers in 12K Target Mean	7320.62
# Individuals	2,578
10K Target	710
12K Target	1132
14K Target	736
Controls	
Experiment Phase	Yes
Day-level Controls	Yes
Demographics (Lasso)	Yes
Activity (Lasso)	No
Exogenous Tag vs. 12K:	
Coef	359.2
Standard Error	279.8
P-value	.199

Notes: This table shows the hypothetical impact of being assigned to the step target chosen by our Tag algorithm relative to being assigned to the 12K Target. This coefficient (473.9) is equal to the coefficient on Low walkers × 10K Target times the fraction of the sample who are Low walkers, plus the coefficient on High walkers × 14K Target times the fraction of the sample who are High walkers. The dependent variable is daily steps measured using the intervention-period pedometer data. The regression interacts Tag target assignment (i.e. dummies for having average baseline activity between 5,500 and 7,500 steps and for having average baseline activity above 7,500 steps, which were the cutoffs for the 10K and 14K Targets in the tag algorithm) with the Fixed step target treatment groups.
Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. Day-level controls include month-year, contract-week, and day-of-week fixed effects. Demographic controls are selected by double-LASSO. Baseline walking is endogenous to treatment assignment in the Tag group, and are therefore excluded.
The sample includes the fixed 10K, 12K, and 14K Target groups only. The omitted category are those who walked between 5,500 and 7,500 steps (i.e. Tag Target is 12K) and were assigned to the 12K Target group. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

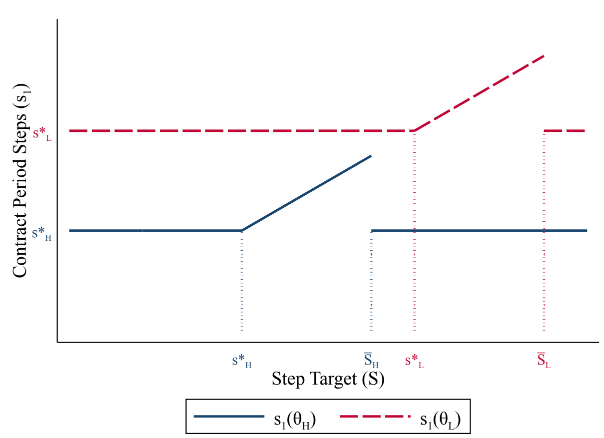
Table 6: Impact of Endogenous Tag Response on Exercise

Dependent Variable:	Daily Steps (Contract Period)
	(1)
Tag	197.2
	[222.6]
Exogenous Tag Mean	7579.44
# Individuals	1,854
10K Target	269
12K Target	273
14K Target	313
Tag	999
Controls	
Experiment Phase	Yes
Day-level Controls	Yes
Demographics (Lasso)	Yes
Activity (Lasso)	No

Notes: This table shows the impact of the endogenous baseline walking response in the Tag group on walking during the contract period. The dependent variable is daily steps measured using the contract-period pedometer data. The omitted group is the Exogenous Tag group: that is, participants in the Fixed target groups who are randomly assigned the step target that the Tag algorithm would have assigned them.

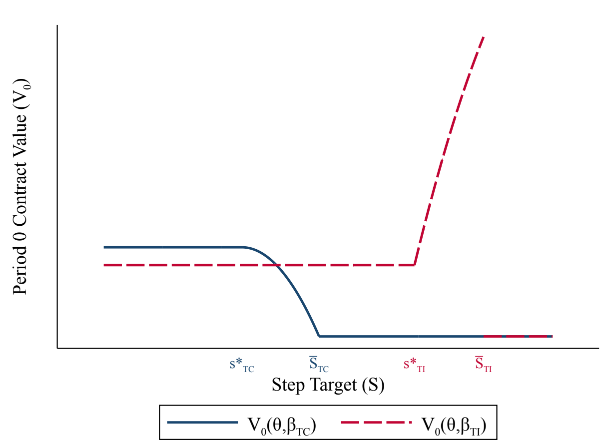
Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. Day-level controls include month-year, contract-week, and day-of-week fixed effects. Demographic controls are selected by double-LASSO. The sample includes the Tag group and individuals in the Fixed step target groups who were assigned the step target that our Tag algorithm would have assigned them based on their baseline activity. To correct for selection due to the larger size of the 12K Target group among the Fixed target groups (and therefore the overrepresentation of individuals whose baseline walking is between 5,500 and 7,500 steps in the Exogenous Tag group), observations are weighted by the inverse probability of treatment (IPTW). Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Figure 1: Walking in the Contract Period for High- and Low-Cost Types



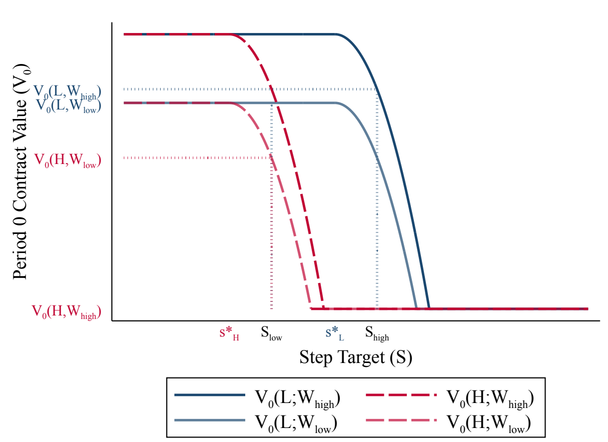
Notes: This figure shows how much a hypothetical high- and low-walking cost type participant would walk in the contract period for different contracts all paying the same incentive W for compliance with the step target. Contract period steps, s_1, are plotted on the y-axis against step targets, S, on the x-axis. The solid blue line shows walking for a high-walking cost type H, and the dashed red line shows walking for a low-walking cost type L.

Figure 2: Contract Valuations in the Pre-Contract Period for Time-consistent and Time-inconsistent Types



Notes: This figure plots contract valuations from the period-0 perspective (y-axis) for different step targets (x-axis) holding the incentive level constant. The solid blue line shows the valuations of a hypothetical time-consistent agent, and the dashed red line shows the valuations of a hypothetical time-inconsistent agent. The time-inconsistent agent would prefer her future self walk $s^+_{TI} > s^*_{TI}$ and therefore places greater value on contracts that serve as commitment devices to bring her future walking closer to s^+_{TI}. The time-consistent agent has no motive to modify her future walking behavior, and places no special value on contracts that increase her future walking.

Figure 3: Valuations for Two Contracts in a Menu that Separates High and Low Types



Notes: This figure plots contract valuations from the period-0 perspective (y-axis) over different step targets (x-axis) for two incentive levels, a "high" incentive level W_{high} and a "low" incentive level W_{low}. The solid blue lines show the valuations of a hypothetical low-walking-cost type agent L, and the dashed red lines show the valuations of a hypothetical high-walking-cost type agent.
We show an example of a separating contract menu. The "high" contract pays W_{high} for the step target S_{high}, and the "low" contract pays W_{low} for the step target S_{low}. The low-walking-cost type prefers the "high" contract

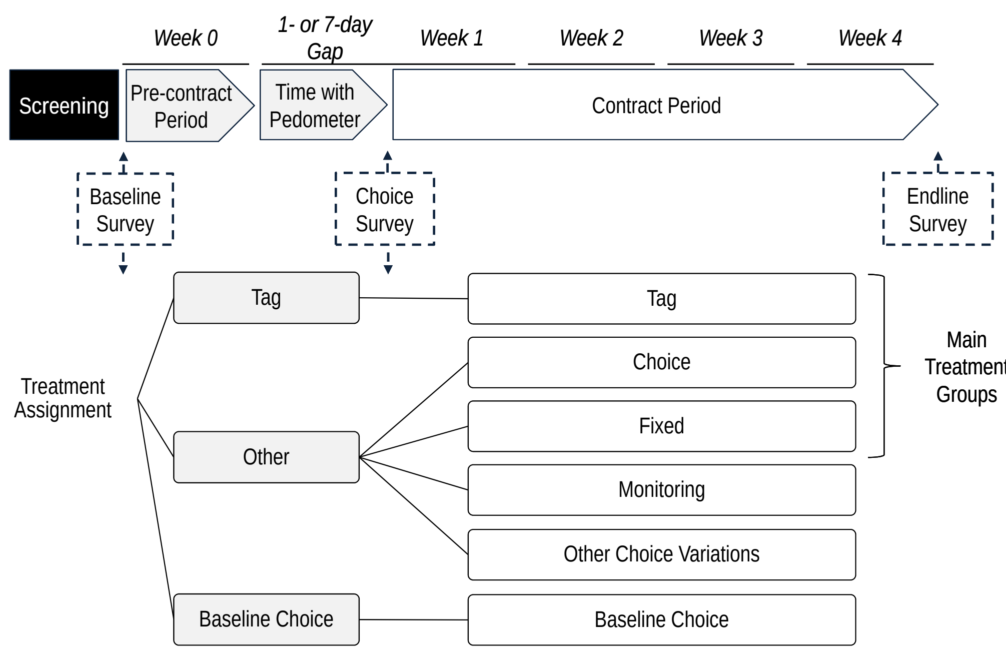
$$V_0(L; (W_{high}, S_{high}) > V_0(L; W_{low}, S_{low}),$$

and the high-walking-cost type prefers the "low" contract

$$V_0(H; Wlow, Slow) > V_0(H; W_{high}, S_{high}).$$

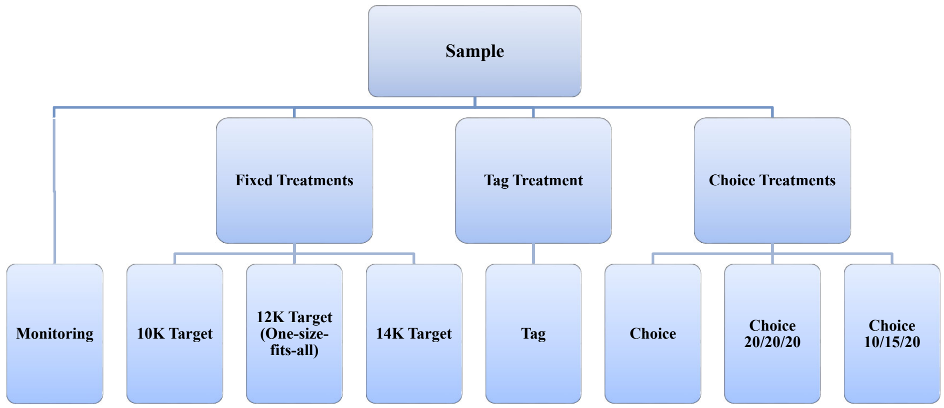
In addition, while the "high" contract is effective for the low-walking-cost type, the "low" contract is effective for the high-walking-cost type. The menu would thus separate the types shown and and be more effective than offering either contract alone.
While both agents represented in the figure are time-consistent (as can be seen from the weakly decreasing valuation curves), the contract menu would also separate time-inconsistent agents with similar regions of step target effectiveness for the two incentive levels.

Figure 4: Experimental Timeline for Sample Participant



Notes: This figure shows an experimental timeline for a participant. Following screening at public camps, we visited interested and eligible individuals to conduct a Baseline survey. After the Baseline, we launched the pre-contract period, during which we measured daily walking with a pedometer. We then visited participants to conduct the Choice survey and launch a four-week contract period. We introduced variation into the timing of contract choice by cross-randomizing the length of the delay between the pre-contract period and the Choice survey (13 days vs. one day). The contract period was exactly four weeks for all participants. After the contract period, we visited participants for an Endline survey and to collect the pedometers. The precise timing of the visits were scheduled according to the participants' availability.
Different treatments were assigned at different times. The Tag and Baseline Choice treatments were assigned at the end of the Baseline survey. All other treatments were assigned at the end of the Choice survey.

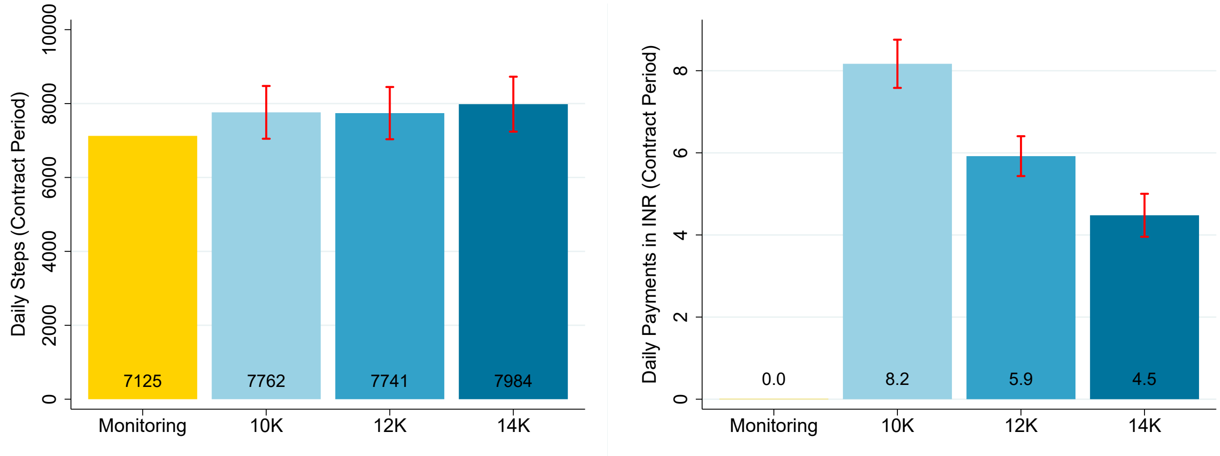
Figure 5: Experimental Design



	Monitoring	10K Target	12K Target (One-size-fits-all)	14K Target	Tag	Choice	Choice 20/20/20	Choice 10/15/20
Payment Amount (INR)								
10K Target	N/A	20	N/A	N/A	20	16	20	10
12K Target	N/A	N/A	20	N/A	20	18	20	15
14K Target	N/A	N/A	N/A	20	20	20	20	20
When Assigned	Choice	Choice	Choice	Choice	Baseline	Choice	Choice	Choice
Experiment Phase	1-3	1-3	1-3	1-3	1-3	1-3	3	1-3
Sample Size (Baseline)	191	854	1320	869	1025	912	398	37
Sample Size (Endline)	167	727	1158	754	872	763	317	27

Notes: This figure compares the different treatment groups. The main comparison group for the Tag and Choice groups is the 12K Target Group, and additional comparisons are made by re-weighting members of all three Fixed step target groups. The Monitoring group is a secondary comparison group, allowing us to measure the overall impact of incentives. The two main personalization treatment groups are Tag and Choice. The Choice 20/20/20 and Choice 10/15/20 groups allow us to collect incentive compatible preferences over the 20/20/20 and 10/15/20 Menus.
"Payment Amount" shows the incentive paid for compliance with each step target in each treatment. "When Assigned" indicates whether the treatment was made known to the participant and the surveyor at the end of the Baseline survey (before the pre-contract period) or at the end of the Choice survey (after the pre-contract period). The underlying randomization occurs before the Baseline but is known only to the researchers, not the field team, until the time of treatment assignment. "Experiment Phase" indicates in which of the three chronological phases of the randomized experiment the treatment group was implemented. "Sample Size (Baseline)" is the number of participants who completed the Baseline survey, and "Sample Size (Endline)" is the number who completed the Endline survey. Attrition is due to withdrawal, largely between the pre-contract period and the choice survey

Figure 6: Incentives for Fixed Step Targets Increase Average Walking and Payments

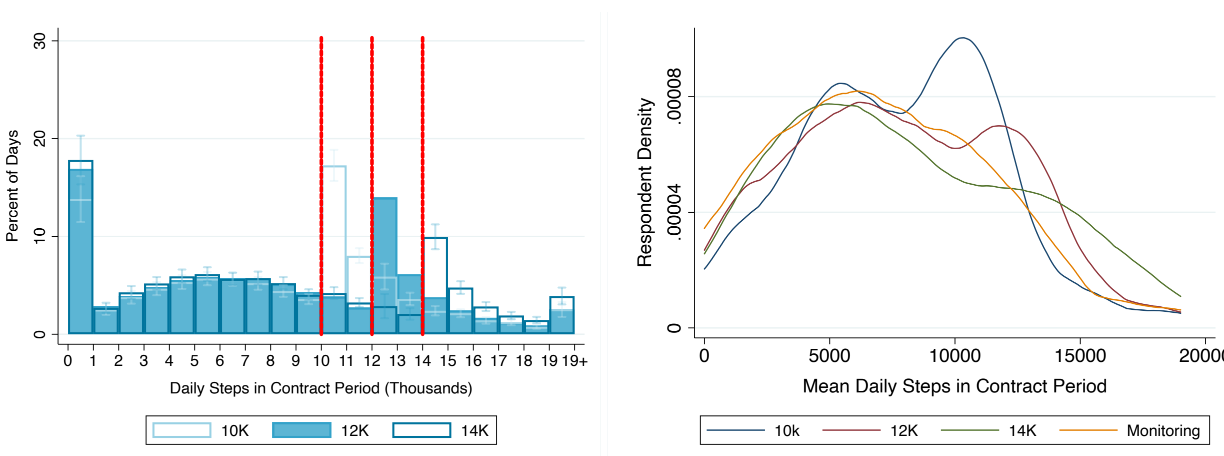


(a) Daily Steps

(b) Daily Payments (INR)

Notes: The figure displays the impact incentives for each of the randomly assigned Fixed step target groups (i.e., the 10K, 12K, and 14K Target groups) on walking outcomes during the contract period. The confidence interval bars represent tests of equality between the step target group and the Monitoring group with the same double-Lasso-selected controls as Table 2 at the 95% confidence level. Panel (a) shows the average daily steps walked during the contract period; Panel (b) shows average daily incentive payments delivered to participants in Indian Rupees (INR) during the contract period.

Figure 7: Different Fixed Step Targets Differently Influence Walking Distribution

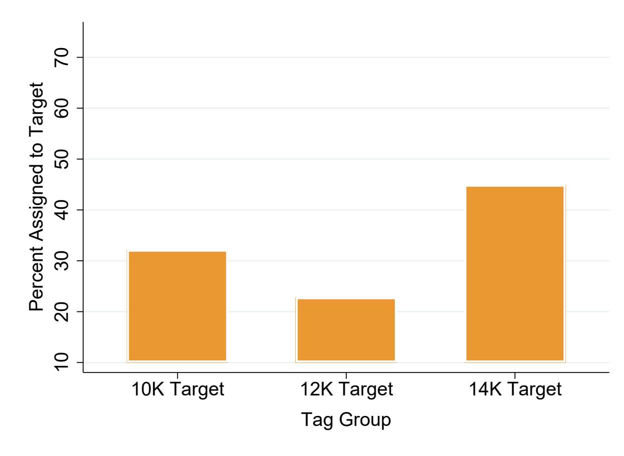


(a) Daily steps bunch just above step targets

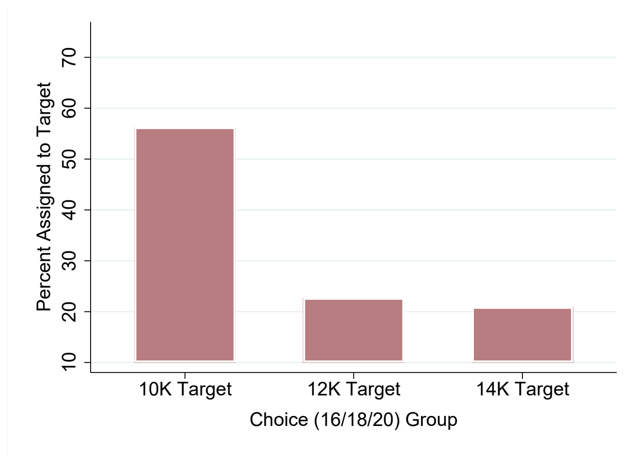
(b) Average daily steps bunch near step targets

Notes: The figure compares walking during the contract period under each of the three randomly assigned Fixed step target groups (i.e., the 10K, 12K, and 14K Target Groups). Panel (a) displays histograms of daily steps. The vertical red lines are drawn at 10,000, 12,000, and 14,000 steps, respectively. The confidence interval bars represent tests of equality for the probability of daily walking being in each step-bin between the step-target group and the 12K Target group with the same controls as Table 2 at the 95% confidence level (the 12K step target is the omitted group). Panel (b) displays kernel density plot of average daily steps across the four-week contract period for the Fixed target groups and the Monitoring group.

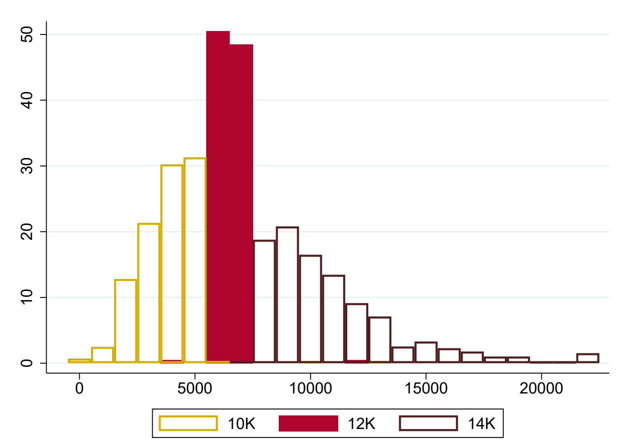
Figure 8: Personalization Sorts Participants between the Three Step Targets



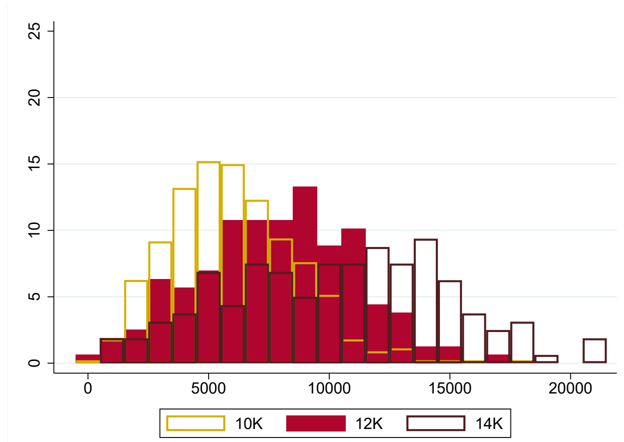
(a) Step Target Distribution in Tag



(b) Step Target Distribution in Choice



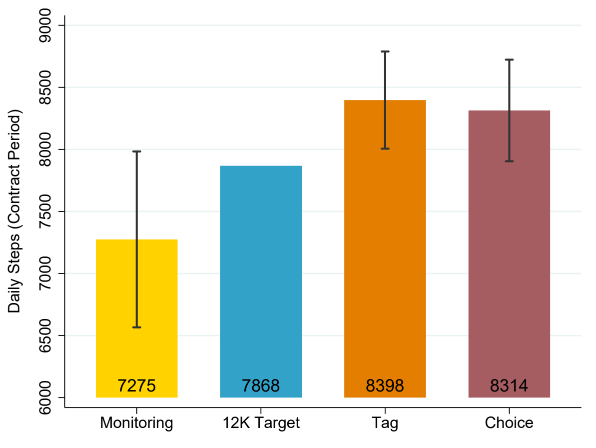
(c) Baseline Activity by Step Target: Tag



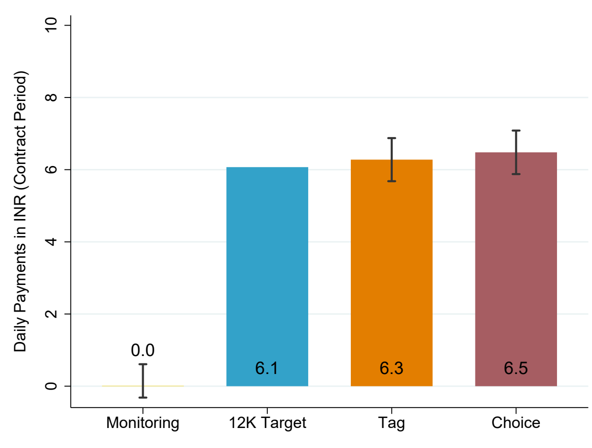
(d) Baseline Activity by Step Target: Choice

Notes: The figure displays the distribution of step targets, and the baseline activity levels within each step target, in each of our primary personalization groups, Tag and Choice. Panel (a) shows the percentage of Tag participants assigned to each of the three step targets; Panel (b) shows the percentage of Choice participants who choose each of the three targets from a menu where 14K pays 20 INR/day achieved, 12K pays 18 INR, and 10K pays 16 INR. Panel (c) shows overlapping histograms of baseline activity for participants assigned each of the three step targets in the Tag group, and Panel (d) shows overlapping histograms of baseline activity for participants who selected each of the three step targets in the Choice group. Note that he bars in Panels (c) and (d) represent the fraction of participants *among* those assigned a given step target in each treatment group who had baseline activities in each baseline step bin, rather than within the entirety of each treatment group overall.

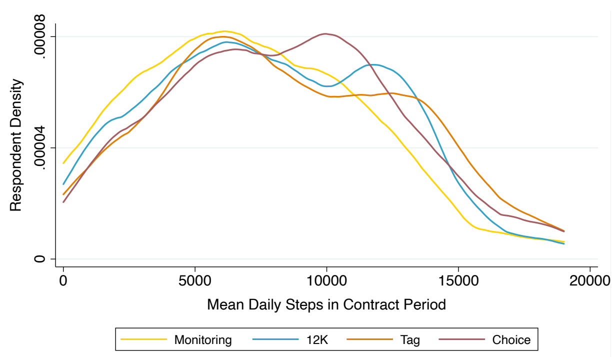
Figure 9: Personalization Increases Walking with Small Impacts on Incentive Payments



(a) Daily Steps



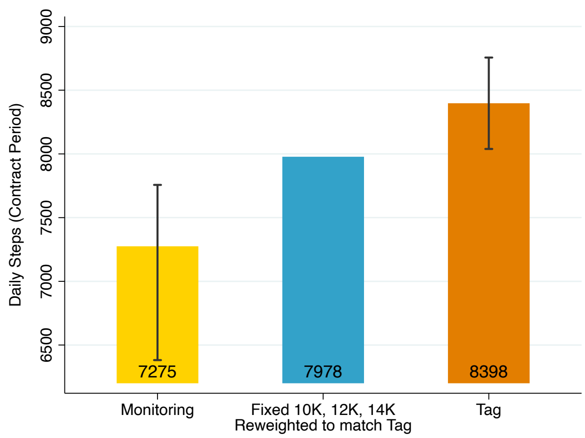
(b) Daily Payments



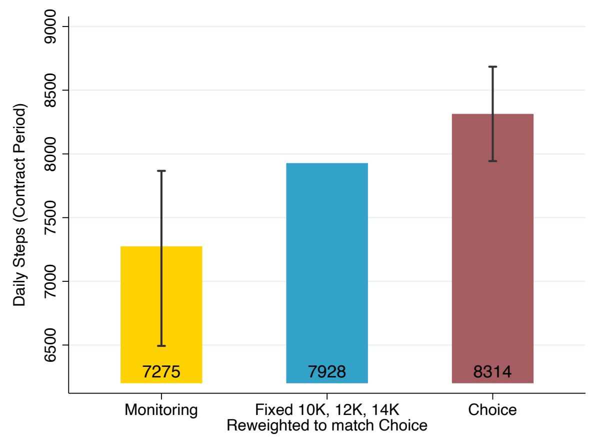
(c) Average Daily Steps

Notes: The figure displays the impact of each of our primary Personalization strategies, Tag and Choice, relative to our pre-specified "one-size-fits-all" treatment, the 12K Target group. Panel (a) shows average daily steps walked during the intervention period; Panel (b) shows the average daily incentive payments delivered to participants during the intervention period in INR. The confidence interval bars represent the test of equality between the personalized groups and the 12K Target group with the same controls as Column (1) of Table 3 at the 95% confidence level. Panel (c) plots kernel densities of average daily steps over the contract period, a way of showing the distribution of walking outcomes across participants (the kernel densities are not residualized to controls).

Figure 10: Personalization Effects Not Just Reshuffling Step Targets



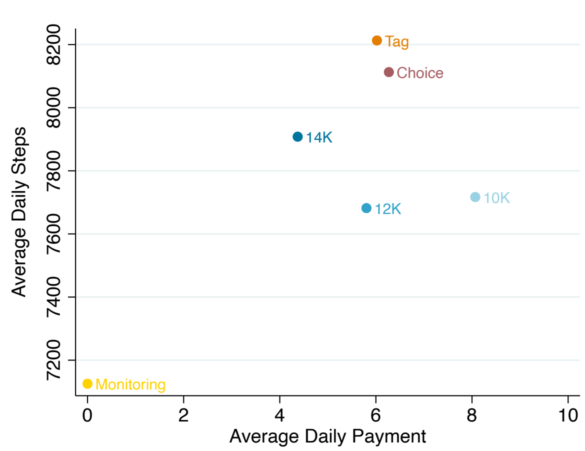
(a) Tag Treatment



(b) Choice Treatment

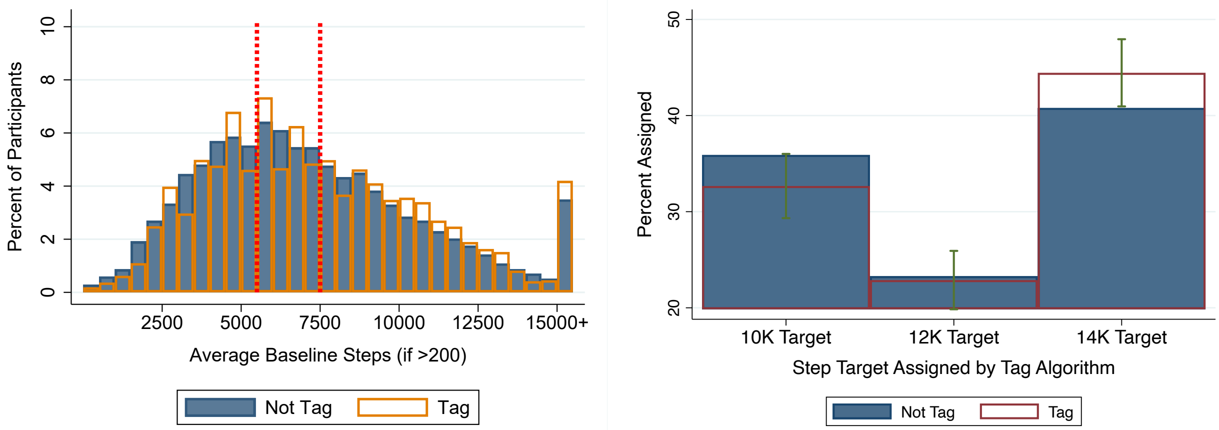
Notes: The figure displays the impact of each of our primary personalization strategies, Tag and Choice, relative to individuals in the three Fixed step targets re-weighted to represent the respective personalized group and the Monitoring group. The bars show average daily steps walked during the intervention period. Panel (a) shows the Tag group relative to the Fixed step targets reweighted in the proportion realized by the Tag group, and Panel (b) shows the Choice group relative to the Fixed step targets reweighted in the proportion realized by the Choice. The confidence interval bars represent the test of equality between the personalized groups and the reweighted Fixed Target groups with the same controls as Column (1) of Table 3 at the 95% confidence level.

Figure 11: Personalization Treatments are Cost-Effective



Notes: The figure plots the average daily steps taken among participants in each of the main treatments against the average daily payments made. Both average daily steps and average daily steps are residualized using controls selected by double-LASSO. The two personalization treatments improve average without substantially increasing incentive costs.

Figure 12: Baseline Steps Endogenously Respond to the Tag Algorithm

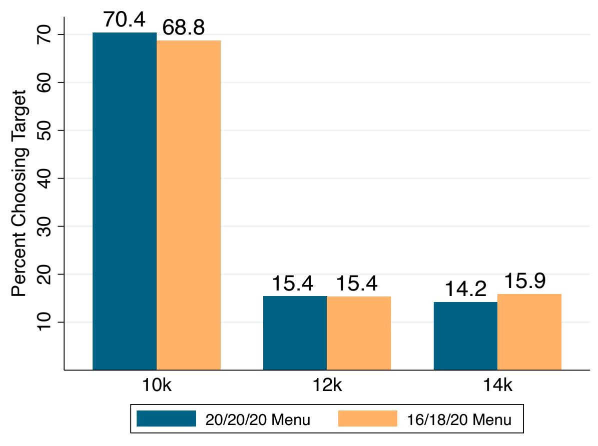


(a) Tag Group Increases Baseline Activity

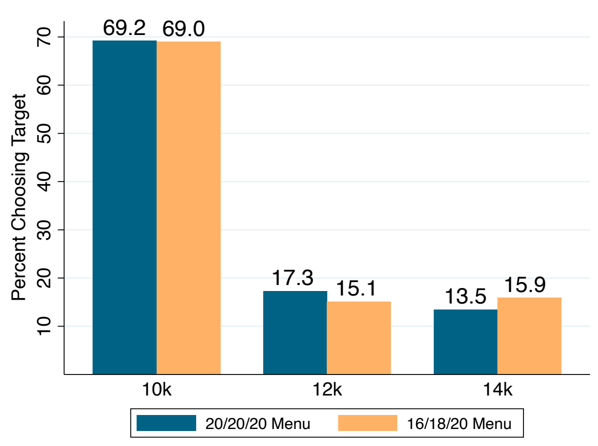
(b) Increased Activity Raises Step Targets

Notes: The figure displays how being assigned to the Tag group influences Baseline activity. Panel (a) shows the distribution of average baseline activity among the Tag group compared to all other groups (except Baseline Choice). Baseline activity is measured as the average daily steps taken during the pre-contract period, excluding days on which fewer than 200 steps were recorded. This is the measure we used to assign step targets in the Tag group. Panel (b) shows how step target assignment in the Tag group differs from how target assignment would have looked without the endogenous response to the Tag. The confidence interval bars represent tests of equality between the likelihood individuals are assigned to each step target with the same controls as Column (1) of Table 3 at the 95% confidence level. Overall, the endogenous response leads to more participants receiving higher step targets: the fraction receiving a 14K Target increases by 2.9 pp, while the fraction receiving the 10K and 12K Target decrease by -.2 pp and -2.6 pp, respectively.

Figure 13: Many Participants Choose Dominated Commitment Contracts



(a) Menu Choices



(b) First-Choice Menu Choices

Notes: The figure displays the share of participants who selected each step target from the menu of contracts where 10K, 12K, and 14K step targets all paid 20 INR for compliance and the menu where 10K step targets paid 16 INR, 12K step targets paid 18 INR, and 14K step targets paid 20 INR. The 12K and 14K step targets are weakly dominated in the 20/20/20 Menu. Choices are shown from the third phase of the experiment. In Panel (a), choices are shown regardless of which menu selection was made first. In Panel (b), only choices from the first menu selection are shown.

Appendices

This section contains all upfront appendix tables, figures, and sections. It also contains Appendix B. Any references to online appendices (tables, figures, or sections C – G) can be found at: https://faculty.chicagobooth.edu/-/media/faculty/rebecca-dizon-ross/research/customizingincentives$_o$nlineapp.pdf

A Appendix Tables and Figures

Appendix Table A.1: Pedometer Sharing and Step Misreporting: Summary Statistics

	Count		Share	
	Incentives	Monitoring	Incentives	Monitoring
	(1)	(2)	(3)	(4)
Shared Fitbit ever*	1	0	0.001	0.000
Suspended for cheating	100	N/A	0.015	N/A
Terminated for cheating	149	N/A	0.022	N/A
Total:	6,665	191	0.97	0.03

Notes: This table reports statistics on cheating and step misreporting among participants. Shared Fitbit ever indicates that when we visited participants unannounced, we found that their pedometer was being worn by another person. Suspended for cheating indicates that the participant was found to be over-reporting steps once, in which case their contract temporarily suspended for one week; terminated for cheating indicates that the participant was found to be over-reporting steps twice or was found to be sharing the pedometer, in which case their contract was terminated.
*Statistics for Fitbit sharing are calculated among the 1705 participants for whom we conducted surprise audits at their homes or workplaces, either randomly or because of suspicious step reporting behavior.

Appendix Table A.2: Pedometer Wearing and Step Misreporting: Treatment Effects

Variable type:	Pedometer Steps	Reported vs. Pedometer Steps		
Dependent variable:	Wore Pedometer	Over- or under-reported	Over-reported by at least 10%	Under-reported by at least 15%
	(1)	(2)	(3)	(4)
A. Pooled incentives				
Incentives	0.015	-0.040*	-0.026	-0.014
	[0.02]	[0.02]	[0.02]	[0.01]
B. Unpooled incentives				
Tag	0.0195	-0.0452*	-0.0224	-0.0228*
	[0.0223]	[0.0248]	[0.0233]	[0.0124]
Choice	0.0206	-0.0479*	-0.0405*	-0.00736
	[0.0223]	[0.0247]	[0.0231]	[0.0125]
Choice Variations	0.0146	-0.0340	-0.0196	-0.0144
	[0.0216]	[0.0239]	[0.0225]	[0.0120]
10K Target	0.0279	-0.0595**	-0.0443*	-0.0152
	[0.0225]	[0.0248]	[0.0231]	[0.0127]
12K Target	0.00855	-0.0368	-0.0246	-0.0122
	[0.0221]	[0.0243]	[0.0228]	[0.0123]
14K Target	-0.00131	-0.0263	-0.0120	-0.0143
	[0.0227]	[0.0251]	[0.0237]	[0.0125]
Monitoring mean	0.860	0.281	0.185	0.096
Exp. Phase Controls	Yes	Yes	Yes	Yes
Day-level Controls	No	No	No	No
P-value for 12K Target vs.				
Tag	0.31	0.51	0.85	0.05
Choice	0.29	0.40	0.18	0.43
Choice Variations	0.54	0.81	0.64	0.67
10K Target	.083	.077	.087	.634
14K Target	.4	.434	.315	.71
# Individuals	5,669	5,585	5,585	5,585
#Observations	154,581	128,533	128,533	128,533

Notes: This table reports how pedometer wearing and reporting behaviors differ by treatment group. Panel A shows coefficient estimates from regressions of the form

$$y_{itk} = \alpha + \beta_1 \times \text{Incentives Group}_i + \mu_k + \varepsilon_{it},$$

and Panel B shows coefficient estimates from regressions of the form

$$y_{itk} = \alpha+\beta_1\text{Tag}_i+\beta_2\text{Choice}_i+\beta_3\text{Choice Variations}_i+\beta_4\text{10K Target}_i+\beta_5\text{12K Target}_i+\beta_6\text{14K Target}_i+\mu_k+\varepsilon_{it},$$

including the full sample.

The dependent variable in column (1) is an indicator for whether daily steps exceed 200, measured using the intervention-period pedometer data. The dependent variable in column (2) is an indicator for whether daily steps were over-reported by more than 10% or under-reported by more than 15% on a given day, column (3) is an indicator for over-reporting, and column (4) is an indicator for under-reporting. Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. The sample includes all participants enrolled in the experiment. The omitted category in all columns is the Monitoring group. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.3: Variables used in Double-LASSO Selection Method

	Regressions with either Tag or Baseline Choice Group	Regressions with neither Tag nor Baseline Choice Group
	(1)	(2)
A. Self-Reported at Baseline		
Gender	X	X
Age	X	X
Diagnosed with diabetes	X	X
Diagnosed with hypertension	X	X
Excersized yesterday	X	X
Days exercised last week	X	X
Mental health index	X	X
Wealth index	X	X
B. Measured at Baseline		
Weight	X	X
Height	X	X
BMI	X	X
Systolic BP	X	X
Diastolic BP	X	X
Waist circumference	X	X
C. Estimated using Baseline Variables		
Predicted pre-contract period steps	X	
Predicted pre-contract period steps (deciles)	X	
D. Measured during Pre-contract Period		
Pre-contract period steps		X
Pre-contract period steps (deciles)		X
E. Other Variables		
Dummies for Missing	X	X
Square of Continuous Baseline Variables	X	X
All Two-Way Variable Interactions	X	X

Notes: This table lists the variables from which we selected covariates using the double-LASSO selection method of Belloni et al. (2014). The variables in Panel A were self-reported at the Baseline survey, or are indices of standardized self-reported variables. The variables in Panel B were directly measured at Baseline. The variables in Panel C are predictions from a cross-validated LASSO model of pre-contract period walking with ten cross-validation folds. The left-hand-side variable is average daily steps taken during the pre-contract period, and the right-hand side variables include all the variables from Panels A, B, and E. The estimation sample excludes the Tag and Baseline Choice groups, for whom pre-contract period steps are endogenous to treatment assignment. We estimate the model and calculate predictions using the cvlasso command in Stata developed by (Ahrens et al., 2018). The variables in Panel D are measured during the pre-contract period. Panel E shows that we included dummies for any missing variables and interactions of all variables in all double-LASSO estimation.

Appendix Table A.4: Robustness of Impacts of Incentives for Fixed Step Targets to Different Controls

Dependent Variable:	Daily Steps (Contract Period)		
	(1)	(2)	(3)
10K Target	622.4*	637.8*	660.5**
	[375.2]	[364.3]	[286.5]
12K Target	624.2*	616.2*	481.4*
	[372.0]	[360.5]	[282.5]
14K Target	814.8**	859.3**	673.0**
	[389.0]	[379.0]	[298.9]
Monitoring Mean	7125.39	7125.39	7125.39
# Individuals	2,762	2,762	2,762
Monitoring	164	164	164
10K Target	714	714	714
12K Target	1,141	1,141	1,141
14K Target	743	743	743
Controls			
Experiment Phase	Yes	Yes	Yes
Day-level Controls	No	Yes	Yes
Demographics (Lasso)	No	Yes	Yes
Predicted Activity (Lasso)	No	Yes	No
Activity (Lasso)	No	No	Yes

Notes: The dependent variable is daily steps measured using the intervention-period pedometer data. The columns show coefficient estimates from regressions based on Equations 7. Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. Day-level controls include month-year, contract-week, and day-of-week fixed effects. Demographic controls are selected by double-LASSO from characteristics measured at baseline (Panels A and B of Appendix Table A.3, and Panel C in column (2)), including all two-way interactions. Activity controls are selected by double-LASSO from measures of average daily pre-contract period walking (Panel D of Appendix Table A.3), and interactions with all demographic controls. Specifically, selected controls in column (2) are: predicted average daily pre-contract period steps, its interaction with age, and the baseline mental health index interacted with a dummy for whether weight is missing. In column (3), selected controls are: actual average daily pre-contract period steps, its interaction with height, its interaction with age, and the baseline mental health index interacted with a dummy for whether weight is missing. The sample includes the fixed 10K, 12K, and 14K Target groups and the monitoring group. The sample includes the fixed 10K, 12K, and 14K Target groups and the monitoring group. The omitted category in all columns is the monitoring group. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.5: Robustness of Impacts of Personalization on Exercise to Different Controls

Dependent Variable:	Daily Steps (Contract Period)		
	(1)	(2)	(3)
Gameable Target Group	545.3***	529.4***	
	[206.5]	[199.7]	
Choice	481.9**	445.8**	414.2**
	[215.4]	[208.7]	[178.1]
10K Target	-3.277	21.29	173.4
	[199.3]	[193.3]	[160.6]
14K Target	183.0	227.4	182.3
	[224.6]	[219.0]	[182.4]
Monitoring	-573.0	-593.3	-485.4*
	[370.8]	[361.3]	[282.7]
12K Target Mean	7,868	7,868	7,868
# Individuals	4,349	4,349	3,506
10K Target	714	714	714
12K Target	1,141	1,141	1,141
14K Target	743	743	743
Tag	843	843	
Choice	744	744	744
Monitoring	164	164	164
Controls			
Experiment Phase	Yes	Yes	Yes
Day-level Controls	No	Yes	Yes
Demographics (Lasso)	No	Yes	Yes
Predicted Activity (Lasso)	No	Yes	No

Notes: The dependent variable is daily steps measured using the intervention-period pedometer data. Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. Day-level controls include month-year, contract-week, and day-of-week fixed effects. Demographic controls are selected by double-LASSO from characteristics measured at baseline (Panels A and B of Appendix Table A.3), including all two-way interactions. Predicted activity controls are selected by double-LASSO from measures of predicted average daily pre-contract period walking (Panel C of Appendix Table A.3), and interacted with all demographic controls. Activity controls are selected by double-LASSO from measures of average daily pre-contract period walking (Panel D of Appendix Table A.3), and interacted with all demographic controls. The sample includes the fixed 10K, 12K, and 14K Target groups, the gameable target group, the choice group, and the monitoring group. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.6: Higher Step Targets Improve Walking More for those with Higher Baseline Activity Levels

Dependent Variable:	Daily Steps (Contract Period)			
	(1)	(2)	(3)	(4)
10K Target × Baseline Activity	-0.087	0.0041		
	[0.094]	[0.048]		
14K Target × Baseline Activity	0.15**	0.16***		
	[0.059]	[0.050]		
Step Target (1000s) × Baseline Activity			0.059**	0.039***
			[0.024]	[0.013]
Baseline Activity	0.62***	-0.013	-0.075	-0.42***
	[0.039]	[0.072]	[0.30]	[0.15]
12K Target Mean	7868	7868	7868	7868
# Individuals	2,578	2,578	2,578	2,578
10K Target	710	710	710	710
12K Target	1132	1132	1132	1132
14K Target	736	736	736	736
Controls				
Experiment Phase	Yes	Yes	Yes	Yes
Day-level Controls	No	Yes	No	Yes
Demographics (Lasso)	No	Yes	No	Yes
Predicted Activity (Lasso)	No	No	No	No
Activity (Lasso)	No	Yes	No	Yes

Notes: This table shows the interaction between average baseline activity levels and step target assignment in encouraging exercise. Columns (1)–(2) show coefficient estimates from regressions of the form:

$$y_{itk} = \alpha + \beta_1 \times \text{10K Target}_i \times y_i^{BL} + \beta_2 \times \text{14K Target}_i \times y_i^{BL} + \beta_3 \times y_i^{BL} + \textbf{Fixed Target}_i'\delta + \boldsymbol{X}_{it}'\lambda + \mu_k + \varepsilon_{it},$$

The omitted category in Columns (1)–(2) is the 12K Target group. The notation is defined in Equation 8. Columns (3)–(4) show coefficient estimates from regressions of the form:

$$y_{itk} = \alpha + \beta_1 \times \text{Step Target}_i \times y_i^{BL} + \beta_2 \times y_i^{BL} + \textbf{Fixed Target}_i'\delta + \boldsymbol{X}_{it}'\lambda + \mu_k + \varepsilon_{it},$$

Baseline activity levels are measured as the average daily steps taken during the first six days of the pre-contract period, ignoring days with fewer than 200 steps on which it is unlikely that the participant wore the pedometer. The dependent variable is daily steps measured using the intervention-period pedometer data. Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. Day-level controls include month-year, contract-week, and day-of-week fixed effects. Demographic controls are selected by double-LASSO from characteristics measured at baseline (Panels A and B of Appendix Table A.3, and Panel C in column (2)), including all two-way interactions. Activity controls are selected by double-LASSO from measures of average daily pre-contract period walking (Panel D of Appendix Table A.3), and interactions with all demographic controls. The sample includes the Fixed Target groups only. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.7: Without Manipulation, Tag Algorithm Better than Random Target Assignment

	Dependent Variable: Daily Steps (Contract Period)	
	(1)	(2)
Exogenous Tag	348.2**	347.3**
	[165.3]	[152.8]
Not Exogenous Tag: Mean	7849.61	7849.61
# Individuals	2,578	2,578
10K Target	710	710
12K Target	1132	1132
14K Target	736	736
Controls		
Tag Target Assignment	Yes	Yes
Experiment Phase	Yes	Yes
Day-level Controls	Yes	Yes
Demographics (Lasso)	Yes	Yes
Activity (Lasso)	No	Yes

Notes: This table shows the impact of randomly being assigned to the step target chosen by our Tag algorithm relative to being assigned to any other step target. The dependent variable is daily steps measured using the intervention-period pedometer data. All specifications control for Tag algorithm assignment (i.e. we control for dummies for having average baseline activity between 5,500 and 7,500 steps and for having average baseline activity above 7,500 steps, which were the cutoffs for the 10K and 14K Targets in the tag algorithm.) Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. Day-level controls include month-year, contract-week, and day-of-week fixed effects. Demographic controls are selected by double-LASSO from characteristics measured at baseline (Panels A and B of Appendix Table A.3, and Panel C in column (1)), including all two-way interactions. Activity controls are selected by double-LASSO from measures of average daily pre-contract period walking (Panel D of Appendix Table A.3), and interactions with all demographic controls. The sample includes the Fixed 10K, 12K, and 14K Target groups only. The omitted category are those who were not assigned the step target that the the Tag algorithm would have assigned them. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.8: Self-Reported Changes to Pre-Contract Period Walking

	Not Tag Group			Tag Group		
	Walked more	Walked less	Walked same	Walked more	Walked less	Walked same
	(1)	(2)	(3)	(4)	(5)	(6)
A. Walking in Pre-Contract Period Compared to Normal						
Self Report	0.21	0.04	0.75	0.22	0.03	0.75
B. Reason Given (Fraction of Column)						
Wanted higher/lower/correct step target	0.00	0.00	0.00	0.10	0.05	0.04
Schedule more/less free	0.15	0.74	0.03	0.13	0.73	0.05
Weather		0.02	0.00		0.05	0.00
Felt motivated	0.86			0.82		
Health issue		0.24	0.00		0.14	0.01
To improve health	0.04		0.00	0.02		0.00
Following instructions			0.23			0.29
Other reason	0.01	0.02	0.02	0.01	0.05	0.04

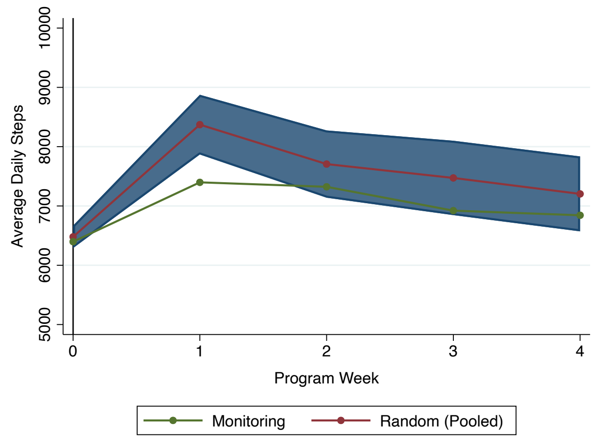
Notes: Panel A shows the fraction of participants who reported walking more, less, or the same as usual during the pre-contract period in all but the Tag group (columns 1-3) and in the Tag group (columns 4-6). Panel B shows the fraction of participants within each group who reported walking more, less, or the same as usual during the pre-contract period giving each reason for doing so. Participants could give more than one reason.

Appendix Table A.9: Lower Incentive Payments Reduce Excercise in the Contract Period

Dependent Variable:	Daily Steps (Contract Period)			Step-target Compliance (Contract Period)		
	(1)	(2)	(3)	(4)	(5)	(6)
Payment Amount (INR)	78.08	65.15	54.21	0.00206	0.00150	0.000480
	[91.97]	[72.18]	[47.08]	[0.00827]	[0.00666]	[0.00489]
Random Group Mean	7886.35	7886.35	7892.86	0.37	0.37	0.37
# Individuals	2,849	2,849	2,835	2,849	2,849	2,835
# Choice 10/15/20	23	23	23	23	23	23
# Choice 16/18/20	988	988	982	988	988	982
# 10K Group	709	709	707	709	709	707
# 12K Group	1129	1129	1123	1129	1129	1123
Menu Choice X Target Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Day-level Controls	No	Yes	Yes	No	Yes	Yes
BL Walking Controls	No	No	Yes	No	No	Yes

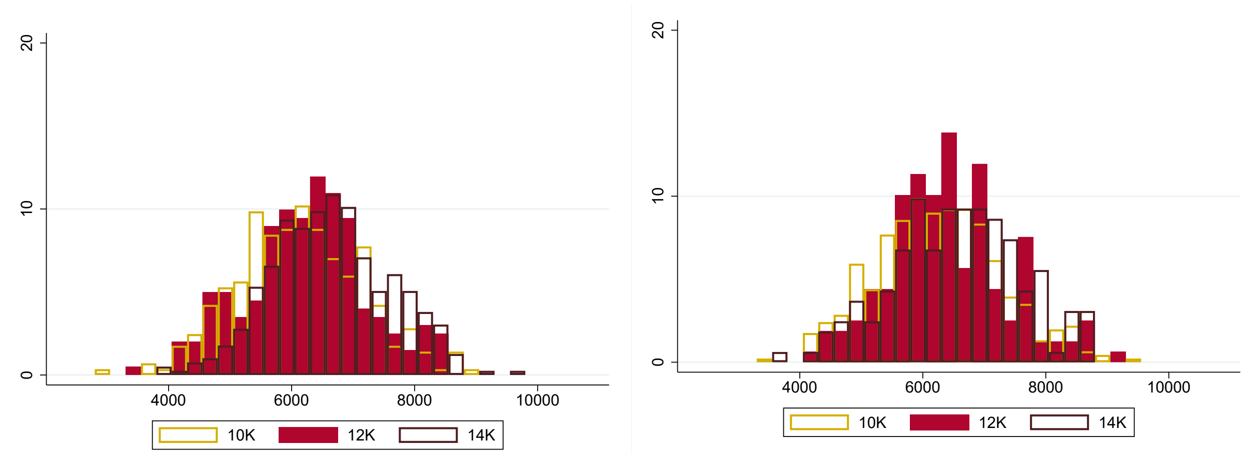
Notes: This table shows the impact of being assigned a higher incentive level on exercise during the contract period, all else equal. We exploit incentive level variation between those who were assigned to the Random Group (where all contracts paid 20 INR) and those who were assigned to the Choice group (wo received their selection from the 16/18/20 Menu) or Choice 10/15/10 group, controlling for a full set of interactions between contract menu choices and step target assignment. The dependent variable for Columns 1-3 is daily steps, and for Columns 4-6 is a dummy for achieving the assigned step target, measured using the intervention-period pedometer data. Experiment phase controls include dummies for changes to the experimental design over time as treatment groups were added. Demographic controls include gender and second order polynomials of age, weight, and height, and a dummy for the randomly assigned duration between the Baseline and Choice surveys. Day-level controls include month-year, contract-week, and day-of-week fixed effects. Walking controls include quintiles of average daily walking during the pre-contract period. The sample includes individuals in the Random, Choice and Choice 20/20/20 groups who were assigned the 10K or 12K step target. Observations are weighted by the inverse probability of treatment (IPTW). Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Figure A.1: Treatment effects persist throughout the intervention



Notes: The figure compares average daily steps taken in the first six days of the pre-contract period (Week 0) and during each week of the intervention period (Weeks 1-4) in the Monitoring group and in the Fixed 10K, 12K, and 14K Target groups (pooled). The confidence intervals represent tests of equality between the pooled fixed step target groups and Monitoring groups for a single week of data with the same controls as Table 2 (the Monitoring group is the ommitted group).

Appendix Figure A.2: Personalization Sorts Participants according to Predicted Baseline Activity

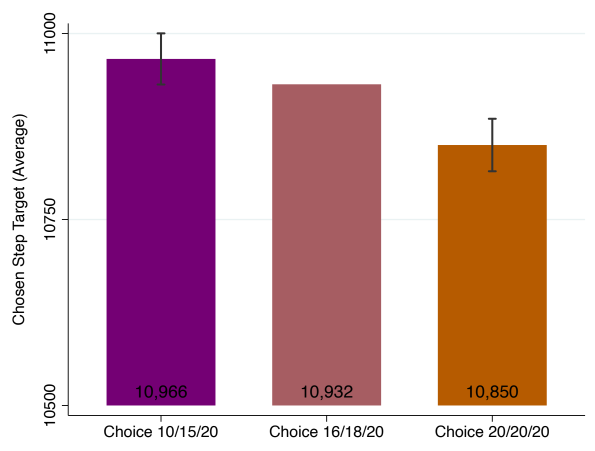


(a) Predicted Baseline Activity by Target in Tag

(b) Predicted Baseline Activity by Target in Choice

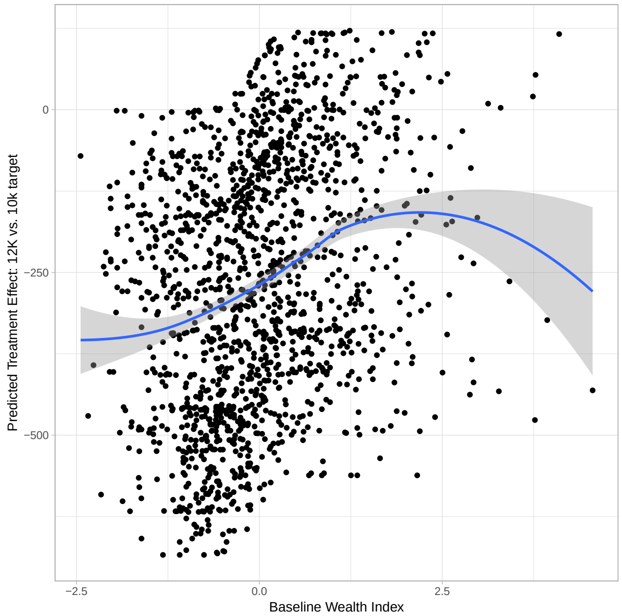
Notes: The figure displays predicted baseline activity levels within each step target in each of our primary personalization groups, Tag and Choice. Panel (a) shows overlapping histograms of predicted baseline activity for participants assigned each of the three step targets in the Tag group, and Panel (b) shows overlapping histograms of predicted baseline activity for participants who selected each of the three step targets in the Choice group. The bars in Panels (c) and (d) represent the fraction of participants *within* those assigned a given step target who had baseline activities in each baseline step bin, rather than within the entirety of the treatment group. Baseline activity is predicted using a cross-validated LASSO with ten folds to select predictors among hundreds of baseline covariates. The cross-validated LASSO is fit from baseline and pre-contract period data from all treatment groups except for the Tag and Baseline Choice. groups.

Appendix Figure A.3: Participant Choices Shift Away from Lower Step Targets when they Pay Less

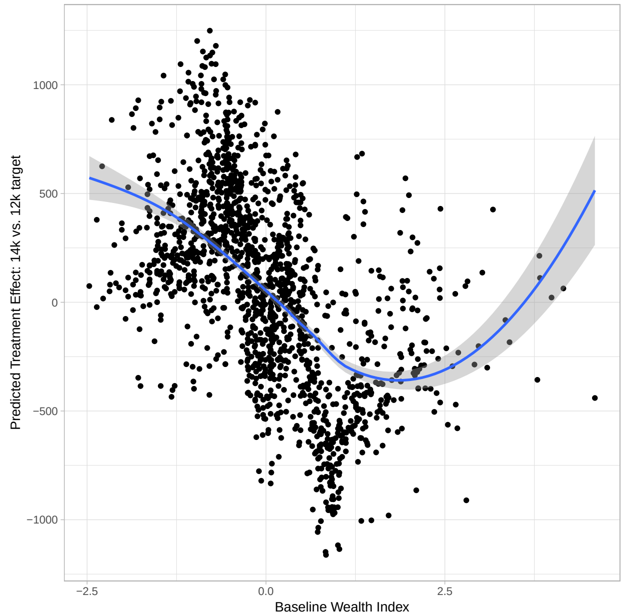


Notes: The figure displays the average step target chosen (in number of steps) in incentive-compatible menus of contracts. The left bar shows the average steps chosen for the step target in the Choice 10/15/20 Menu, where the 10K, 12K, and 14K targets respectively paid 10,15, and 20 INR for compliance, respectively. The middle bar represents the choices in the Choice 16/18/20 menu where the targets paid 16, 18, and 20 INR for compliance, respectively. Finally, the bar on the right represents the step target choices from the Choice 20/20/20 menu where all targets paid 20 INR for compliance. The confidence interval bars represent a test of equality between the average step target chosen in the respective menu and the 16/18/20 menu at a 95% confidence level. Choices shown are from the third phase of the experiment, which is the only phase where participants made incentive compatible choices over all three menus.

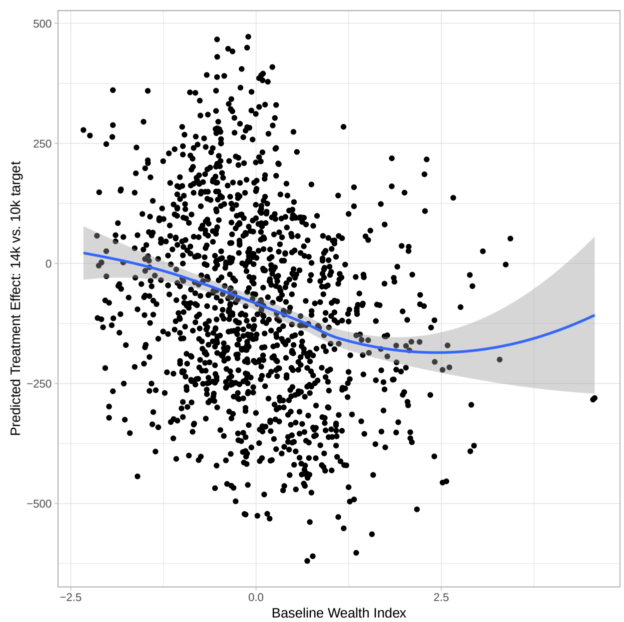
Appendix Figure A.4: Heterogenous Treatment Effects Vary with Baseline Wealth



(a) 12K vs. 10K Target



(b) 14K vs. 12K Target



(c) 14K vs. 10K Target

Notes: The figures plot heterogeneous treatment effects pairwise between the three Fixed Target groups. Heterogeneous treatment effects are estimated using causal forests. The causal forest is fit using all variables in Panels A and B Appendix Table A.3, as well as average pre-contract period steps. The figures show that the 14K step target tends to be better for those with lower baseline wealth levels, while the 12K Target is worse than the 10K Target for those with lower baseline wealth levels. This suggests that poorer individuals have a lower willingness to accept the highest step target.

B Tag and Choice Design

In this section, we discuss how we designed our two personalization interventions. Our personalization interventions both sorted participants into three step targets: 10,000, 12,000, or 14,000 steps per day. Because the primary goal of personalization is to increase contract-period walking, we aimed to choose three step targets that would each be most effective (i.e., have the largest treatment effects on contract-period steps) for some portion of our sample.

In order to choose the step targets, we used the results from a previous evaluation of a similar incentive program for walking which was implemented in the same setting (Aggarwal et al., 2020).[60] The previous evaluation involved six days of pre-contract period walking followed by a 12-week contract period where participants were paid 20 INR for achieving a daily 10,000 step target. The details of the present study's setting, recruitment, pre-contract period, and contract period closely follow Aggarwal et al. (2020),[61] with the primary difference being that we shortened the intervention from 12 to four weeks, and that we offered multiple step targets instead of only a 10,000 step target.

Our process for selecting the three step targets also led to our Tag algorithm. First, we estimated each persons' "most effective" step target as a function of their average baseline activity (i.e. the average of their daily pre-contract period steps). To do so, we assumed that, for the payment amount of 20 INR, each person's most effective step target was a fixed number of steps above their baseline activity. We used a linear regression to model the treatment effect of a 10,000 step target as a quadratic function of baseline activity,[62] and found that treatment effect heterogeneity took an inverted U-shape. The peak of the inverted U, or the maximum treatment effect a daily 10,000 step target, occurred at a baseline activity of 4,500 daily steps. Therefore, we estimated that each individual's most effective step target was 5,500 steps above their average baseline activity. Second, with our mapping from baseline activity to most-effective targets in hand, we chose three round-number step targets such that each would be relatively more effective for approximately one third of our sample.[63] Our mapping from baseline activity level bins to our estimate of the most effective step target is shown in the short table below, and also forms the basis of our Tag algorithm.

[60] The authors of the present study are co-authors of the evaluation by Aggarwal et al. (2020).

[61] Both samples include people living with or at risk of lifestyle disease recruited through in public screenings in the city of Coimbatore. The samples had slightly different characteristics: while Aggarwal et al. (2020) recruited only diabetics and individuals with elevated blood sugar, this study also includes hypertensives and individuals with elevated blood pressure.

[62] We did not have power to non-parametrically model heterogeneous treatment effects.

[63] We assumed baseline activity in our sample would closely resemble Aggarwal et al. (2020).

Average Daily Steps Pre-contract Period	Most Effective Step Target (Estimate)
<5,500	10,000 steps
5,500-7,500	12,000 steps
>7,500	14,000 steps

In order to select the contract menu that we would offer the Choice participants, we conducted a small pilot study. Pilot participants were given a pedometer for six days, and then asked which contract they would prefer among menus with step targets of 10,000, 12,000, and 14,000 steps where lower step targets had lower payments. The 16/18/20 Menu appeared to strike a good balance between separating walking types and maintaining incentive payments that we had found to be effective in the previous evaluation.