Online Supplement for:

Mechanism design for personalized policy:
A field experiment incentivizing exercise

Rebecca Dizon-Ross
University of Chicago

Ariel Zucker
UC Santa Cruz

Table of Contents

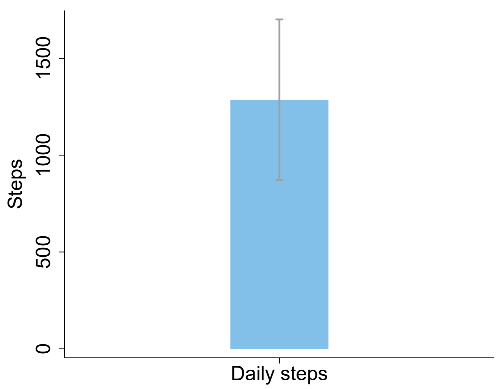
F Impacts of Exercise Incentives Measured in Aggarwal et al. (2024)

Figure F.1 summarizes some of the evidence from Aggarwal et al. (2024) showing the effectiveness of a very similar step target incentive program. That program offered incentives of 20 INR to participants for walking 10,000 steps per day, with subtreatments varying contract parameters such as the frequency of payment.

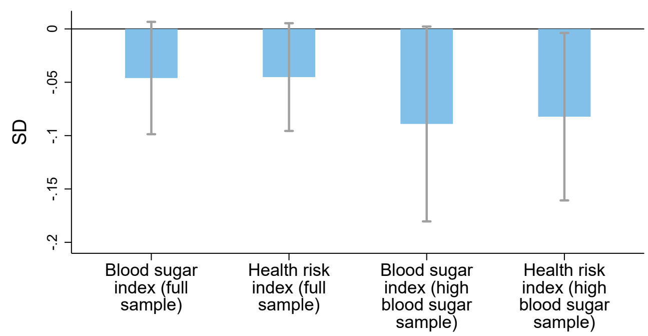
Panel (a) of Figure F.1 shows that the incentive program substantially increased steps relative to a Monitoring group that received a pedometer but no incentives. The point estimate represents a 20% increase in steps.

Panel (b) of Figure F.1 shows that the incentive program also improved health relative to both the Monitoring group and a pure Control group that received neither pedometers nor incentives. It meaningfully decreased a blood sugar index of RBS and HbA1c, as well as an overall health risk index that also incorporated blood pressure, body mass index (BMI), and waist circumference (impacts are statistically significant at the 10% level). The improvements in health are larger among those with higher blood sugar at baseline (impacts are statistically significant at the 5% level).

Figure F.1: Treatment Effects of Fixed Step Target Incentives (from Aggarwal et al. 2024)



(a) Treatment Effect on Daily Steps



(b) Treatment Effects on Health Risk

Notes: The figure displays the impact of non-personalized incentives for hitting a 10,000 daily step target. Panel (a) shows the treatment effect on average daily steps walked during the contract period from receiving a step target incentive; the confidence interval bar represents a test of equality between the step target incentive group and the Monitoring group (whose steps were monitored but received no incentive) at the 95% confidence level. Panel (b) shows the treatment effect of the step target incentive program on various health outcomes during the contract period relative to a "pure" Control group, who received neither incentive nor step monitoring; the confidence interval bars represent tests of equality between the step target incentive group and the Control group. The results are shown separately for two health measures: a blood sugar index which is the average of two standardized measures of blood sugar (Random Blood Sugar, a shorter-run measure, and HbA1c, a longer-term measure), and a health risk index, which additionally incorporates Body Mass Index, Blood Pressure, and waist circumference. The results are also shown separately in the full sample and the subsample with above-median values of the blood sugar index at baseline ("high blood sugar sample").

Finally, Table F.1 estimates the marginal health returns to steps during the contract period using all the variation induced by the incentive subtreatments. Specifically, we instrument for average daily steps with a full set of subtreatment dummies. Results include the pure Control group for statistical power; we impute Control group steps using the average of Monitoring group steps. We see substantial and statistically significant improvements in blood sugar and the health risk index in the full sample (Panel A): for each additional 1,000 steps per day improves blood sugar control by 0.03 standard deviations. The estimated marginal returns to steps in terms of blood sugar control are nearly twice as large among those with above-median blood sugar at baseline (Panel B).

Table F.1: Health Impacts of Marginal Steps (from Aggarwal et al. 2024)

	Blood sugar				Other health outcomes		
Dependent variable:	Blood sugar index	Random blood sugar	HbA1c	Health risk index	Mean arterial BP	BMI	Waist circum-ference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Full sample							
Average steps (000's)	-0.030*	-5.26**	-0.015	-0.033**	0.0062	-0.028	-0.13
	[0.017]	[2.19]	[0.048]	[0.016]	[0.27]	[0.026]	[0.17]
Control & Monitoring mean	0.0	195.1	8.4	0.0	103.4	26.5	94.5
# Individuals	3,038	3,038	3,037	3,039	3,027	3,029	3,030
Panel B. Above-median blood sugar sample							
Average steps (000's)	-0.056**	-8.61**	-0.078	-0.046*	0.16	0.026	-0.10
	[0.028]	[3.68]	[0.077]	[0.024]	[0.38]	[0.038]	[0.22]
Control & Monitoring mean	0.5	242.0	9.8	0.4	103.9	26.2	94.6
# Individuals	1,527	1,527	1,526	1,528	1,526	1,525	1,522

Notes: This table shows the effect of average steps per day (in 1000s) during the contract period on health outcomes from an IV specification. Panel A includes the full sample; Panel B includes those with above-median baseline RBS. Control steps are imputed using the Monitoring average. The blood sugar index is the average of RBS and HbA1c standardized, and the health risk index is the average of RBS, HbA1c, mean arterial BP, BMI and waist circumference standardized. Instruments are dummies for each treatment group (Base Case, 4-Day, 5-Day, Daily, Monthly, Small Payment, SMS); Monitoring and Control are omitted. Controls are selected by double-Lasso from those in Panels A, B, and F in Table A.5 column 3; the baseline values of average steps, RBS, HbA1c, and the outcome (if not already included), and their missing dummies. Standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

G The Role of Information About Type: Learned Information and the Nudge

In this section, we explore the impact of varying the information that participants have when they choose their contracts on the menu. Section G.1 examines the impact of varying the information that participants learn about themselves via experience with a pedometer. Section G.2 investigates the impacts of the randomized Nudge given to the Choice + Nudge group.

G.1 Learned Information

To test whether learning about walking costs could improve choice, we varied the amount of time participants had with a pedometer before making their selections on the Base Menu. As described earlier, the Baseline Choice group made their selection before even receiving their pedometer. Among the other non-Baseline Choice participants, we randomly varied whether the Choice survey visit happened immediately after the six day pre-contract period or if we waited an additional week before returning, giving participants seven extra days to learn with a pedometer before making their choices. Thus, we have random variation in whether selections from the Base Menu were made after zero, six, or thirteen days of walking with a pedometer.

To test whether the amount of time that participants had to walk with the pedometer influences choices, we estimate:

$$y_{ik}^{j} = \alpha + \beta_1 \times \text{Choice after 6 Days}_i + \beta_2 \times \text{Choice after 13 Days}_i + \boldsymbol{X}_i'\gamma + \mu_k + \varepsilon_{it}, \quad (4)$$

where the outcome y_{ik}^{j} is an indicator for whether participant i chose contract $j \in \{$10K at 16 INR, 12K at 18 INR, 14K at 20 INR$\}$. Choice after 6 Days and Choice after 13 Days are indicators for making the Base Menu choice after 6 and 13 days with the pedometer, respectively. The sample includes the Baseline Choice, Choice, Monitoring, and Fixed groups.[78] The Baseline Choice group is the omitted group. The coefficients of interest, β_1 and β_2, represent the additional probability of choosing each step target after 6 and 13 days of walking with a pedometer, respectively.

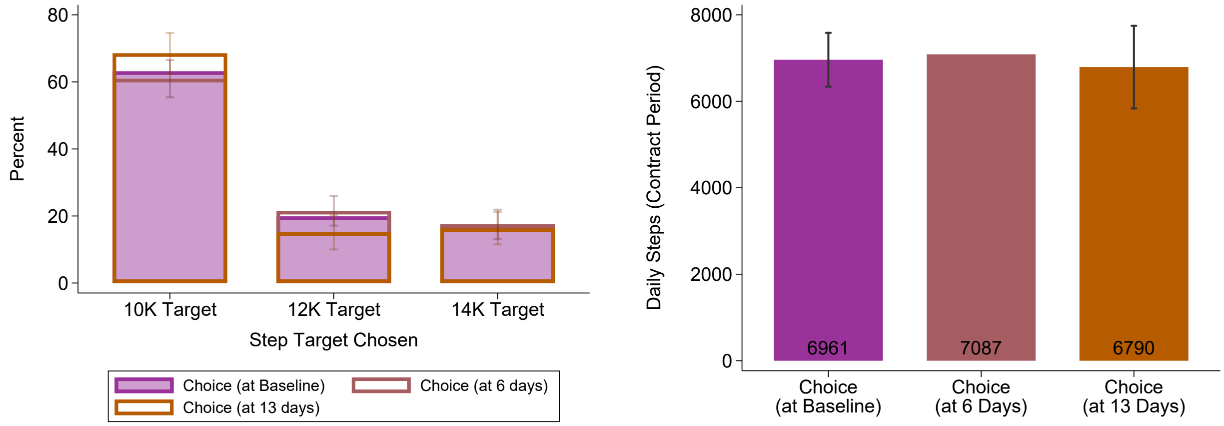
The results are shown graphically in Panel (a) of Figure G.1. Surprisingly, the timing of choice has no detectable effect on contract choice. Participants who have never worn a pedometer or recorded their daily steps prior to selecting their step target are just as likely to select each step target as participants who have had 13 days with the pedometer.

Even though learning does not impact the distribution of step targets selected, it may still impact sorting and therefore effectiveness. We examine this possibility in two ways. First, Table G.1 shows that there is no significant difference in sorting based on baseline steps (specifically, the association between chosen step targets and baseline steps) between those who had their Choice survey after 6 days and those who had it after 13 days.[79] Second, we analyze whether there are differences in contract period walking between those who chose at baseline, after 6 days, and after 13 days. To do so, we estimate a regression of the same

[78] We introduced the Baseline Choice treatment in experiment phase 2.

[79] See the coefficient coefficient on 13 Days Between × Baseline Steps. Note that in either case, we define baseline steps as steps incurred during the pre-contract period (the first 6 days), as we do throughout the paper. We omit the Baseline Choice group from this analysis since their baseline steps were endogenous to treatment.

Figure G.1: Impacts of Choice Timing



(a) Choice Timing Has Small Effects on Distribution of Chosen Targets

(b) Choice Timing has no Discernible Effect on Walking

Notes: Panel (a) shows the distribution of step targets chosen on the Base Menu at the Baseline survey, after 6 days of wearing the pedometer, and after 13 days of wearing the pedometer. The sample in Panel (a) includes the Baseline Choice, Choice, Flat Choice, Monitoring, and Fixed groups, but excludes all who received the Nudge. Panel (b) displays walking patterns during the contract period among each of these three groups. The sample in Panel (b) includes the Baseline Choice and Choice groups. Controls in Panel (a) are selected by double-Lasso for the Medium (12K) Target group from the list of potential controls in column 3 of Table A.5. The selected controls are then used in the regressions for the Low (10K) and High (14K) Target groups. Controls in Panel (b) are selected by double-Lasso from the list of controls in column 1 of Table A.5. All regressions also control for experiment phase, time between Baseline and Choice surveys, and year-month fixed effects (in Panel (a) the year-month fixed effects are for the date of the Baseline survey). The confidence interval bars in Panel (a) represent a test of equality between the percentage choosing a given step target relative to those choosing at baseline, and in Panel (b) the average daily steps walked relative to the Fixed Medium Target group at a 95% confidence level.

basic form as equation 4, but with daily contract period steps as the outcome variable. The results are shown in Panel (b) of Figure G.1. Once again, choice timing has minimal impact on walking.[80]

Taken together, we find no evidence that learning plays a significant role in our context. This may be surprising given that participants likely did not have prior experience with pedometers and step counts. However, based on qualitative interviews, most participants appear to know if they are high, medium, or low walkers. As a result, the fact that the choice menu had 3 levels (which could be interpreted as high, medium, and low) may have provided sufficient guidance to enable good choices. Indeed Woerner et al. (2024) argue that being familiar with a task (as our participants are with walking) may be important for Choice mechanisms to work well.

G.2 Told Information

We now analyze the effect of our randomized Nudge to assess whether Choice can be improved by sharing the principal's information with agents. Recall that, before they made contract menu choices, we provided a nudge to participants in the Choice + Nudge group (along with a random subset of participants in the Monitoring and Fixed groups) by telling them which contract we thought would be most effective. We nudged them towards the target that our Tag algorithm (shown in Table B.1) indicated was best, based on their baseline

[80]These results are confirmed in Table 2. The coefficients on Baseline Choice and Choice are very similar (p-value for equality 0.724).

Table G.1: Heterogeneity in Sorting Between Those with 6 and 13 Days between Baseline and Choice Surveys

Dependent Variable:	Step Target Chosen (Value)
	(1)
13 Days Between	
$\times$ Baseline Steps	-0.0284
	[0.0243]
Baseline Steps	0.203***
	[0.0169]
13 Days Between	156.1
	[232.9]
6 Days Between Mean	11,211
# Individuals	957

Notes: This table shows the differences in step target choices on the Base Menu by time between the Baseline and Choice surveys (either 6 or 13 days). 6 days between is the omitted group. The dependent variable is the value of step target (i.e., 10,000; 12,000; or 14,000). The sample includes the Choice group only. We control for experiment phase, year-month fixed effects, and additional controls selected by double-Lasso from the controls shown in column 3, Panel A, B, E, and F of Table A.5. Robust standard errors are shown in parentheses. Significance levels: * 10%, ** 5%, *** 1%.

steps.[81] These participants were treated identically to the non-Nudge-treated participants in all other regards. Thus, we have random variation in whether participants had access to an expert guess on the contract that was likely to encourage them to walk the most.

Figure G.2 shows the additional probability that participants who were given the Nudge chose each step target, conditional on baseline step level. The red lines indicate the thresholds for baseline steps at which participants were instructed to choose a new target (i.e., at the first red line, the Nudge switched from the Low (10K) target to the Medium (12K) target, and at the second, it switched from the Medium to the High (14K)). The Nudge affected some participants in the intended direction; for example, there appears to be excess mass in nudged participants choosing the Medium target for those with baseline steps between 5,500 and 7,500—the range of baseline steps which mapped to that target.

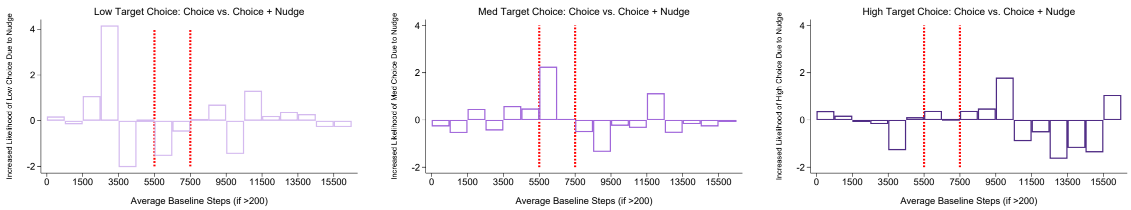
However, our nudge did not work uniformly across the distribution of baseline steps—there appears to be a range of baseline steps for which the Nudge "backfired," decreasing the share of participants who chose the nudged target. Specifically, as shown in Panel (c) of Figure G.2, participants with steps above 10,500 appear less likely to choose the High target, even though that is the target they were nudged towards. Indeed, for participants with steps greater than 10,500, the Nudge decreased the chosen step target by 630 steps, p-value < 0.001,[82] despite the fact that those participants had all been nudged to the highest target.

To more formally test for the presence of heterogeneous Nudge impacts by baseline

[81] Specifically, the surveyor said: "you have walked an average of ⟨BASELINE STEPS⟩ steps in the first 6 days of the 'pre-program period'. People are different and you might know what's best for you. While we can't be completely sure, based on our previous study and your walking levels in the past week we think that you will probably walk the most under a program with a ⟨TARGET⟩ step target."

[82] Column 3 of Table G.2 further substantiates this conclusion, showing that the Nudge decreased the likelihood of choosing the recommended target in that range.

Figure G.2: Impact of the Nudge on Choices



(a) Low (10K) Target (b) Medium (12K) Target (c) High (14K) Target

Notes: The figure shows the impact of the Nudge treatment on chosen step targets. Panels (a), (b), and (c) show the differences in the likelihood of selecting the Low, Medium, and High targets between those in the Choice + Nudge group and the Choice group, by bins of baseline steps.
In each panel, the x-axis represents baseline steps, with each bin except the last one being an interval of 1,000 steps and the last bin including all participants with baseline steps > 15500. For participants in each bin of baseline steps, the y-axis plots the difference in the percentage of participants who selected a given step target among those who were and were not given the Nudge (i.e., the Choice + Nudge relative to Choice group). The two vertical red lines show the cutoffs for which step target nudged participants were told was the "best" for people like them: those with average baseline steps to the left of the first vertical line were told that the Low target was best, those with steps between the two lines were told that the Medium target was best, and those with steps to the right of the second vertical line were told that the High target was best. The sample includes the Choice and Choice + Nudge groups and is limited to the first two phases of the experiment (we did not assign the Nudge treatment in the third phase).

Table G.2: Impact of the Nudge on Choices

Omitted Group:	Choice		
Dependent Variable:	Fraction Choosing Best-Guess Target		
Sample:	All	Baseline Steps < 10,500	Baseline Steps ≥ 10,500
	(1)	(2)	(3)
Choice + Nudge	0.0351	0.0847**	-0.157**
	[0.0335]	[0.0376]	[0.0718]
Choice Mean	0.44	0.39	0.58
# Individuals	907	717	190
Choice + Nudge	540	435	105
Choice	367	282	85

Notes: This table shows the effect of the Nudge on the fraction of participants who chose the target that we recommended. The sample includes the Choice and Choice + Nudge groups in phases 1 and 2. Controls are selected separately for each column using double-Lasso from the list of controls in Table A.5. We also control for experiment phase, time between Baseline and Choice surveys, and year-month fixed effects for the date of the Baseline survey. Robust standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

steps without arbitrarily choosing baseline step ranges for analysis, we use machine learning (specifically, a 4-leaf policy tree, following Athey and Wager (2021), with baseline steps as the predictor) to identify three potential "cut points" where the effect of the Nudge on chosen step target may have switched between positive and negative.[83] The method identified 5,300, 10,500, and 15,100 as the three cut points. Reassuringly, 5,300 is very close to the

[83] We run this analysis in the Choice and Choice + Nudge groups.

actual 5,500 cut point at which the Nudge switched from recommending the lowest step target (which should have had a negative effect on the chosen step target) to recommending the medium step target (which would likely have a positive effect). Moreover, 10,500 is exactly where visual inspection of Figure G.2(c) suggested that the effect of the Nudge on the likelihood of choosing the High target turned negative.

As shown in Table G.3, we then estimate treatment effects of the Nudge on chosen targets (Panel A) and contract-period steps (Panel B) separately for each of the 4 ranges defined by the cut points. Since we are interested in the effects on contract period steps, we limit to the Choice and Choice + Nudge groups; receiving the Nudge thus maps one to one with a dummy for being in the Choice + Nudge group.

Table G.3: Heterogeneity in Nudge Impacts by Baseline Steps

Omitted Group:	Choice			
Baseline Step Range:	0–5,300	5,300 –10,500	10,500 –15,100	> 15,100
	(1)	(2)	(3)	(4)
A. Dependent variable: Chosen Step Target				
Choice + Nudge	-205	295**	-1113***	643
	[166]	[147]	[234]	[398]
Choice Mean	10,849	11,109	12,928	13,167
# Observations	281	431	151	34
# Individuals	281	431	151	34
Choice	106	175	69	12
Choice + Nudge	175	256	82	22
B. Dependent variable: Daily Steps				
Choice + Nudge	-9	-802**	-1127	-1871
	[428]	[338]	[694]	[1466]
Choice Mean	5,511	8,814	11,670	15,986
# Observations	7,286	11,549	3,971	834
# Individuals	268	421	145	32
Choice	100	171	66	11
Choice + Nudge	168	250	79	21

Notes: This table shows the effect of the Nudge treatment on the chosen step target and daily steps separately in four sub-samples of the Choice and Choice + Nudge groups defined by a policy tree prediction. Each column shows the regression result in one sub-sample, where each sub-sample is given by a leaf-node in the policy tree as all individuals in the same leaf-node are assigned the same treatment. This table is based on leaf-node assignment with baseline steps as the only covariate, and the leaf-node definition is the baseline step range shown above the column numbers. Controls are selected by double-Lasso separately for each column from the controls shown in column 1 of Table A.5. Clustered standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

Intuitively, for those below the lowest 5,300 step cutoff, the effect of the Nudge is negative (although insignificant, see column 1) since the Nudge always recommended the lowest target. For those between 5,300 and 10,500 steps at baseline, for whom the Nudge recommended targets that were generally higher than average choices in the no-Nudge sample, the Nudge intuitively increases chosen targets (p-value < 0.05, column 2).

Again, the surprising results come between 10,500 and 15,100 steps.[84] Although all participants were nudged towards the highest step target (and so we would expect the Nudge to positively impact their choices), the effect of the Nudge is robustly negative. It also results in fewer contract period steps, with the Choice + Nudge group walking 1,100 steps less on average (p-value 0.104).

This large, negative impact on contract period walking for this group of high baseline walkers helps explain why the Choice + Nudge group performed much worse than the Choice group (Table 2). The Choice + Nudge group appears to perform just as well (and we cannot reject better performance) as the Choice group for participants with low baseline steps, as shown in column 1 of Panel B of Table G.3.

Despite the fact that the Nudge resulted in higher choices for those with baseline steps between 5,300 and 10,500, it appears to have decreased contract period walking in that group (column 2 of Panel B of Table G.3). This may be due to psychological factors. Similar to reactance (discussed in Section H), participants might have felt pressured to choose the target we suggested to them but may have regretted that choice later. Alternatively, individuals may have known that the target we suggested was not the best for them due to private information, but they may not have felt confident enough in their assessment to disregard our expert nudge, hence choosing a worse target for themselves and decreasing the impact of the Nudge. We also have a negative coefficient on walking among the highest group of walkers (column 4), but this group is very small, with only 34 participants, making it difficult to draw firm conclusions.

[84]The final region, above 15,100 steps, is too small for robust conclusions with only around 40 people. We include it in the table for completeness.

H The Role of Autonomy Effects in Choice's Performance

In this appendix, we investigate an alternative mechanism that might drive the impacts of Choice: autonomy. Because incentive programs aim to influence behavior, non-personalized incentives may lead people to feel that their autonomy is being threatened in a phenomenon called reactance. With reactance, individuals take actions to assert their own freedom in an effort to "reclaim" their autonomy (e.g., Whitehead and Russell, 2004). In our context, this may result in participants asserting their freedom by not complying with their step target contracts. Choice should alleviate this phenomenon and could thus have a positive impact not through sorting but through decreasing reactance and allowing autonomy.

The ideal experiment for identifying autonomy effects isolates the impact of having the power to choose one's own contract from being assigned the contract of one's choice. Specifically, the ideal treatment group would be the Choice and Flat Choice groups, and the ideal control group would include those in the Fixed groups who were assigned their chosen contract by chance, rather than by choice. Note that by "assigned their chosen contract," we mean that both the step target and payment level of the contract they received in the Fixed group would match the contract they chose on the menu.

We approximate this experiment in Table H.1. We do not find any evidence of significant autonomy effects, but our standard errors are relatively large.[85]

[85]In our setting, the ideal treatment and control groups are relatively small: the control (Fixed) groups all received contracts with 20 INR, and so we can only use portions of the "treatment" (Choice and Flat Choice) groups that chose contracts offering 20 INR, which excludes everyone who chose the 10K or 12K contracts on the Base Menu. Hence, we only have an ideal experiment for two groups (a) those who chose the 14K Target on the Base Menu and were randomly assigned to either the Choice group (treatment group) or the Fixed 14K group (control), and (b) those who chose any contract from the Flat Menu (which all paid 20 INR for all step targets) and were randomly assigned to either the Flat Choice group (treatment group) or the Fixed group with the same target they chose (control).

Table H.1: No Significant Autonomy Effects

Omitted Group:	Fixed Targets (Pooled)			
Dependent Variable:	Daily Steps			
Step Targets in Sample:	Pooled	10,000 Steps	12,000 Steps	14,000 Steps
	(1)	(2)	(3)	(4)
Treatment	69	-15	209	262
	[289]	[344]	[806]	[561]
Fixed Target Mean	8,335	6,341	8,104	10,446
# Observations	26,403	12,365	3,445	10,593
# Individuals	991	464	128	399
Flat Choice	618	297	82	239
Fixed Groups	373	167	46	160

Notes: This table shows the treatment effect of being assigned by a "Choice" mechanism among individuals who were assigned their chosen contract from the Flat Menu either randomly (Fixed groups) or by choice (Flat Choice group or Choice group with a High (14K) step target). The dependent variable is daily steps measured using the contract-period pedometer data. The sample includes Flat Choice participants and individuals in the Fixed groups who were randomly assigned the same target they had chosen on the Flat Menu, as well as individuals who chose a High step target on the Base Menu and were assigned that target either randomly or by being in the Choice group. The omitted category is the Fixed groups. We restrict the analysis to the third phase of the experiment, as it was the only phase where Flat Menu choices were incentive-compatible. All columns control for the chosen step target, along with experiment phase, time between Baseline and Choice surveys, and year-month fixed effects. Additional column-specific controls for columns 2–4 are selected by double-Lasso from the list of controls in column 1 of Table A.5. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

I Details on the Back-of-Envelope Calculations of the Value of Steps in Appendix E.1

This section provides details on the calculations behind the four estimates of the cost-savings from exercise displayed in Table E.1 in Appendix E.1.

Estimate 1: Johnson et al. (2015) Johnson et al. (2015) examines a pedometer-based program focuses on adults with type-2 diabetes in Canada. Participants were allocated into intervention and control groups using an interrupted time series design, with the intervention group participating in a pedometer-based walking program.

Over 6 months, the intervention group increased their physical activity by 919 steps/day compared to the control group. According to their estimates, the patient's direct health care costs (including costs of physician services, and in- and out-patient admissions) from a public payer perspective were \$238 lower in the intervention than the control group. (Note that this estimate is not statistically significant due to small sample size of 186 people but we use it for benchmarking purposes.)

Health care costs for the intervention group were \$1176 over 6 months, resulting in direct cost savings of 20.24% ($\frac{\$238}{\$1176}$). Assuming this percent change applies to the total health costs, the total cost savings per 100 additional steps are $20.24\% \times \frac{100}{919} = 2.20\%$. Using the previously calculated daily total cost of 176.91 INR per diabetic patient, the marginal private and public cost savings per 100 additional steps were estimated at 3.90 INR (176.91 INR $\times$ 2.20%) and 2.12 INR (96.37 INR $\times$ 2.20%), respectively.

Estimate 2: Anokye et al. (2018) Anokye et al. (2018) presents an experimental evaluation of the cost-effectiveness of pedometer-based walking interventions over one year in the UK. Participants are randomized into a control group and two intervention groups, both of which received pedometers and various encouragements to exercise (e.g., a walking diary with individualized 12-week walking plan).[86] The interventions were conducted over a yearlong period, and the study estimates the impacts of the interventions on steps and on combined public and private health costs. The interventions (pooled) increased average steps by 660 and decreased total costs by 16.62%, which (assuming linearity in a small region) suggests that the cost savings per 100 steps is $16.62\% \times \frac{100}{660} = 2.52\%$. We then apply this cost saving estimates to the estimates of the public and private costs of care for diabetics in India to estimate that the private and public cost savings per 100 steps walked are 4.46 and 2.43 INR, respectively.

A limitation of this study for our purposes is that its sample is not restricted to diabetic patients—it includes adults aged 45–75 years—but cost savings from a diabetic population should be higher than the general population.

Estimate 3: Di Loreto et al. (2005) This approach estimates cost savings from exercise among individuals with diabetes. Di Loreto et al. (2005) conducted a post hoc analysis on diabetic subjects to investigate the impact of increased energy expenditure on per capita healthcare costs, comparing changes in health costs over two years between groups with higher increments in physical activity to a reference group with no physical activity change.

According to Di Loreto et al. (2005), walking 3 miles per day (at 3 mph) reduces the

[86] One group additionally received nurse-supported consultations covering physical activity only, while the other groups did not. We pool the two groups for power.

total healthcare costs for diabetic patients by \$1,651 over two years, out of their annual total healthcare costs of \$3,990 (USD in 2000). This translates to an annual cost reduction of $(\$1,651/2)/\$3,990 = 20.69\%$. Assuming a linear relationship between exercise and healthcare costs for diabetics, the cost saving of 1-mile walking per day is therefore 20.69% / 3 = 6.90%. Applying the 6.90% reduction to the estimated daily private cost of 176.91 INR and public cost of 96.37 INR yields a marginal private cost saving of 12.20 INR and a public cost saving of 6.65 INR per diabetic for each additional mile of walking (176.91 INR $\times$ 6.90% and 96.37 INR $\times$ 6.90%). Assuming an average walking speed of 2252 steps per mile (Hoeger et al., 2008), we estimate private and public cost savings of 0.54 INR and 0.30 INR per 100 steps (12.20 INR $\times \frac{100}{2252}$ and 6.65 INR $\times \frac{100}{2252}$).

Estimate 4: Yates et al. (2014) This method estimates the reduction in cardiovascular disease (CVD) events from exercise among diabetics. Yates et al. (2014) use prospective data (observational study) to examine the relationship between changes in ambulatory activity (measured by pedometer) over one year and the risk of future cardiovascular events in individuals with impaired glucose tolerance.

Yates et al. (2014) estimates a 5% reduction in CVD event risk with an additional 1,000 steps per day, among high-risk individuals with impaired glucose tolerance. This 5% estimate was also used in Sangarapillai et al. (2021) (cited in Section 6.2) to estimate CVD cost savings (as the product of reduction in CVD incidence $\times$ cost of a CVD event). Such an "incidence-based approach" is commonly applied in cost-of-illness (COI) studies, which provides a projection of potential savings if a preventive measure is implemented (Jo, 2014). Assuming a linear relationship between exercise and CVD prevention, this corresponds to a marginal reduction of 0.5% in CVD event risk per 100 additional steps per day. If we assume that the 0.5% reduction in CVD event risk applies to total healthcare costs, the estimated reduction in private and health healthcare costs per 100 steps would be 0.88 INR and 0.48 INR for a diabetic patient (176.91 INR $\times$ 0.5% and 96.37 INR $\times$ 0.5%).

Online Supplement References

Aggarwal, S., R. Dizon-Ross, and A. D. Zucker (2024). Designing incentives for impatient people: An RCT promoting exercise to manage diabetes. *NBER Working Paper*, No. 27079.

Anokye, N., J. Fox-Rushby, S. Sanghera, D. G. Cook, E. Limb, et al. (2018). Short-term and long-term cost-effectiveness of a pedometer-based exercise intervention in primary care: A within-trial analysis and beyond-trial modelling. *BMJ open 8*(10), e021978.

Athey, S. and S. Wager (2021). Policy learning with observational data. *Econometrica 89*, 133–161.

Di Loreto, C., C. Fanelli, P. Lucidi, G. Murdolo, A. De Cicco, et al. (2005). Make your diabetic patients walk: Long-term impact of different amounts of physical activity on type 2 diabetes. *Diabetes care 28*(6), 1295–1302.

Hoeger, W. W., L. Bond, L. Ransdell, J. M. Shimon, and S. Merugu (2008). One-mile step count at walking and running speeds. *ACSM's Health & Fitness Journal 12*(1), 14–19.

Jo, C. (2014). Cost-of-illness studies: Concepts, scopes, and methods. *Clinical and molecular hepatology 20*(4), 327.

Johnson, S., D. Lier, A. Soprovich, C. Mundt, and J. Johnson (2015). How much will we pay to increase steps per day? Examining the cost-effectiveness of a pedometer-based lifestyle program in primary care. *Preventive Medicine Reports 2*, 645–650.

Sangarapillai, T., M. Hajizadeh, S. S. Daskalopoulou, and K. Dasgupta (2021). Cost-comparison analysis of a physician-delivered step-count prescription strategy. *CJC open 3*(8), 1043–1050.

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

Woerner, A., G. Romagnoli, B. M. Probst, N. Bartmann, J. N. Cloughesy, and J. W. Lindemans (2024). Should individuals choose their own incentives? Evidence from a mindfulness meditation intervention. *Management Science*.

Yates, T., S. M. Haffner, P. J. Schulte, L. Thomas, K. M. Huffman, et al. (2014). Association between change in daily ambulatory activity and cardiovascular events in people with impaired glucose tolerance (navigator trial): A cohort analysis. *The Lancet 383*(9922), 1059–1066.