Customizing Incentive Policies to Improve Cost-Effectiveness
Rebecca Dizon-Ross, University of Chicago Booth School of Business
Ariel Zucker, University of California Berkeley

Abstract
Policymakers are increasingly providing financial incentives to encourage beneficial behaviors (e.g., exercise, job-search). However, a policy design challenge is that policymakers only want to pay beneficiaries for “marginal” behavior (i.e., behavior they would not have engaged in absent the program), which is difficult when “inframarginal” behavior (i.e., what would have occurred absent the program) varies across beneficiaries. This project evaluates two methods for addressing this challenge by customizing incentives so that beneficiaries with higher inframarginal behavior must hit higher targets to earn incentives. The first method uses machine learning to predict inframarginal behavior. The second creates a menu of contracts and allows participants to choose. We evaluate how these methods affect the cost-effectiveness of an incentive program for exercise among diabetics. The findings should translate directly to policy: based on promising findings from our previous research, policymakers in Tamil Nadu, India are considering scaling up this program to their full population.

1. Overview
Policies providing financial incentives to encourage behavioral change are increasingly common: policymakers pay people to exercise, to study more, to maintain tree cover, and to search for jobs. These policies encourage behaviors that are under-provided by the market due to either externalities or “internalities” (e.g., present bias), and can be very effective in changing behavior.

Designing simple, cost-effective financial incentive contracts is difficult in heterogeneous populations. The challenge is that the policymaker only wants to pay a beneficiary for her “marginal” behavior (i.e., behavior she would not have engaged in absent the program), which is difficult when levels of “inframarginal” behavior (i.e., what would have occurred even absent the program) vary meaningfully across beneficiaries.

A promising strategy to improve “cost-effectiveness” (i.e., increase the level of induced behavior relative to the value of incentives paid out) is to customize the contracts based on individual levels of inframarginal behavior, or “types”. A policymaker who observed beneficiary types could customize incentives to each individual, requiring those with higher inframarginal levels of behavior to meet higher behavioral targets in order to receive incentives (analogous to the perfect price discrimination solution of a monopolist). However, types are often not observable. We propose two main approaches in that case. First, if the policymaker has observable proxies for types, she can customize the incentive contract based on those proxies. This is analogous to the “third degree price discrimination” solution of a monopolist; here we will refer to it as the Targeting solution. For example, she could use machine learning to predict levels of inframarginal behavior and assign targets based on predicted levels of behavior. However, while machine learning is becoming widespread for targeting programs (not just incentives), it may not be effective in settings where the set of predictive variables available to the policymaker is small. A second option that could work better in those cases is that the policymaker could offer a menu of differentiated incentive contracts and allow individuals to choose. The goal is to get agents with higher levels of counterfactual or inframarginal behavior to choose contracts with higher behavioral targets in exchange for higher payments. This solution is analogous to “second degree price discrimination”; here we refer to it as the incentive-compatible choice or IC choice method.

We are conducting a randomized field experiment to evaluate the efficacy of these methods of customizing incentives to improve cost-effectiveness. We will compare these methods against each other as well as against a non-customized approach. Since the right approach likely varies across settings, we will also test predictions about when each method will work better.
2. Setting and intervention

Diabetes and other lifestyle-related non-communicable diseases are exploding problems worldwide, and promoting exercise among diabetics is an important policy objective. Our context is an incentive program for walking among diabetics in Tamil Nadu, India. The program gives participants fitbits that measure their walking and pays them incentives for each day they meet a daily step target. In a previous experiment (Aggarwal et al., 2018), we show that this program is very effective at generating exercise: providing incentives of 20 INR or roughly 0.30 USD per day walked increases the percent of days that participants met a 10,000 step target by 20 percentage points, from 30% to 50%. This translates to an increase in daily walking of 1,300 steps or roughly 13 minutes of brisk walking. As a result of these promising impacts, the Government of Tamil Nadu, a state in India with roughly 10 million diabetics, is considering scaling up this program as a publicly-funded program covering their full population. However, they are concerned about program cost-effectiveness and do not want to “waste” incentive payments paying for inframarginal behavior. In particular, many individuals in the program would have reached the step target on many days even without payment, but are compensated for doing so. This project aims to reduce this “waste” by tailoring the incentive contract to agents’ inframarginal walking (i.e. counterfactual walking in the absence of the program), assigning higher exercise targets to those with higher inframarginal walking.

3. Experimental design

We will implement the incentive program in largely the same way as we did in our previous experiment (Aggarwal et al., 2018). Participants will receive pedometers that measure the number of steps they walk every day, and receive small incentives (20 INR or 0.30 USD), paid in the form of cell-phone minutes, if they meet a daily step target. We will recruit subjects through screenings in public spaces, like markets.

The goal of our experimental design is to evaluate strategies to assign higher daily step targets to participants who would walk more in the absence of the program (i.e., who have higher “baseline walking” or “inframarginal walking”). If baseline walking were observable, we would assign higher step targets to those with higher baseline walking. However, because baseline walking is not observable to the policymaker, we will assign participants to 3 treatment groups, which use different methods to customize the program for greater cost-effectiveness.

1. **Gameable targeting:** In the first treatment group, we will attempt to directly measure baseline walking by recording steps for one week before the program begins. We will then tailor the incentive program by assigning higher step targets to those who walk more at baseline. If there were no potential for gaming, this would be the most cost-effective tailoring method. However, the baseline walking measurement is gameable (i.e., individuals can adjust their baseline walking downwards to try to get assigned a lower step target), which may lower its efficacy. Our next two treatment groups attempt to mitigate gaming.

2. **Non-gameable targeting:** We will use machine learning and data from our previous experiment to predict baseline (counterfactual or inframarginal) walking at the individual level based on non-gameable baseline data (e.g., gender, age, BMI). We will then assign agents in this treatment group to incentive contracts (step targets) based on their individual-level predicted walking, assigning agents with higher predicted walking to higher step targets.

3. **IC choice:** We will offer a menu of contracts that vary the step target level and the incentive amount. The aim is to achieve a separating equilibrium wherein those who walk more at baseline

---

1 An alternative would be to offer lower or no payments to higher walkers. We view changing the step targets as a better solution here since there is no evidence of diminishing marginal health benefits for higher levels of walking – i.e., even those who walk a lot at baseline should generate social benefits by walking more.

2 Note that this statement assumes the only heterogeneity in incentive responses is due to heterogeneity in inframarginal behavior.
select into contracts with higher step targets; to achieve this, we would pay higher incentive amounts for contracts that have higher step targets. To develop this menu of contracts, we are currently measuring how demand for various contracts varies with baseline walking.

We benchmark these customization methods against the following two treatment groups:

4. **“One-size-fits-all” incentives:** In this group, we will randomly assign participants to the incentive contracts used in treatment groups 1-3. Comparing the other approaches to this group allows us to see how large the gains to customizing the incentives are.

5. **Monitoring (no incentives):** This group will receive pedometers but no incentives. It will play the role of a control group so that we can measure the overall efficacy of incentives.

**Overview of procedures:** We will begin with a baseline measurement period where we give participants pedometers, tell them to walk like they normally do, and measure their steps. At the beginning of this period, the Gameable targeting group will be told that their steps will affect their incentive contract assignment; other treatment groups will not be told this. At the end of the measurement period, respondents will be informed of their incentive contract. The incentives intervention will then run for 4 weeks during which we will measure walking with the pedometers.

**Analysis:** Our primary analysis will compare the levels of exercise (and the cost per exercise generated) in the intervention period across treatments to see which is most effective at generating exercise. We will also use heterogeneity analysis to see how the impacts of the methods vary across the distribution of counterfactual walking. In addition, we will assess how much people “game” their walking when assignments depend on walking by comparing baseline walking between Gameable targeting and the other groups. Finally, we will use heterogeneity analysis to test predictions for when the various methods might work better. For example, relative to Non-gameable targeting, IC choice should be more effective when beneficiaries have better information about their own levels of inframarginal behavior or in populations where the machine learning prediction of behavior is less accurate.

**4. Academic Contribution**

In addition to directly informing policy in India and beyond, the results of this project will make several academic contributions. While several papers have outlined the potential challenges associated with cost-effectively providing incentives in heterogeneous populations (e.g., Jack and Jayachandran, 2018), our experiment will be the first to propose and test methods to address these challenges, and to characterize the types of settings in which the various methods should work better. Our work also relates to a large literature examining how to adjust public program design to affect selection into the programs at the extensive margin (e.g., Jack 2013, Deshpande and Li 2017); in contrast, we study how to adjust program design at the intensive margin (conditional on enrollment).
References


